Instruments Measuring Prospective Memory: A Systematic and Meta-Analytic Review

Running title: Instruments Measuring Prospective Memory

Geoffrey Blondelle, Mathieu Hainselin*, Yannick Gounden, Véronique Quaglino

CRP-CPO, EA 7273, Université de Picardie Jules Verne, Amiens, France.

* Corresponding author at: CRP-CPO, EA 7273, Université de Picardie Jules Verne, Amiens, 80025 France. Tel.: +33 (0)3 22 82 89 22 ; E-mail address : mathieu.hainselin@u-picardie.fr

Supplementary material (4 Tables and 4 Figures):

- **Supplementary Table S1.** Reviewed studies for available PM test batteries
- **Supplementary Table S2.** Reviewed studies for available PM single-trial procedures
- **Supplementary Table S3.** Reviewed studies for available PM questionnaires
- **Supplementary Table S4.** Reviewed studies for available PM experimental procedures
- **Supplementary Figure S1.** PM test batteries: funnel plot of standard error showing study dispersion
- **Supplementary Figure S2.** Single-trial PM procedures: funnel plot of standard error showing study dispersion
- **Supplementary Figure S3.** Self-reported PM questionnaires: funnel plot of standard error showing study dispersion
- **Supplementary Figure S4.** PM experimental procedures: funnel plot of standard error showing study dispersion

Abstract

Objective: To identify the available measures to assess prospective memory (PM) abilities, to describe their content and to quantitatively summarize the effects of various diseases on PM depending on the type of assessment.

Method: Three databases (PsycInfo, PsycArticles and PubMed) were searched up to June 2019 to identify the existing prospective memory measures. The identified PM measures were classified according to the type of assessment: test batteries, single-trial procedures, questionnaires and experimental procedures. The characteristics and psychometric properties were assessed. PM performance were compared between patients with various diseases and controls depending on the type of assessment.

Results: In total, 16 measures were identified. Most measures evaluated both event- and time-based tasks, were linked to functional outcomes, showed empirical evidences regarding validity and reliability and provided parallel versions. To a slightly lesser extent, few measures provided normative data, translations/adaptation into another language, cutoff scores for diagnostic purposes, qualitative scoring, parallel version and external aids during the test. Compared to healthy controls, patients had significantly poorer performances when PM was assessed with experimental procedures. Subgroup analyses indicated consistent PM impairments for patients relative to controls for three test batteries. PM complaints did not differ between patients and controls and the scores were homogeneous for the Comprehensive Assessment of Prospective Memory.

Conclusions: This work contributes in inventorying the existing PM measures both for research and clinical purposes. We suggest some future directions based on these findings to enhance clinical applicability of PM assessment instruments.

Keywords: Learning and Memory; Assessment; Meta analysis; Everyday functioning

Introduction

Everyone forms intentions that are not executed immediately (e.g., taking medication) but are instead scheduled for another moment or context (e.g., at 8 pm or during diner). The term *prospective memory* (PM), or *realization of delayed intentions* (Ellis, 1996), is used to define memory for activities to be performed in the future (Einstein & McDaniel, 1990). It is commonly distinguished from retrospective memory, which refers rather to the ability to remember past information (e.g., remembering the activities we did during the last holidays).

PM can be defined as a multicomponential process which involves a prospective component to remember *that* something has to be done (intent) and a retrospective component to remember *what* and when has to be done (content) (Einstein & McDaniel, 1990, 1996). Another distinction has been made according to the nature of the cue that triggers the retrieval of the delayed intention: event- or time-based. In conditions requiring time-based PM tasks, the intention execution is auto-initiated by the person in a specific temporal frame (e.g., taking medication *at 8 pm*). In event-based PM tasks, the intention execution is prompted when an external cue occurs (e.g., taking medication during diner). These two theoretical distinctions are supported by neuropsychological studies highlighting dissociation between the two PM components (e.g., Hainselin et al., 2011; Umeda, Nagumo, & Kato, 2006), as well as between time- and event-based tasks (e.g., Yang, Zhong, Qiu, Cheng, & Wang, 2015). According to McDaniel and Einstein (2007), a typical PM task requires to respect several features: 1) the action must not be fulfilled immediately, there must be a delay between the encoding and the retrieval phases; 2) it should be embedded into another task (named "ongoing task") in which the PM cue represents a part of the situation; 3) the time period during which the action can be performed must be established; 4) the time required to perform the PM task must be established (e.g., deadline for taking medication) and 5) the to-be-

performed action must be formulated consciously without remaining constantly in mind otherwise, it would become a vigilance task.

The prevalence of PM lapses is estimated between 50% to 80% of everyday memory problems (Crovitz & Daniel, 1984; Terry, 1988), such that PM has a significant impact on autonomy, especially for professional difficulties and medication non-adherence (e.g., Mathias & Mansfield, 2005; Zogg, Woods, Sauceda, Wiebe, & Simoni, 2012). Several meta-analyses have reported a PM impairment in normal aging, especially after 70+ (Henry, MacLeod, Phillips, & Crawford, 2004; Ihle, Hering, Mahy, Bisiacchi, & Kliegel, 2013), and in a wide range of clinical groups, including neurodegenerative (Ramanan & Kumar, 2013; van den Berg, Kant, & Postma, 2012), neurodevelopmental disorders (Landsiedel, Williams, & Abbot-Smith, 2017), neurological injuries (Wong Gonzalez, 2015) and psychiatric syndromes (Wang et al., 2009; Zhou et al., 2017). The PM impairments in these clinical groups reflects a multiprocess model which argues that successful performance not only requires the episodic retrieval of delayed intentions based on a (bottom-up) automatic-associative memory system (McDaniel & Einstein, 2000; Moscovitch, 1994), but also a strategic monitoring system (Smith, 2003; Smith & Bayen, 2004). This latter system operates for complex PM tasks, requiring attention allocation and executive control for relevant environmental cues to activate delayed intentions. This strategic monitoring system has been found to be associated with activation in a predominantly frontoparietal network including lateral Broadman area 10, Broadman area 40, insula and anterior cingulate (for a review, see McDaniel, Umanath, Einstein, & Waldum, 2015), often impaired in normal aging, patients with mild cognitive impairment, autism spectrum disorders, traumatic brain injury or schizophrenia. Therefore, assessing PM performance in addition to retrospective episodic memory, would provide great benefit for patients for which functional outcomes and daily issues are not identified by traditional memory tasks used by clinicians, and for detecting

individuals who are at risk of developing dementia (Rabin et al., 2014; Troyer & Murphy, 2007).

However, PM assessment is barely used in the daily neuropsychologists' clinical practice. In their 747 neuropsychologists survey, Rabin, Barr and Burton (2005) showed that the Wechsler Adult Intelligence Scale-Revised (Wechsler, 1987) and the Wechsler Memory Scale-Third Edition (Wechsler, 1997), two batteries without PM subtests, were ranked in 1st position, endorsed by 70.80% of the respondents. The single test with partial PM assessment, the Rivermead Behavioural Memory Test (Wilson, Cockburn, & Baddeley, 1985), was ranked in 19th position and endorsed by only 6.40% of respondents.

The 10-year follow-up study, Rabin, Paolillo and Barr (2016) reported the same top 5 answers, and not a single PM assessment was mentioned as the respondents' daily clinical practice. However, at the same time, many papers developed several measures to assess PM abilities (i.e., paper-and-pencil psychological PM measure, single-trial procedures, questionnaires and experimental procedures). Rabin et al. (2016) also reported ongoing challenges encountered by neuropsychologists, these included the lack of: adequate normative data, ecological validity, reliability, diagnostic accuracy, parallel version, translation into another language, intercultural adaptation and population-specific assessment instruments. In the current review, we show that PM measures do not cover all of these factors, which in turn render clinical neuropsychologist reluctant to use time-consuming PM assessment in clinical practice, as it also requires trained personnel. Anecdotally, the first normative ecological PM measure, the Cambridge Test of Prospective Memory (CAMPROMPT; Wilson, Emslie, Foley, Shiel, Watson, Hawkins, & Groot, 2005), was published and available for clinicians in 2005.

The goal of the current paper was (1) to identify through a systematic review the available measures to assess PM abilities by describing their content and (2) to use a meta-

analytical approach to quantitatively summarize the effects of various diseases on PM depending on the type of assessment.

Methods

Protocol and Registration

To our knowledge, there is currently no review protocol to examine the objectives of the current study. To objectively assemble and screen the literature in search of PM assessment tools, we selected empirical studies that met the criteria according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines (PRISMA; Gates & March, 2016; Liberati et al., 2009).

Eligibility Criteria

Criteria for inclusion were: (1) peer-reviewed journal articles, paper presented in scientific conferences or dissertations (2) published in English language and (3) studies that have assessed PM abilities. Studies were excluded if they were (1) primarily focused on another area than PM (2) PM studies with training/rehabilitation purposes (3) focused on non-human populations (3) single case studies and (5) review articles, systematic reviews and meta-analyses. All studies that met the inclusion and exclusion criteria were included in the review. To maximize the identification of existing measures to assess PM, the literature searches were not limited by the age and neurological status of individuals.

Information Sources

A systematic search of published studies was conducted by the first author using PubMed, PsycArticles and PsycInfo databases. No search was conducted after June 12, 2019.

Literature Search

The initial search was conducted on PubMed, PsycArticles and PsycInfo databases and included the following terms in abstract, titles or keywords: “prospective memory”, “standardization”, “test”, “questionnaire”. In an effort to identify studies investigating PM assessment tools psychometric properties, we added appropriate search filters query proposed by Terwee, Jansma, Riphagen, and De Vet (2009). Therefore, we added the most sensitive (i.e., “valid*” and “reliab*” with percentages of 39.70% and 37.90%) and specific (i.e., “internal consistency”, “psychometrics” and “validation studies” with a precision percentage of 100,00%, 42.30% and 35.70%, respectively) terms in abstract, titles or keywords to the final search query. For all articles found, titles, abstracts and keywords were screened for eligibility and the Abstrackr machine learning tool (Wallace, Small, Brodley, Lau, & Trikalinos, 2012) was used for screening of eligibility.

Data Items

Data collected from each reviewed study contained the names of the tests used, country of publication (ISO 3166) with language, age range with mean and standard deviation, education range with mean and standard deviation, samples size with the male/female ratio, normative data or mean and standard deviations of PM performance as a dependent variable (i.e., the proportion of PM cues correctly responded or the proportion of PM complaints for questionnaires) for patients and healthy controls, total duration of the test with the length of the retention interval, number of PM item, and study reference. We developed an algorithm for each study to clarify whether the identified PM measures meet the frequent challenges associated with selection of neuropsychological instruments reported in Rabin et al.' study (2016): language translation, cross-cultural adaptation, validity assessed (irrespective of quality), reliability assessed (irrespective of quality), normative data collected, diagnostic value (irrespective of quality), parallel version provided, linked to functional

outcome measure, qualitative scoring. This initial algorithm has also been extended to three other key variables of interest specific to the field of PM, namely event-based tasks, time-based tasks and the use of external aids during the test. Each study was assigned one or a combination of these 12 key variables depending on whether it has endeavored these criteria.

Summary Measures, Synthesis of Results, and Risk Bias Across Studies

The identified PM measures were assigned to four distinct types of assessment: test batteries, single-trial measures, questionnaires and experimental measures. This classification is based on the distinction commonly made in the field of neuropsychological assessment between objective and subjective measures, as well as the number of PM items included in the test (e.g., Kinsella, Pike, Cavuoto, & Lee, 2018). The number of studies included in the current systematic review and meta-analysis was calculated, as well as the number of PM measures identified for each category, mean age and education (in years) of samples. An excel sheet was created for each assessment type to facilitate the identification of studies, as well as to reduce the risk of counting duplicates. Our 3 stages analysis included 1) identifying key variables that met the criterion for each study, 2) aggregating each key variable occurrence meeting the criterion for each of the identified PM measures, and 3) calculating the percentages of criteria met for each key variable according to the number of measures assigned to the concerned category.

Meta-Analytic Approach

We used MedCalc 19.0.3 software (Mariakerke, Belgium) to analyze the data. We used the random effects model for all analyses in order to provide a more realistic approach when combining data from various methodology and sample characteristics compared to the fixed effects model (Borenstein, Hedges, Higgins, & Rothstein, 2010; Cheung & Vijayakumar, 2016).

Given the small sample sizes of some of the included studies, we calculated the Hedges' *g* as a standardized mean difference method to estimate study effect sizes, as recommended by the Cochrane Collaboration (Higgins & Green, 2011). Effect sizes were considered as small, medium and large when $g \geq .20$, .50 and .80 respectively (cf. Cohen, 1988).

Six studies reported data on multiple, but distinct PM measures. The main feature of these studies is that the *same* participant provided data on at least two different PM tasks, including both objective (i.e., test batteries or experimental procedures) and subjective (i.e., questionnaires) measures. In fact, such cases are problematic in meta-analyses because we cannot treat the different outcomes as though they were independent as this would lead to misleading estimate the variance for the overall effect (cf. Senn, 2009). In an effort to improve the reliability of our analyses, and given that the administration of objective and subjective measures is a traditional approach used by the authors to assess PM, Hedges's *g*s of individual studies were pooled to a mean effect size according to the type of assessment (i.e., test batteries, single-trial measures, questionnaires and experimental measures). Pooled effect sizes in the negative direction indicated that PM performance was lower for the patients compared to healthy controls.

The homogeneity of the effect sizes between the samples was measured using the *Q* statistic. A significant *Q* index indicates that the variance of effect sizes in the population is greater than expected as compared to the sampling error. We also calculated the I^2 statistic, which refers to the percentage of variation across studies that is due to heterogeneity rather than chance (Higgins & Thompson, 2002; Higgins, Thompson, Deeks, & Altman, 2003). A I^2 value of 0% indicates that there is no heterogeneity, while a larger percentage indicates an increase in heterogeneity. The heterogeneity was assumed to be low, moderate and high when I^2 value was 25%, 50% and 75% respectively (Higgins et al., 2003). In the case where

heterogeneity estimates indicated a substantial difference between individual studies, we conducted planned subgroup analyses for all measures included in the assigned assessment category to further examine the source of the heterogeneity. Subsequent analyses were conducted only for measures that were used in at least two different studies.

A common limitation of meta-analyses is the existence of a publication bias or a file-drawer effect (Rosenthal, 1979). A publication bias is characterized by a trend to publish more studies showing statistically significant results than studies with non-significant results. This causes a Type I publication bias error and results in a spurious effect of the parameter under consideration. Typically, the existence of a publication bias is characterized by an asymmetrical funnel shape, with a Egger test $p < .05$. To overcome this bias, the influence of unpublished studies should be taken into consideration. Therefore, we used funnel plots and Egger's tests to examine whether asymmetry due to publication bias was present in the study and we also applied Rosenthal's (1979) fail-safe *N* formula to estimate the number of unpublished studies with null findings.

Results

The initial literature search of the 3 databases generated a total of 326 references (63 in PubMed, 4 in PsycArticles and 259 in PsycInfo) and 15 additional studies were identified in the reference lists of these articles and other studies known to the first author through previous readings were also considered for inclusion. From the 341 references, 50 duplicate records were excluded, 105 were excluded based on titles and abstracts and consequently, 186 full texts of articles were retained. After having removed duplicates, reviewed the entire full content of articles and applied the exclusion criteria, the number of studies that met the inclusion criteria was 52 and 23 for the literature review and the meta-analysis respectively. The flowchart showing the selection process is depicted in Figure 1.

The literature review identified a total of 16 PM measures including:

- Five test batteries (see Table 1 for an overview of criteria met and Supplementary Table S1 for the characteristics of the studies included in this section): the Rivermead Behavioural Memory Test (Wilson et al., 1985) the Cambridge Behavioural Prospective memory Test (Kime, Lamb, & Wilson, 1996), the Cambridge Test of Prospective Memory (Wilson et al., 2005), the Memory for Intention Screening Test (Raskin, 2004), the Royal Prince Alfred Prospective Memory Test (Radford, Lah, Say, & Miller, 2011).
- Three single-trial procedures (see Table 2 for an overview of criteria met and Supplementary Table S2 for the characteristics of the studies included in this section): the envelope task (Huppert, Johnson, & Nickson, 2000), the prompt card task (Delprado et al., 2012) and the telephone test (Hsu, Huang, Tu, & Hua, 2014).
- Four questionnaires (see Table 3 for an overview of criteria met and Supplementary Table S3 for the characteristics of the studies included in this section): the Prospective Memory Questionnaire (Hannon et al., 1990), the Prospective and Retrospective Memory Questionnaire (Smith, Del Sala, Logie, & Maylor, 2000), the Comprehensive Assessment of Prospective Memory Questionnaire (Waugh, 1999) and the Brief Assessment of Prospective Memory questionnaire (Man, Fleming, Hohaus, & Shum, 2011).
- Four experimental procedures (see Table 4 for an overview of criteria met and Supplementary Table S4 for the characteristics of the studies included in this section): the Prospective Remembering Video Procedure (Titov & Knight, 2001), the Test Écologique de Mémoire Prospective (Potvin, Rouleau, Audy,

Charbonneau, & Giguère, 2011), the Virtual Week and the Actual Week (Rendell & Craik, 2000).

***** insert Figure 1 about here *****

Test Batteries of PM

Rivermead Behavioural Memory Test (RBMT). The RBMT (Wilson et al., 1985, commercially distributed) is an 11 sub-tests ecological everyday memory test with 3 separate event-based tasks (e.g., remembering to ask the experimenter for the next appointment time when an alarm sounds) among the 11 sub-tests making up the battery. The original RBMT has been translated into fourteen languages (Wilson, 2009). The test was subsequently standardized for older adults (Cockburn & Smith, 1989) and adapted for both adolescents (Wilson, Forester, Bryant, & Cockburn, 1990) and young children as the Rivermead Behavioural Memory Test for Children (RBMT-C, commercially available; Wilson, Ivani-chalian, Besag, & Bryant, 1993). The RBMT-3 is the last commercially published version of the test and provides a general memory index based that follows the basic principles of standardized IQ score but does not provide a standardized PM score (Wilson et al., 2008).

The validity of the RBMT has been assessed on the basis of therapists' observations of 80 brain-damaged patients (35 hours of observation per patient; range 16–55 hours) suffering from everyday memory failures (Wilson, Cockburn, Baddeley, & Hiorns, 1989). Wilson (1991) showed that the standardized profile scores obtained at the RBMT were good predictors of functional independence (e.g., having a paid job) for patients who experienced severe head injury, although other authors (Mathias & Mansfield, 2005; Mills et al., 1997) could not replicate this result. The limited number of items (3 PM items only), the lack of

TBPM tasks and long-term naturalistic task and a ceiling effect (Mathias & Mansfield, 2005) reduce the validity of this measure. Wilson herself (2009) argues that the RBMT "is not sufficient on its own. It can highlight some of the areas that one might want to tackle in a treatment program but it does not specify with sufficient precision the nature and extent of the everyday problems in such a way that we can set appropriate goals" (p. 46). Indeed, one study reported that even older adults without cognitive impairment and functional difficulties failed PM tasks of the RBMT, especially for the *Appointment* and *Belonging* sub-tests, but not for the *Message* sub-test (Martin, Kliegel, & McDaniel, 2003).

Cambridge Behavioral Prospective Memory Test (CBPMT) and the Cambridge Test of Prospective Memory (CAMPROMPT). The CBPMT was initially in a study of a patient with severe amnesia (Kime et al., 1996) and adapted in an extended 40-minutes version, including 4 time-based and 4 event-based PM tasks, for people with brain injury and controls to specifically assess the construct of PM. Despite the lack of validation or normative data, the CBPMT is sufficiently sensitive to identify variations in PM performance between brain-damaged patients and healthy controls (Groot, Wilson, Evans, & Watson, 2002). The CBPMT is the first assessment to allow the participants to take notes to help them in remembering PM tasks. Interestingly, note takers performed better than non-note takers, regardless of brain injury presence or absence (Groot et al., 2002).

Wilson et al. (2005) improved the scoring of the CBPMT and created the CAMPROMPT (*commercially published*). Contrary to the CBPMT, the CAMPROMPT provides normative data based on age and IQ. The test includes six (3 time- and event-based tasks) tasks and requires 25 to 30 minutes for completion. Furthermore, participants are proposed to perform a set of distractor tasks comprising word-finder puzzles or a general knowledge quiz during a 20-minutes delay prior to performing the PM tasks.

The initial validation and normative data of the CAMPROMPT were collected on 72 patients (mainly traumatic brain-injured patients and patients with degenerative neurological conditions) and 212 healthy controls, ranging from 16 to 92 years old. Wilson et al. (2005) found a moderate correlation of .38 between the total profile score of the Rivermead Behavioural Memory Test (RBMT) and both the CAMPROMPT total score and the event-based PM score sub-scale (r = .47 for each), but not between the total profile score of the RBMT and the event-based PM score sub-scale. Because of the lack of time-cues PM task in the RBMT, and the wide range of cognitive abilities it encompasses, it might not be still considered as the PM evaluation gold standard. The CAMPROMPT is sensitive enough to distinguish control participants from smokers (Heffernan, O'Neill, & Moss, 2010a), amnestic mild cognitive impairment (Delprado et al. 2012) and young binge-drinkers (Heffernan & O'Neill, 2012). Patients with spina bifida meningomyelocele had also poorer performances than controls (Dennis, Nelson, Jewell, & Fletcher, 2010). The authors also noted that patients took fewer notes than controls (50.00% vs 82.35%) which was inconsistent with the hypothesis of internal control mechanisms problems.

The Memory for Intentions Screening Test (MIST). The MIST (*commercially published,* Raskin, 2004) provides comprehensive scoring system for omissions (e.g., loss of content or time) and commission errors (e.g., task substitutions). These variables have proved to be relevant in clinical research (see Woods, Twamley, Dawson, Narvaez, & Jeste, 2007 for patients with schizophrenic disorders; but also with HIV-infected individuals Carey et al., 2006; Woods, Iudicello, et al., 2008a). The MIST includes a total of 8 PM tasks (4 time- and event-based tasks), with two parallel versions, norms on 736 participants from 18 to 94 and education percentiles. The MIST includes a more ecological (optional) task where participants have to leave a phone message to the clinician 24-hour after the testing.

Woods et al. (2008b) later published the psychometric characteristics of the MIST collected on 67 healthy adults, ranging from 19 to 74 years old, but no clinical group was enrolled. The correlation analyses showed an acceptable split-half reliability (.70; Spearman-Brown coefficient) and an excellent inter-rater reliability (.99). However, the poor internal consistency for the eight PM tasks (Cronbach's α: .48) might be due to the particularly high level of education of the participants and the restricted range of scores observed in this sample. The authors also showed that the call-back PM task was not linked to any other MIST measures and demographic characteristics. Indeed, unlike the other MIST items, the participants could use strategies such as taking notes, but did not receive specific advice. However, the authors neither recorded nor published data concerning the number and the type of strategies that may have been used, limiting the conclusions that could support the psychometric properties of this long-term PM task. Carey et al. (2006) showed deficits in time- and event-based tasks, as well as more failure on the 24-hour delay PM task and substitution errors for HIV patients compared to controls. A Receiver Operating Characteristic (ROC) analysis highlighted a high discriminative power for the MIST (acceptable sensitivity and specificity with a coefficient of the area under the curve of .83) in predicting global neuropsychological impairment with acceptable sensitivity (.73) and specificity (.74) coefficients. The MIST demonstrated a good ecological validity via significant relationships with the Instrumental Activities of Daily Living (IADL; Lawton & Brody, 1969) scale (Woods, Iudicello, et al., 2008a).

Although the MIST and the CAMPROMPT integrate in their design some useful and relevant indicators which allow a more comprehensive measurement of PM skills in various samples, their long administration time (30–40 minutes on average) is a technical limitation for incorporation into a classical two-hour neuropsychological assessment.

Royal Prince Alfred Prospective Memory Test (RPA-ProMem). The design of the RPA-ProMem renders it easier to be incorporated into a classical neuropsychological assessment since it takes less than 15 minutes for administration and does not include classical distractor tasks (i.e., "filler" tasks such as puzzles or questionnaires), thereby reducing additional cognitive demand for patients. The RPA-ProMem also includes a PM task to be carried outside the laboratory with a longer period (1 week after testing) compared to the 24-hour delay ecological PM task of the Memory for Intentions Screening Test (Raskin, 2004).

The validation of the test was conducted by Radford et al. (2011) with 20 patients presenting various brain disorders (ranging from 18 to 63 years old) and 20 healthy control participants (ranging from 21 to 64 years old). The RPA-ProMem proved to be sensitive enough to identify patients' PM deficits compared to healthy controls. Radford et al. (2011) did not show any correlation between the RPA-ProMem and the Memory for Intentions Screening Test (MIST; Raskin, 2004) in the control group. According to the authors, this could be partly due to the fact that MIST does not allow participants to use external aids in contrast to the RPA-ProMem. Concluding on these elements is thus difficult given that Radford et al. (2011) did not pay attention to the possible use of other external aids not part of the RPA-ProMem. Rabin et al. (2014) showed that those with amnestic mild cognitive impairment had worst performance than controls on the RPA-ProMem for time- and event-based PM tasks, as well as for both short- and long-term delays. Patients with subjective cognitive decline also scored lower than controls on long-term and naturalistic subtask. The authors also reported a strong inter-rater reliability (coefficient of intraclass correlation of .97) and a good alternate form of reliability (Rho = .71). Notably, the RPA-ProMem total scores were negatively correlated with informant reports, the Comprehensive Assessment of Prospective Memory (CAPM; see below) and the Instrumental Activities of Daily Living (IADL; Lawton & Brody, 1969), which demonstrates the ecological validity of the tool.

Beyond the interest of questionnaires in assessing functional difficulties in everyday life, the RPA-ProMem may be interesting when these measures cannot be obtained from informants.

***** insert Table 1 about here *****

Single-Trial Procedures

The envelope Task (Huppert et al., 2000), the prompt card task (Delprado et al., 2012) and the Telephone Test (Hsu, Huang, Tu, & Hua, 2014) have been developed to assess PM under a shorter time.

Envelope Task. The envelope task is a single-trial event-based PM task used by Huppert et al. (2000) to establish the prevalence of PM impairment in an elderly population. Despite the lack of formal standardization, the envelope task was administered to 11,956 individuals aged 65 years and above. The clinician tells the participant that he/she will have to write a given name and address ("John Brown, 42 West Street, Bedford") on an envelope when it is shown, to add their own initials, seal it and return it back to the clinician. The test allows to assess both the prospective and retrospective components of PM. To assess the PM prospective and retrospective components, the participant has to remember to do something after receiving the envelope within about 5 to 10 seconds. The clinician gives a prompt if the participant did not do so within the proper time or performed only one action (i.e., just seal the envelope or just write initials on the back). Responses are coded as follows: 2 (correct action without prompt), 1 (correct action with prompt) and 0 (the participant did not remember the action, even when he/she was prompted). The clinician also scored 2, 1 or 0 point for correct action following a prompt in order to assess the retrospective memory component. These instructions are followed by a 10-minute interval in which the participant has to perform a set of cognitive

tasks. Huppert and colleagues (2000) reported that decrement of PM performance was linearly and strongly correlated with age. They also showed that 54% of individuals aged 65 successfully performed the task without prompt, compared to 19% for the elderly over 90 and 8% for individuals with probable dementia (n = 388). The envelope task is also sensitive to spot patients with amnestic mild cognitive impairment (Lee et al., 2016). The authors administered a single-item subjective rating scale for which the participants were asked to assess the effectiveness of their memory on a daily basis compared to individuals of the same age. Compared to controls, their results showed that patients performed poorly the envelope task compared to controls. Their results also showed that the envelope task achieved a better level of discrimination compared to the subjective memory rating (area under the curve coefficient of .83 and .76, respectively). The specificity of the envelope task in detecting a group difference was good (91.9%), although its specificity was low (64.3%).

Prompt Card Task. This single-trial event-based prompt card task starts with writing details about the next appointment on a card that the participant is supposed to give to the clinician at the end of the session. The diagnostic value of the prompt card task seems interesting, and showed poorer PM performance for patients with amnestic mild cognitive impairment patients compared to healthy participants (Delprado et al., 2012). A ROC analysis was performed on a verbal retrospective episodic memory measure (CVLT-II; Delis, Kramer, Kaplan, & Ober, 2000) and the three measures of PM to determine their diagnostic value. For the case of PM, the CVLT-II has shown to bear the highest discriminative power in identifying patients from healthy participants with a coefficient of the area under the curve of .93; followed by the envelope task (.85), the prompt card task (.77) and the Cambridge Test of Prospective Memory (.76 for both time- and event-based sub-tests). It was quite predictable that the CVLT-II has been found to be the best measure to distinguish patients from healthy

participants because a similar retrospective memory screening measure was used prior to the investigation to diagnose patients with amnestic mild cognitive impairment. In conclusion, the envelope task seems to be a decent PM measure to identify patients with mild cognitive impairment, and the best tool when it is compared with the prompt card task the Cambridge Test of Prospective Memory (Wilson et al. 2005).

Telephone Test. The telephone test is the only single-trial procedure that allows to assess time-based PM (Hsu et al., 2014). Participants are requested to remind the clinician to make a phone call to the counter 5 minutes after the instruction. Like the envelope task, the telephone test allows to measure both prospective and retrospective PM components. A prompt is given to participants if no action is triggered within the 60 seconds following the 5-minutes delay. For the prospective component, 2 points are given when the participant reminds the experimenter that something needs to be done within the 60 seconds following the 5-minutes delay, 1 point if the reminder is given after this delay and 0 point is scored if the participant does not perform the expected action. For the retrospective component, 2 points are given if the content of the action is correctly recalled, 1 point if the participant does not remember the content of the action but remembers that something needs to be done with the telephone or the counter. Combining the telephone test and the envelope task scores, the authors showed poorer performance for patients with dementia compared to healthy controls and negative correlations between informant rating of both prospective and retrospective sub-scales of the Prospective and Retrospective Memory Questionnaire (see below) and combined PM scores (r = -.57; -.58). In another study with patients with subjective cognitive decline, Hsu, Huang, Tu, and Hua (2015) found poorer performance on the telephone test for compared to healthy controls, especially for the prospective component.

***** insert Table 2 about here *****

Questionnaires

Prospective Memory Questionnaire (PMQ). The PMQ (Hannon et al., 1990) was the first self-report PM measure available for assessing PM failures as well as the frequency use of memory aids.

The initial validation study (Hannon et al., 1990) included 361 individuals (291 healthy students, 19 brain-injured students, 14 brain-injured patients from a rehabilitation center and 37 alcohol-dependent patients). It initially included 74 items of situations requiring PM (e.g., item 1 "I missed appointments I had scheduled") rated on a 9-point Likert scale. The PMQ assesses several dimensions of PM with 5 sub-scales: Long-Term Episodic, Short-Term Habitual, Internally Cued Scale and Techniques to Remember. The latest published version of the PMQ includes 52 items (Hannon, Adams, Harrington, Fries-Dias, & Gipson, 1995). The authors have confirmed this initial factor analysis with another factor analysis using *varimax* rotation in healthy younger adults and older adults, but also with brain-injured patients. The internal consistency coefficient of the PMQ was high (.92) and ranged from .78 to .90 for the sub-scales.

Brain-injured patients and age-matched healthy older adults performed poorer than younger adults on three PM measures including short- and long-term ecological tasks with an alpha coefficient of .76. Moreover, groups differed only on one dimension of the PMQ, namely the Short-Term Habitual sub-scale. Hannon et al. (1995) also reported negative relationships between scores on the short-term tasks and total scores obtained on the PMQ (r = -.17), but also for the 3 sub-scales of the PMQ, namely Long-Term Episodic, Short-Term Habitual and Internally Cued sub-scales (r = -.19; -.25; -.22). The PMQ has a good test-retest

reliability with a coefficient of .88 for the PMQ among 72 participants of the sample 10 to 14 days after the initial administration.

Heffernan, O'Neill and Moss (2013) used the PMQ with a video-based procedure to assess PM (*Prospective Remembering Video Procedure*; PRVP, see below) and showed no difference between smokers and controls on the questionnaire, despite poorer performances on the PRVP were reported for smoker individuals, suggesting a lack of self-awareness of such PM deficits.

Prospective and Retrospective Memory Questionnaire (PRMQ). The PRMQ (Smith, Del Sala, Logie, & Maylor, 2000) is one of the most widely used questionnaire designed to provide a self- and informant rating of memory complaints for both prospective and retrospective failures (8 items for each) in everyday life context. To our knowledge, the PRMQ has been translated into 5 languages (Gondo et al., 2010; Hsu & Hua, 2011; Piauilino et al., 2010; Rönnlund, Mäntylä, & Nilsson, 2008; Wong Gonzalez, 2015). Each item of the questionnaire can be categorized along three dimensions: (1) assessing retrospective episodic memory by (2) self- or external cues (i.e., time- and event-based tasks) and (3) requiring long- or short-term delay. For example, the item 1 ("Do you decide to do something in a few minutes' time and then forget to do it?") is defined as measuring prospective, short-term and self-cued memory, while the item 2 ("Do you fail to recognise a place you have visited before?") is defined as retrospective, long-term and environmental-cued.

The validation and standardization of the PRMQ included 551 healthy individuals aged between 17 to 94 years (Crawford, Smith, Maylor, Della Sala, & Logie, 2003). The latent structure of the tool was studied using confirmatory factor analysis. The model was composed of a tripartite structure including a general memory factor (all items included) plus two orthogonal factors specific to prospective and retrospective memory with acceptable

Cronbach's alpha coefficients of .89, .84 and .80, respectively. However, the confirmatory factor analysis suggests that the classical distinction between self- and environmental cues does not explain the pattern of covariance among items. While this factorial structure was confirmed by 3 studies (Hsu & Hua, 2011; Piauilino et al., 2010; Rönnlund et al., 2008), this was not the case for the Spanish version of the PRMQ (González-Ramírez & Mendoza-González, 2011) compared with the original study (Smith et al., 2000).

PM failures were rated as more frequent than retrospective failures for both Alzheimer's disease and healthy older adults groups (Smith et al., 2000). Moreover, PM failures of Alzheimer's disease patients were rated as more frustrating for informants than retrospective memory failures. Using questionnaires completed by participants and their spouse, self- and informant ratings did not differ, suggesting a relative coherence of these measures. Several studies reported no significant difference in self-reported PM failures on the PRMQ for both smokers individuals (Heffernan et al., 2010a), young binge-drinkers (Heffernan, Clark, Bartholomew, Ling, & Stephens, 2010b; Heffernan & O'Neill, 2012) compared to healthy controls individuals; while their performance on the Cambridge Test of Prospective Memory as an objective measure of PM were poorer than controls. Thompson et al. (2015) reported a similar pattern of result by showing that self-reported PM failures on the PRMQ did not differ across patients with MCI, patients with dementia and controls. The authors highlighted that informants tended to rate higher PM failures for patients with dementia than those presenting MCI and control participants. Moreover, the reports of patients presenting MCI and healthy controls were not linked to informant reports. These results suggest that informant reports represent a more valid diagnostic indicator, notably for individuals with dementia, but not for patient with a lesser degree of impairment. Other studies led to similar conclusion for patients with amnestic mild cognitive impairment (Lee et al., 2016) and Alzheimer's disease (Hsu et al., 2014).

Comprehensive Assessment of Prospective Memory (CAPM). This questionnaire is specifically devoted to brain-injured individuals (Roche, Fleming, & Shum, 2002). The CAPM is divided into three sections to evaluate frequency of PM failures (Section A, 39 items), degree of concern (Section B, same 39 items) and reasons for each PM failure (Section C, 15 items). What distinguishes Section A of the CAPM from other questionnaires is the nature of its two subscales, which refer to the type of daily living activity that is remembered. The principal components analysis conducted by Waugh (1999) indicated that the Section A of the CAPM was defined by two components: (1) common memory failures referring to Instrumental Activities of Daily Living (IADL; Item 1 "Forgetting to buy an item at the grocery store") and (2) uncommon failures referring to basic activities of daily living (BADL; Item 6 "Not locking the door when leaving home").

The initial validation study using the CAPM was conducted among 525 healthy participants aged 17 to 91 years (M = 48.00; SD = 23.07) (Waugh, 1999). The internal consistency of these two sub-scales showed acceptable alpha coefficients of .92 and .79, respectively. In addition, the CAPM proved to be sensitive enough to discriminate age groups. Fleming et al. (2009) have shown that CAPM self-reports scores were not correlated with neither the Cambridge Test of Prospective Memory (CAMPROMPT; Wilson et al., 2005) nor the Memory for Intentions Screening Test (MIST; Raskin, 2004). However, the informants' reports on the IADL sub-scale and total scores of the CAPM were negatively correlated to the CAMPROMPT and the MIST. This result highlights the usefulness of the Section A of the CAPM to assess PM failures with brain-injured patients.

The reliability and normative data of the CAPM on 95 healthy individuals with an age range of 15 to 60 years showed more failures for younger adults (15–30 years) than the healthy older adults (31–60 years) (Chau, Lee, Fleming, Roche, & Shum, 2007). This result is congruent with the age-PM-paradox showing an age-related benefit in naturalistic PM tasks

while deficits were observed in laboratory-based PM tasks (Rendell & Thomson, 1999). The older age group being relatively young compared to other studies (e.g., Waugh, 1999) might explain this result. Both internal consistency and test-rest reliability coefficients of the CAPM were good and similar to those reported for the PMQ.

While no difference was found in self-rating condition between brain-injured patients and controls for the Section A of the CAPM, ratings from informants showed that brain-injured patients had more frequent PM failures compared to controls (i.e., patients tended to underestimate the frequency of PM failures compared to informants) (Roche et al., 2002). The authors suggested that impaired self-awareness could be a factor affecting the accuracy of self-ratings brain-injured patients when using the CAPM.

Brief Assessment of Prospective Memory (BAPM). As the Section A of the Comprehensive Assessment of Prospective Memory (CAPM; Waugh, 1999), the BAPM includes both IADL and BADL sub-scales (8 items for each) into a 16 items short form test (Man et al., 2011). The authors assessed the validity of the BAPM from 3 samples. The first sample was a group 527 healthy participants included Waugh' study (1999), while second and the third samples were 95 healthy participants and 45 brain-injured patients who participated in Fleming et al.' study (2009). The authors also reported acceptable internal consistency and test-retest reliability for both IADL and BADL sub-scales for all samples with coefficient ranging between .66 and .98. Like for the CAPM, the correlations between self-reports on the BAPM and the CAMPROMPT were not significant, suggesting a poor concurrent validity of the BAPM. Results also showed that BAPM scores correlated with Sydney Psychosocial Reintegration Scale (Tate, Hodgkinson, Veerabangsa, & Maggiotto, 1999), indicating a good predictive validity of the questionnaire.

***** insert Table 3 about here *****

Experimental Procedures

Prospective Remembering Video Procedure (PRVP). The PRVP is a video-based method in which participants watch a 12-minutes video recorded at a shopping precinct and have to recall future intentions (e.g., remembering to buy a soccer ball) in response to event-based PM cues appearing during the movie (Titov & Knight, 2001). Each item of the PRVP assess both prospective and retrospective components of PM.

Their results supported the inter-item reliability (Cronbach's alpha of .79 for the first list and .67 for the second list), as well as the alternate form of reliability (.65). The authors also found that familiarity, assessed with a 10-point Likert scale, enhanced recall and that pre-exposure to a video of unfamiliar stimuli could attenuate this effect. Moreover, evidence for the concurrent validity of the PRVP was found by showing relationship between participants' total scores and their performance on comparable PM tasks performed in natural settings (coefficient of .71). The PRVP was also sensitive enough to distinguish healthy control participants from young binge-drinkers (Heffernan et al., 2010b) and smokers (Heffernan et al., 2013).

Test Écologique de Mémoire Prospective (TEMP). Inspired by the PRVP (Titov & Knight, 2001), the TEMP (Potvin et al., 2011) is a 20-minute movie that displays several areas (i.e., commercial, residential and industrial) of a city. It includes 15 tasks (10 event-based and 5 time-based tasks) simulating real activities of daily living (e.g., reserving train tickets). The TEMP provides two versions for test-retest (no significant differences between the two versions) and assess both PM components (prospective and retrospective), the 3 main phases

(encoding, storage and retrieval) and both time- and event-based aspects of PM. The test-retest reliability of the TEMP was found to be high with a coefficient of .93.

Brain-injured patients showed poorer performance patients compared to healthy controls for the encoding phase and when retrieving intentions at the right context (i.e., prospective component), especially for time-based tasks (Potvin et al., 2011). Correlational analyses indicated that retrospective memory measures were linked to both prospective and retrospective PM components. Furthermore, the prospective component was specifically correlated with attentional processes and executive functions. Moreover, the authors found a correlation of -.51 between the TEMP total scores and the informant's reports on the CAPM. However, there was no significant correlation between TEMP total scores and participant' results on the CAPM (r = .06). Finally, the significant correlation between TEMP scores and those obtained on the envelope task (r =.47) provides good evidence of convergent validity of the TEMP, at least for event-based PM tasks.

Virtual Week. The Virtual Week (Rendell & Craik, 2000, Experiment 1) is a computerized PM task which simulates daily life activities on a virtual board game. As participants move around the board, they make decisions about daily activities and are asked to perform lifelike activities as PM tasks. The full version of the Virtual Week board game provides PM assessment on 1-week simulation (from Monday to Sunday) and takes approximately one hour to be completed. For each virtual day, participants perform 10 tasks, including 4 regular activities (e.g., remembering to take asthma medication at breakfast and dinner), 4 irregular activities (e.g., remembering to return a book to the library at 4 pm.) and 2 regular time-check activities (i.e., remembering to do a lung test at 2 minutes and 4 minutes on a chronometer placed on the screen). Half of the regular and irregular activities are time- and event-based PM tasks. Overall, regular tasks performances were better than both irregular tasks and time-

check tasks. The young participants (M = 21.30; age range = 19–24) performed better than young-old participants (M = 67.83; age range = 61–73) for the time-check and irregular tasks, and better than old-old participants (M = 78.84; age range = 75–84) for regular, irregular and time-check tasks. To date, the Virtual Week has been translated and adapted in 2 different languages (Italian version: Mioni, Stablum, Biernacki, & Rendell, 2015; Polish version: Niedźwieńska, Rendell, Barzykowski, & Leszczyńska, 2016).

The Virtual Week has also proved to be valid and consistently sensitive to impairment in various clinical groups including substance abuse (Leitz, Morgan, Bisby, Rendell, & Curran, 2009), schizophrenics (Henry, Rendell, Kliegel, & Altgassen, 2007), Parkinson's disease (Foster, Rose, McDaniel, & Rendell, 2013), mild cognitive impairment and dementia (Thompson et al., 2015), multiple sclerosis (Rendell, Jensen, & Henry, 2007) and brain damage (Mioni et al., 2013).

The reliability of the computerized version of the Virtual Week, the most common version, showed an acceptable Spearman-Brown split-half reliability for both young (.64) and older adults (.93) (Rose, Rendell, McDaniel, Aberle, & Kliegel, 2010), as well as in various clinical groups including schizophrenia (Henry et al., 2007), multiple sclerosis (Rendell et al., 2012), Parkinson's disease (Foster et al., 2013) and traumatic brain injury (Mioni, Rendell, Henry, Cantagallo, & Stablum, 2013) versus controls. Overall, split-half reliability coefficients ranged from .74 to .89 for these studies. Furthermore, the authors reported poorer PM performances on the Virtual Week for older adults compared to their younger counterparts and for individuals from clinical groups compared to healthy controls for all the previously mentioned studies. Test-retest reliability of the Virtual Week was also examined among healthy participants (Mioni, Rendell, Stablum, Gamberini, & Bisiacchi, 2014). In experiment 1, when using the same version A, the older adults showed lower performance compared to their younger counterparts and a high test-retest reliability coefficient was found

for older adults (r = .80), while the young adults had moderate test-retest coefficient (r = .61). In the second experiment, the authors created a parallel version (version B) in which they varied the content of the PM actions. The study only included an older adult sample assigned to one of the two experimental conditions (version A and B at retest or vice versa) and showed no differences in performance between the two versions with a moderate test-retest reliability coefficient (r = .68).

Actual Week. The Actual Week is a Virtual Week adaptation in naturalistic settings (Rendell & Craik, 2000, Experiment 2). Targets are the same for all participants but different from VW because they are adapted to the situations encountered by participants in everyday life. Participants are also requested to return one daily sheet per day to the experimenter without checking any sheet after completion. Participants are also asked to record via a micro-recorder full or partial fulfilment of each task. Older adults outperformed the young adults, congruently with the well-known and intriguing pattern of age-related (i.e. *age-PM-paradox*).

Recently, Au, Vandermorris, Rendell, Craik and Troyer (2017) adapted the Actual Week (Rendell & Craik, 2000) to assess healthy older adults (age range: 50–90 years) PM performance in naturalistic settings. The time-check tasks were removed because participants reported during the pre-test phase that these tasks were too difficult to recall and did not clearly reflect real-world PM demands. In addition to 4 regular and 4 irregular tasks, participants had to remember to really perform each day an irregular call-back task in which they were requested to send the experimenter a voicemail message. Participants were encouraged to use all the techniques commonly used to remember everyday tasks and to report the time of completion of each task in the appropriate daily log sheet. Their results partly replicated those of their previous study (Rendell & Craik, 2000, Experiment 2), in that event-based tasks were better recalled than time-based tasks. However, performance on

irregular tasks was better than on regular tasks. According to the authors, this result can be explained by the procedural differences mentioned above, in particular by the novelty effect of the irregular tasks.

Au et al.'s results (2017) also provided evidence for the reliability of this adaptation of the Actual Week by showing a high internal consistency for both time completion of tasks (α = .93) and accuracy (α = .95). The test-retest reliability coefficient of the Actual Week showed that performance was stable over the time (r = .76). First day performance was correlated to the remaining days with correlation coefficients ranging from .73 to 83, suggesting that the administration of a single day was sufficient to ensure the reliability of the measure. These authors found a pattern of significant correlations with measures of convergent (i.e., memory of strategy use and verbal episodic memory with coefficients ranging from .27 to .46) and not for divergent validity (i.e., the health-promoting lifestyle behaviors and both positive and negative emotions experienced in the last week with coefficients ranging from .01 to .18). They also observed that 82% of assigned voicemail messages were totally in concordance with the self-reported time completion of tasks and actual completion of the call-back task, which support the ecological validity of the Actual Week.

***** insert Table 4 about here *****

Outcomes of the Assessed Criteria

Among the 16 PM measures identified, 81.25% (N = 13) allowed to measure event-based tasks while 62.50% (N = 10) assessed time-based tasks. Results also indicated that 87.50% (N = 14) of the identified PM measures were linked to functional outcomes, 68.75% (N = 11) showed empirical evidences regarding the validity, 75.00% (N = 12) for reliability.

To a slightly lesser extent, 43.75% ($N = 7$) provided parallel versions and 31.25% ($N = 5$) were normed, translated, and showed evidence for diagnostic value. Finally, 18.75% of the measures ($N = 3$) allowed the use of external helps while only 12.50% ($N = 2$) were adapted to a respective culture and provided qualitative scoring system. Tables 5–8 present the percentages of PM measures that tested the criteria according to the type of assessment.

***** insert Table 5 about here *****

***** insert Table 6 about here *****

***** insert Table 7 about here *****

***** insert Table 8 about here *****

Meta-Analyses Results

Data from 3,136 different (nonoverlapping) participants (1,194 patients and 1,942 controls) were analyzed across the 22 studies included to compute a summary weighted mean effect (see Table 9). Averaged mean age and education were 53.06 years (range = 18.64–80.78) and 13.21 years (range = 11.62–14.49) respectively. Participants characteristics were very different and included outpatient psychiatric or substance use treatment, patients with subjective cognitive decline, patients diagnosed with mild cognitive impairment, Alzheimer's disease and other dementias.

***** insert Table 9 about here *****

Effect of Administering Test Batteries on PM Performance. Compared to healthy controls, patients had significant impairment in PM summary scores, with a mean effect size ranging from -2.93 to -0.35 across studies, SMD = -1.49; SE = 0.24; 95% CI [-1.96, -1.01], $p < .001$. However, as was evident in the forest plot, there was a large significant heterogeneity between studies, $Q(10) = 126.35$, $p < .001$; $I^2 = 92.09\%$, 95% CI [87.82, 94.86], suggesting the need for a more in-depth analysis of study subgroups.

We created a funnel plot and used the Egger's regression intercept in order to identify the possible presence of asymmetry due to publication bias. Visual inspection of the figure (see Supplementary Figure S1) shows asymmetry and the Egger's regression intercept suggests there is a publication bias ($p = .04$). Rosenberg's fail-safe *N* suggests that 1,465 additional studies with null results would be required to yield a non-significant effect of PM summary scores for the administration of PM test batteries.

Effect of Administering Single-Trial Procedures on PM Performance. The effect of single-trial procedures showed that patients had lower PM summary scores compared to controls, with a mean effect size of and ranging from -2.33 to -1.21 across studies, SMD = -2.16; SE = 0.52; 95% CI [-3.18, -1.14], $p < .001$. Once again, there was a very large heterogeneity between studies, $Q(2) = 31.24$, $p < .001$; $I^2 = 93.60\%$, 95% CI [84.67, 97.33], also indicating the need for planned subgroup analysis.

Visual inspection of the funnel plot revealed potential asymmetry (see Supplementary Figure S2) but the Egger's regression intercept was not statistically significant ($p = .33$), suggesting the absence of a publication bias. Rosenberg's fail-safe *N* suggests that 268

additional studies with null results would be required to yield a non-significant overall effect of PM summary scores for measures classified as single-trial procedures.

Effect of Administering Self-Reported Questionnaires on PM Performance. The effect of administering questionnaires showed that there were no significant differences for self-reported PM failures between patients and controls, with a mean effect size of and ranging from -2.26 to 3.68 across studies, SMD = 0.18; SE = 0.45; 95% CI [-0.72, 1.06], $p = .70$. Heterogeneity estimates were statistically significant and the effect was very large, $Q(11) = 591{,}62$, $p < .001$; $I^2 = 98.14\%$, 95% CI [97.58, 98.57], also indicating the need for planned subgroup analysis.

Visual inspection of the funnel plot revealed potential asymmetry (see Supplementary Figure S3) but the Egger's regression intercept was not statistically significant ($p = .38$), suggesting the absence of a publication bias. Rosenberg's fail-safe *N* suggests that 45 additional studies with null results would be required to yield a non-significant overall effect of PM summary scores for measures classified as questionnaires.

Effect of Administering Experimental Procedures on PM Performance. The effect of administering experimental procedures indicated that patients had lower PM summary scores than controls, with a mean effect size of and ranging from -1.44 to -0.41 across studies, SMD = -0.79; SE = 0.18; 95% CI [-1.14, -0.44], $p < .001$. The distribution of scores was homogeneous across the individual studies, with moderate nonsignificant variation, $Q(4) = 8.48$, $p = .08$; $I^2 = 52.84\%$, 95% CI [0.00, 82.65],suggesting that this result is a consistent finding.

A symmetrical funnel plot was observed (see Supplementary Figure S4) and the Egger's regression intercept showed that there was no publication bias ($p = .74$).

Rosenberg's fail-safe *N* suggests that 71 additional studies with null results would be required to yield a non-significant overall effect of PM summary scores for measures classified as experimental procedures.

Planned Subgroup Analyses. The data on the prompt card task used in Delprado et al.' study (2012) were excluded from the analyses because it was the only one we found that met the inclusion criteria in the current review. Therefore, planned subgroup analyses were only performed for the envelope task in this category of measurement. In accordance with Richardson, Garner, and Donegan (2019), the results of the planned subgroup analyses were considered statistically significant when the p-value was less than 0.1.

As displayed in Table 10, patients had lower PM summary scores in comparison to controls when the Rivermead Behavioural memory Test (RBMT), the Cambridge Test of Prospective Memory (CAMPROMPT), the Memory for Intentions Screening Test (MIST), the Royal Prince Alfred Prospective Memory (RPA-ProMem) test and the envelope task were administered. Although the distribution of scores was homogeneous for the RBMT, the MIST, the RPA-ProMem and the Comprehensive Assessment of Prospective Memory Questionnaire with small and moderate nonsignificant variations, this was not the case for the CAMPROMPT and the envelope task, as well as for the Prospective Memory Questionnaire and the Prospective and Retrospective Memory Questionnaire, for which the heterogeneity coefficients were higher.

***** insert Table 10 about here *****

Discussion

This paper is the first attempt to review systematically the literature regarding the existing measure to assess PM and quantitatively summarize the effects of various diseases according to the type of assessment. Fifty-two studies were included to examine the characteristics of the identified PM measures and 22 studies were retained to summarize the effect of diseases on PM. Among the 16 identified measures, we found 5 psychological tests (Rivermead Behavioural memory Test, Cambridge Behavioural Prospective Memory Test, Cambridge Test of Prospective Memory, Memory for Intentions Screening Test, Royal Prince Alfred Prospective Memory Test), 3 single-trial procedures (envelope task, prompt card task and telephone Test), 4 questionnaires (Prospective Memory Questionnaire, Prospective and Retrospective Memory Questionnaire, Comprehensive Assessment of Prospective Memory Questionnaire Brief Assessment of Prospective Memory Questionnaire) and 4 experimental procedures (Prospective Remembering Video Procedure, Test Écologique de Mémoire Prospective, Virtual Week, Actual Week). These results now reduce the ambiguity regarding the existing measures devoted to PM assessment in the literature. The findings of the current study also showed that the use of specific measures may be of interest to identify PM impairments among various clinical groups. In the following sections, we outlined opportunities and research gaps in this arena and made recommendations to integrate the assessment of PM into clinical practice of clinical neuropsychologists.

Before PM assessment can be expanded in day-to-day clinical practice of psychologists, it is important to know whether PM measures meet some of the frequent challenges associated with the selection of neuropsychological instruments (Rabin et al., 2016). We showed that more than 50% of the identified PM measures were associated with functional outcome measures (e.g., information provided by informants, self-reported dependence in instrumental activities of daily living), showed empirical evidences regarding

validity and reliability and measured both the event- and time-based PM tasks. However, it appears that some of the challenges encountered by psychologists in their clinical practice have received relatively little attention by memory researchers. This includes the lack of normative data, test translations/adaptations, available cutoff scores for diagnostic purposes, qualitative scoring, parallel versions for test-retest and specific instructions for use of external aids during the test. These results suggest that, like classical neuropsychological assessment instruments, PM measures suffer from these pitfalls, which may ultimately limit their utility in clinical settings (Rabin et al., 2005, 2016). Together, these findings encourage researchers to respond to these challenges to extend the clinical utility of the PM measures. We believe that such an endeavor will answer the frequent assessment referral questions raised by psychologists and contribute to the determination of diagnosis and rehabilitation.

The preliminary results from the meta-analyses indicated a trend toward lower PM performance for clinical groups compared to non-clinical groups when test batteries, single-trial procedures and experimental procedures are administered. However, PM performance were heterogenous across studies, except for experimental procedures for which the effect sizes were moderate and homogeneous. Results also showed null effect of group for the administration of PM questionnaires and effect sizes were heterogeneous across studies. Subsequent planned subgroup analyses indicated consistent differences for three test batteries (Rivermead Behavioural memory Test, Memory for Intentions Screening Test, Royal Prince Alfred Prospective Memory Test) with high effect sizes. This suggests that like experimental procedures, the use of the Rivermead Behavioural memory Test (RBMT; Wilson et al., 1985), the Memory for Intentions Screening Test (MIST; Raskin, 2004) and the Royal Prince Alfred Prospective Memory Test (RPA-ProMem; Radford et al., 2011) are relevant to identify variations in PM performance in many clinical groups, especially for patients with subjective cognitive decline, mild cognitive impairment, brain damage and schizophrenia. However, it

was not possible to estimate a group effect of test batteries administration on the Cambridge Test of Prospective Memory (CAMPROMPT; Wilson et al., 2005) measure of PM due to the high level of heterogeneity (85.38%) across studies.

The planned subgroup analyses for questionnaires showed no significant difference in self-reported PM failures between groups. Effect sizes remained heterogeneous for studies using the Prospective Memory Questionnaire (Hannon et al., 1990), and the Prospective and Retrospective Memory Questionnaire (Smith et al., 2000), except for the Comprehensive Assessment of Prospective Memory Questionnaire (Waugh, 1999). (It should be noted that we excluded the Brief Assessment of Prospective Memory Questionnaire from the analyses due to insufficient data for estimating effect sizes.) Compared to objective measures of PM, this result corroborates previous empirical findings revealing that PM questionnaires were not able to differentiate healthy participants from clinical groups, whereas these clinical groups showed poorer PM performance on PM objective measures. These results are congruent with Uttl and Kibreab's meta-analysis (2011) highlighting a lack of validity for PM self-report measures. Taken together, PM questionnaires might not be used for diagnosis but have an interest to understand everyday difficulties and explore patient/informant discrepancies.

Beside the theoretical relevance of PM measures, their integration into a classical neuropsychological assessment in neuropsychologists' day-to-day clinical practice is, to date, limited. As an example, the average duration of a French standard neuropsychological assessment in most memory centers is estimated between 90 and 120 minutes. The specific PM measures like the CAMPROMPT and the MIST take about 30 to 40 minutes to be administered. Therefore, the use of such tests may be complicated. This is partly due to the fact that these have their own set of distractor tasks. In such a situation, the use of a more flexible measure like the RPA-ProMem, which takes only 15 minutes to be administered, could be an alternative solution to overcome this limitation and simplify its administration in

day-to-day clinical practice. However, further studies are needed to establish normative data before expanding the use of RPA-ProMem in clinical practice. Although the literature review showed that the use of single-trial procedures is interesting in assessing patients' ability to remember to carry out intended actions in a shorter period of time, other studies with various clinical groups are required to ascertain their contribution to diagnosis, especially for patients who are at risk of developing dementia. Indeed, it appears that some traditional retrospective memory measures appear to be more effective in identifying patients with mild cognitive impairment PM tests batteries like the CAMPROMPT or even single-trial procedure like the envelope test (Delprado et al., 2012). This seems to corroborate the results of a previous study that indicated that the reduced reliability of PM tasks was associated with a small number of trials, which remains an issue for most of the measures identified in this study (see Kelemen, Weinberg, Alford, Mulvey, & Kaeochinda, 2006). In this context, the use of experimental procedures such as the Virtual Week may be relevant because it makes it possible to propose a larger number of PM trials and it has also proved to be effective in detecting PM impairment in a wide range of clinical groups. Face to these exciting lines of work, future research should also focus on the development of revised, shortened versions of PM to extend their clinical applicability in different languages and cultures. Such projects are often organized at a cross-country level and should agree with standard international guidelines for test development such as those provided by the International Test Commission (2010).

The current study has several limitations including a relatively small number of studies that have been selected for inclusion in the meta-analytical review (n = 22), limited range of test administered and variability in study populations with a lack of potential relevant demographic data in some studies. Indeed, the studies included a wide range of populations including normal aging, Alzheimer's disease, mild cognitive impairment, multiple sclerosis, brain injury, substance abuse, schizophrenia, HIV and spina bifida. Moreover, the features of

the identified PM were relatively different from each other (e.g., number of items, retention intervals, administration time), even for the same type of assessment so it is difficult to suggest the use of a unique measure of PM that can be an appropriate candidate for all assessment situations, particularly to distinguish between clinical and non-clinical groups. Strengths of the review included the use of PRISMA guidelines to identify the existing PM measures, which involved the establishment of criteria for inclusion and exclusion of studies, analysis of publication bias, and the use of planned subgroup analyses.

This review of the available PM measures should provide a useful and valuable information to guide therapists who work with patients with various neuropathologies towards the choice of the appropriate PM assessment, taking all due account of their clinical requirement. Our work should also guide future research to ultimately extend the clinical applicability of PM assessment instruments and the understanding of PM functioning.

Funding

This work was supported by the Regional Council of Hauts-de-France and the European Reginal Development Fund (FEDER) [REG16026]. The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication. GB drafted the protocol with MH, VQ and YG, and GB assessed the eligibility of the studies for inclusion and extracted data. All authors contributed to the interpretation of the findings. GB drafted the manuscript, to which all authors contributed.

Conflict of interests

All authors declare that they have no conflicts of interest.

Acknowledgments

The authors thank Barbara Wilson and Peter Watson (MRC Cognition and Brain Sciences Unit, University of Cambridge, United Kingdom) for their valuable and useful information about the Rivermead Behavior Memory Test. The authors would also like to thank Béatrice Desgranges (INSERM-EPHE Neuropsychologie et Imagerie de la Mémoire Humaine, Université de Caen Normandie, France) for her precious comments on a previous version of this manuscript.

References

Au, A., Vandermorris, S., Rendell, P. G., Craik, F. I. M., & Troyer, A. K. (2018). Psychometric properties of the Actual Week test: a naturalistic prospective memory task. *The Clinical Neuropsychologist*, *32*(6), 1068–1083. https://doi.org/10.1080/13854046.2017.1360946

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*, *1*(2), 97–111. https://doi.org/10.1002/jrsm.12

Carey, C. L., Paul Woods, S., Rippeth, J. D., Heaton, R. K., Grant, I., & the HIV Neurobehavioral Research Ce. (2006). Prospective Memory in HIV-1 Infection. *Journal of Clinical and Experimental Neuropsychology*, *28*(4), 536–548. https://doi.org/10.1080/13803390590949494

Chau, L. T., Lee, J. B., Fleming, J., Roche, N., & Shum, D. (2007). Reliability and normative data for the comprehensive assessment of prospective memory (CAPM). *Neuropsychological Rehabilitation*, *17*(6), 707–722.

https://doi.org/10.1080/09602010600923926

Cheung, M. W.-L., & Vijayakumar, R. (2016). A Guide to Conducting a Meta-Analysis. *Neuropsychology Review*, *26*(2), 121–128. https://doi.org/10.1007/s11065-016-9319-z

Cockburn, J., & Smith, P. T. (1989). *The Rivermead Behavioural Memory Test supplement three: Elderly people* (Thames Val). Bury St. Edmunds, U.K.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. NY: Routledge Academic.

Crawford, J., Smith, G., Maylor, E., Della Sala, S., & Logie, R. (2003). The Prospective and Retrospective Memory Questionnaire (PRMQ): Normative data and latent structure in a large non-clinical sample. *Memory*, *11*(3), 261–275. https://doi.org/10.1080/09658210244000027

Crovitz, H. F., & Daniel, W. F. (1984). Measurements of everyday memory: Toward the prevention of forgetting. *Bulletin of the Psychonomic Society*, *22*(5), 413–414. https://doi.org/10.3758/BF03333861

Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (2000). CVLT-II: California verbal learning test: adult version.

Delprado, J., Kinsella, G., Ong, B., Pike, K., Ames, D., Storey, E., … Rand, E. (2012). Clinical Measures of Prospective Memory in Amnestic Mild Cognitive Impairment. *Journal of the International Neuropsychological Society*, *18*(02), 295–304. https://doi.org/10.1017/S135561771100172X

Dennis, M., Nelson, R., Jewell, D., & Fletcher, J. M. (2010). Prospective memory in adults with spina bifida. *Child's Nervous System*, *26*(12), 1749–1755. https://doi.org/10.1007/s00381-010-1140-z

Einstein, G. O., & McDaniel, M. A. (1990). Normal aging and prospective memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(4), 717–726. https://doi.org/10.1037/0278-7393.16.4.717

Einstein, G. O., & McDaniel, M. A. (1996). Retrieval processes in prospective memory: Theoretical approaches and some new empirical findings. In M. Brandimonte, G. O. Einstein, & M. A. McDaniel (Eds.), *Prospective memory: Theory and applications* (pp. 115-141). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Ellis, J. (1996). Prospective memory or the realization of delayed intentions: A conceptual framework for research. In M. A. Brandimonte, G. O. Einstein, & M. A. Mcdaniel (Eds.), *Prospective Memory: Theory and Applications* (pp. 1–22). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Fleming, J., Kennedy, S., Fisher, R., Gill, H., Gullo, M., & Shum, D. (2009). Validity of the Comprehensive Assessment of Prospective Memory (CAPM) for Use With Adults With Traumatic Brain Injury. *Brain Impairment*, *10*(1), 34–44. https://doi.org/10.1375/brim.10.1.34

Foster, E. R., Rose, N. S., McDaniel, M. A., & Rendell, P. G. (2013). Prospective memory in Parkinson disease during a virtual week: Effects of both prospective and retrospective demands. *Neuropsychology*, *27*(2), 170–181. https://doi.org/10.1037/a0031946

Gates, N. J., & March, E. G. (2016). A Neuropsychologist's Guide To Undertaking a Systematic Review for Publication: Making the most of PRISMA Guidelines. *Neuropsychology Review*, *26*(2), 109–120. https://doi.org/10.1007/s11065-016-9318-0

Gondo, Y., Renge, N., Ishioka, Y., Kurokawa, I., Ueno, D., & Rendell, P. (2010). Reliability and validity of the Prospective and Retrospective Memory Questionnaire (PRMQ) in young and old people: A Japanese study. *Japanese Psychological Research*, *52*(3), 175–

185. https://doi.org/10.1111/j.1468-5884.2010.00433.x

González-Ramírez, M. T., & Mendoza-González, M. E. (2011). Spanish Version of the Prospective and Retrospective Memory Questionnaire (PRMQ-S). *The Spanish Journal of Psychology*, *14*(1), 385–391. https://doi.org/10.5209/rev_SJOP.2011.v14.n1.35

Groot, Y. C., Wilson, B. A., Evans, J., & Watson, P. (2002). Prospective memory functioning in people with and without brain injury. *Journal of the International Neuropsychological Society*, *8*(5), 645–654. https://doi.org/10.1017/S1355617702801321

Hainselin, M., Quinette, P., Desgranges, B., Martinaud, O., Hannequin, D., De La Sayette, V., … Eustache, F. (2011). Can we remember future actions yet forget the last two minutes? Study in transient global amnesia. *Journal of Cognitive Neuroscience*, *23*(12), 4138–4149. https://doi.org/10.1162/jocn_a_00076

Hannon, R., Adams, P., Harrington, S., Fries-Dias, C., & Gipson, M. T. (1995). Effects of brain injury and age on prospective memory self-rating and performance. *Rehabilitation Psychology*, *40*(4), 289–298. https://doi.org/10.1037/0090-5550.40.4.289

Hannon, R., Gipson, M. T., Rebmann, M., Keneipp, J., Sattler, J., Lonero, P., … Bolter, J. F. (1990). Self-rating of prospective memory by normal, brain-injured and alcoholic individuals. In *Paper presented at the meeting of the National Academy of Neuropsychology*. Reno, NV.

Heffernan, T., Clark, R., Bartholomew, J., Ling, J., & Stephens, S. (2010b). Does binge drinking in teenagers affect their everyday prospective memory? *Drug and Alcohol Dependence*, *109*(1–3), 73–78. https://doi.org/10.1016/j.drugalcdep.2009.12.013

Heffernan, T. M., O'Neill, T. S., & Moss, M. (2013). Smoking-related prospective memory deficits observed on naturalistic everyday memory task. *Irish Journal of Psychological Medicine*, *30*(1), 21–27.

https://doi.org/http://dx.doi.org.ez.statsbiblioteket.dk:2048/10.1017/ipm.2012.4

Heffernan, Thomas, & O'Neill, T. (2012). Time based prospective memory deficits associated with binge drinking: evidence from the Cambridge Prospective Memory Test (CAMPROMPT). *Drug and Alcohol Dependence*, *123*(1–3), 207–212. https://doi.org/10.1016/j.drugalcdep.2011.11.014

Heffernan, Thomas, O'Neill, T., & Moss, M. (2010a). Smoking and everyday prospective memory: A comparison of self-report and objective methodologies. *Drug and Alcohol Dependence*, *112*(3), 234–238. https://doi.org/10.1016/j.drugalcdep.2010.06.012

Henry, J. D., MacLeod, M. S., Phillips, L. H., & Crawford, J. R. (2004). A Meta-Analytic Review of Prospective Memory and Aging. *Psychology and Aging*, *19*(1), 27–39. https://doi.org/10.1037/0882-7974.19.1.27

Henry, J. D., Rendell, P. G., Kliegel, M., & Altgassen, M. (2007). Prospective memory in schizophrenia: Primary or secondary impairment? *Schizophrenia Research*, *95*(1–3), 179–185. https://doi.org/10.1016/j.schres.2007.06.003

Higgins, J., & Green, S. (2011). Cochrane Handbook for Systematic Reviews of Interventions | The Cochrane Collaboration.

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, *21*(11), 1539–1558. https://doi.org/10.1002/sim.1186

Higgins, J. P. T. (2003). Measuring inconsistency in meta-analyses. *BMJ*, *327*(7414), 557–560. https://doi.org/10.1136/bmj.327.7414.557

Hsu, Y.-H., & Hua, M.-S. (2011). Taiwan Version of the Prospective and Retrospective Memory Questionnaire: Latent Structure and Normative Data. *Archives of Clinical*

Neuropsychology, *26*(3), 240–249. https://doi.org/10.1093/arclin/acr012

Hsu, Yen-Hsuan, Huang, C.-F., Tu, M.-C., & Hua, M.-S. (2014). The Clinical Utility of Informants' Appraisals on Prospective and Retrospective Memory in Patients with Early Alzheimer's Disease. *PLoS ONE*, *9*(11), 1–8. https://doi.org/10.1371/journal.pone.0112210

Hsu, Yen-Hsuan, Huang, C.-F., Tu, M.-C., & Hua, M.-S. (2015). Prospective Memory in Subjective Cognitive Decline: A Preliminary Study on the Role of Early Cognitive Marker in Dementia. *Alzheimer Disease & Associated Disorders*, *29*(3), 229–235. https://doi.org/10.1097/WAD.0000000000000060

Huppert, F. A., Johnson, T., & Nickson, J. (2000). High prevalence of prospective memory impairment in the elderly and in early-stage dementia: Findings from a population-based study. *Applied Cognitive Psychology*, *14*(7), S63–S81. https://doi.org/10.1002/acp.771

Ihle, A., Hering, A., Mahy, C. E. V, Bisiacchi, P. S., & Kliegel, M. (2013). Adult age differences, response management, and cue focality in event-based prospective memory: a meta-analysis on the role of task order specificity. *Psychology and Aging*, *28*(3), 714–720. https://doi.org/10.1037/a0033653

International Test Commission. (2010). Guidelines for translating and adapting tests. Retrieved from https://www.intestcom.org/files/guideline_test_use.pdf

Kelemen, W. L., Weinberg, W. B., Alford, H. S., Mulvey, E. K., & Kaeochinda, K. F. (2006). Improving the reliability of event-based laboratory tests of prospective memory. *Psychonomic Bulletin & Review*, *13*(6), 1028–1032. https://doi.org/10.3758/BF03213920

Kime, S. K., Lamb, D. G., & Wilson, B. A. (1996). Use of a comprehensive programme of external cueing to enhance procedural memory in a patient with dense amnesia. *Brain*

Injury, *10*(1), 17–26. https://doi.org/10.1080/026990596124683

Kinsella, G. J., Pike, K. E., Cavuoto, M. G., & Lee, S. D. (2018). Mild cognitive impairment and prospective memory: translating the evidence into neuropsychological practice. *Clinical Neuropsychologist*, *32*(5), 960–980. https://doi.org/10.1080/13854046.2018.1468926

Landsiedel, J., Williams, D. M., & Abbot-Smith, K. (2017). A Meta-Analysis and Critical Review of Prospective Memory in Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, *47*(3), 646–666. https://doi.org/10.1007/s10803-016-2987-y

Lawton, M. P., & Brody, E. M. (1969). Assessment of Older People: Self-Maintaining and Instrumental Activities of Daily Living. *The Gerontologist*, *9*(3), 179–186. https://doi.org/10.1093/geront/9.3_Part_1.179

Lee, S., Ong, B., Pike, K. E., Mullaly, E., Rand, E., Storey, E., … Kinsella, G. J. (2016). The Contribution of Prospective Memory Performance to the Neuropsychological Assessment of Mild Cognitive Impairment. *The Clinical Neuropsychologist*, *30*(1), 131–149. https://doi.org/10.1080/13854046.2015.1135983

Leitz, J. R., Morgan, C. J. A., Bisby, J. A., Rendell, P. G., & Curran, H. V. (2009). Global impairment of prospective memory following acute alcohol. *Psychopharmacology*, *205*(3), 379–387. https://doi.org/10.1007/s00213-009-1546-z

Man, D. W. K., Fleming, J., Hohaus, L., & Shum, D. (2011). Development of the Brief Assessment of Prospective Memory (BAPM) for use with traumatic brain injury populations. *Neuropsychological Rehabilitation*, *21*(6), 884–898. https://doi.org/10.1080/09602011.2011.627270

Martin, M., Kliegel, M., & McDaniel, M. A. (2003). The involvement of executive functions

in prospective memory performance of adults. *International Journal of Psychology*, *38*(4), 195–206. https://doi.org/10.1080/00207590344000123

Mathias, J. L., & Mansfield, K. M. (2005). Prospective and declarative memory problems following moderate and severe traumatic brain injury. *Brain Injury*, *19*(4), 271–282. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15832873

McDaniel, M. A., & Einstein, G. O. (2000). Strategic and automatic processes in prospective memory retrieval: a multiprocess framework. *Applied Cognitive Psychology*, *14*(7), 127–144. https://doi.org/10.1002/acp.775

McDaniel, M. A., & Einstein, G. O. (2007). Prospective Memory: An Overview and Synthesis of an Emerging Field. Thousand Oaks, CA: SAGE Publications. https://doi.org/http://dx.doi.org/10.4135/9781452225913

McDaniel, M. A., Umanath, S., Einstein, G. O., & Waldum, E. R. (2015). Dual pathways to prospective remembering. *Frontiers in Human Neuroscience*, *9*(July), 1–12. https://doi.org/10.3389/fnhum.2015.00392

Mills, V., Kixmiller, J. S., Gillespie, A., Allard, J., Flynn, E., Bowman, A., & Brawn, C. M. (1997). The correspondence between the Rivermead Behavioral Memory Test and ecological prospective memory. *Brain and Cognition*, *35*(3), 322–325.

Mioni, G, Rendell, P. G., Henry, J. . D., Cantagallo, A., & Stablum, F. (2013). An investigation of prospective memory functions in people with traumatic brain injury using Virtual Week. *Journal of Clinical and Experimental Neuropsychology*, *35*(6), 617–630. https://doi.org/10.1080/13803395.2013.804036

Mioni, Giovanna, Rendell, P. G., Stablum, F., Gamberini, L., & Bisiacchi, P. S. (2014). Test–retest consistency of Virtual Week: A task to investigate prospective memory. *Neuropsychological Rehabilitation*, *25*(3), 1–21.

https://doi.org/10.1080/09602011.2014.941295

Mioni, Giovanna, Stablum, F., Biernacki, K., & Rendell, P. G. (2015). Virtual Week: Translation and adaptation for the Italian population. *Neuropsychological Rehabilitation*, 1–21. https://doi.org/10.1080/09602011.2015.1103758

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., … Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Journal of Clinical Epidemiology*, *62*(10), e1–e34. https://doi.org/10.1016/j.jclinepi.2009.06.006

Moscovitch, M. (1994). Memory and Working with Memory: Evaluation of a Component Process Model and Comparisons with Other Models. In S. Daniel & E. Tulving (Eds.), *Memory Systems* (pp. 369–394). Cambridge, MA : Massachusetts Institute of Technology.

Niedźwieńska, A., Rendell, P. G., Barzykowski, K., & Leszczyńska, A. (2016). Virtual Week: Validity and psychometric properties of a Polish adaptation. *Revue Européenne de Psychologie Appliquée/European Review of Applied Psychology*, *66*(2), 79–84. https://doi.org/10.1016/j.erap.2016.02.003

Piauilino, D. C., Bueno, O. F. A., Tufik, S., Bittencourt, L. R., Santos-Silva, R., Hachul, H., … Pompéia, S. (2010). The Prospective and Retrospective Memory Questionnaire: A population-based random sampling study. *Memory*, *18*(4), 413–426. https://doi.org/10.1080/09658211003742672

Potvin, M.-J., Rouleau, I., Audy, J., Charbonneau, S., & Giguère, J.-F. (2011). Ecological prospective memory assessment in patients with traumatic brain injury. *Brain Injury*,

25(2), 192–205. https://doi.org/10.3109/02699052.2010.541896

Rabin, L. A., Barr, W. B., & Burton, L. A. (2005). Assessment practices of clinical neuropsychologists in the United States and Canada: A survey of INS, NAN, and APA Division 40 members. *Archives of Clinical Neuropsychology*, *20*(1), 33–65. https://doi.org/10.1016/j.acn.2004.02.005

Rabin, L. A., Chi, S. Y., Wang, C., Fogel, J., Kann, S. J., & Aronov, A. (2014). Prospective memory on a novel clinical task in older adults with mild cognitive impairment and subjective cognitive decline. *Neuropsychological Rehabilitation*, *24*(6), 868–893. https://doi.org/10.1080/09602011.2014.915855

Rabin, L. A., Paolillo, E., & Barr, W. B. (2016). Stability in Test-Usage Practices of Clinical Neuropsychologists in the United States and Canada over a 10-Year Period: A Follow-Up Survey of INS and NAN Members. *Archives of Clinical Neuropsychology*, *31*(3), 206–230. https://doi.org/10.1093/arclin/acw007

Radford, K. A., Lah, S., Say, M. J., & Miller, L. A. (2011). Validation of a new measure of prospective memory: The Royal Prince Alfred Prospective Memory Test. *The Clinical Neuropsychologist*, *25*(1), 127–140. https://doi.org/10.1080/13854046.2010.529463

Ramanan, S., & Kumar, D. (2013). Prospective Memory in Parkinson's Disease: A Meta-analysis. *Journal of the International Neuropsychological Society*, *19*(10), 1109–1118. https://doi.org/10.1017/S1355617713001045

Raskin, S. A. (2004). Memory for intentions screening test [abstract]. *Journal of the International Neuropsychological Society*, *10*.

Rendell, P. G., & Craik, F. I. M. (2000). Virtual Week and Actual Week : Age-related Differences in Prospective Memory. *Applied Cognitive Psychology*, *14*(7), 43–62. https://doi.org/10.1002/acp.770

Rendell, P. G., Henry, J. D., Phillips, L. H., de la Piedad Garcia, X., Booth, P., Phillips, P., & Kliegel, M. (2012). Prospective memory, emotional valence, and multiple sclerosis. *Journal of Clinical and Experimental Neuropsychology*, *34*(7), 738–749. https://doi.org/10.1080/13803395.2012.670388

Rendell, P. G., Jensen, F., & Henry, J. D. (2007). Prospective memory in multiple sclerosis. *Journal of the International Neuropsychological Society*, *13*(03), 410–416. https://doi.org/10.1017/S1355617707070579

Rendell, P. G., & Thomson, D. M. (1999). Aging and Prospective Memory: Differences Between Naturalistic and Laboratory Tasks. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *54B*(4), 256–269. https://doi.org/10.1093/geronb/54B.4.P256

Richardson, M., Garner, P., & Donegan, S. (2019). Interpretation of subgroup analyses in systematic reviews: A tutorial. *Clinical Epidemiology and Global Health*, *7*(2), 192–198. https://doi.org/10.1016/j.cegh.2018.05.005

Roche, N. L., Fleming, J. M., & Shum, D. H. K. (2002). Self-awareness of prospective memory failure in adults with traumatic brain injury. *Brain Injury*, *16*(11), 931–945. https://doi.org/10.1080/02699050210138581

Rönnlund, M., Mäntylä, T., & Nilsson, L. G. (2008). The Prospective and Retrospective Memory Questionnaire (PRMQ): Factorial structure, relations to global subjective memory ratings, and Swedish norms: *Scandinavian Journal of Psychology*, *49*(1), 11–18. https://doi.org/10.1111/j.1467-9450.2007.00600.x

Rose, N. S., Rendell, P. G., McDaniel, M. A., Aberle, I., & Kliegel, M. (2010). Age and individual differences in prospective memory during a “Virtual Week”: The roles of

working memory, vigilance, task regularity, and cue focality. *Psychology and Aging*, *25*(3), 595–605. https://doi.org/10.1037/a0019771

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638–641. https://doi.org/10.1037/0033-2909.86.3.638

Senn, S. J. (2009). Overstating the evidence – double counting in meta-analysis and related problems. *BMC Medical Research Methodology*, *9*(1), 1–7. https://doi.org/10.1186/1471-2288-9-10

Smith, G., Del Sala, S., Logie, R. H., & Maylor, E. A. (2000). Prospective and retrospective memory in normal ageing and dementia: A questionnaire study. *Memory*, *8*(5), 311–321. https://doi.org/10.1080/09658210050117735

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

Smith, R. E., & Bayen, U. J. (2004). A Multinomial Model of Event-Based Prospective Memory. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, *30*(4), 756–777. https://doi.org/10.1037/0278-7393.30.4.756

Tate, R., Hodgkinson, A., Veerabangsa, A., & Maggiotto, S. (1999). Measuring Psychosocial Recovery after Traumatic Brain Injury: Psychometric Properties of a New Scale. *Journal of Head Trauma Rehabilitation*, *14*(6), 543–557. https://doi.org/10.1097/00001199-199912000-00003

Terry, W. S. (1988). Everyday Forgetting - Data from a Diary Study. *Psychological Reports*,

62(1), 299–303. https://doi.org/10.2466/pr0.1988.62.1.299

Terwee, C. B., Jansma, E. P., Riphagen, I. I., & De Vet, H. C. W. (2009). Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Quality of Life Research*, *18*(8), 1115–1123. https://doi.org/10.1007/s11136-009-9528-5

Thompson, C. L., Henry, J. D., Rendell, P. G., Withall, A., & Brodaty, H. (2015). How Valid Are Subjective Ratings of Prospective Memory in Mild Cognitive Impairment and Early Dementia? *Gerontology*, *387380*(3), 251–257. https://doi.org/10.1159/000371347

Titov, N., & Knight, R. G. (2001). A video-based procedure for the assessment of prospective memory. *Applied Cognitive Psychology*, *15*(1), 61–83. https://doi.org/10.1002/1099-0720(200101/02)15:1<61::AID-ACP689>3.0.CO;2-Y

Troyer, A. K., & Murphy, K. J. (2007). Memory for intentions in amnestic mild cognitive impairment: Time- and event-based prospective memory. *Journal of the International Neuropsychological Society*, *13*(02), 365–369. https://doi.org/10.1017/S1355617707070452

Umeda, S., Nagumo, Y., & Kato, M. (2006). Dissociative Contributions of Medial Temporal and Frontal Regions to Prospective Remembering. *Reviews in the Neurosciences*, *17*(1–2), 267–278. https://doi.org/10.1515/REVNEURO.2006.17.1-2.267

Uttl, B., & Kibreab, M. (2011). Self-report measures of prospective memory are reliable but not valid. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *65*(1), 57–68. https://doi.org/10.1037/a0022843

Van Den Berg, E., Kant, N., & Postma, A. (2012). Remember to Buy Milk on the Way Home! A Meta-analytic Review of Prospective Memory in Mild Cognitive Impairment

and Dementia. *Journal of the International Neuropsychological Society*, *18*(04), 706–716. https://doi.org/10.1017/S1355617712000331

Wallace, B. C., Small, K., Brodley, C. E., Lau, J., & Trikalinos, T. A. (2012). Deploying an interactive machine learning system in an evidence-based practice center. In *Proceedings of the 2nd ACM SIGHIT symposium on International health informatics - IHI '12* (p. 819). New York, New York, USA: ACM Press. https://doi.org/10.1145/2110363.2110464

Walter, S. D., & Yao, X. (2007). Effect sizes can be calculated for studies reporting ranges for outcome variables in systematic reviews. *Journal of Clinical Epidemiology*, *60*(8), 849–852. https://doi.org/10.1016/j.jclinepi.2006.11.003

Wang, Y., Cui, J., Chan, R. C. K., Deng, Y., Shi, H., Hong, X., … 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

Waugh, N. (1999). *Self-report of the young, middle-aged, young-old and old-old individuals on the prospective memory functioning*. Unpublished honours thesis, Griffith University, Brisbane, Queensland, Australia.

Wechsler, D. (1987). *Wechsler Memory Scale-Revised manual*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1997). *Wechsler Memory Scale-Third Edition: Administration and scoring manual*. San Antonio, TX: The Psychological Corporation.

Wilson, B. A. (1991). Long-term prognosis of patients with severe memory disorders. *Neuropsychological Rehabilitation*, *1*(2), 117–134. https://doi.org/http://dx.doi.org/10.1080/09602019108401386

Wilson, B. A. (2009). Memory rehabilitation: Integrating theory and practice. New York, NY,

US: Guilford Press.

Wilson, B. A., Cockburn, J., & Baddeley, A. D. (1985). *The Rivermead Behavioural Memory Test*. Bury St. Edmunds, U.K: Thames Valley Test Company.

Wilson, B. A., Cockburn, J., Baddeley, A., & Hiorns, R. (1989). The development and validation of a test battery for detecting and monitoring everyday memory problems. *Journal of Clinical and Experimental Neuropsychology*, *11*(6), 855–870. https://doi.org/10.1080/01688638908400940

Wilson, B. A., Emslie, H., Foley, J., Shiel, A., Watson, P., Hawkins, K., & Groot, Y. C. (2005). *Cambridge Test of Prospective Memory (CAMPROMPT)*. San Antonio, TX: Harcourt Assessment.

Wilson, B. A., Forester, S., Bryant, T., & Cockburn, J. (1990). Performance of 11–14 year olds on the Rivermead Behavioural Memory Test. *Clinical Psychology Forum*, *30*, 8–10.

Wilson, B. A., Greenfield, E., Clare, L., Baddeley, A. D., Cockburn, J., Watson, P., … Crawford, J. (2008). *Rivermead Behavioural Memory Test - Third Edition (RBMT-3)*. San Antonio, TX: Pearson/PsychCorp.

Wilson, B. A., Ivani-chalian, R., Besag, F. M. C., & Bryant, T. (1993). Adapting the rivermead behavioural memory test for use with children aged 5 to 10 years. *Journal of Clinical and Experimental Neuropsychology*, *15*(4), 474–486. https://doi.org/10.1080/01688639308402572

Wong Gonzalez, D. (2015). *Prospective Memory Following Traumatic Brain Injury : A Meta-Analysis*. University of Windsor.

Woods, S. P., Iudicello, J. E., Moran, L. M., Carey, C. L., Dawson, M. S., & Grant, I. (2008a). HIV-associated prospective memory impairment increases risk of dependence in everyday functioning. *Neuropsychology*, *22*(1), 110–117.

https://doi.org/10.1037/0894-4105.22.1.110

Woods, S. P., Moran, L. M., Dawson, M. S., Carey, C. L., Grant, I., & HIV Neurobehavioral Research Center (HNRC) Group. (2008b). Psychometric Characteristics of The Memory for Intentions Screening Test. *Clinical Neuropsychology*, *22*(5), 864–878. https://doi.org/10.1080/13854040701595999.Psychometric

Woods, S. P., Twamley, E. W., Dawson, M. S., Narvaez, J. M., & Jeste, D. V. (2007). Deficits in cue detection and intention retrieval underlie prospective memory impairment in schizophrenia. *Schizophrenia Research*, *90*(1–3), 344–350. https://doi.org/10.1016/j.schres.2006.11.005

Yang, J., Zhong, F., Qiu, J., Cheng, H., & Wang, K. (2015). Dissociation of event-based prospective memory and time-based prospective memory in patients with prostate cancer receiving androgen-deprivation therapy: A neuropsychological study. *European Journal of Cancer Care*, *24*(2), 198–204. https://doi.org/10.1111/ecc.12299

Zhou, F.-C., Wang, Y.-Y., Zheng, W., Zhang, Q., Ungvari, G. S., Ng, C. H., … Zaza, S. (2017). Prospective memory deficits in patients with depression: a meta-analysis. *Journal of Affective Disorders*, *5*(0), 183–190. https://doi.org/10.1016/j.jad.2017.05.042

Zogg, J. B., Woods, S. P., Sauceda, J. A., Wiebe, J. S., & Simoni, J. M. (2012). The role of prospective memory in medication adherence: A review of an emerging literature. *Journal of Behavioral Medicine*, *35*(1), 47–62. https://doi.org/10.1007/s10865-011-9341-9

Tables

Table 1. Overview of the criteria met for each of the identified PM batteries and their main characteristics

Measures	T	CA	V	R	N	EB PM	TB PM	DV	EA	PV	QS	LF O	N of criteria met/N of criteria (max. 12)	N of PM items	Retention interval	Duration (mn)
Rivermead Behavioural memory Test	✓		✓	✓	✓	✓				✓	✓	✓	8	3	20 mn	30
Cambridge Behavioural Prospective Memory Test						✓	✓		✓				3	8	3, 15 and 20 mn	40
Cambridge Test of Prospective Memory	✓		✓	✓	✓	✓	✓	✓	✓	✓		✓	10	6	7, 13 ,20 mn and 24 hours	25—30
Memory for Intentions Screening Test			✓	✓	✓	✓	✓	✓		✓	✓	✓	9	8	2, 15 mn and 24 hours	30—40
Royal Prince Alfred Prospective Memory Test			✓	✓		✓	✓			✓		✓	6	4	15 mn, total duration of the session, when arrived at home and 1 week after the end of the session	15

Notes: T = translation; CA = cross-cultural adaptation; V = validity assessed; R = reliability assessed; N = normative data collected; EBPM = event-based prospective memory; TBPM = time-based prospective memory; DV = diagnostic value; EA = use of external aids; PV = parallel versions; QS = qualitative scoring; LFO = linked to functional outcome measure.

Table 2. Overview of the criteria met for each of the identified single-trial procedures and their main characteristics

Measures	T	CA	V	R	N	EBPM	TBPM	DV	EA	PV	QS	LFO	N of criteria met/N of criteria (max. 12)	Retention interval
Envelope task						✓		✓				✓	3	10 mn
Prompt card task						✓		✓					2	total duration of the session
Telephone test							✓					✓	2	5 mn

Notes: T = translation; CA = cross-cultural adaptation; V = validity assessed; R = reliability assessed; N = normative data collected; EBPM = event-based prospective memory; TBPM = time-based prospective memory; DV = diagnostic value; EA = use of external aids; PV = parallel versions; QS = qualitative scoring; LFO = linked to functional outcome measure.

Table 3. Overview of the criteria met for PM questionnaires and their main characteristics

Measures	T	CA	V	R	N	EBPM	TBPM	DV	QS	LFO	N of criteria met/N of criteria (max. 10)	N of PM items	Duration (mn)
Prospective Memory Questionnaire			✓	✓		✓	✓			✓	5	52	15-17
Prospective and Retrospective Memory Questionnaire	✓	✓	✓	✓	✓	✓	✓	✓		✓	9	8	3-5
Comprehensive Assessment of Prospective Memory	✓		✓	✓	✓					✓	5	39	13-15
Brief Assessment of Prospective Memory			✓	✓						✓	3	16	5-7

Notes: T = translation; CA = cross-cultural adaptation; V = validity assessed; R = reliability assessed; N = normative data collected; EBPM = event-based prospective memory; TBPM = time-based prospective memory; DV = diagnostic value; QS = qualitative scoring; LFO = linked to functional outcome measure. The key variables “EA” and “PV” were excluded from the analyses due to their irrelevance for PM questionnaires.

Table 4. Overview of the criteria met for each of the identified PM experimental procedures and their main characteristics

Measures	T	CA	V	R	N	EB PM	TB PM	DV	EA	PV	QS	LF O	N of criteria met/N of criteria (max. 12)	N of PM item	Duration (mn)
Prospective Remembering Video Procedure			✓	✓		✓				✓		✓	5	18 and 21	12
Test Écologique de Mémoire Prospective			✓	✓		✓	✓			✓		✓	6	15	20
Virtual Week	✓	✓	✓	✓		✓	✓			✓		✓	8	10 /day[a]	60 (full version)[a]
Actual Week			✓	✓		✓	✓		✓				5	10 /day[a]	5 and 7-day version

Notes: T = translation; CA = cross-cultural adaptation; V = validity assessed; R = reliability assessed; N = normative data collected; EBPM = event-based prospective memory; TBPM = time-based prospective memory; DV = diagnostic value; EA = use of external aids; PV = parallel versions; QS = qualitative scoring; LFO = linked to functional outcome measure.

[a]The number of days and items may differ between studies.

Table 5. Percentages of criteria met for PM test batteries

Criteria met (n)	Percentages of criteria met	Related tests
T (2)	40.00	RBMT, CAMPROMPT
CA (0)	0.00	–
V (4)	80.00	RBMT, CAMPROMPT, MIST, RPA-ProMem
R (4)	80.00	RBMT, CAMPROMPT, MIST, RPA-ProMem
N (3)	60.00	RBMT, CAMPROMPT, MIST
EBPM (5)	100.00	RBMT, CBPMT, CAMPROMPT, MIST, RPA-ProMem
TBPM (4)	80.00	CBPMT, CAMPROMPT, MIST, RPA-ProMem
DV (2)	40.00	CAMPROMPT, MIST
EA (2)	40.00	CBPMT, CAMPROMPT
PV (4)	80.00	RBMT, CAMPROMPT, MIST, RPA-ProMem
QS (2)	40.00	RBMT, MIST
LFO (4)	80.00	RBMT, CAMPROMPT, MIST, RPA-ProMem

Notes: T = translation; CA = cross-cultural adaptation; V = validity assessed; R = reliability assessed; N = normative data collected; EBPM = event-based prospective memory; TBPM = time-based prospective memory; DV = diagnostic value; EA = use of external aids; PV = parallel versions; QS = qualitative scoring; LFO = linked to functional outcome measure. RBMT = Rivermead Behavioural memory Test; CBPMT = Cambridge Behavioural Prospective Memory Test; CAMPROMPT = Cambridge Test of Prospective Memory; MIST = Memory for Intentions Screening Test; RPA-ProMem = Royal Prince Alfred Prospective Memory Test.

Table 6. Percentages of criteria met for single-trial PM procedures

Criteria met (n)	Percentages of criteria met	Related tests
T (0)	0.00	–
CA (0)	0.00	–
V (0)	0.00	–
R (0)	0.00	–
N (0)	0.00	–
EBPM (2)	66.67	Envelope task, Prompt card task
TBPM (1)	33.33	Telephone test
DV (2)	66.66	Envelope task, Prompt card task
EA (0)	0.00	–
PV (0)	0.00	–
QS (0)	0.00	–
LFO (2)	66.67	Envelope task, Telephone test

Notes: T = translation; CA = cross-cultural adaptation; V = validity assessed; R = reliability assessed; N = normative data collected; EBPM = event-based prospective memory; TBPM = time-based prospective memory; DV = diagnostic value; EA = use of external aids; PV = parallel versions; QS = qualitative scoring; LFO = linked to functional outcome measure.

Table 7. Percentages of criteria met for self-reported PM questionnaires

Criteria met (n)	Percentages of criteria met	Related tests
T (2)	50.00	PRMQ, CAPM
CA (1)	25.00	PRMQ
V (4)	100.00	PMQ, PRMQ, CAPM, BAPM
R (4)	100.00	PMQ, PRMQ, CAPM, BAPM
N (2)	50.00	PRMQ, CAPM
EBPM (2)	50.00	PMQ, PRMQ
TBPM (2)	50.00	PMQ, PRMQ
DV (1)	25.00	PRQM
QS (0)	0.00	–
LFO (4)	100.00	PMQ, PRMQ, CAPM, BAPM

Notes: T = translation; CA = cross-cultural adaptation; V = validity assessed; R = reliability assessed; N = normative data collected; EBPM = event-based prospective memory; TBPM = time-based prospective memory; DV = diagnostic value; QS = qualitative scoring; LFO = linked to functional outcome measure. The key variables "EA" and "PV" were excluded from the analyses due to their irrelevance for PM questionnaires. PMQ = Propective Memory Questionnaire; PRMQ = Prospective and Retrospective Memory Questionnaire; CAPM = Comprehensive Assessment of Prospective Memory; BAPM = Brief Assessment of Prospective Memory questionnaire.

Table 8. Percentages of criteria met for PM experimental procedures

Criteria met (n)	Percentages of criteria met	Related tests
T (1)	25.00	Virtual Week
CA (1)	25.00	Virtual Week
V (4)	100.00	PRVP, TEMP, Virtual Week, Actual Week
R (4)	100.00	PRVP, TEMP, Virtual Week, Actual Week
N (0)	0.00	–
EBPM (4)	100.00	PRVP, TEMP, Virtual Week, Actual Week
TBPM (3)	75.00	TEMP, Virtual Week, Actual Week
DV (0)	0.00	–
EA (1)	25.00	Actual Week
PV (3)	75.00	PRVP, TEMP, Virtual Week
QS (0)	0.00	–
LFO (4)	100.00	PRVP, TEMP, Virtual Week, Actual Week

Notes: T = translation; CA = cross-cultural adaptation; V = validity assessed; R = reliability assessed; N = normative data collected; EBPM = event-based prospective memory; TBPM = time-based prospective memory; DV = diagnostic value; EA = use of external aids; PV = parallel versions; QS = qualitative scoring; LFO = linked to functional outcome measure. PRVP = Prospective Remembering Video Procedure; TEMP = Test Écologique de Mémoire Prospective.

Table 9. Characteristics of the studies included in the meta-analyses

Study	Sample size	Population characteristics	Mean age in years (SD)	Mean education in years (SD)	PM measures administered
Carey et al., 2006	71	HIV, healthy	44.28 (10.13)	14.15 (2.54)	MIST
Delprado et al., 2012	168	aMCI, healthy	74.82 (6.11)	13.13 (2.92)	CAMPROMPT, Envelope task, Prompt card task
Dennis et al., 2010	102	SBM, healthy	31.60 (13.80)	-	CAMPROMPT
Fleming et al., 2009	72	brain injury, healthy	30.02 (11.46)	-	CAPM
Groot et al. 2002	62	patients with various neurological conditions, healthy	35.43 (10.97)	13.11 (2.65)	CBPMT
Hannon et al., 1995	129	brain injury, healthy	-	-	PMQ
Heffernan & O'Neil, 2012	56	substance abuse, healthy	24.20 (5.38)	-	CAMPROMPT, PRMQ
Heffernan et al., 2010a	40	substance abuse, healthy	23.66 (4.92)	-	CAMPROMPT, PRMQ
Heffernan et al., 2010b	50	substance abuse, healthy	18.64 (0.47)	-	PRMQ, PRVQ
Heffernan et al., 2013	78	substance abuse, healthy	20.85 (2.39)	-	PMQ, PRVP
Henry et al., 2007	69	schizophrenia, healthy	36.72 (10.54)	13.80 (2.70)	Virtual week
Lee et al., 2016	154	aMCI, healthy	73.63 (2.55)	13.38 (1.61)	Envelope task, PRMQ
Man et al., 2011	667	TBI, healthy	44.98 (22.14)	-	BAPM
Mathias & Mansfield, 2005	50	TBI, healthy	28.50 (9.85)	11.90 (1.86)	RBMT
Mioni et al., 2013	36	TBI, healthy	31.86 (10.08)	12.11 (3.20)	Virtual Week
Rabin et al., 2014	257	naMCI, MCI, healthy	80.78 (5.57)	14.49 (3.44)	RPA-ProMem
Radford et al., 2011	40	patients with various neurological conditions, healthy	38.45 (15.38)	14.05 (2.49)	RPA-ProMem

Rendell et al., 2012	60	MS, healthy	47.05 (9.86)	14.15 (2.95)	Virtual Week
Smith et al., 2000	397	AD, healthy	73.21 (8.51)	12.65 (3.50)	PRMQ
Thompson et al., 2015	138	dementia, healthy	78.62 (5.15)	11.62 (3.61)	PRMQ, Virtual Week
Wilson et al., 1989	294	brain injury, healthy	43.10 (11.24)*	-	RBMT
Woods et al., 2007	82	schizophrenia, healthy	46.85 (10.38)	13.4 (1.78)	MIST

Notes: AD = Alzheimer disease; aMCI = amnestic mild cognitive impairment; HIV = human immunodeficiency virus; MS = multiple sclerosis; naMCI = non-amnestic mild cognitive impairment; SBM = spina bifida meningomyelocele; TBI = traumatic brain injury. BAPM = Brief Assessment of Prospective Memory questionnaire; CAMPROMPT = Cambridge Test of Prospective Memory; CAPM = Comprehensive Assessment of Prospective Memory questionnaire; CBPMT = Cambridge Behavioural Prospective Memory Test; MIST = Memory for Intentions Screening Test; PMQ = Prospective Memory Questionnaire; PRMQ = Prospective and Retrospective Memory Questionnaire; PRVP = Prospective Remembering Video Procedure; RBMT = Rivermead Behavioural memory Test; RPA-ProMem = Royal Prince Alfred Prospective Memory Test. * we estimated standard deviations for both patients and controls using a tabulated conversion *f* (see Walter & Yao, 2007 for a description of the statistical method used) because they were not provided in Wilson et al.' study (1989).

Table 10. Planned subgroup analyses

Type of PM assessment	Tests	*N* of studies	*N* of participants	SMD	SE	*p*	*Q*	*p*	I^2 (%)
Test batteries	RBMT	2	344	-0.64	0.21	.002	2.08	.14	52.02
	CAMPROMPT	4	366	-1.61	0.34	<.001	20.52	<.001	85.38
	MIST	2	153	-2.86	0.23	<.001	0.09	.76	0.00
	RPA-ProMem	2	297	-0.85	0.33	.01	3.45	.06	70.98
Single-trial measures	Envelope task	2	321	-1.77	0.56	<.001	17.57	<.001	94.31
Questionnaires	PMQ	2	207	-1.03	1.22	.40	41.26	<.001	97.58
	*PRMQ	7	1032	0.44	0.67	.51	438.46	<.001	98.63
	CAPM	2	134	-0.30	0.26	.24	2.20	.14	54.60

Notes: CAMPROMPT = Cambridge Test of Prospective Memory; CAPM = Comprehensive Assessment of Prospective Memory Questionnaire; MIST = Memory for Intention Screening Test; PMQ = Prospective Memory Questionnaire; PRMQ = Prospective and Retrospective Memory Questionnaire; RBMT = Rivermead Behavioural memory Test; RPA-ProMem = Royal Prince Alfred Prospective Memory Test; SMD = standard mean difference; SE = standard error. *The data from the retrospective memory subscale of the PRMQ were excluded from the analyses.

Legends to figures

Figure 1. Flow chart depicting the study selection process through the phases of the systematic review and meta-analysis