scientific reports

Check for updates

OPEN

Relations between suggestibility, working memory and response inhibition in middle childhood

Alessandro Santirocchi[1,3]✉, Pietro Spataro[2,3]✉, Maria Chiara Pesola[1], Aicia Naser[1], Vincenzo Cestari[1] & Clelia Rossi-Arnaud[1]

Individual difference in children's interrogative suggestibility should be theoretically associated with working memory and response inhibition skills. Yet, previous studies have reached mixed conclusions regarding the role of cognitive factors. In the present research, we re-examined this question in a sample of 203 children attending primary or middle school. They were administered the Gudjonsson Suggestibility Scale, a verbal span task (to measure the capacity of the phonological loop), and two inhibitory tasks (the Stroop Color and Word Test and the Global–Local Task). In line with our predictions, children having higher span scores were less suggestible than children having lower span scores (although the effect did not survive after controlling for class differences). On the other hand, the associations with inhibitory measures were null or even in contrast with our expectations, showing lower suggestibility indices in children who were less efficient in inhibiting the elaboration of the semantic meaning of words in the Stroop test. We conclude that individual differences in working memory and inhibitory skills have a limited impact on interrogative suggestibility, at least during middle childhood.

Keywords Interrogative suggestibility, Working memory, Span task, Response Inhibition, Stroop task, Global–local task

Starting from the last decades of the twentieth century, there has been a substantial growth in the number of studies addressing the role of individual differences in predicting children's suggestibility[1]. Before reviewing these studies, it is important to clarify that different experimental paradigms have been used to operationalize the concept of suggestibility and each of them can be associated with different psychological mechanisms. Bruck and Melnyk (2004)[1], for example, drew a differentiation among four components of suggestibility: *interrogative suggestibility* (defined as "the extent to which, within a closed social interaction, people come to accept messages communicated during formal questioning, as the result of which their subsequent behavioural response is affected"[2,3]), *misinformation effects* (defined as the incorporation of false information, provided from external sources or confabulated, into later reports about a specific event[4,5]), *source misattribution* (defined as the inability to determine whether recalled details about an event actually occurred or were simply imagined or suggested by others[6]), and *false event creation* (defined as the implantation of memories about events that never occurred through suggestive techniques[7]). The present study was specifically focused on the examination of the roles of working memory (WM) capacity and response inhibition in modulating children's interrogative suggestibility.

To this purpose, we used the *Gudjonsson Suggestibility Scale*[8], in which participants are firstly required to listen to a short story, whose content may be forensic (a robbery: GSS-1) or non-forensic (a couple saving a boy from an accident: GSS-2). Their memory is later assessed via two free-recall tasks, administered immediately and after a 50-min delay, and two cued-recall tasks in which they must answer a series of misleading questions, before or after receiving a negative feedback about their performance. Previous studies investigating the factorial structure of the GSS have clearly demonstrated that this scale assesses two types of suggestibility – namely, the tendency to give in to misleading questions (*Yield*) and the tendency to change responses after the negative feedback (*Shift*)[9,10]. Furthermore, clear developmental trends have been demonstrated, indicating that primary- and middle-school children are less suggestible than preschoolers[11,12]. Gudjonsson, Vagni, Maiorano and Pajardi (2016)[13], for instance, found that yield and shift scores decreased significantly in children aged 10–12

[1]Department of Psychology, Sapienza University, Via dei Marsi 78, 00185 Rome, Italy. [2]Faculty of Society and Communication Sciences, University of the System of the Italian Chambers of Commerce, Rome, Italy. [3]Alessandro Santirocchi and Pietro Spataro contributed equally to this work. ✉email: alessandro.santirocchi@uniroma1.it; pietro.spataro@unimercatorum.it

and 13–16 years, compared to children aged 7–9 years. Taken together, these properties make the GSS an ideal instrument to investigate interrogative suggestibility in middle childhood.

Gudjonsson and Clark (1986)[2] proposed the leading theoretical framework to explain how vulnerable individuals come to be suggestible during an interview. The model assumes that naïve participants start the interview with a general cognitive set (or mindset), which influences the way in which they respond to the novel interview situation, as well as to the presence of the interviewer, and may determine their behavioural response to the GSS questions – that is, whether they resist or yield to the suggestive questions. It is further proposed that several factors modulate individual differences in suggestibility, including the uncertainty about the correct answers to the questions, the expectations of success (participants may feel that the interviewer expects them to know the correct answer), the use of negative feedback during the interview, and the establishment of a good trust/rapport between the interviewer and the participant[14]. Although the Gudjonsson and Clarke model provides a relatively comprehensive explanation of the development of interrogative suggestibility during the GSS procedure, the cognitive mechanisms underlying performance in this task are less well understood and deserve further empirical research.

More specifically, individual differences in memory skills are central in determining the degree of uncertainty about the correct answers to suggestive questions. In fact, it has been repeatedly demonstrated that suggestibility correlates negatively and significantly with memory recall[3,13]. In other words, the poorer the participant's recall of the original story, the more likely he/she is to yield to suggestive questions. Along the same line, it has been proposed that a high WM capacity may be helpful during the encoding of the original story, since it facilitates the storage of larger units of meaningful information, thereby increasing the comprehension of the overall episode and improving long-term memory[3,15]. Other researchers have likewise contended that children with higher WM capacity may be better able to understand questions, to maintain the suggested information in short-term memory while trying to retrieve the correct answer from their own long-term memory, and to accurately compare self-generated and suggested answers[16].

In the present study, WM was measured with a verbal span task. According to the classical model proposed by Baddeley and Hitch (1974)[17] and later amended by Baddeley (2000)[18], WM comprises an attentional component called 'central executive' and two separate slave systems called 'phonological loop' (involved in the processing of auditory–verbal information) and 'visuospatial sketchpad' (involved in the processing of visual–spatial information). In addition, a fourth component, called 'episodic buffer', allows the temporary storage of information held in a multimodal code. We chose to use verbal span measures for two reasons. First, because in the GSS children are asked to listen to a brief story and auditory information is thought to be processed by the phonological loop[19]. Second, because there is empirical evidence indicating that children older than 7 years (i.e., the target of the present study) use primarily the phonological loop for the short-term remembering of both verbal and visual information, whereas children before the age of 7 use preferentially the visuospatial sketchpad[20–22].

Previous studies have sought to determine the relation between WM capacity and suggestibility, but the results were not homogenous. On the one hand, the review performed by Bruck and Melnyk (2004)[1] concluded that children's suggestibility was not related to memory competence. These authors were able to locate eight studies assessing the relationships between traditional measures of memory (e.g. span, paired associates learning, memory for digits, sentences or pictures) and suggestibility. The results showed that only five of the resulting 65 correlation coefficients were significant; furthermore, these correlations were only obtained in three studies. On the other hand, research focusing more specifically on phonological span measures provided more positive evidence in support of a negative association with suggestibility[16,23–26]. For example, Lee (2004)[26] found that, in a sample of 65 children and adolescents, performance in the backward digit span was negatively associated with the GSS yield and shift scores, although only the latter correlation reached the significance level. Bottoms et al. (2007)[16] investigated 6- and 7-year-olds' memory for games and activities performed with a confederate babysitter. They concluded that children with lower WM capacity, as measured with a modified counting span task, were more suggestible to misleading questions, but only in a non-supportive condition in which the interviewer was intimidating (no correlation emerged in the standard supportive condition). Finally, Caprin et al. (2016)[23] showed that the overall digit span performance of a sample of 372 children from 6 to 11 years (including both forward and backward span measures) was negatively and significantly associated with yield and shift scores. However, these authors used a self-devised instrument to assess suggestibility (rather than the GSS) and no significant correlations emerged with a different measure of auditory WM.

Besides WM capacity, individual differences in inhibitory control might likewise influence the GSS performance. This is because, from a theoretical perspective, resistance to misleading questions should rely on the children's ability to keep track of the original events while simultaneously inhibiting misleading information. That is, memories of the original event should be maintained in WM along with the suggested information regarding incorrect details, which should be subsequently inhibited[1]. In agreement, Roebers and Schneider (2005)[15] contended that children with well-developed executive skills should be better able to "first stop and think" – i.e., to inhibit the dominant responses provided by simply agreeing with the interviewer's questions, compare the suggested details with the content of their own memory and then give a reasoned response. To test the role of response inhibition in the modulation of interrogative suggestibility, in the present study children were asked to perform the Stroop Color and Word Test (SCWT) and the Global–Local Task (GLT). The SCWT requires participants to focus on one dimension of the stimulus, while simultaneously ignoring another dimension[27]. They are in fact presented with words displayed in a color that may be equal or different from the one that they actually designate (e.g., the word *red* shown in red or blue font). The typical outcome is that naming aloud the color that a word is printed in takes longer when the word denotes a different color (incongruent trials) than when it denotes the same color (congruent trials). Similarly, the GLT makes use of hierarchical compound stimuli, such as global shapes (or letters) composed of smaller local shapes that can be either congruent or incongruent with the global stimuli (e.g., a global circle composed of smaller circles or squares)[28–30]. Participants are required to identify either the global or local shape. The usual finding is that the identification of the local (or global) shape is slower and less accurate when the global (or local) shape is incongruent, than when it is congruent, although the magnitude of the global-to-local interference is almost always greater than that of local-to-global interference[31]. The simultaneous use of verbal and visual inhibitory tasks was justified by the fact that we were interested in determining whether children's suggestibility was dependent on a general ability of response inhibition, irrespectively of the nature of task materials (verbal or visual).

The literature assessing the relation between inhibitory control and suggestibility in children has not reached unequivocal conclusions. Alexander et al. (2002)[32] used the Day/Night task in a study in which 51 children between 3 and 7 years were interviewed about an inoculation after a delay of two weekends from the original event. Performance in the Day/Night task was inversely related to total incorrect units provided in response to free-recall prompts, as well as to omission errors made to misleading yes/no questions. Thus, children with poorer inhibitory skills exhibited higher levels of suggestibility, even when controlling for age[33,34]. However, later studies failed to confirm these findings[15,35]. In the study by Roberts and Powell (2005)[35], for example, 125 children aged 5 to 7 years participated in a staged event, were suggestively interviewed and were later given a recognition test. Of the four interference measures used in this research, only retroactive inhibition exhibited a significant relation with children's suggestibility (i.e., children with higher-than-average verbal retroactive inhibition skills were more resistant to suggestions). In contrast, no associations were observed with proactive inhibition and two conflict tasks (day/night and tapping). Similarly, Roebers and Schneider (2005)[15] examined recognition errors due to prior misinformation and resistance to false suggestions in 4-year-old children. Executive function, assessed through a composite score of five different inhibitory measures (including go/no-go, movement suppression, conflict hand movements, and Stroop tasks), showed no significant relation with suggestibility and misinformation scores.

To recap, the present study aimed at examining the relations between suggestibility (investigated with the GSS-2), WM capacity (investigated with a verbal span task) and response inhibition (investigated with SCWT and GLT) in primary- and middle-school children ranging in age from 8 to 14 years. We focused on these ages because most of the studies available in literature included children younger than 7 years[15,32,35]. As suggested by Bruck and Melnyk (2004, p.968–969)[1], this problem raises the possibility that WM and executive function skills were not significantly associated with suggestibility because of the relative immaturity of the cognitive systems of young children. If this were the case, then the associations might be stronger when assessed in older children. Several predictions could be advanced based on the evidence briefly reviewed above:

(i) First, we expected to observe classical age effects in the GSS2, as well as in the span and response inhibition tasks. That is, older children attending middle school should recall more correct story elements in the immediate and delayed free-recall tasks and should be less suggestible than younger children attending primary school[13]. Furthermore, middle-school children should have higher span scores and a better performance in the incongruent conditions of the GLT and SCWT tasks (in terms of accuracy and RTs), compared to primary-school children, suggesting that the former group have higher WM capacity and are more efficient in inhibiting conflictual responses.

(ii) Second, we expected children's span scores to be positively associated with the number of correct story elements reported in the GSS2, but negatively associated with the number of false assents to misleading questions[23,26];

(iii) Third, for response inhibition, we expected that children having higher interference ability should be less likely to yield to misleading questions and less suggestible[32–34,36].

Method
Participants
Data were collected in different waves, due to the forced suspension of face-to-face activities following the COVID-19 pandemic. The entire sample comprised 203 children, of which 106 boys and 97 girls. The overall mean age was 10.66 years (SD = 1.98). They were divided in two subsamples, according to whether they attended primary school (N = 118, of which 60 girls and 58 boys; age: M = 8.94, SD = 0.99) or middle school (N = 85, of which 37 girls and 48 boys; age: M = 12.45, SD = 0.84).

All the children performed the GSS2. However, due to small procedural differences across successive waves of data collection, the two span tasks were administered to 145 children (78 males and 67 females; 60 primary-school and 85 middle-school children; age: M = 11.40, SD = 1.93), whereas the GLT and SCWT were administered to 172 children (90 males and 83 females; 88 primary-school and 85 middle-school children; age: M = 10.66, SD = 1.93). The two subsamples had 115 children in common.

Materials
Gudjonsson suggestibility scale
The Italian version of the Gudjonsson Suggestibility Scale 2 (GSS2)[9] was used. It consists of a short paragraph describing a non-criminal story, divided into 40 items, and a set of 20 questions, of which five were true questions concerning details and events that occurred in the story (control questions), while the other 15 were misleading questions concerning details and events that were not presented in the original story. The procedure followed the instructions described in the manual[37]. At the beginning, each child was read the story with the instruction to listen carefully ("*I would like you to listen to a short story. Listen carefully because when I finish I would like you to recall everything you can remember*"). After an immediate free recall of the story, there was a delay of 50 min, during which children performed the three WM tasks described below. At the end of this period, they were asked to provide a delayed recall of the story. For both the immediate and delayed free recall tasks, no time

limits were set, and the instructions were as follows: "*Now please recall everything you remember from the story*". Next, children were asked to answer the 20 questions a first time ("*I'm going to ask you some questions about the story. Try to be as accurate as possible*"). Finally, regardless of performance, they were given a negative feedback stating that many of their responses were incorrect and that they had the possibility to answer the questions for a second time ("*You have made a number of errors. It is therefore necessary to go through the questions once more, and this time try to be as accurate as possible*"). The 20 questions were then repeated, and the answers recorded on an appropriate response sheet.

Verbal span tasks
For the verbal span tasks, the stimuli were taken from the PROMEA (*Prove di Memoria e Apprendimento per l'Età Evolutiva*)[38], a standardized test recommended for children from age 5 and older. Specifically, we used the span tasks based on high- and low-frequency quadrisyllabic words, which were judged to be appropriate for children attending primary- and middle-school classes. For example, in Italian, '*bicicletta*' (bicycle) is a high-frequency word, while '*cartolina*' (postcard) is a low-frequency word. Span levels from 2 to 8 words are provided in the manual and each level includes a maximum of 5 different trials. The two span tasks were administered according to the PROMEA instructions[38]. For both high- and low-frequency words, the task began with level 2. Children were read each sequence with the instruction to repeat the words in the same order in which they were pronounced by the experimenter ("*Now I'm going to speak some words. Try to repeat them in the same order in which I pronounce them. Are you ready?*"). If the child repeated correctly at least three (out of five) sequences, the experimenter proceeded to the next level. The task was discontinued if the child made errors in three or more sequences: in this case, the assigned span corresponded to the level immediately preceding the one in which they failed.

Inhibitory tasks
In the GLT, the stimuli were black outlines of small squares or circles measuring approximately 0.7 cm^2. These small shapes were used to construct large squares, circles, or Xs (for the neutral condition) that measured about 6.5 cm^2 and were comparable to those used by Cottini et al. (2015)[29] and Bialystok (2010)[28]. All the stimuli were presented on a white background. The procedure (borrowed from Cottini et al., 2015[29]) consisted of two different conditions (global and local) including a total of 72 trials: each condition comprised three blocks of 12 congruent trials, 12 incongruent trials, and 12 neutral trials. All trials began with a fixation cross at the center of the screen for 500 ms, followed by a shape stimulus which remained on the screen until the child's response and a blank screen pause for 2000 ms. In the congruent trials, the global and local shapes matched (e.g., a global circle formed by small circles), whereas in the incongruent trials the two levels conflicted (e.g., a global circle formed by small squares, or vice versa). In the neutral trials, the unattended dimension included a series of Xs (e.g., a global circle composed of small Xs or a global X composed of small circles). Children were instructed to identify either the global or the local shape. Responses were given by pressing one of two large buttons in a response pad (RB-844; Cedrus Corporation). To make the buttons easily distinguishable, they were covered by white patches depicting either a circle or a square. A 12-trial practice block preceded each condition. The order of the global and local conditions was counterbalanced across participants; in addition, the order of the congruent, incongruent, and neutral trials in the three blocks was randomized anew for each participant. Response RTs were collected by the software Superlab 4.0.

In the SCWT, the stimuli were five color words (blue, yellow, green, red, and purple) and a string of seven asterisks. Both the color names and the asterisks were written in bold, uppercase letters in the Arial font (size: 72 points) and could be presented in each of the above-mentioned colors on a black background. Thus, for example, the word 'blue' could be presented in blue (congruent condition) or in yellow, green, red, and purple (incongruent conditions). The procedure was modeled on that illustrated by Miyake et al. (2000)[39]. Specifically, the task included 60 trials with a string of seven asterisks printed in one of five colors (red, green, blue, yellow, and purple), 60 trials with a color word printed in a different color (e.g., 'blue' printed in red), and 20 trials with a color word printed in the same color (e.g., 'blue' in blue), with the different trial types being randomly mixed (i.e., non-blocked). The experimental phase was preceded by a short block of 24 practice trials (8 trials of each type, randomly mixed). Each trial began with a fixation cross at the center of the screen for 500 ms, followed by a colored letter or a colored string of asterisks which remained on the screen until the child's response and a blank screen pause for 2000 ms. The instructions were to identify the colors in which the stimuli were printed, by pressing one of five buttons in a response pad (RB-844; Cedrus Corporation). To make the buttons easily distinguishable, they were covered by colored patches. Response RTs were collected by the software Superlab 4.0.

Procedure
Children participated in a single session lasting about 75 min, which was conducted in a quiet room of their schools. In this session, they were administered the GSS2 and the three cognitive tasks by a trained experimenter. As stated above, the cognitive tasks were performed during the 50-min interval between the immediate and delayed recall phases of the GSS and their order was fixed for all children (the two span tasks were firstly administered, followed by the GLT and SCWT tasks). The entire procedure was reviewed and approved by the Ethical Committee of Sapienza University (Protocol N. 0,001,293). Informed consent for each child was obtained from parents. All methods were performed in accordance with the relevant guidelines and regulations.

Data coding
For the GSS2, the total number of correct and confabulated elements reported by each child were coded and tallied[40]. The GSS manual provides detailed examples about how correct elements in the free recall tasks should be coded. It is recommended that a score of 1 should be assigned whenever the general concept is clearly

reported, even if the exact wording is different from that used in the story. When the elements are made up of two different details, the manual recommends assigning 0.5 point if the child retrieves only one correct detail and 1 if he/she retrieves both details. At the end, this procedure resulted in children being assigned a single score (0, 0.5 or 1) for each of the 40 key elements mentioned in the story.

Detailed instructions are likewise provided about how to code confabulated elements, which include distorted and fabricated details. Distorted elements were given 1 point when the concept originally reported in the story underwent substantial changes. On the other hand, fabricated details refer to elements that were never cited in the story.

The same process was applied to the coding of the "Yield" ("yes") responses given by children to the 15 misleading questions included in the cued-recall tasks. Following the GSS manual, responses such as "it's possible", "I think yes", "maybe" were considered as yield responses (i.e., as "yes" responses) and were awarded 1 point. In contrast, responses such as "I don't remember", "this was not mentioned in the story", "I don't know", "I'm not sure" were not considered as yield responses and were therefore assigned 0 points. This procedure resulted in children being assigned a score of 0 or 1 for each of the 15 leading questions. Finally, as concerns shifting scores, the change in the responses provided before and after the negative feedback had to be clear-cut to be taken as positive evidence. For example, changes from "yes" to "no", from "don't know" to "yes", or from "don't remember" to "yes" were considered as substantial changes and were therefore awarded 1 point. In contrast, changes from "don't remember" to "no", from "maybe" to "no", or from "it's possible" to "yes" were not sufficiently strong to be regarded as response shifts and were assigned 0 points. Again, the outcome of this procedure was a response sheet in which children were assigned either 0 or 1 points to each of the 20 questions asked to them.

For the GLT and SCWT tasks, RTs for correct responses and accuracy scores were recorded in each condition. Interference indices were computed from these raw data[41,42]. Specifically, for the GLT, the *local-to-global interference index* represented the difference between the RTs for incongruent global trials and the RTs for neutral global trials: thus, it reflected the children's tendency to incorrectly report the identity of the smaller shapes when instructed to identify the larger shapes[42]. Likewise, the *global-to-local interference index* represented the difference between the RTs for incongruent local trials and the RTs for neutral local trials: thus, it reflected the children's tendency to incorrectly report the identity of the larger shapes when instructed to identify the smaller shapes. For the SCWT, the interference index reflected the difference between the RTs for incongruent trials and the RTs for neutral trials. Similar indices were also computed for accuracy. In this case the *local-to-global interference index* represented the difference between accuracy for neutral global trials and accuracy for incongruent global trials, whereas the *global-to-local interference index* represented the difference between accuracy for neutral local trials and accuracy for incongruent local trials. For the SCWT, the interference index was simply given by the difference in accuracy between neutral and incongruent trials. In all cases, the rationale underlying the computation of these indices is that smaller interference scores should reflect better inhibitory skills, because strong inhibition of the irrelevant stimulus feature should reduce the costs (in terms of RTs or accuracy) arising from the conflict in incongruent trials[43].

Results

GSS2

The mean number of correct story elements remembered by children in the immediate and delayed recall tasks (see Fig. 1) were submitted to a mixed ANOVA, considering Time (immediate vs. delayed recall) as the within-subject factor and Group (primary vs. middle school) as the between-subject factor. Results showed a significant main effect of Time [$F(1, 201) = 39.25$, $p < 0.001$, $\eta^2 = 0.16$], indicating that children recalled more correct elements in the immediate ($M = 14.48$) than in delayed ($M = 13.33$) task, and a significant main effect of Group [$F(1, 201) = 20.55$, $p < 0.001$, $\eta^2 = 0.09$], indicating that middle-school children ($M = 15.57$) recalled more correct elements than primary-school children ($M = 12.23$). The Time x Group interaction was not significant [$F(1, 201) = 1.62$, $p = 0.21$, $\eta^2 = 0.01$].

The mean number of confabulated (fabricated plus distorted) elements reported in the free recall tasks (see Fig. 2) were also submitted to a mixed Time × Group ANOVA. Results revealed a significant main effect of Time [$F(1, 201) = 5.24$, $p = 0.023$, $\eta^2 = 0.03$], indicating that children recalled more confabulated elements in the

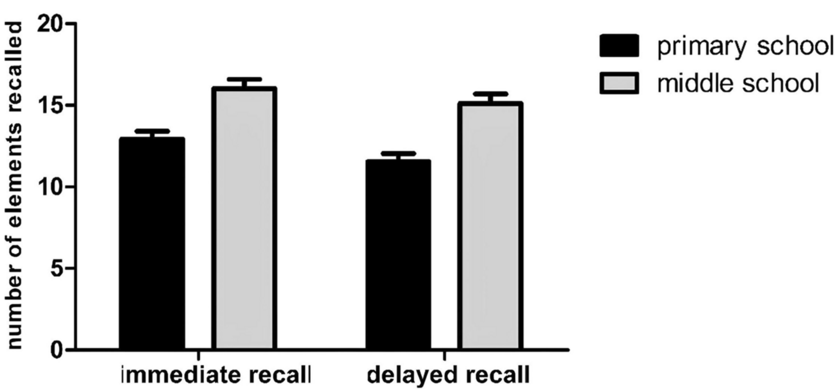

Fig. 1. Mean number of elements correctly recalled in the GSS2. Bars represent standard errors.

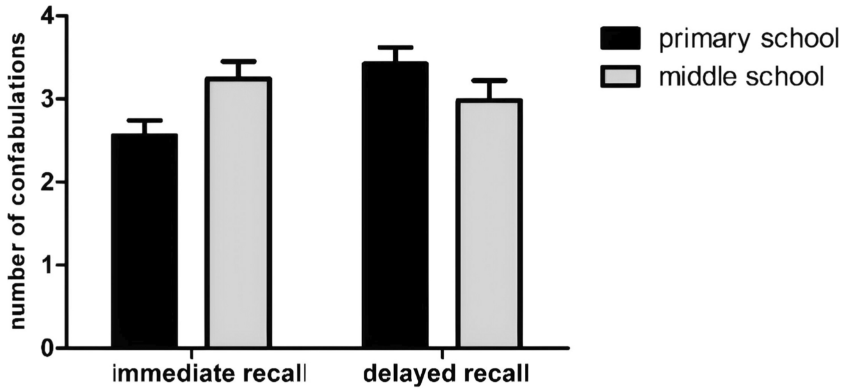

Fig. 2. Mean number of confabulations produced in the GSS2. Bars represent standard errors.

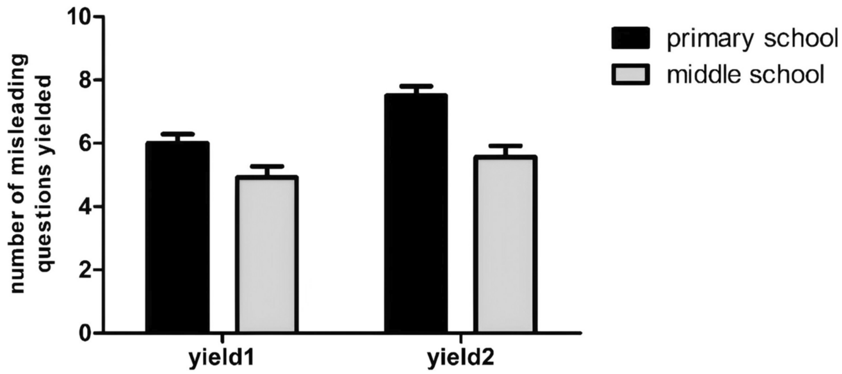

Fig. 3. Mean number of misleading questions to which children yielded in the GSS2. Bars represent standard errors.

delayed ($M = 3.20$) than in immediate ($M = 2.90$) task, and a significant interaction between Time and Group [$F(1, 201) = 18.28$, $p < 0.001$, $\eta^2 = 0.08$]. A follow-up analysis of simple effects indicated that the effect of Time was significant in primary-school children [$F(1, 201) = 25.74$, $p < 0.001$, $\eta^2 = 0.11$], but not in middle-school children [$F(1, 201) = 1.69$, $p = 0.19$]. If anything, the numerical trend was reversed for the latter group (i.e., they recalled more confabulated elements in the immediate than in the delayed task). The main effect of Group was not significant [$F(1, 201) = 0.20$, $p = 0.65$].

The mean number of yield responses (i.e., the mean number of misleading questions to which children provided an incorrect affirmative response; see Fig. 3) was analyzed through a mixed ANOVA, considering Feedback (before vs. after the negative feedback) as the within-subject factor and Group (primary vs. middle school) as the between-subject factor. Results showed a significant main effect of Feedback [$F(1, 201) = 48.52$, $p < 0.001$, $\eta^2 = 0.19$], indicating that yield responses increased after children received the negative feedback ($M = 6.53$ vs $M = 5.46$), and a significant main effect of Group [$F(1, 201) = 11.71$, $p = 0.001$, $\eta^2 = 0.06$], indicating that younger primary-school children ($M = 6.75$) provided more incorrect yield responses than older middle-school children ($M = 5.24$). In addition, the Feedback × Group interaction was also significant [$F(1, 201) = 7.95$, $p = 0.005$, $\eta^2 = 0.04$]. A follow-up analysis of simple effects indicated that the effect of Feedback was significant in both primary-school [$F(1, 201) = 57.18$, $p < 0.001$, $\eta^2 = 0.22$] and middle-school children [$F(1, 201) = 7.38$, $p = 0.007$, $\eta^2 = 0.04$], although it was stronger in the former group.

Lastly, shift scores and total suggestibility scores were analyzed with two *t*-tests for independent samples, considering Group (primary vs. middle school) as the between-subject factor. The results revealed that primary-school children ($M = 4.78$) were more likely to change their responses after receiving the negative feedback, as compared to middle-school children ($M = 3.59$) [$t(201) = 3.21$, $p = 0.002$, Cohen's $d = 0.46$]. In addition, primary-school children were more suggestible than middle-school children ($M = 10.79$ vs. $M = 8.53$) [$t(201) = 3.19$, $p = 0.002$, Cohen's $d = 0.45$].

Span tasks

A mixed ANOVA, considering Frequency (low vs. high-frequency) as the within-subject factor and Group (primary vs. middle school) as the between-subject factor, revealed a significant main effect of Frequency [$F(1, 143) = 7.10$, $p = 0.009$, $\eta^2 = 0.05$], indicating that span scores were slightly greater for high- than for low-frequency words ($M = 4.07$ vs $M = 3.89$), and a significant main effect of Group [$F(1, 143) = 34.28$, $p < 0.001$,

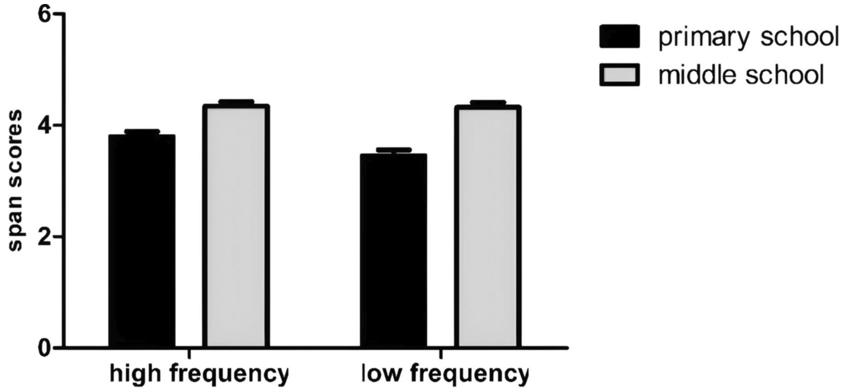

Fig. 4. Mean span scores in the PROMEA. Bars represent standard errors.

	Primary school		**Middle school**	
	Global	**Local**	**Global**	**Local**
Reaction times				
Congruent trials	1001 (343)	1064 (347)	734 (251)	734 (188)
Incongruent trials	1114 (457)	1184 (400)	775 (240)	807 (192)
Neutral trials	1009 (325)	1097 (338)	739 (205)	762 (191)
Interference	105 (266)	87 (227)	36 (164)	45 (104)
Accuracy				
Congruent trials	0.98 (0.04)	0.98 (0.03)	0.99 (0.02)	0.99 (0.02)
Incongruent trials	0.90 (0.16)	0.92 (0.17)	0.93 (0.11)	0.95 (0.09)
Neutral trials	0.96 (0.10)	0.98 (0.05)	0.98 (0.04)	0.99 (0.02)
Interference	0.06 (0.13)	0.06 (0.16)	0.05 (0.11)	0.04 (0.08)

Table 1. Reaction times, accuracy scores and interference indices in the GLT. Standard deviations are reported in parenthesis.

$\eta^2 = 0.19$], indicating that span scores were greater for middle-school than for primary-school children ($M = 4.33$ vs $M = 3.62$). In addition, the Frequency × Group interaction was also significant [$F(1, 143) = 6.21$, $p = 0.014$, $\eta^2 = 0.04$]. A follow-up analysis of simple effects indicated that the effect of Frequency was significant in primary-school children [$F(1, 143) = 11.34$, $p = 0.001$, $\eta^2 = 0.07$], but not in middle-school children [$F(1, 143) = 0.02$, $p = 0.89$]. Thus, only primary-school children exhibited a memory advantage for high-frequency words (see Fig. 4).

Global–local task

RTs in the global–local task (see Table 1) were submitted to a mixed ANOVA, considering Condition (global vs. local) and Type of Trial (congruent, incongruent and neutral) as the within-subject factors, and Group (primary vs. middle-school children) as the between-subject factor. The results revealed a significant main effect of Condition [$F(1, 169) = 18.15$, $p < 0.001$, $\eta^2 = 0.09$], indicating that RTs were faster in the global ($M = 896.61$ ms) than in the local condition ($M = 942.53$ ms), a significant main effect of Group [$F(1, 169) = 57.43$, $p < 0.001$, $\eta^2 = 0.25$], indicating that RTs were faster for middle-school ($M = 759.82$ ms) than for primary-school children ($M = 1079.31$ ms), and a significant main effect of Type of Trial [$F(2, 338) = 39.51$, $p < 0.001$, $\eta^2 = 0.19$]. For the latter effect, the post-hoc comparisons (with Bonferroni adjustment) indicated that RTs were slower in the incongruent ($M = 971.16$ ms) than in both the congruent ($M = 884.52$ ms, $p < 0.001$) and neutral conditions ($M = 903.02$ ms, $p < 0.001$); the congruent and neutral conditions did not differ between them ($p = 0.21$). In addition, the two-way interactions between Condition and Group [$F(1, 169) = 6.52$, $p = 0.011$, $\eta^2 = 0.04$] and between Type of Trial and Group [$F(2, 338) = 5.24$, $p = 0.006$, $\eta^2 = 0.03$] were also significant. A follow-up analysis of simple effects on the first interaction showed that the effect of Condition was significant in primary-school children [$F(1, 169) = 23.36$, $p < 0.001$, $\eta^2 = 0.12$], but not in middle-school children [$F(1, 169) = 1.44$, $p = 0.23$]: thus, only younger children exhibited a RT advantage in the global condition. A similar follow-up analysis on the second interaction showed that the effect of Type of Trial was significant for both primary-school [$F(2, 168) = 29.90$, $p < 0.001$, $\eta^2 = 0.26$] and middle-school children [$F(2, 168) = 7.08$, $p = 0.001$, $\eta^2 = 0.08$], although the size of the interference effect was more pronounced in the former group.

Accuracy scores were likewise submitted to a mixed ANOVA, considering Condition (global vs. local) and Type of Trial (congruent, incongruent and neutral) as the within-subject factors, and Group (primary vs. middle-school children) as the between-subject factor. The analysis found a significant main effect of Condition [F(1,

169) = 5.48, $p = 0.020$, $\eta^2 = 0.03$], indicating that accuracy was higher in the local ($M = 0.97$) than in the global condition ($M = 0.96$), a significant main effect of Group [$F(1, 169) = 4.39$, $p = 0.037$, $\eta^2 = 0.02$], indicating that accuracy was higher for middle-school ($M = 0.97$) than for primary-school children ($M = 0.95$), and a significant main effect of Type of Trial [$F(2, 338) = 57.53$, $p < 0.001$, $\eta^2 = 0.25$]. For the latter effect, the post-hoc comparisons (with Bonferroni adjustment) indicated that accuracy was lower in the incongruent ($M = 0.92$) than in both the congruent ($M = 0.99$, $p < 0.001$) and neutral conditions ($M = 0.98$, $p < 0.001$); although small in size, the difference between the congruent and neutral conditions was significant ($p = 0.027$). None of the two-way and three-way interactions reached the significance level [$F < 2.25$, $p > 0.11$, $\eta^2 < 0.03$].

In addition to these analyses, the *local-to-global* and *global-to-local interference indexes* were also computed, considering both RTs and accuracy scores (see Table 1). For the RT-based indexes, a mixed ANOVA, considering Interference Type (global-to-local vs. local-to-global) as the within-subject factor and Group (primary vs. middle-school children) as the between-subject factor revealed only a main effect of Group [$F(1, 169) = 5.93$, $p = 0.016$, $\eta^2 = 0.03$], confirming that interference scores were higher in primary-school ($M = 96.00$ ms) than in middle-school ($M = 40.25$ ms) children. Thus, as expected, primary-school children were less efficient in inhibiting the elaboration of the unattended perceptual features (compared to middle-school children). For the accuracy-based indexes, the same ANOVA as above showed no significant effects [$F < 1.51$, $p > 0.22$, $\eta^2 < 0.01$].

Stroop task

RTs in the Stroop task (see Table 2) were submitted to a mixed ANOVA, considering Type of Trial (congruent, incongruent and neutral) as the within-subject factor and Group (primary vs. middle-school children) as the between-subject factor. The results revealed a significant main effect of Group [$F(1, 169) = 51.62$, $p < 0.001$, $\eta^2 = 0.23$], indicating that RTs were faster for middle ($M = 964.55$ ms) than for primary-school children ($M = 1189.71$ ms), and a significant main effect of Type of Trial [$F(2, 338) = 76.54$, $p < 0.001$, $\eta^2 = 0.31$]. For the latter effect, the post-hoc comparisons (with Bonferroni adjustment) indicated that RTs were slower in the incongruent ($M = 1165.98$ ms) than in both the congruent ($M = 1035.25$ ms, $p < 0.001$) and neutral trials ($M = 1030.16$ ms, $p < 0.001$); the congruent and neutral trials did not differ between them ($p = 1.00$). In addition, the Type of Trial × Group interaction was also significant [$F(2, 338) = 5.97$, $p = 0.003$, $\eta^2 = 0.03$]. However, a follow-up analysis of simple effects showed that the effect of Type of Trial was significant in both primary-school [$F(2, 168) = 46.38$, $p < 0.001$, $\eta^2 = 0.36$] and middle-school children [$F(2, 168) = 20.77$, $p < 0.001$, $\eta^2 = 0.19$], although the size of the interference effect was again greater in the younger group.

For accuracy scores, the same ANOVA found a significant main effect of Type of Trial [$F(2, 342) = 39.75$, $p < 0.001$, $\eta^2 = 0.19$], confirming that accuracy was lower in the incongruent ($M = 0.91$) than in both the congruent ($M = 0.98$ ms, $p < 0.001$) and neutral trials ($M = 0.97$, $p < 0.001$). The main effect of Group was marginal [$F(1, 171) = 3.30$, $p = 0.071$, $\eta^2 = 0.02$], whereas the Trial × Group interaction did not reach the significance level [$F(2, 342) = 1.17$, $p = 0.31$, $\eta^2 = 0.01$].

As for the global–local task, we also computed an *interference index*, reflecting the difference between the RTs for incongruent trials and the RTs for neutral trials. A *t*-test for independent samples revealed that the interference index was higher for primary-school ($M = 178.35$ ms) than for middle-school children ($M = 92.01$ ms) [$t(170) = 2.75$, $p = 0.007$], confirming the fact that the latter group had a better inhibitory efficiency. Like for the global–local task, no effect emerged when the interference index was computed on accuracy [$t(171) = -0.59$, $p = 0.55$].

Relationships between GSS2 measures and span scores

Pearson's correlations between the GSS2 measures and span scores are illustrated in Table 3. As can be noted, span scores for high-frequency and low-frequency words were positively and significantly associated with both immediate and delayed recall. In addition, span scores were negatively associated with yield1 and yield2, as well as with total suggestibility (only for high-frequency words). Thus, children having a higher verbal span recalled more correct elements from the GSS2 story, were less likely to provide incorrect 'yes' responses to misleading questions and were less suggestible.

	Primary school	Middle school
Reaction times		
Congruent trials	1140 (223)	923 (194)
Incongruent trials	1298 (333)	1027 (221)
Neutral trials	1120 (180)	934 (145)
Interference	178 (250)	93 (145)
Accuracy		
Congruent trials	0.98 (0.03)	0.98 (0.03)
Incongruent trials	0.92 (0.12)	0.89 (0.19)
Neutral trials	0.98 (0.02)	0.97 (0.02)
Interference	0.06 (0.12)	0.08 (0.19)

Table 2. Reaction times, accuracy scores and interference indices in the SCWT. Standard deviations are reported in parenthesis.

	Rec_Imm	Conf_Imm	Rec_Del	Conf_Del	Yield1	Yield2	Shift	Sugg_tot
Span_hf	0.35**	0.16†	0.31**	0.16*	−0.16*	−0.25**	−0.14†	−0.18*
Span_lf	0.32**	0.14	0.27**	−0.01	−0.15†	−0.24**	−0.08	−0.13
Glo_interf_RT	−0.01	−0.12	0.04	−0.06	−0.09	−0.08	−0.06	−0.08
Loc_interf_RT	0.03	−0.02	0.09	0.07	−0.02	0.06	−0.01	−0.02
Stroop_interf_RT	0.08	−0.09	0.14†	−0.17*	−0.12	−0.16*	−0.06	−0.11
Glo_interf_acc	−0.08	0.09	−0.09	0.14†	0.01	−0.07	−0.07	−0.03
Loc_interf_acc	−0.11	−0.01	−0.12	0.06	−0.02	0.06	0.04	0.01
Stroop_interf_acc	−0.08	−0.10	−0.14†	−0.09	0.06	0.05	0.05	0.07

Table 3. Pearson's correlations between the GSS2 indices and cognitive measures.

Predicted Measure		Predictors	*β*	*t*-test	ΔR^2	*F* Change
Immediate recall	Step 1	Class	0.35	$t=4.26$**	0.19	$F=34.16$**
	Step 2	Span_hf	0.21	$t=2.27$*	0.05	$F=4.50$**
		Span_lf	0.04	$t=0.43$		
Immediate	Step 1	Class	0.27	$t=3.01$**	0.08	$F=12.65$**
confabulation	Step 2	Span_hf	0.08	$t=0.81$	0.00	$F=0.34$
		Span_lf	−0.03	$t=-0.30$		
Delayed recall	Step 1	Class	0.37	$t=4.50$**	0.19	$F=33.88$**
	Step 2	Span_hf	0.18	$t=1.91^{\dagger}$	0.03	$F=2.64^{\dagger}$
		Span_lf	0.01	$t=0.06$		
Delayed	Step 1	Class	−0.01	$t=-0.15$	0.00	$F=0.00$
confabulation	Step 2	Span_hf	0.27	$t=2.59$**	0.05	$F=3.36$*
		Span_lf	−0.16	$t=-1.52$		
Yield1	Step 1	Class	−0.10	$t=-1.12$	0.03	$F=3.63^{\dagger}$
	Step 2	Span_hf	−0.09	$t=-0.94$	0.01	$F=1.04$
		Span_lf	−0.04	$t=-0.42$		
Yield2	Step 1	Class	−0.22	$t=-2.47$**	0.08	$F=13.41$**
	Step 2	Span_hf	−0.14	$t=-1.44$	0.03	$F=2.22$
		Span_lf	−0.06	$t=-0.53$		
Shift	Step 1	Class	−0.20	$t=-2.19$*	0.04	$F=6.40$**
	Step 2	Span_hf	−0.13	$t=-1.21$	0.01	$F=0.74$
		Span_lf	0.09	$t=0.78$		
Total	Step 1	Class	−0.17	$t=-1.88^{\dagger}$	0.04	$F=6.52$**
Suggestibility	Step 2	Span_hf	−0.13	$t=-1.28$	0.01	$F=1.00$
		Span_lf	0.01	$t=0.16$		

Table 4. Hierarchical regressions GSS2 measures from class and span scores. $^{\dagger}0.05 < p < 0.10$. $^{*}p \leq 0.05$. $^{**}p \leq 0.01$. $^{***}p \leq 0.001$.

To further investigate the role of WM capacity, we performed a series of hierarchical regressions, entering children's class in the first step and span scores (for both high and low frequency words) in the second step. The results are illustrated in Table 4. It can be seen that high-frequency span scores positively predicted immediate recall, delayed recall ($p=0.058$) and delayed confabulation. Thus, having higher span scores for high-frequency words predicted a greater recall in the immediate and delayed tests, but also a higher confabulation rate in the delayed test.

Relationships between GSS2 measures and interference indices

For the GLT and SCWT tasks, correlations were computed with the RT-based and accuracy-based interference indexes, because these were the measures that best reflected inhibitory efficiency[42].

Overall, only 2 correlations (out of 48: 4.1%) reached the significance level (see Table 3). Specifically, we found that children who had higher RT-based interference scores in the SCWT produced less confabulations in the delayed test of the GSS2 and were less likely to yield to misleading questions after receiving negative feedback.

As for span measures, we performed a series of hierarchical regressions, entering children's class in the first step and the RT-based, or accuracy-based, interference scores in the second step. The analyses, reported in Tables 5 and 6, showed that having higher RT-based interference scores in the SCWT predicted a higher recall of correct elements in the delayed test, a lower production of confabulated items in the delayed recall test, a lower tendency to yield to misleading questions before and after the negative feedback, and a lower total suggestibility.

Predicted Measure		Predictors	β	t-test	ΔR^2	F Change
Immediate recall	Step 1	Class	0.29**	t = 3.82**	0.06	F = 11.94**
	Step 2	glo_interf_RT	0.02	t = 0.22	0.02	F = 3.95*
		loc_interf_RT	0.05	t = 0.71		
		Stroop_interf_RT	0.13†	t = 1.80†		
Immediate	Step 1	Class	0.11	t = 1.36	0.02	F = 3.44†
confabulation	Step 2	glo_interf_RT	−0.11	t = −1.38	0.01	F = 1.57
		loc_interf_RT	0.01	t = −0.06		
		Stroop_interf_RT	−0.07	t = −0.94		
Delayed recall	Step 1	Class	0.34**	t = 4.50**	0.07	F = 13.55**
	Step 2	glo_interf_RT	0.06	t = 0.91	0.06	F = 6.30**
		loc_interf_RT	0.12	t = 1.63		
		Stroop_interf_RT	0.19**	t = 2.67**		
Delayed	Step 1	Class	−0.19*	t = −2.37*	0.02	F = 2.88†
confabulation	Step 2	glo_interf_RT	−0.09	t = −1.28	0.06	F = 3.29**
		loc_interf_RT	0.06	t = 0.89		
		Stroop_interf_RT	−0.22**	t = −2.79**		
Yield1	Step 1	Class	−0.19*	t = −2.46*	0.02	F = 3.26†
	Step 2	glo_interf_RT	−0.13†	t = −1.66†	0.04	F = 2.66*
		loc_interf_RT	0.00	t = 0.04		
		Stroop_interf_RT	−0.16*	t = −2.09*		
Yield2	Step 1	Class	−0.29**	t = −3.81**	0.05	F = 9.47**
	Step 2	glo_interf_RT	−0.11	t = −1.56	0.06	F = 5.16**
		loc_interf_RT	0.04	t = 0.61		
		Stroop_interf_RT	−0.21**	t = −2.82**		
Shift	Step 1	Class	−0.27**	t = −3.53**	0.05	F = 9.54**
	Step 2	glo_interf_RT	−0.10	t = −1.36	0.03	F = 3.49**
		loc_interf_RT	−0.01	t = −0.10		
		Stroop_interf_RT	−0.12	t = −1.53		
Total	Step 1	Class	−0.27**	t = −3.43**	0.04	F = 7.63**
Suggestibility	Step 2	glo_interf_RT	−0.14†	t = −1.79†	0.04	F = 2.54†
		loc_interf_RT	0.00	t = 0.00		
		Stroop_interf_RT	−0.16*	t = −2.14*		

Table 5. Hierarchical regressions predicting GSS2 measures from class and interference scores in the GLT and SCWT. glo = global-to-local, loc = local-to-global, interf = interference. †0.05 < p < 0.10. *$p \leq 0.05$. **$p \leq 0.01$.

In addition, having higher accuracy-based interference scores in the global condition of the GLT predicted a higher production of confabulated items in the delayed recall test.

Discussion

We examined developmental differences in suggestibility, WM capacity, and response inhibition in primary and middle-school children between 8 and 14 years of age, as well as the associations between these measures. The first important result was that typical age effects were observed in all cases. For the GSS2, older children attending middle school remembered more correct story elements, provided less assents to misleading questions, were less likely to change their responses after receiving the negative feedback and were less suggestible than younger children attending primary school. These results are in line with those available in the literature. Gudjonsson et al. (2016)[13], for instance, conducted a study in which children aged 7–9, 10–12 and 13–16 years were administered the GSS2. Results showed that immediate and delayed recall performance increased linearly across the three age bands, whereas yield2, shift and total suggestibility scores decreased from younger to older children.

For the cognitive tasks, older children had higher verbal span scores, and were faster and more accurate in both the GLT and SCWT (compared to younger children). Classical interference effects were obtained in the latter tasks, with RTs being higher and accuracy being lower in the incongruent trials, as compared to neutral and congruent trials. Moreover, the analysis of the interactions confirmed that middle-school children were more efficient than primary-school children in inhibiting unwanted responses during incongruent trials. For the GLT, similar results (i.e., significant decreases in RTs and parallel increases in accuracy) have been previously reported by Scherf et al. (2009)[42], when comparing children (8–13 years), adolescents (14–17 years) and adults (20–30 years), and by Huizinga et al. (2010)[41], when comparing 7-year-olds, 11-year-olds and young adults. Interestingly, we found that children exhibited the classical pattern of global precedence in RTs, since they were faster in identifying global than local shapes. In line with the orthogenetic principle, which states that global processing reaches adult-like levels earlier in life[41], this global precedence effect was stronger in primary-school

Predicted Measure		Predictors	β	t-test	ΔR^2	F Change
Immediate recall	Step 1	Class	0.24**	t = 3.32**	0.06	F = 11.79**
	Step 2	glo_interf_acc	− 0.07	t = − 1.00	0.03	F = 3.98**
		loc_interf_acc	− 0.08	t = − 1.16		
		Stroop_interf_acc	− 0.07	t = − 1.03		
Immediate	Step 1	Class	0.16*	t = 2.06*	0.02	F = 3.59†
confabulation	Step 2	glo_interf_acc	0.11	t = 1.56	0.03	F = 2.03†
		loc_interf_acc	0.02	t = 0.34		
		Stroop_interf_acc	− 0.12	t = − 1.62		
Delayed recall	Step 1	Class	0.27**	t = 3.61**	0.07	F = 13.47**
	Step 2	glo_interf_acc	− 0.07	t = − 1.06	0.04	F = 5.08**
		loc_interf_acc	− 0.08	t = − 1.16		
		Stroop_interf_acc	− 0.13†	t = − 1.72†		
Delayed	Step 1	Class	− 0.11	t = − 1.46	0.02	F = 2.90†
confabulation	Step 2	glo_interf_acc	0.16*	t = 2.07*	0.03	F = 2.38*
		loc_interf_acc	0.07	t = 0.95		
		Stroop_interf_acc	− 0.12	t = − 1.59		
Yield1	Step 1	Class	− 0.13	t = − 1.74†	0.02	F = 2.73
	Step 2	glo_interf_acc	− 0.01	t = −0.16	0.01	F = 1.01
		loc_interf_acc	− 0.04	t = −0.60		
		Stroop_interf_acc	0.08	t = 1.07		
Yield2	Step 1	Class	− 0.23**	t = − 3.17**	0.06	F = 9.92**
	Step 2	glo_interf_acc	− 0.08	t = − 1.15	0.01	F = 3.06*
		loc_interf_acc	0.03	t = 0.45		
		Stroop_interf_acc	0.07	t = 0.94		
Shift	Step 1	Class	− 0.23**	t = − 3.03**	0.05	F = 8.95**
	Step 2	glo_interf_acc	− 0.09	t = − 1.22	0.01	F = 2.81*
		loc_interf_acc	0.01	t = 0.22		
		Stroop_interf_acc	0.07	t = 0.96		
Total	Step 1	Class	− 0.21**	t = − 2.69**	0.04	F = 6.78**
Suggestibility	Step 2	glo_interf_acc	− 0.05	t = −0.75	0.01	F = 2.13†
		loc_interf_acc	− 0.02	t = −0.27		
		Stroop_interf_acc	0.09	t = 1.18		

Table 6. Hierarchical regressions predicting GSS2 measures from class and interference scores in the GLT and SCWT. glo = global-to-local, loc = local-to-global, interf = interference. †$0.05 < p < 0.10$. *$p \le 0.05$. **$p \le 0.01$.

than in middle-school children (in fact, the advantage in the identification of global shapes was only significant in the former group). Our conclusions are consistent with those obtained by Huizinga et al. (2010)[41], who found a global advantage effect that decreased with age in children aged 7 and 11 years, whereas they are incongruent with the local advantage reported by Scherf et al. (2009)[42] in children and adolescents (note, however, that in this study a global advantage was instead exhibited by young adults). Turning to the SCWT, many previous studies have shown a significant and gradual reduction of the interference effect with age, reflecting the protracted development of inhibition and executive processes during middle childhood[44–46].

In addition to these developmental patterns, the main contribution of the present study comes from the analysis of the correlations between the GSS2 and cognitive measures. In terms of span scores, our results match those illustrated by Caprin et al. (2016)[23] in a sample of 372 Italian children between the ages of 6 to 11 years. These authors found that digit span scores were positively associated with free and cued recall performance, as well as with the number of distortions reported by children; furthermore, negative correlations were reported between digit span scores and both yield and shift measures. Although Caprin et al. (2016)[23] did not use the GSS (making a direct comparison difficult), our findings are very similar, showing robust positive associations between span scores and free recall (both immediate and delayed) and negative associations between span scores and yield1, yield2 and total suggestibility. More generally, significant relations between WM, eyewitness memory and suggestibility have been often reported in studies involving middle childhood and adolescents[24,26]. As stated above, when children need to decide whether to endorse or not a misleading question containing false details, having a higher WM capacity might allow them to retrieve the original details from long-term memory and evaluate the discrepancy between the two types of information with greater accuracy: this process would make them less likely to accept misleading questions[15]. Similar conclusions have been obtained with other paradigms: Zhu et al. (2010)[47], for example, found a negative relation between adults' WM capacity and the propensity to integrate misleading post-event information in free recall, and Finnilä, Mahlberg, Santtila, Sandnabba, & Niemi (2003)[48] proposed that age differences in suggestibility might be due to a shorter memory span in younger than in older children. Yet, we must also note that, when class differences were controlled for in the follow-up regression analyses, the role of WM capacity remained significant for immediate recall and delayed confabulation, whereas it fell below the significance level for yield1, yield2 and total suggestibility. This pattern of results suggests that the negative correlations between span scores and suggestibility might be driven by aspecific developmental (i.e., age-related) effects, rather than by pure memory effects.

Regarding cognitive inhibition, our findings were mixed. On the one hand, we showed that children who exhibited higher accuracy-based interference in the global condition of the GLT produced more confabulated items in the delayed recall test of the GSS2. Such a result is in line with theoretical hypotheses, since children having high local-to-global interference scores are less efficient in inhibiting the elaboration of local shapes when requested to identify global shapes. If inhibitory skills are involved in reducing the production of confabulated items, then children with a lower inhibitory efficiency should be more likely to produce incorrect elements (compared to children with higher efficiency) – as it occurred in our results. On the other hand, the results obtained with the RT-based interference indexes are in contrast with expectations. Here, we found that children having higher interference scores in the SCWT (i.e., children who were less efficient in inhibiting the elaboration of the semantic meaning of the words when asked to identify the colors in which they were printed) produced more correct elements and less confabulated items in the delayed test, were less likely to yield to misleading questions and were less suggestible. If high levels of suggestibility in children are explained, at least in part, by a lower inhibitory efficiency, then the direction of the above associations should be just the opposite of that reported in our analyses (i.e., children which higher interferences scores in the SCWT should be more likely to yield to misleading questions, and therefore more suggestible than children having lower interference indices). Moreover, in both cases, the results obtained with a given interference measure did not generalize to the other measures, raising doubts about the idea that children's suggestibility may reflect the involvement of a general ability of response inhibition. Thus, when viewed in their entirety, the conclusions arising from the analysis of the interference scores in the present study are in line with those reached by Bruck and Melnyk (2004)[1], who argued that the review of the studies using executive function tasks provided little illumination on the cognitive mechanisms of children's suggestibility.

It should be noted that the inhibitory tasks used in the present study were designed to have different congruency proportions. Specifically, in the SCWT the number of congruent trials was much lower (1/3) than the number of incongruent trials, whereas in the GLT the number of congruent trials was equivalent to that of incongruent trials. Previous studies indicate that the congruency proportion in the Stroop task represents a relevant factor which may affect the degree of attention devoted to the word and color dimensions and hence the size of the inhibitory effect (e.g., Spinelli & Lupker, 2023). In our dataset, the two inhibitory tasks showed very similar patterns of correlations with the GSS measures and most results were non-significant, making it difficult to evaluate the impact of these methodological differences. However, in future studies the congruency proportions of the inhibitory tasks should be carefully matched to ascertain whether this factor could explain the different associations with suggestibility measures. In addition, the two inhibitory tasks used different materials – the GLT employed visual stimuli (geometrical shapes), whereas the SCWT employed verbal stimuli (words). As noted in the introduction, we used different materials because we wanted to determine whether children's suggestibility was dependent on a general ability of response inhibition. Yet, the interference indices obtained from the GLT and SWCT were not significantly correlated, suggesting that they did measure a common underlying construct. ($r(196) < 0.13$, $p > 0.095$ for RT-based interferences indices and $r(196) < 0.14$, $p > 0.071$ for accuracy-based interferences indices.) This problem has been recently discussed by Rouder et al. (2023), who found that trial noise in 24 different inhibitory tasks was much greater than individual variability, resulting in massive attenuation in correlations. Lastly, the GLT and SCWT were administered to different subsamples of children, which were only partially overlapping. This imbalance might have altered the conclusions of the analyses, although developmental trends were fully consistent with our expectations.

The present study has several limitations that must be acknowledged. First, WM was investigated with a single span task. Considering the most recent version of the WM model proposed by Baddeley (2000)[18], we only examined the processes mediated by the phonological loop. While this choice was justified by the verbal nature of the GSS2 story, previous studies have reported results similar to the present ones using the Corsi span task[49], which taps primarily the visuospatial sketchpad. Thus, both tasks, or complex span tasks (which may be more appropriate for older children[16]), should be used and compared in future studies. Second, in the present study we did not discriminate between different types of resistant responses to leading questions. In fact, responses such as "I don't remember", "this was not mentioned in the story", "I don't know", "I'm not sure" were all coded as non-yield responses. Recent studies have however shown that different types of "resistant behavioral responses" (RBRs) call for a differential involvement of source monitoring abilities (e.g., responses such as "I don't remember" may be driven by specific source monitoring difficulties: Gudjonsson et al., 2024). Thus, in future studies it would be helpful to investigate the associations between RBRs and executive functions and working memory. Third, our conclusions were based on correlations analyses, which did not allow us to determine whether individual differences in WM capacity or executive skills were causally related to suggestibility in the GSS2. Training studies could help reveal whether improvements in children's WM capacity or inhibitory abilities are accompanied by decreases in their suggestibility. Fourth, we only investigated interrogative suggestibility in typically developing children. As stated above, there is mounting evidence suggesting that suggestibility can be studied by adopting different experimental paradigms and that each type of suggestibility can be associated with different psychological mechanisms. A meta-analytical approach might be useful to determine whether the associations with WM processes can be generalized to different forms of suggestibility. It would be also interesting to extend the present analyses to children with mild intellectual disabilities and borderline intellectual functioning, who produce a greater number of confabulations and are more vulnerable to yield to leading questions and to shift responses after the negative feedback (Giostra & Vagni, 2024). Despite these problems, we believe that the present study contributes useful results to the ongoing debate regarding the impact of individual differences in cognitive factors on the suggestibility of primary- and middle-school children – two groups that have been underrepresented in previous research[1].

Data availability

The datasets generated and analysed in the current study are available from the corresponding authors on reasonable request.

Received: 18 September 2024; Accepted: 28 February 2025
Published online: 18 March 2025

References

1. Bruck, M. & Melnyk, L. Individual differences in children's suggestibility: A review and synthesis. *Appl. Cogn. Psychol.* **18**(8), 947–996. https://doi.org/10.1002/acp.1070 (2004).
2. Gudjonsson, G. H. & Clark, N. K. Suggestibility in police interrogation: A social psychological model. *Soc. Behav.* **1**(2), 83–104 (1986).
3. Henry, L. A. & Gudjonsson, G. H. Eyewitness memory, suggestibility, and repeated recall sessions in children with mild and moderate intellectual disabilities. *Law Hum. Behav.* **27**(5), 481–505. https://doi.org/10.1023/A:1025434022699 (2003).
4. Memon, A., Zaragoza, M., Clifford, B. R. & Kidd, L. Inoculation or antidote? The effects of cognitive interview timing on false memory for forcibly fabricated events. *Law Hum. Behav.* **34**(2), 105–117. https://doi.org/10.1007/s10979-008-9172-6 (2010).
5. Rossi-Arnaud, C. et al. Long-lasting positive effects of collaborative remembering on false assents to misleading questions. *Acta Psychol.* https://doi.org/10.1016/j.actpsy.2019.102986 (2020).
6. Ceci, S. J., Loftus, E. F., Leichtman, M. D. & Bruck, M. The possible role of source misattributions in the creation of false beliefs among preschoolers. *Int. J. Clin. Exp. Hypn.* **42**(4), 304–320. https://doi.org/10.1080/00207149408409361 (1994).
7. Pezdek, K. & Hodge, D. Planting false childhood memories in children: The role of event plausibility. *Child Dev.* **70**(4), 887–895. https://doi.org/10.1111/1467-8624.00064 (1999).
8. Gudjonsson, G. H. A new scale of interrogative suggestibility. *Personal. Indiv. Differ.* **5**(3), 303–314. https://doi.org/10.1016/0191-8869(84)90069-2 (1984).
9. Bianco, A. & Curci, A. Measuring interrogative suggestibility with the Italian version of the Gudjonsson Suggestibility Scales (GSS): Factor structure and discriminant validity. *Personal. Indiv. Differ.* **82**, 258–265. https://doi.org/10.1016/j.paid.2015.03.035 (2015).
10. Gignac, G. E. & Powell, M. B. A psychometric evaluation of the Gudjonsson Suggestibility Scales: Problems associated with measuring suggestibility as a difference score composite. *Personal. Indiv. Differ.* **46**(2), 88–93. https://doi.org/10.1016/j.paid.2008.09.007 (2009).
11. Bruck, M. & Ceci, S. Forensic developmental psychology: Unveiling four common misconceptions. *Curr. Direct. Psychol. Sci.* **13**(6), 229–232. https://doi.org/10.1111/j.0963-7214.2004.00314.x (2004).
12. Hünefeldt, T., Lucidi, A., Furia, A. & Rossi-Arnaud, C. Age differences in the interrogative suggestibility of children's memory: Do shift scores peak around 5–6 years of age?. *Personal. Indiv. Differ.* **45**(6), 521–526. https://doi.org/10.1016/j.paid.2008.06.005 (2008).
13. Gudjonsson, G., Vagni, M., Maiorano, T. & Pajardi, D. Age and memory related changes in children's immediate and delayed suggestibility using the Gudjonsson Suggestibility Scale. *Personal. Indiv. Differ.* **102**, 25–29. https://doi.org/10.1016/j.paid.2016.06.029 (2016).
14. Gudjonsson, G. H. *The Psychology of Interrogations and Confessions: A Handbook* (Wiley, 2003).
15. Roebers, C. M. & Schneider, W. Individual differences in young children's suggestibility: Relations to event memory, language abilities, working memory, and executive functioning. *Cogn. Dev.* **20**(3), 427–447. https://doi.org/10.1016/j.cogdev.2005.05.006 (2005).
16. Bottoms, B. L., Quas, J. A. & Davis, S. L. The influence of the interviewer-provided social support on children's suggestibility, memory, and disclosures. In *Child Sexual Abuse: Disclosure, Delay, and Denial* (ed. Bottoms, B. L.) 135–157 (Lawrence Erlbaum Associates Publishers, 2007).
17. Baddeley, A. D. & Hitch, G. J. Working memory. In *The Psychology of Learning and Motivation* Vol. 8 (ed. Bower, G. H.) 47–89 (Academic Press, 1974).
18. Baddeley, A. The episodic buffer: A new component of working memory?. *Trends Cogn. Sci.* **4**(11), 417–423. https://doi.org/10.1016/S1364-6613(00)01538-2 (2000).
19. Baddeley, A. Working memory: Theories, models, and controversies. *Annu. Rev. Psychol.* **63**, 1–29. https://doi.org/10.1146/annurev-psych-120710-100422 (2012).
20. Cestari, V., Lucidi, A., Pieroni, L. & Rossi-Arnaud, C. Memory for object location: A span study in children. *Can. J. Exp. Psychol./Revue Canadienne de Psychologie Expérimentale* **61**(1), 13–20. https://doi.org/10.1037/cjep2007002 (2007).
21. Pickering, S. J. The development of visuo-spatial working memory. *Memory* **9**(4–6), 423–432. https://doi.org/10.1080/09658210143000182 (2001).
22. Pickering, S. J., Gathercole, S. E. & Peaker, S. M. Verbal visuospatial short-term memory in children: Evidence for common and distinct mechanisms. *Mem. Cogn.* **26**(6), 1117–1130. https://doi.org/10.3758/BF03201189 (1998).
23. Caprin, C. et al. True and false memories in middle childhood: The relationship with cognitive functioning. *Psychol. Crime Law* **22**(5), 473–494. https://doi.org/10.1080/1068316X.2016.1168420 (2016).
24. Clarke-Stewart, K. A., Malloy, L. C. & Allhusen, V. D. Verbal ability, self-control, and close relationships with parents protect children against misleading suggestions. *Appl. Cogn. Psychol.* **18**(8), 1037–1058. https://doi.org/10.1002/acp.1076 (2004).
25. Danielsdottir, G., Sigurgeirsdottir, S., Einarsdottir, H. R. & Haraldsson, E. Interrogative suggestibility in children and its relationship with memory and vocabulary. *Personal. Indiv. Differ.* **14**(3), 499–502. https://doi.org/10.1016/0191-8869(93)90323-U (1993).
26. Lee, K. Age, neuropsychological, and social cognitive measures as predictors of individual differences in susceptibility to the misinformation effect. *Appl. Cogn. Psychol.* **18**(8), 997–1019. https://doi.org/10.1002/acp.1075 (2004).
27. MacLeod, C. M. The Stroop task: The "gold standard" of attentional measures. *J. Exp. Psychol. Gen.* **121**(1), 12–14. https://doi.org/10.1037/0096-3445.121.1.12 (1992).
28. Bialystok, E. Global–local and trail-making tasks by monolingual and bilingual children: Beyond inhibition. *Dev. Psychol.* **46**(1), 93–105. https://doi.org/10.1037/a0015466 (2010).
29. Cottini, M. et al. Feature binding and the processing of global–local shapes in bilingual and monolingual children. *Mem. Cogn.* **43**(3), 441–452. https://doi.org/10.3758/s13421-014-0467-1 (2015).
30. Navon, D. Forest before trees: The precedence of global features in visual perception. *Cogn. Psychol.* **9**(3), 353–383. https://doi.org/10.1016/0010-0285(77)90012-3 (1977).
31. Kimchi, R. Primacy of wholistic processing and global/local paradigm: A critical review. *Psychological Bulletin* **112**(1), 24–38. https://doi.org/10.1037/0033-2909.112.1.24 (1992).

32. Alexander, K. W. et al. The role of attachement and cognitive inhibition in children's memory and suggestibility for a stressful event. *Journal of Experimental Child Psychology* **83**(4), 262–290. https://doi.org/10.1016/S0022-0965(02)00149-2 (2002).
33. Schaaf, J. M., Alexander, K. W. & Goodman, G. S. Children's false memory and true disclosure in the face of repeated questions. *J. Exp. Child Psychol.* **100**(3), 157–185. https://doi.org/10.1016/j.jecp.2007.09.002 (2008).
34. Melinder, A., Endestad, T. & Magnussen, S. Relations between episodic memory, suggestibility, theory of mind, and cognitive inhibition in the preschool child. *Scand. J. Psychol.* **47**(6), 485–495. https://doi.org/10.1111/j.1467-9450.2006.00542.x (2006).
35. Roberts, K. P. & Powell, M. B. The relation between inhibitory control and children's eyewitness memory. *Appl. Cogn. Psychol.* **19**(8), 1003–1018. https://doi.org/10.1002/acp.1141 (2005).
36. Quas, J. A. & Schaaf, J. M. Children's memories of experienced and nonexperienced events following repeated interviews. *J. Exp. Child Psychol.* **83**(4), 304–338 (2002).
37. Gudjonsson, G. H. *The Gudjonsson Suggestibility Scales Manual* (Psychology Press, 1997).
38. Vicari, S. *Promea, prove di memoria e apprendimento per l'età evolutiva*. Giunti OS Organizzazioni Speciali, Florence. (2007).
39. Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H. & Howerter, A. The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cogn. Psychol.* **41**(1), 49–100 (2000).
40. Rossi-Arnaud, C., Spataro, P., Bhatia, D. & Cestari, V. Collaborative remembering reduces suggestibility: A study with the Gudjonsson Suggestibility Scale. *Memory* **27**(5), 603–611 (2019).
41. Huizinga, M., Burack, J. A. & Van der Molen, M. W. Age-related change in shifting attention between global and local levels of hierarchical stimuli. *J. Cogn. Dev.* **11**(4), 408–436. https://doi.org/10.1080/15248371003700031 (2010).
42. Scherf, K. S., Behrmann, M., Kimchi, R. & Luna, B. Emergence of global shape processing continues through adolescence. *Child Dev.* **80**(1), 162–177. https://doi.org/10.1111/j.1467-8624.2008.01252.x (2009).
43. Rey-Mermet, A., Gade, M. & Oberauer, K. Should we stop thinking about inhibition? Searching for individual and age differences in inhibition ability. *J. Exp. Psychol. Learn. Mem. Cogn.* **44**(4), 501–526. https://doi.org/10.1037/xlm0000450 (2018).
44. Adleman, N. E. et al. A developmental fMRI study of the Stroop color-word task. *Neuroimage* **16**(1), 61–75. https://doi.org/10.1006/nimg.2001.1046 (2002).
45. Bellaj, T., Salhi, I., Le Gall, D. & Roy, A. Development of executive functioning in school-age Tunisian children. *Child Neuropsychol.* **22**(8), 919–954. https://doi.org/10.1080/09297049.2015.1058349 (2015).
46. Schroeter, M. L., Zysset, S., Wahl, M. & von Cramon, D. Y. Prefrontal activation due to Stroop interference increases during development—An event-related fNIRS study. *Neuroimage* **23**(4), 1317–1325. https://doi.org/10.1016/j.neuroimage.2004.08.001 (2004).
47. Zhu, B. et al. Individual differences in false memory from misinformation: Cognitive factors. *Memory* **18**(5), 543–555. https://doi.org/10.1080/09658211.2010.487051 (2010).
48. Finnilä, K., Mahlberg, N., Santtila, P., Sandnabba, K. & Niemi, P. Validity of a test of children's suggestibility for predicting responses to two interview situations differing in their degree of suggestiveness. *J. Exp. Child Psychol.* **85**(1), 32–49. https://doi.org/10.1016/S0022-0965(03)00025-0 (2003).
49. Angelini, G. & Rossi-Arnaud, C. Accuratezza e completezza nella testimonianza dei bambini: Relazioni con la memoria di lavoro. *Rassegna Di Psicologia* **34**(3), 67–81 (2017).

Author contributions
C.R.A., A.S. and P.S conceived the present study. A.S., M.P. and A.N. collected the data. A.S., PS. analyzed the results. P.S., A.S., C.R.A and V.C., wrote the manuscript, and all.

authors including V.C. and C.R.A., reviewed and edited the manuscript.

Funding
This research was supported by Sapienza Research Grant n. (RP123188E81EFE36) PI: Clelia Rossi-Arnaud.

Declarations

Competing interests
The authors declare no competing interests.

Ethics approval
The research was approved by the Ethical Committee of the authors' institution (Protocol N. 0001293).

Consent to participate
Informed consent was obtained from all individual participants included in the study.

Additional information
Correspondence and requests for materials should be addressed to A.S. or P.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.