The development of prospective memory in adolescence: A systematic review

Esther Bijl (2043437)

Psychology

Bachelorthesis Developmental and Lifespan Psychology

Department of Developmental Psychology, Tilburg University

dr. Nicola Ballhausen

dr. Charlotte Out

13th of February 2022

Abstract

Prospective memory (PM) refers to the essential ability to carry out future intentions at a certain moment in time or after a specific event. So far, research has shown contradictory findings on how this ability develops, especially when focusing on adolescents. Therefore, the current systematic review aimed to answer the following question: ‘What are the differences in prospective memory performances between children, adolescents and young adults?’ Based on previous neurological and behavioral research, it was hypothesized that adolescents would outperform school-aged children, and young adults would exceed adolescents. Studies would be included in if they were peer-reviewed articles, in English, investigated a healthy population, focused on prospective memory (development), and compared children’s and/or young adults PM performance to that of adolescents. When articles were other forms of publication than peer-reviewed articles, were systematic reviews or meta-analyses, or included a clinical population, they were excluded. In November 2021, a literature search was performed using Web of Science, SAGE Journals Online, PsycArticles, PsycInfo and the Psychology and Behavioral Sciences Collection. The search led to a total of *n* = 1579 studies, of which 14 of cross-sectional nature were eligible. The risk of bias of each study was subsequently assessed using AXIS. The synthesis of all findings showed that PM performance improves from childhood to adolescence, and further, though less forceful, to young adulthood. However, factors such as executive functions and task characteristics seem to play an important role in the extent to which these age group differences emerge.

Keywords: prospective memory, adolescence, systematic review, development, age differences

The development of prospective memory in adolescence: A systematic review

An essential ability in our day-to-day lives is prospective memory (PM). This entails the ability to carry out future intentions at a certain moment in time (time-based PM) or after a certain event occurs (event-based PM; Altgassen et al., 2017). Examples such as remembering to take your medication on time or to pass on a message when you see a colleague, illustrate the importance of this ability and its development in our everyday lives (Bowman et al., 2015). This process of ‘remembering to remember’ involves many cognitive processes and factors that each play a part in the development of PM. For instance, executive functions (EFs) are considered to be a major driving force in PM development (Mahy et al., 2014). EFs refer to higher-order cognitive processes involved in monitoring and controlling thought and action, such as attention regulation, planning and working memory (Carlson et al., 2005). Additionally, an important feature of PM tasks is that one has to perform the task while simultaneously executing a competing activity, the so-called ongoing task (Mahy et al., 2015).

The development of prospective memory has led to an increasing amount of research in the last few decades, measuring PM performance reliably from children as young as 3 years old (Kliegel & Jäger, 2007) to older adults of 97 years old (Varley et al., 2021). Research seems to have reached consensus on the decline of PM performance in older adulthood (Ballhausen et al., 2017; Maylor, 1996; Zuber & Kliegel, 2020). These findings are often linked to the simultaneous age-related decreases in specific EFs such as working memory and shifting (Hering et al., 2014). This has serious consequences: Successful performance on PM tasks has shown to be a critical determinant of whether an older adult is able to live independently and it influences the capacity to execute functional everyday activities (Henry et al., 2021).

Researchers also seem to agree that PM performance progressively increases throughout childhood (Kerns, 2000; Kliegel & Jäger, 2007; Voigt et al., 2014). This is often attributed to the parallel development of EFs and thus the development of the prefrontal cortex (Altgassen et al., 2017). Children's developing PM is essential as they are expected to act more and more on their own initiative, such as remembering to pack their favorite toy before going on holiday.

However, the existing literature is inconclusive about the development of PM in the gap between childhood and older adulthood, especially when focused on adolescence. Some studies found young adults outperforming adolescents on PM (Wang et al., 2006; Zöllig et al., 2007). Neurological measures found these differences connected to the relative late maturation of prefrontal cortical areas related to EF development, which seems to take place in the early 20s (Blakemore & Choudhury, 2006), and differential recruitment of underlying neurocognitive processes (Zöllig et al., 2007). The prefrontal cortex has been repeatedly linked to PM performance, both directly (e.g. Momennejad & Haynes, 2013; Okuda et al., 2007) and indirectly through EF (e.g. Burgess et al., 2008; He et al., 2017). Other researchers found that adolescents and young adults disposed of similar PM performance levels (Ward et al., 2005; Zimmermann & Meier, 2006), which was ascribed to the use of a focal PM task. The multiprocess framework of McDaniel and Einstein (2000), a key theory in the field of PM research, explains this: Focal PM cues are in the focus of attention of the ongoing task, which enables one to retrieve the intended action more automatically and thus requires relatively few cognitive resources. Non-focal PM cues, on the other hand, lie outside the focus of attention of the ongoing task. Therefore, the intention has to be self-initiated, which demands more cognitive resources, such as EFs. When comparing age groups who are at different levels of EF development, age differences in PM performance are more likely to appear when using a non-focal PM cue.

The contradictory findings of significant (Wang et al., 2006; Zöllig et al., 2007) vs. non-significant (Ward et al, 2005; Zimmermann & Meier, 2006) age group differences in PM performance seem surprising, as PM has shown to be a critical aspect in developing autonomy and independence in adolescence (Hering et al., 2020). Furthermore, orientation towards the future has shown to mature during adolescence, which enhances PM function not only in adolescence, but also further in the lifespan (Altgassen et al., 2017). Thus, the development of PM in adolescence seems to be essential for the ability to act independently in adolescence as well as for day-to-day functioning.

Given this importance, the current systematic review aims to create a comprehensive overview of what is currently known about the normative development of PM in adolescence. By integrating the findings on this subject to date, empirically supported conclusions can be drawn about where research stands on this topic today. This will dissolve the fog that hangs over this important, however neglected, area of research. Moreover, it will be a useful resource for future researchers on the topic. They will be able to zoom in onto specific aspects of and unanswered questions around PM in adolescence to elaborate the existing literature.

In order to create a clear, thorough review of the developmental process of PM in adolescence, the findings on PM in the adjacent life-phases will be included to compare the findings on PM in adolescence. That is, the empirical evidence concerning school age children and young adulthood. The main research question is thus formulated as: ‘What are the differences in prospective memory performance between children, adolescents and young adults?’

As previously stated, the development of EFs has consistently shown to play an important role in the development of PM between childhood and adulthood (Altgassen et al., 2017; Mahy et al., 2014). The prefrontal cortex, on which the development of EFs rely and which is correlated with PM performance, continues to develop throughout adolescence (Wang et al., 2011). Moreover, some structural development, such as myelination and synaptic proliferation, does not happen until late adolescence or young adulthood (Wang et al., 2006). This neurological line of evidence is further supported by behavioral research. For example, it seems that, as they grow towards adolescence, children engage in more adaptive, proactive time monitoring strategies than younger children (Voigt et al., 2014). This has shown to be a significant predictor of PM performance among pre- adolescents. Collectively, these findings led to the hypothesis that young adults outperform adolescents, who, in turn, outperform children on PM.

While investigating this hypothesis, the possible moderating factors on age differences in PM will be explored and summarized in order to interpret the findings in a correct and valid manner.

Method

This systematic review was conducted in line with the Preferred Reporting Items for Systematic review and Meta-Analyses guidelines (PRISMA; Page et al., 2021). PRISMA provides of a checklist and a flow diagram for researchers to ensure their transparency and integrality when reporting their findings in systematic review and meta-analyses (for details and elaboration on PRISMA, see Page et al., 2021).

Selection criteria

In line with the aims of this systematic review, studies were included if the following criteria were applicable: (a) Peer-reviewed articles, (b) in English, (c) investigated a healthy population, (d) focused on PM, and (e) compared children’s and/or young adults’ PM performance to that of adolescents. Studies were excluded when they (a) were other forms of publication than peer-reviewed articles, such as e-books and dissertations, (b) were systematic reviews or meta-analyses, or (c) included a clinical population. These criteria were applied both during the search strategy and during the process of selecting studies.

Databases and search strategy

In order to identify relevant studies, the following bibliographic databases were used: Web of Science, SAGE Journals Online and EBSCOhost. The latter of which allowed to search through multiple databases simultaneously. Here, PsycArticles, PsycInfo and the Psychology and Behavioral Sciences Collection were selected. All databases were chosen for their wide scope of the social sciences literature. A search strategy was designed using PICO, a widely used tool in the health sciences that helps researchers to identify the relevant aspects of their research question (Cajal et al., 2020; Davies, 2011). The population or participants (P) consisted of adolescents between 12 and 17 years of age. Adolescents' PM performance was adopted as intervention (I). The comparison (C) consisted of school-aged children of 6 to 11 years old or 18 to 30 year old young adults. Lastly, prospective memory performance functioned as the outcome (O).

Using these components, search strings were designed for prospective memory, adolescents, school-age children and young adults. For an overview of the complete search string, see Appendix A. Every search contained the search strings of prospective memory and adolescents. Additionally, the search string of school-age children was added in the first search of each database, in order to find articles comparing adolescents' and children's prospective memory performance. In the second search, the search string of children was exchanged for that of young adults. The Boolean operator 'AND' was applied to combine the search strings, so every search result would contain each PICO-component. In order to keep the synonyms within each search string distinct, the operator 'OR' was employed. Lastly, the search results of both searches were combined and collectively analyzed.

Study selection

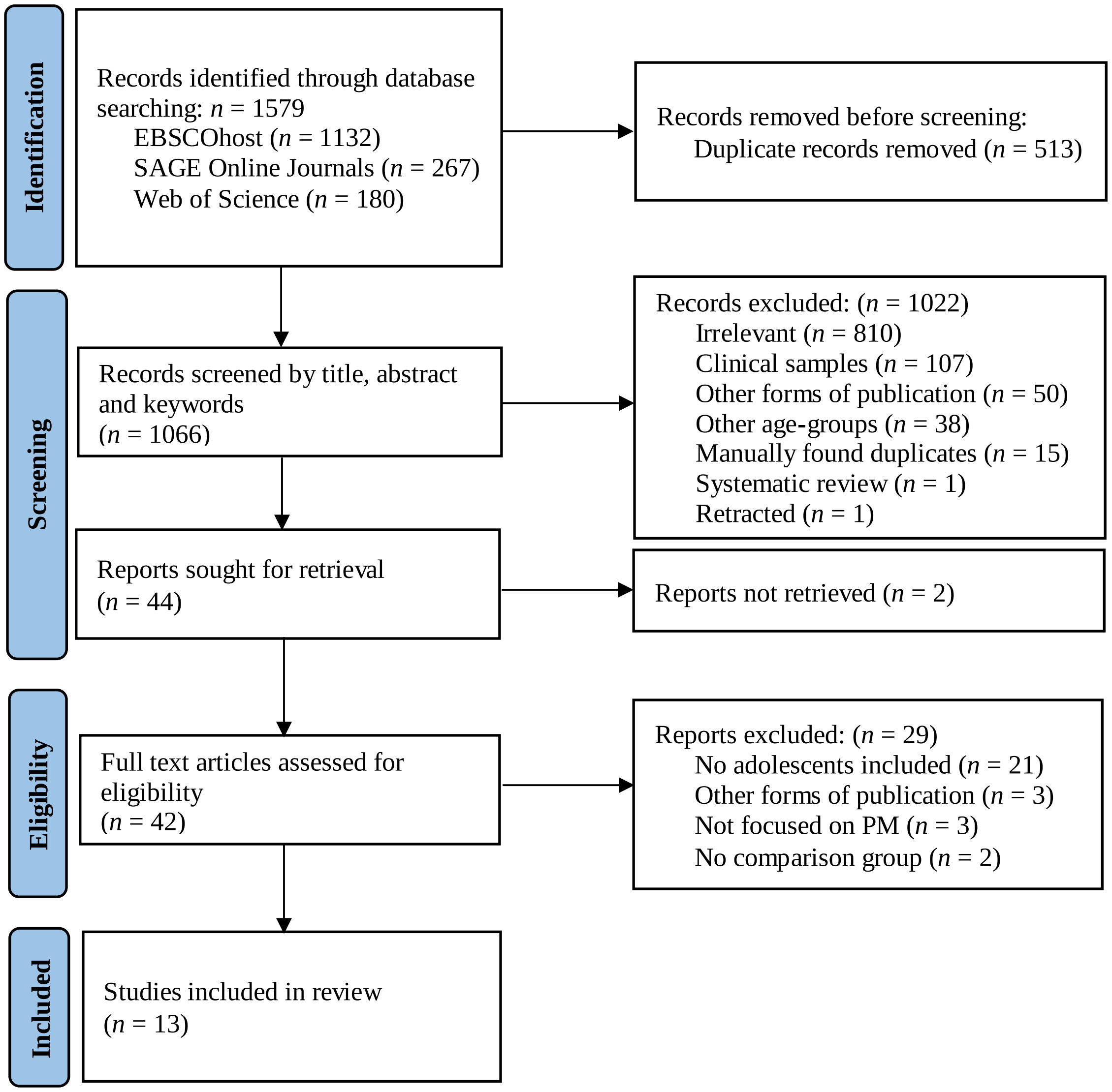
After performing the literature search, all search results were exported to EndNote, an online reference manager that facilitates the organization of reports and referencing (King et al., 2011). EndNote evaluated all records to find duplicates which were subsequently removed. Next, all articles were screened by title, abstract and keywords to identify the studies that match the PICO model. In order to reduce selection bias, a colleague reviewer replicated the screening process for articles marked as a good fit and for articles deemed ambiguous. During a peer consultation the screening processes were compared and discussed. Here, full articles were retrieved for doubtful cases and jointly matched to the inclusion and exclusion criteria. Any remaining uncertainty was discussed and resolved with the supervisor. Subsequently, all other studies were sought for retrieval. Articles that could not be retrieved, were excluded from this review. The last step of the bibliographic search consisted of the assessment of full text articles for eligibility according to the inclusion and exclusion criteria. All reasons for exclusion of articles were reported in the PRISMA flow diagram in Figure 1. The remaining articles were included in this systematic review.

Data extraction

In order to extract and collect the relevant data from each article, a data extraction table was designed. As this systematic review focuses on the development of PM over different phases of life, the mean age, standard deviation and, when available, age ranges were included in the sample characteristics together with the gender ratio. The age groups to which adolescents were compared were reported as well. Some included studies also investigated middle-aged and/or older adults, but these data were irrelevant for the current review thus were not included in the data extraction. Furthermore, types of PM measures were reported in order to take different task demands into account when discussing the findings of these studies. Main findings contained statistical test results and, when provided, effect sizes of age comparisons in PM performance along with mediating effects. In order to accurately compare different effect sizes, the webpage of the MRC Cognitive and Brain Sciences (https://imaging.mrc-cbu.cam.ac.uk/statswiki/FAQ/effectSize) was consulted, which follows suggestions of e.g. Cohen (1988).

Figure 1

PRISMA flow diagram



Note. Adapted from "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews," by Page, M. J. et al., 2021, *British Medical Journal*, *372*, 1-9. https://doi.org/10.1136/bmj.n71

Quality assessment

Each included study was evaluated for the methodological rigor the researchers applied to minimize risk of bias. This systematic review employed AXIS, a critical appraisal tool assessing study design, quality and bias of cross-sectional studies across disciplines (Downes et al., 2016). AXIS was selected due to its ability to assess cross-sectional research in systematic reviews and its rigorous developmental process. As suggested by the National Toxicology Program, a short version of AXIS was employed. A list of these items was added in Appendix B. It consists of 8 items evaluated by low, unclear or high risk of bias. The more items have low risk of bias, the better the methodological quality of the study. This assessment was carried out by the author and a colleague reviewer independently. In comparing the separate assessments and coming to a consensus, disagreements were solved by choosing the most critical judgement as arguments for this choice of judgement risked to be overlooked by the other.

Results

Study selection

In November 2021, the literature search was performed. This resulted in a total of n = 1579 articles. All articles were exported to EndNote, which identified 513 duplicates that were subsequently removed. Screening of title, abstract and keywords by the author and a colleague reviewer led to the exclusion of 1022 articles. The majority of these studies were excluded due to the examination of irrelevant topics. All reasons for exclusion are stated in the PRISMA flow diagram (see Figure 1). The remaining articles were sought for retrieval, where two studies could not be encountered and therefore had to be excluded. Full text articles of 42 studies were assessed for eligibility, 29 of which did not suffice according to the criteria. This left 13 articles to include in this systematic review. Amongst there was one article (Kretschmer-Trendowicz et al., 2021) reporting two studies, each with their own research goals and sampling. Therefore, these studies were analyzed and presented separately, consequently adding a 14th study.

Before moving on to the results of this systematic review, it is important to note that four sub-samples from different studies were excluded in the synthesis of the findings. Kretschmer-Trendowicz et al. (2016) and Zimmermann and Meier (2006) included 4- and 5-year-olds in their children samples. These participants were recruited from child care centers and kindergartens next to schools, thus not all counting as school-aged participants. Also, Kretschmer-Trendowicz et al. (2021a, 2021b) included samples of older children aged between 11 and 13 years old. As this group is on the borderline between children and adolescents, these were also excluded from this review. Lastly, five studies involved middle-aged and/or older adults in their samples, which were not incorporated in the analysis of the findings. All other (sub-)samples were included and analyzed.

Study characteristics

Among the 14 included studies, nine investigated the age group differences in PM performance between adolescents and young adults. One study compared adolescents and children, and the three remaining studies included children, adolescents and young adults. The sample sizes ranged from 28 to 341 participants. These were all healthy participants who aged between 5 to 10 years for children, 12 to 17 years for adolescents and 17 to 30 years for young adults. Of the studies that reported gender characteristics, the majority had slightly more female (50,5% to 65,9%) than male participants, with one outlier having females representing 82% of the sample.

Eight out of 14 studies used a word- or letter-based task to measure prospective memory performance, for instance, participants in the study of Altgassen (2014) were instructed to indicate which of the two words had more vowels. Three studies adopted a picture-based task, one a shapes-based task and one a statement-based task where participants were asked to indicate which personality descriptions applied to them. All studies used a unique prospective component, such as pressing predefined key as quickly as possible whenever a PM target was presented.

Quality assessment

In Table 2 an overview of the quality assessment using AXIS is presented. The main limitation among the included studies concerned the justification of sample sizes (Q1), which none of the studies defended. Especially for studies using small sample sizes (e.g. Zöllig et al., 2007), this information would have been of value. The second most common limitations were the approach to non-responders (Q3) and declaration of conflict of interest (Q8). These were both missing in most of the studies. Only Zhao et al. (2019) mentioned non-responders, but did not undertake measures to avoid non-responders bias. However, this study was the only one, together with Altgassen et al. (2017), that assured no conflict of interest.

Furthermore, half of the included studies failed to recruit a sample representative of the target population (Q2). Altgassen et al. (2014) and Bowman et al. (2015) had small age ranges for their (adolescent and) young adult sub-sample and, additionally, recruited these participants through a convenience sample. Hering et al. (2016) and Zöllig et al. (2007) selected their participants similarly, with the additional disadvantage of having a small sample size. Magis-Weinberg et al. (2020), Wang et al. (2006) and Zimmermann and Meier (2006) had their weak spots, such as small age ranges. However, as they also had their strengths, such as large sample sizes (Wang et al., 2006), these studies received an unknown risk of bias for Q2. Lastly, Zimmermann and Meier (2006) received an unknown risk of bias for the justification of their conclusions (Q5), since the data of some age groups were not discussed in their conclusions.

Table 2

Quality assessment using AXIS

	Altgassen et al. (2017)	Altgassen et al. (2014)	Bowman et al. (2015)	Hering et al. (2020)	Hering et al. (2016)	Kretschmer-Trendowicz et al. (2026)	Kretschmer-Trendowicz et al. (2021a)	Kretschmer-Trendowicz et al. (2021b)	Magis-Weinberg et al. (2020)	Wang et al. (2006)	Ward et al. (2005)	Zhao et al. (2019)	Zimmermann & Meier (2006)	Zöllig et al. (2007)
Q1	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Q2	+	-	-	+	-	+	+	+	?	?	+	+	?	-
Q3	?	?	?	?	?	?	?	?	?	?	?	-	?	?
Q4	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Q5	+	+	+	+	+	+	+	+	+	+	+	+	?	+
Q6	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Q7	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Q8	+	?	?	?	?	?	?	?	?	?	?	+	?	?

Note. The green plus-icons indicate a low risk of bias, yellow question marks an unknown risk of bias and the red minus-icons a high risk of bias. Keywords per item: Q1 = Justification sample size, Q2 = Selection process, Q3 = Non-responders, Q4 = Research aims, Q5 = Justification conclusions, Q6 = Description methods, Q7 = Research instruments, Q8 = Conflict of interest.

Main findings

Below follows a synthesized description of the main findings of all included studies. An overview of these findings is provided in Table 3.

PM performance

Children vs. adolescents. Out of the four studies, all except Kretschmer-Trendowicz (2021a) found better PM performance in adolescents than children. The effect sizes presented by Kretschmer-Trendowicz et al. (2021b) and Zhao et al. (2019) indicated a large effect of age for these differences. However, it should be noted that Zhao et al. (2019) only found an effect for the time-based PM task, not for the event-based tasks. These researchers also found significant mediating effects of interference control (β = -.49) and response inhibition (β = .23) on the PM performances of their adolescent participants. This points to the significance of cognitive control resources in PM performance, especially when a focal PM cue is employed as only these researchers did. Looking further into the one study that did not find significant results, the methodology of Kretschmer-Trendowicz et al. (2021a) was largely similar to those of the other studies. The authors themselves suggested that differences in the development of cognitive control functions or differences in specific task demands could undermine possible age effects.

With regards to PM reaction time, the two studies by Kretschmer-Trendowicz et al. (2021a, 2021b) reported that adolescents were significantly faster than children in their PM responses. Both results were paired with large effect sizes, but, as mentioned, each included a sub-sample that was not included in this review. Nonetheless, these sub-samples included 11- to 13-year-old children, thus these participants were on the borderline between childhood and adolescence. The fact that both studies found significant and large effects of age with these age groups relatively close to each other indicates developmental progress in PM reaction time already towards early adolescence.

Adolescents vs. young adults. Among the 12 studies, five reported no significant age differences in PM performance. Looking closer at these studies, it is noteworthy that three of them employed a focal PM task. As previously explained, focal cues initiate intention retrieval more automatically, whereas non-focal cues rely more on executive processes and working memory. Since those functions likely develop parallel to PM, age differences could be more pronounced when a non-focal PM task is used as opposed to a focal PM task. This is also notable when looking at the studies that did find significant age differences in PM performance: Five out of seven studies employed a non-focal task. This indicates that the use of a non-focal or focal task, thus whether the task is more or less resource-demanding, influences the extent to which age differences in PM performance emerge.

When looking further into the seven studies that reported significant age differences, age was paired with medium (Magis-Weinberg et al., 2020) to mostly large effect sizes. However, it must be noted that two of the large effect sizes (Hering et al., 2020; Zöllig et al., 2007) referred to age differences among adolescents, young adults and older adults. Furthermore, one of the studies (Altgassen et al., 2014) also found mediating effects of theory of mind (r = .32, p = .04) and switching (r = .-33, p = .04) on adolescents' PM performance, even after controlling for (non-)verbal abilities. Zhao et al. (2019) confirmed the findings with regards to switching, although only for adolescents' focal event-based PM task (β = -.23, p = .02). However, Zhao et al. (2019) reported that working memory updating predicted PM performances of young adults, while Altgassen et al. (2014) did not confirm this. This difference could be the result of the wider age range of young adults and larger sample size Zhao et al. (2019) had relative to Altgassen et al. (2014).

With regards to PM reaction time, three out of four studies reported that young adults were significantly faster than adolescents in their PM responses. Age had a medium (Bowman et al., 2015) to large effect (Hering et al., 2016; Zöllig et al., 2007) on these results. Interestingly, Bowman et al. (2015) and Hering et al. (2016) were two of the five studies that did not find such age differences in PM performance. It could have been, as suggested by Bowman et al. (2015), that this inconsistency in age differences appeared due to a speed/accuracy trade-off in PM performance. However, since only few studies reported findings on PM reaction time, this remains to be speculation. Only Kretschmer-Trendowicz et al. (2016) found no differences between adolescents' and young adults' reaction time. This could be due to their use of a focal PM task, whereas the other three studies employed a non-focal PM task. As discussed, age differences appear more likely when using a non-focal task. Another possibility is that their mediating effect of cue saliency for young adults ($p < .001$, $\eta_p^2 = .22$) could have influenced possible age group differences.

OT performance

Children vs. adolescents. All three studies reported that adolescents outperformed children on the ongoing task. Moreover, the studies also agree that there was a large effect of age for these differences. These results indicated that ongoing task performance shows significant developmental progress from childhood to adolescence. Nonetheless, it should be taken into account that only three of all included studies examined this, which offered a limited amount of data to make solid statements.

A similar pattern was observed regarding reaction time. The two studies (Kretschmer-Trendowicz et al., 2021a, 2021b) agreed that adolescents responded significantly faster on the ongoing task than children. Also here, age was paired with a large effect size in both studies, indicating that reaction speed improves from childhood to adolescence. However, the same caution as with ongoing task performance should be applied when interpreting these results since the available amount of information was limited.

Adolescents vs. young adults. Out of four studies, three reported that adolescents were exceeded by young adults in ongoing task performance. The accompanying effect sizes indicated a medium (Altgassen et al. 2017) to large effect of age (Bowman et al., 2015; Hering et al., 2016). The slight outlier by Altgassen et al. (2017) could be the result of the

Table 3

Summary of the development of prospective memory in adolescence

Study	Participants	PM measures	Main findings		
			PM performance	OT performance	Mediators
Altgassen et al. (2017)	N = 109 AD: n = 49 (21♂) M_{age} = 14.43 SD_{age} = 0.71 YA: n = 60 (18♂) M_{age} = 21.18 SD_{age} = 2.38	OT: Picture-based two-back working-memory task. PM: Press a highlighted button when 1 of 4 specific pictures appeared.	AD < YA PM accuracy (p = .001, η_p^2 = .12)	AD < YA PM accuracy (p = .002, η_p^2 = .09)	Encoding condition (**p = .006, η_p^2 = .11**)
Altgassen et al. (2014)	N = 83 AD: n = 42 (10♂) M_{age} = 13.55 SD_{age} = 0.5 Range: 13-14 YA: n = 41 (5♂) M_{age} = 19.44 SD_{age} = 0.5 Range: 19-20	OT: Press left/right key depending on which of the noun pairs has more vowels. PM: Press space bar as quickly as possible when one of the words was a verb.	AD < YA PM accuracy (p = .003, d = .70)	AD < YA PM accuracy (p = .001, η_p^2 = .29)	AD Theory of mind (**r = .32, p = .04**) Switching (**r = -.33, p = .04**) Updating (r = .16, p = .30) Inhibition (r = -.22, p = .17) YA All measures (rs < .14)
Bowman et al. (2015)	N = 93 AD: n = 65 (22♂) Range: 12-17 YA: n = 28 (10♂) Range: 18-19	OT: Indicate whether the letter string was a word or a non-word. PM: Withhold the OT response and press a target key whenever the PM cue appeared.	AD = YA PM accuracy (p = .14) AD > YA (p < .05) PM RT (ω^2 = .06)	AD < YA (p < .01) OT accuracy (ω^2 = .20) AD > YA (p < .001) OT RT (ω^2 = .21)	n.a.

Hering et al. (2020)	$N = 45$ AD: $n = 25$ (9♂) $M_{age} = 13.5$ $SD_{age} = 0.82$ Range: 12-15 YA: $n = 20$ (10♂) $M_{age} = 25.5$ $SD_{age} = 2.70$ Range: 21-30	OT: Decide if two words belonged to the same or different categories. PM: Encode and store the combination of letters and colors until the retrieval trial occurred.	AD < YA ($p < .01$) PM accuracy ($\eta_p^2 = .246$[a])	AD < YA ($p < .001$) OT accuracy ($\eta_p^2 = .456$)	Encoding time and intention load ($p > .108$)
Hering et al. (2016)	$N = 52$ AD: $n = 26$ $M_{age} = 12.98$ $SD_{age} = 0.79$ Range: 12-14 YA: $n = 26$ $M_{age} = 23.04$ $SD_{age} = 2.63$ Range: 19-28	OT: Decide if two words belonged to the same or different categories. PM: Encode and store the combination of letters and colors until the retrieval trial occurred.	AD = YA ($p_{adj} = .202$) PM accuracy AD > YA ($p_{adj} < .001$) PM RT ($\eta_p^2 = .240$)	AD < YA ($p_{adj} < .001$) OT accuracy ($\eta_p^2 = .380$) AD > YA ($p_{adj} < .001$) OT RT ($\eta_p^2 = .196$)	Distinctiveness of PM cue ($p = .606$)
Kretschmer-Trendowicz et al. (2016)	$N = 40$ AD: $n = 20$ (7♂) $M_{age} = 13.6$ $SD_{age} = 1.05$ Range: 12-15 YA: $n = 20$ $M_{age} = 22.6$ $SD_{age} = 1.19$ Range: 21-25	OT: Indicate if both pictures belonged to the same category or not. PM: Press a predefined key when one of the specific PM targets was presented.	AD = YA PM accuracy ($ps = 1$) PM RT ($ps > .09$)	AD = YA OT accuracy ($ps > .19$) OT RT ($p = 1$)	Salience of PM cue (**$p < .001$, $\eta_p^2 = .22$**) for YA.
Kretschmer-Trendowicz et al. (2021a)	$N = 77$ CH: $n = 30$ (10♂) $M_{age} = 9.1$ $SD_{age} = 0.31$ Range: 9-10 AD: $n = 47$ (22♂)	OT: Indicate whether the two letter strings were equal or not.	CH = AD PM accuracy CH > AD ($p < .001$) PM RT ($\eta_p^2 = .24$[a])	CH < AD ($p < .001$) OT accuracy ($\eta_p^2 = .15$) CH > AD ($p < .001$) OT RT ($\eta_p^2 = .26$)	Implementation intentions and importance of PM task ($ps \geq .44$)

	M_{age} = 15.06 SD_{age} = 0.39 Range: 14-16	PM: Press a specific key whenever one of the PM cues was presented.			
Kretschmer-Trendowicz et al. (2021b)	N = 83 CH: n = 37 (15♂) Range: 8-10 AD: n = 46 (20♂) Range: 14-16	OT: Classify pictures as living vs. non-living or small vs. big. PM: Press a specific key whenever one of two specific line-drawings was presented.	CH < AD ($p < .001$) PM accuracy ($\eta_p^2 = .14^a$) CH > AD ($p < .001$) PM RT ($\eta_p^2 = .34^a$)	CH < AD ($p < .001$) OT accuracy ($\eta_p^2 = .26$) CH > AD ($p < .001$) OT RT ($\eta_p^2 = .26$)	Implementation intentions ($ps \geq .11$)
Magis-Weinberg et al. (2020)	N = 47 AD: n = 28 (13♂) M_{age} = 14.6 SD_{age} = 1.4 Range: 12-17 YA: n = 19 (10♂) M_{age} = 27.1 SD_{age} = 1.9 Range: 22-30	OT: Indicate whether the triangle shape was located left/right of the other shape. PM: Press a specific key if the shapes are the same color or one chess knight move away.	AD < YA ($p = .005$) PM accuracy^b ($\eta_G^2 = .06$)	n.a.	n.a.
Wang et al. (2006)	N = 341 AD: n = 122 (53♂) M_{age} = 14.47 SD_{age} = 0.58 Range: 13-16 YA: n = 219 (54♂) M_{age} = 20.54 SD_{age} = 0.63 Range: 19-22	OT: Make a √ behind the statements that represent their personality or emotional states. PM: Make a second √ behind the statements that include any negative word.	AD < YA PM accuracy ($p < .001$)	n.a.	n.a.

Ward et al. (2005)	$N = 90$ (43♂) CH: $n = 30$ $M_{age} = 8.6$ $SD_{age} = 1.19$ Range: 7-10 AD: $n = 30$ $M_{age} = 14.57$ $SD_{age} = 1.30$ Range: 13-16 YA: $n = 30$ $M_{age} = 19.07$ $SD_{age} = 1.14$ Range: 18-21	OT: Indicate whether the letter string was a word or a non-word. PM: Press a specific key when the letter string appeared in italic.	CH < AD ($p < .01$) PM accuracy AD = YA PM accuracy	n.a.	Importance of PM task ($p > .05$)
Zhao et al. (2019)	$N = 326$ CH: $n = 108$ (46♂) $M_{age} = 8.07$ $SD_{age} = 0.59$ Range: 7-9 AD: $n = 112$ (49♂) $M_{age} = 13.35$ $SD_{age} = 0.75$ Range: 12-14 YA: $n = 106$ $M_{age} = 19.66$ $SD_{age} = 1.17$ Range: 17-23	EBPM OT: Decide whether the current letter shown was the same as the letter presented two trials back. Focal PM: Press the space bar instead of responding according to the OT. Non-focal PM: Press the space bar when a letter is surrounded by a red or yellow frame. TBPM OT: Decide whether the letters on the second and fourth position in the letter string were identical or not. TBPM PM: Press a specific key when the participant thought the target time had passed.	CH = AD EBPM AD < YA ($ps < .001$) EBPM (focal $\eta^2 = .25$; non-focal $\eta^2 = . 09$) CH < AD < YA ($ps < .001$) TBPM ($\eta^2 = . 32$)	CH < AD < YA ($ps < .001$) EBPM (focal $\eta^2 = .41$; non-focal $\eta^2 = .27$) TBPM ($\eta^2 = .46$)	AD for focal EBPM Switching (**$\beta = -.23$, $p = .02$**) Working memory updating (**$\beta = .29$, $p = .02$**) CH for TBPM All EFs together ($p = .02$)

Zimmermann & Meier (2006)	$N = 80$ AD: $n = 40$ M_{age}: 13.3 $SD_{age} = 0.46$ Range: 13-14 YA: $n = 40$ $M_{age} = 21.2$ $SD_{age} = 1.75$ Range: 19-26	OT: Indicate whether the two pictures are identical or not. PM: Press a specific key whenever a target picture was presented.	AD = YA PM accuracy	n.a.	n.a.
Zöllig et al. (2007)	$N = 28$ AD: $n = 14$ (7♂) $M_{age} = 12.8$ $SD_{age} = 0.6$ YA: $n = 14$ (7♂) $M_{age} = 22.5$ $SD_{age} = 1.4$	OT: Decide whether the two words belonged to the same or different semantic category. PM[c]: Press a specific key when a word in the prospective color appears the second time.	AD < YA ($p < .05$; $p < .01$) PM accuracy ($d = 1.05$[a]) PM RT[b] ($\eta^2 = .28$[a])	AD < YA ($p < .01$) OT accuracy ($d = 1.34$)	n.a.

Note. Effect sizes and p-values next to the age group comparisons (e.g. AD < YA) refer to the overall age difference, whereas the ones after the specific measurement (i.e. PM accuracy) refer to the age difference between the specific age groups. Bold values highlight significant predictors for age differences in PM performance. Abbreviations: CH = children, AD = adolescents, YA = young adults, PM = prospective memory, OT = ongoing task, RT = reaction time, n.a. = not available or not applicable.

[a] Overall age effects refer to differences among age groups of which some (e.g. young children and older adults) are not included in this systematic review.

[b] Values were calculated for both prospective memory and ongoing task aspects.

[c] Due to the research goals of this study, three different prospective memory tasks were designed. The task described was selected, because it measured prospective memory performance without including manipulating factors that were irrelevant for this systematic review, though could influence age differences in the participants' performance.

reported interaction between age and encoding condition. Ongoing task performance was measured across conditions, while the age groups responded differently to the two experimental conditions. This could have undermined the effect of age the other two studies found, as encoding condition mediated the effect of age on ongoing task performance. The other, more striking, outlier (Kretschmer-Trendowicz et al., 2016) reported no age differences between adolescents and young adults in their OT performance. A possible explanation is that their sample size of $N = 40$ could have lowered the power of this study. However, this is merely speculation as other studies with similar sample sizes (e.g. Hering et al., 2020; Zöllig et al., 2007) have found significant age differences.

With regards to ongoing task reaction time, three out of four studies agree that young adults were significantly faster in their OT responses than adolescents. The exception was, once more, presented by Kretschmer-Trendowicz et al. (2016), who did not find any age differences in ongoing task reaction time. Reasons for this divergence remain to be speculations, however, it is notable that Kretschmer-Trendowicz et al. (2016) is the only study using a picture-based task in comparing adolescents and young adults.

Discussion

The aim of this systematic review was to answer the following question: ‘What are the differences in prospective memory performance between children, adolescents and young adults?’ Additionally, the contribution of various mediators has been explored and summarized. It was hypothesized that young adults would outperform adolescents, who in turn would exceed school-aged children.

Overall, PM performance seems to show a careful developmental trend. Most of the included articles agree that adolescents’ PM performances outperform those of children.

Among the studies that compare PM performances between adolescents and young adults, there is less coherence: Five out of 12 studies found similar PM performance levels for these age groups, whereas the other seven studies reported that young adults outperformed adolescents. The reported effect sizes indicate that age often explained a large part of these differences. However, several cognitive and executive functions, such as theory of mind and response inhibition, also play a significant role in performance differences between age groups. These functions contributed to especially adolescents’ PM performance, which suggests that cognitive and executive functions are still developing in adolescence and may partially account for their performance level. When looking further into the studies that did not find significant differences between adolescents and young adults, it is noticeable that most of them employed a focal PM task, whereas the majority of the studies with significant results used a non-focal PM task. This suggests that when a PM task is resource-demanding, age differences emerge more than when a PM task requires few cognitive resources. With regards to PM reaction time, a similar developmental trend emerges: Most studies report the older age group exceeding the younger in giving their PM response, with age having a medium to mostly large effect on these age differences. In ongoing task performance, the improvement with age seems to be even more pronounced. Apart from one exception (Kretschmer-Trendowicz et al., 2016), all studies agree that young adults are better and faster than adolescents, who in turn outperform children. Taken together, the results of this systematic review seem to confirm the hypothesis that PM performance improves from childhood (6 to 11 years old) to adolescence (12 to 17 years old) as well as from adolescence to young adulthood (18 to 30 years old).

As mentioned, the age differences in this review are consistent in comparing children and adolescents, whereas there is more disagreement among studies comparing adolescents and young adults. Because the majority of the included studies still shows developmental progress, this disagreement could be due to the declining effect of age further in the lifespan. This is in line with the review of Zuber and Kliegel (2020), showing an inverted U-shaped curve of PM performance over the lifespan. That is, PM performance improves with age until it reaches a peak at the beginning of young adulthood, after which it declines as people get older. The rationale behind this is the developmental trajectory of the prefrontal cortex. As previously discussed, the prefrontal cortex is both directly (Momennejad & Haynes, 2013; Okuda et al., 2007) and indirectly through EFs (Burgess et al., 2008; He et al., 2017) linked with PM. The prefrontal cortex and EFs develop most rapidly throughout childhood (Mattli et al., 2014). They further develop, though less quickly, in adolescence and reach their peak in late adolescence or young adulthood. Therefore, differences between adolescents and young adults are much less pronounced than when comparing children and adolescents. Especially when the PM task is low resource-demanding and invokes few EFs, adolescents' and young adults' PM performance levels are more likely to be similar and age differences between these age groups are less likely to emerge.

Another explanation for the studies reporting no significant age differences in PM is offered by the multiprocess framework of McDaniel and Einstein (2000). The theory states that type of PM and particularly the use of a focal vs. non-focal cue influence whether the PM cue is detected more automatically or more resource demanding. The detection of focal cues, that four out of six studies with non-significant age differences used, relies on more automatic processes and thus should require less resources. Therefore, these show less age-effects than when non-focal cues are employed, where the memory cue lies outside of the focus of attention while working on the ongoing task. Six out of eight studies that used a non-focal cue found significant age differences. Moreover, Zhao et al. (2019) reported a clear developmental effect for their time-based task. In contrast to event-based tasks, time-based tasks do not provide external cues, so the intention must be self-initiated and it thus requires more resources (Zuber et al, 2019). All in all, age differences in PM performance are more likely to be found when a the PM task is more resource-demanding, thus when a non-focal PM task is employed.

Diving deeper in such resources, some studies examined executive functions as possible mediating factors for age differences in PM performance. As Mahy et al. (2014) stated, executive functions are assumed to play a part in PM performance, at least for children. Zhao et al. (2019) partially confirmed this by reporting that their measured EFs collectively contributed to the time-based, but not event-based, PM performance of children. For their adolescent participants, together with Altgassen et al. (2014), more specific EFs predicted their PM performance, namely switching, working memory updating and response inhibition. This is in line with previously discussed neurological evidence, showing that the prefrontal cortex is still under development throughout adolescence (Wang et al., 2011). However, the fact that not all EFs contributed to every type of adolescents' PM performances highlights the importance of individual differences in EF developmental trajectories (Rose et al., 2011; Zuber et al., 2019) as well as the fact that there is still much unknown about the extent to which PM task invoke which specific EFs.

In addition, some task characteristics have been examined by several included studies. Almost all of these results did not reveal a mediating effect, except for salience of the PM cue for young adults (Kretschmer-Trendowicz et al., 2016). Salient cues are theorized to facilitate shifting and reduce the demand for monitoring and inhibition, which results in better PM performance (Mahy et al., 2014). This finding builds on evidence by Einstein and McDaniel (1990), Mahy et al. (2014) and Kliegel et al. (2013) that found similar effects for several age groups. The other, non-significant results are in contrast with previous studies that showed how several task characteristics can influence age differences in PM performances (e.g. Kliegel et al., 2004; Kliegel et al., 2008; Rendell et al., 2007). However, the same studies as well as the multiprocess framework stress that various factors, and the complex interplay between them, can influence PM performance and age differences within. It is recommended that future research looks more explicitly into such factors as EFs and task characteristics in order to learn more about their precise role in this interplay.

Taken together, this systematic review provides corroborating findings for the widespread assumption that (at least the first half of) the lifespan development of PM follows an inverted U-shape as well as the multiprocess framework of McDaniel and Einstein (2000). It highlights that, while there is an overall improvement of PM performance from childhood through young adulthood, factors such as EFs and task characteristics may determine the extent to which these age differences emerge in individual studies, particularly the use of a focal vs. non-focal PM task. This is also important to keep in mind for (pre-)adolescents and their environments: As they are given more independence and responsibilities growing up, there are many aspects where successfully executing a delayed intention is dependent on. Since PM is an essential ability to act independently in life, it is important to give them the opportunity to develop this ability, but equally so to know when they might not yet be able to be autonomous.

Nevertheless, the results of this review should be interpreted in light of several caveats. First of all, some included studies had a high or unknown risk of bias on a few of the quality assessment items, such as justification for sample sizes as discussed earlier. This has consequences for their reliability and validity, which should not be ignored. Notwithstanding, all studies scored also had their strengths, such as their statistical methodologies. Secondly, there were only four studies that included children in their sample relative to 12 studies that involved young adults. Consequently, there was significantly less information available to base conclusions on when comparing children and adolescents. It is therefore recommended that future research collects more data about the PM performances of these two age groups collectively, in order to draw more powerful conclusions. Thirdly, the reported effect sizes of age sometimes referred to differences among age groups that were not all included in this review. In addition, not every study reported an effect size. Hence, it is hard to determine what portion of the variance in PM performance differences can be ascribed to age. Future research examining age effects should report such effect sizes and preferably look for and report age effect for separate contrasts of age groups instead of only an overall age effect. Fourthly, despite the widespread assumption that EFs are a major driving force in PM development (Mahy et al., 2014), just two of the included studies examined their possible mediating effect on PM in different age groups. Many studies examined their presumed parallel development across childhood (e.g. Voigt et al., 2014; Zuber et al., 2019), but few extend their research to adolescence or further. It would be valuable for future research to fill this gap in order to elucidate the roles of age and executive functions in PM performance differences across age groups.

In conclusion, the present systematic review shows that PM performance improves from childhood to adolescence and further, although less forceful, into young adulthood. The majority of the studies agree that age explains a significant part of these age group differences. Additionally, other factors, such as executive functions and task characteristics, likely influence the extent to which these differences emerge. These findings provide further support for the multiprocess framework (McDaniel & Einstein, 2000) as well as the theory that PM development follows an inverted U-shaped curve over the lifespan. However, due to the limitations of the review, this field of research would benefit of future studies collecting more data on underaged age groups and isolating the effects of various influential factors in order to disentangle their complex interplay.

References

Altgassen, M., Kretschmer, A., & Schnitzspahn, K. M. (2017). Future thinking instructions improve prospective memory performance in adolescents. *Child Neuropsychology, 23*(5), 536-553. https://doi.org/10.1080/09297049.2016.1158247

Altgassen, M., Vetter, N. C., Phillips, L. H., Akgün, C., & Kliegel, M. (2014). Theory of mind and switching predict prospective memory performance in adolescents. *Journal of Experimental Child Psychology*, *127*, 163-175. https://doi.org/10.1016/j.jecp.2014.03.009

Ballhausen, N., Schnitzspahn, K. M., Horn, S. S., & Kliegel, M. (2017). The interplay of intention maintenance and cue monitoring in younger and older adults' prospective memory. *Memory & Cognition*, *45*(7), 1113-1125. https://doi.org/10.3758/s13421-017-0720-5

Blakemore, S.-J., & Choudhury, S. (2006). Development of the adolescent brain: Implications for executive function and social cognition. *Journal of Child Psychology and Psychiatry, 47*(3-4), 296-312. https://doi.org/10.1111/j.1469-7610.2006.01611.x

Bowman, C., Cutmore, T., & Shum, D. (2015). The development of prospective memory across adolescence: An event-related potential analysis. *Frontiers in Human Neuroscience, 9*, 1-15. https://doi.org/10.3389/fnhum.2015.00362

Burgess, P. W., Dumontheil, I., Gilbert, S. J., Okuda, J., Schölvinck, M. L., & Simons, J. S. (2008). On the role of rostral prefrontal cortex (area 10) in prospective memory. In M. Kliegel, M. A. McDaniel, & G. O. Einstein (Eds.), *Prospective memory: Cognitive, neuroscience, developmental, and applied perspectives* (pp. 235-260). Taylor & Francis Group/Lawrence Erlbaum Associates.

Cajal, B., Jiménez, R., Gervilla, E., & Montaña, J. J. (2020). Doing a systematic review in health sciences. *Clínica y Salud, 31*(2), 77-83. https://doi.org/10.5093/clysa2020a15

Carlson, S. M., Davis, A. C., & Leach, J. G. (2005). Less is more: Executive function and symbolic representation in preschool children. *Psychological Science*, *16*(8), 609-618. https://doi.org/10.1111/j.1467-9280.2005.01583.x

Davies, K. S. (2011). Formulating the evidence based practice questions: A review of the frameworks. *Evidence Based Library and Information Practice*, *6*(2), 75-80. https://doi.org/10.18438/B8WS5N

Downes, M. J., Brennan, M. L., Williams, H. C., & Dean, R. S. (2016). Development of a critical appraisal tool to assess to quality of cross-sectional studies (AXIS). *British Medical Journal Open*, *6*(12), 1-7. https://doi.org/10.1136/bmjopen-2016-011458

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

He, W., Fan, C., & Li, L. (2017). Transcranial magnetic stimulation reveals executive control dissociation in the rostral prefrontal cortex. *Frontiers in Human Neuroscience*, *11*, 1-8. https://doi.org/10.3389/fnhum.2017.00464

Henry, J. D., Hering, A., Haines, S., Grainger, S. A., Koleits, N., McLennan, S., Pelly, R., Doyle, C., Rose., N. S., Kliegel, M., & Rendell, P. G. (2021). Acting with the future in mind: Testing competing prospective memory interventions. *Psychology and Aging*, *36*(4), 491-503. https://doi.org/10.1037/pag0000593

Hering, A., Rendell, P. G., Rose, N. S., Schnitzspahn, K. M., & Kliegel, M. (2014). Prospective memory training in older adults and its relevance for successful aging. *Psychological Research*, *78*(6), 892-904. https://doi.org/10.1007/s00426-014-0566-4

Hering, A., Wild-Wall, N., Falkenstein, M., Gajewski, P. D., Zinke, K., Altgassen, M., & Kliegel, M. (2020). Beyond prospective memory retrieval: Encoding and remembering of intentions across the lifespan. *International Journal of Psychophysiology*, *147*, 44-59. https://doi.org/10.1016/j.ijpsycho.2019.11.003

Hering, A., Wild-Wall, N., Gajewski, P. D., Falkenstein, M., Kliegel, M., & Zinke, K. (2016). The role of cue detection for prospective memory development across the lifespan. *Neuropsychologia*, *93*, 289-300. https://doi.org/10.1016/j.neuropsychologia.2016.11.008

Kerns, K. A. (2000). The CyberCruiser: An investigation of development of prospective memory in children. *Journal of the International Neuropsychology Society, 6*, 62-70. https://doi.org/10.1017/S1355617700611074

King, R., Hooper, B., & Wood, W. (2011). Using bibliographic software to appraise and code data in educational systematic review research. *Medical Teacher, 33*(9), 719-723. https://doi.org/10.3109/0142159X.2011.558138

Kliegel, M., & Jäger, T. (2007). The effects of age and cue-action reminders on event-based prospective memory performance in preschoolers. *Cognitive Development, 22*(1), 33-46. https://doi.org/10.1016/j.cogdev.2006.08.003

Kliegel, M., Jäger, T., & Phililips, L. H. (2008). Adult age differences in event-based prospective memory: A meta-analysis on the role of focal versus nonfocal cues. *Psychology and Aging, 23*(1), 203-208. https://doi.org/10.1037/0882-7974.23.1.203

Kliegel, M., Mahy, C. E. V., Voigt, B., Henry, J. D., Rendell, P. G., & Aberle, I. (2013). The development of prospective memory in young schoolchildren: The impact of ongoing task absorption, cue salience, and cue centrality. *Journal of Experimental Child Psychology, 116*(4), 792-810. https://doi.org/10.1016/j.jecp.2013.07.012

Kliegel, M., Martin, M., McDaniel, M., & Einstein, G. O. (2004). Importance effects on performance in event-based prospective memory tasks. *Memory, 12*(5), 553-561. https://doi.org/10.1080/09658210344000099

Kretschmer-Trendowicz, A., & Altgassen, M. (2016). Event-based prospective memory across the lifespan: Do all age groups benefit from salient prospective memory cues? *Cognitive Development*, *39*, 103-112. https://doi.org/10.1016/j.cogdev.2016.04.005

Kretschmer-Trendowicz, A., Kliegel, M., Goschke, T., & Altgassen, M. (2021). 'If-then' but when? Effects of implementation intentions on children's and adolescents' prospective memory. *Cognitive Development*, *57*, 1-16. https://doi.org/10.1016/j.cogdev.2020.100998

Magis-Weinberg, L., Custers, R., & Dumontheil, I. (2020). Sustained and transient processes in event-based prospective memory in adolescence and adulthood. *Journal of Cognitive Neuroscience*, *32*(10), 1924-1945. https://doi.org/10.1162/jocn_a_01604

Mahy, C. E. V., Moses, L. J., & Kliegel, M. (2014). The development of prospective memory in children: An executive framework. *Developmental Review*, *34*(4), 205-326. https://doi.org/10.1016/j.dr.2014.08.001

Mahy, C. E., Voigt, B., Ballhausen, N., Schnitzspahn, K., Ellis, J., & Kliegel, M. (2015). The impact of cognitive control on children’s goal monitoring in a time-based prospective memory task. *Child Neuropsychology*, *21*(6), 823-839. https://doi.org/10.1080/09297049.2014.967202

Mattli, F., Schnitzspahn, K. M., Studerus-Germann, A., Brehmer, Y., & Zöllig, J. (2014). Prospective memory across the lifespan: Investigating the contribution of retrospective and prospective processes. *Aging, Neuropsychology, and Cognition*, *21*(5), 515-543. https://doi.org/10.1080/13825585.2013.837860

Maylor, E. (1996). Age-related impairment in an event-based prospective memory task. *Psychology and Aging*, *11*(1), 74-78. https://doi.org/10.1037/0882-7974.11.1.74

McDaniel, M. A., & Einstein G. O. (2000). Strategic and automatic processes in prospective memory retrieval: A multiprocess framework [Special issue]. *Applied Cognitive Psychology, 14*, S127-S144. https://doi.org/10.1002/acp.775

Momennejad, I., & Haynes, J.-D. (2013). Encoding of prospective tasks in the human prefrontal cortex under varying task loads. *The Journal of Neuroscience, 33*(44), 17342-17349. https://doi.org/10.1523/JNEUROSCI.0492-13.2013

MRC Cognitive and Brain Sciences Unit. (2021, November 30). *Rules of thumb on magnitudes of effect sizes*. https://imaging.mrc-cbu.cam.ac.uk/statswiki/FAQ/effectSize

National Toxicology Program (2020, April 1). *Risk of bias tools*. U.S. Department of Health and Human Services. https://ntp.niehs.nih.gov/go/ohat_tools

Okuda, J., Fujii, T., Ohtake, H., Tsukiura, T., Yamadori, A., Frith, C. D., & Burgess, P. W. (2007). Differential involvement of regions of rostral prefrontal cortex (Brodmann area 10) in time- and event-based prospective memory. *International Journal of Psychophysiology, 64*(3), 233-246. https://doi.org/10.1016/j.ijpsycho.2006.09.009

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *British Medical Journal, 372*, 1-9. https://doi.org/10.1136/bmj.n71

Rendell, P. G., McDaniel, M. A., Forbes, R. D. & Einstein, G. O. (2007). Age-related effects in prospective memory are modulated by ongoing task complexity and relation to target cue. *Aging, Neuropsychology, and Cognition*, *14*(3), 236-256. https://doi.org/10.1080/13825580600579186

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

Varley, D., Henry, J. D., Gibson, E., Suddendorf, T., Rendell, P. G., & Redshaw, J. (2021). An old problem revisited: How sensitive is time-based prospective memory to age-related differences? *Psychology and Aging, 36*(5), 616-625. https://doi.org/10.1037/pag0000625

Voigt, B., Mahy, C. E. V., Ellis, J., Schnitzspahn, K., Krause, I., Altassen, M., & Kliegel, M. (2014). The development of time-based prospective memory in childhood: The role of working memory updating. *Developmental Psychology, 50*(10), 2393-2404. https://doi.org/10.1037/a0037491

Wang, L., Altgassen, M., Liu, W., Xiong, W., Akgün, C., & Kliegel, M. (2011). Prospective memory across adolescence: The effects of age and cue focality. *Developmental Psychology, 47*(1), 226-232. https://doi.org/10.1037/a0021306

Wang, L., Kliegel, M., Yang, Z., & Liu, W. (2006). Prospective memory performance across adolescence. *The Journal of Genetic Psychology: Research and Theory on Human Development*, *167*(2), 179-188. https://doi.org/10.3200/GNTP.167.2.179-188

Ward, H., Shum, D., McKinlay, L., Baker-Tweney, S., & Wallace, G. (2005). Development of Prospective Memory: Tasks Based on the Prefrontal-Lobe Model. *Child Neuropsychology*, *11*(6), 527-549. https://doi.org/10.1080/09297040490920186

Zhao, X., Fu, J., Ma, X., & Maes, J. H. R. (2019). Age Differences in Prospective Memory: A Further Evaluation of the Executive Framework. *Journal of Cognition & Development*, *20*(5), 680-701. https://doi.org/10.1080/15248372.2019.1648268

Zimmermann, T. D., & Meier, B. (2006). The rise and decline of prospective memory across the lifespan. *The Quarterly Journal of Experimental Psychology*, *59*(12), 2040-2046. https://doi.org/10.1080/17470210600917835

Zöllig, J., West, R., Martin, M., Altgassen, M., Lemke, U., & Kliegel, M. (2007). Neural correlates of prospective memory across the lifespan. *Neuropsychologia*, *45*(14), 3299-3314. https://doi.org/10.1016/j.neuropsychologia.2007.06.010

Zuber, S., & Kliegel, M. (2020). Prospective memory development across the lifespan: An integrative framework [Special issue]. *European Psychologist*, *25*(3), 162-173. https://doi.org/10.1027/1016-9040/a000380

Zuber, S., & Kliegel, M. (2020). Prospective memory development across the lifespan: An integrative framework. *European Psychologist*, *25*(3), 162-173. https://doi.org/10.1027/1016-9040/a000380

Zuber, S., Mahy, C. E. V., & Kliegel, M. (2019). How executive functions are associated with event-based and time-based prospective memory during childhood. *Cognitive Development*, *50*, 66-79. https://doi.org/10.1016/j.cogdev.2019.03.001

Appendix A

Table 1

Complete search string

Search terms	EBSCOhost, SAGE Online Journals and Web of Science
Prospective memory	“Prospective memory” OR “Delayed intention*” OR “Memory for intentions” OR “Remembering to remember” OR “Intended action*” OR “Future intention*” OR “Intentional memory”
School-age children	Child* OR Youth* OR Kid* OR “School age*” OR School-age* OR Development* OR Lifespan OR “Life-span” OR “Life span”
Adolescents	Adolescen* OR Teenager* OR Teen*
Young adults	Adulthood OR “Young adult*” OR Development* NOT “older adult*”

Note. The search strings were combined with the Boolean operator ‘AND’ as explained in the method section.

Appendix B

List of items of the quality assessment tool AXIS

Q1. Was the sample size justified?

Q2. Was the selection process likely to select subjects/participants that were representative of the target/reference population under investigation?

Q3. Were measures undertaken to address and categorize non-responders? Does the response rate raise concerns about non-response bias?

Q4. Were the aims/objective of the study clear?

Q5. Were the authors' discussions and conclusions justified by the results?

Q6. Were the methods (including statistical methods) sufficiently described to enable them to be repeated?

Q7. Were the risk factors and outcome variables measured correctly using instruments/measurements that had been trialed, piloted or published previously?

Q8. Were there any funding sources or conflicts of interest that may affect the author's interpretation of the results?