



# Normative data on the n-back task for children and young adolescents

Santiago Pelegrina<sup>1\*</sup>, M. Teresa Lechuga<sup>1</sup>, Juan A. García-Madruga<sup>2</sup>, M. Rosa Elosúa<sup>3</sup>, Pedro Macizo<sup>4</sup>, Manuel Carreiras<sup>5</sup>, Luis J. Fuentes<sup>6</sup> and M. Teresa Bajo<sup>4</sup>

<sup>1</sup> Department of Psychology, University of Jaén, Jaén, Spain, <sup>2</sup> Department of Developmental and Educational Psychology, National University of Distance Education, Madrid, Spain, <sup>3</sup> Department of Basic Psychology, National University of Distance Education, Madrid, Spain, <sup>4</sup> Research Center for Mind, Brain and Behavior, University of Granada, Granada, Spain, <sup>5</sup> Basque Center of Cognition, Brain and Language, San Sebastián-Donostia, Spain, <sup>6</sup> Department of Basic Psychology and Methodology, University of Murcia, Murcia, Spain

## OPEN ACCESS

### Edited by:

Yusuke Moriguchi,  
Joetsu University of Education, Japan

### Reviewed by:

Michael R. Miller,  
Western University, Canada  
Irene C. Mammarella,  
University of Padova, Italy

### \*Correspondence:

Santiago Pelegrina,  
Department of Psychology,  
University of Jaén, Paraje Las  
Lagunillas S/N, 23071 Jaén, Spain  
spelegri@ujaen.es

### Specialty section:

This article was submitted to  
Developmental Psychology,  
a section of the journal  
Frontiers in Psychology

**Received:** 13 July 2015

**Accepted:** 24 September 2015

**Published:** 08 October 2015

### Citation:

Pelegrina S, Lechuga MT,  
García-Madruga JA, Elosúa MR,  
Macizo P, Carreiras M, Fuentes LJ  
and Bajo MT (2015) Normative data  
on the n-back task for children  
and young adolescents.  
Front. Psychol. 6:1544.  
doi: 10.3389/fpsyg.2015.01544

The n-back task is a frequently used measure of working memory (WM) in cognitive neuroscience research contexts, and it has become widely adopted in other areas over the last decade. This study aimed to obtain normative data for the n-back task from a large sample of children and adolescents. To this end, a computerized verbal n-back task with three levels of WM load (1-back, 2-back, and 3-back) was administered to 3722 Spanish school children aged 7–13 years. Results showed an overall age-related increase in performance for the different levels of difficulty. This trend was less pronounced at 1-back than at 2-back when hits were considered. Gender differences were also observed, with girls outperforming boys although taking more time to respond. The theoretical implications of these results are discussed. Normative data stratified by age and gender for the three WM load levels are provided.

**Keywords:** n-back, working memory updating, working memory development, memory development in children, working memory

## Introduction

Working memory (WM) is a system that enables one to actively maintain and regulate a limited amount of task-relevant information (Baddeley and Logie, 1999). It plays an important role in processes like reading comprehension (e.g., García-Madruga et al., 1997; Cain, 2006), reasoning (e.g., Conway et al., 2003; García-Madruga et al., 2007), arithmetic calculations (e.g., Deschuyteneer et al., 2006), mathematical problem solving (e.g., Passolunghi and Pazzaglia, 2004), and academic performance (e.g., Alloway and Alloway, 2010). WM is also closely associated with fluid intelligence (e.g., Friedman et al., 2006).

WM has been traditionally measured using processing-and-storing dual tasks that involve performing a cognitive task while certain information has to be maintained in memory. For instance, in the Reading Span Test (Daneman and Carpenter, 1980), participants have to read increasing series of sentences aloud and eventually recall the last word of each sentence. In the counting span task (Case et al., 1982), subjects are presented with series of sets of figures, and they have to count the number of figures in order to recall this number at the end of the series. In the operation span task (Turner and Engle, 1989) participants are presented with a series of simple arithmetical equations and they have to verify each equation while memorizing a digit for later recall. In the analogy and anaphora span tasks (García-Madruga et al., 2005), participants have to resolve a series of analogies or anaphoric inferences and later recall the results of these inferences.

The n-back is a different kind of WM measure that does not involve a processing-and-storing dual task and in which individuals are not asked to recall any information but to recognize it. The

n-back is a continuous recognition task in which participants must decide whether a stimulus was previously presented in certain conditions. In this task a sequence of items are shown, ranging from letters and drawings to words and numbers. For each item it must be determined whether the last one shown is identical to the stimulus presented “n” trials back. Therefore, at the 1-back level, each item needs to be compared with the one presented immediately before, that is, with the prior letter. At 2-back each stimulus is compared with the one shown two positions prior. At higher n levels the decision is based on stimuli separated each time by that many positions. Participants must respond “yes” or “no” to each trial. Response accuracy and response times are usually recorded.

First introduced by Kirchner (1958), the n-back is a widely used tool for assessing WM in the field of neuroscience. Its popularity in this area has much to do with how well it meets the demands of neuroimaging studies (Owen et al., 2005; Jaeggi et al., 2010). Thus, the responses that participants must provide are far simpler compared with other standard WM tasks; it is possible to record both accuracy and response times, and one can easily manipulate the level of difficulty depending on greater or lesser memory loads as well as the timing of stimulus presentation. Since the 1990s, as neuroimaging techniques have become more numerous, the n-back has grown in use and has been progressively adopted in other fields where other WM tasks were traditionally used.

The reliability of the task varies greatly depending on the study, from  $r = 0.16$  (Van Leeuwen et al., 2007) to  $r = 0.91$  (Friedman et al., 2008). The more complex levels such as 2-back and 3-back produce higher reliability coefficients (see Jaeggi et al., 2010). Correlations with other WM measures, although they differ from one study to another, are surprisingly low and not even significant (Redick and Lindsey, 2013). The variety of WM tasks and the different contents that they include (e.g., spatial, numeric, etc.) may be a reason for these low correlations. Thus, research findings where different types of WM measures (traditional or n-back) have been employed should not be considered as interchangeable (Kane et al., 2007).

Paradoxically, the n-back may show higher correlations with tasks that capture other constructs. In this regard, it is worth noting that relationships with intelligence measures are consistently found, in particular fluid intelligence, such as Raven's Advanced Progressive Matrices (e.g., Kane et al., 2007; Colom et al., 2008; Jaeggi et al., 2010).

These seemingly contradictory findings may be related to the fact that different tasks capture different processes that contribute to WM performance. In this sense, the n-back includes components that have traditionally been considered parts of WM tasks. For example, a 2-back task requires participants to store two elements (e.g., letters) in WM. It further involves updating WM content, given that each time a new item is presented subjects must eliminate the previous item, add the new one, and maintain the presentation order of each item. In short, storage and processing operations that fall under the traditional definition of WM are also present in this task (Baddeley, 2007).

Furthermore, the n-back requires processes that are not found in other WM tasks. Perhaps the most relevant aspect is that n-back relies on recognition processes whereas traditional WM

tasks require the retrieval of information. Hence, in n-back tasks participants may rely on assessments of familiarity, something that is not possible in traditional tasks. Moreover, n-back tasks depend on interference resolution processes in order to resolve the conflicts between familiarity and recollection, as well as other processes such as decision-making, temporal ordering, inhibition of no-longer-relevant information, and creation and updating of bindings between content and temporal context (see Oberauer, 2005; Kane et al., 2007).

The processes involved in the n-back task may also vary depending on the level of load. While all levels require active information maintenance, they differ not only in the load or quantity of information to maintain but also in the degree of content processing and involvement of executive control processes. Thus, at the 1-back level there is no need for content manipulation, whereas 2-back and above will involve constantly adding a new item and dropping the obsolete item. Moreover, content at 1-back is less susceptible to interference than at higher levels. At the 2-back level, participants have to store an element (e.g., a letter) for later use, and at the same time prevent it from influencing the response that they must provide. Thus control and interference resolution is crucial at higher load levels (Schleppen and Jonkman, 2010). On the other hand, it should be considered that the n-back is a recognition task, which may also involve familiarity and recollection. At 1-back, recognition could be based on familiarity. That is, because of its recency, the last presented item would likely have the higher level of activation in WM, which in turn would elicit a higher level of familiarity than other items in WM (see Oberauer, 2005). In contrast, at the 2-back (and higher) level, the decision may not only be based on familiarity given that different items could elicit stronger signals of familiarity than the item presented 2 or more steps back. Then a recollection process is needed in order to access the contextual information (e.g., order of presentation, temporal context) related to the item presented. Finally, the different levels vary according to the need to retrieve information outside the focus of attention in WM (Verhaeghen and Basak, 2005; Oberauer, 2006). Some models of WM (e.g., Cowan, 1995; Oberauer, 2002) differentiate between two embedded components: the focus of attention and an activated region of long-term memory. Oberauer (2002; see also McElree, 2001) assumed that the focus of attention is able to hold only one item that is directly accessible for processing operations. Thus, while at 1-back the necessary information to respond is in the focus of attention, at higher levels this information has to be accessed from outside of the focus. In short, 1-back and higher n-back levels have differing storage, processing, and executive control demands.

Working memory undergoes constant change through childhood and adolescence (e.g., Elosúa et al., 1997; Gathercole et al., 2004; Luna et al., 2004; Luciana et al., 2005), which in turn has an impact on the performance of different complex tasks (e.g., Siegel, 1994). In addition, developmental changes in the effectiveness with which information in WM is updated has been reported (e.g., Huizinga et al., 2006; Lechuga et al., 2006; Belachi et al., 2010; Lendínez et al., 2015). In recent years, some developmental studies using the n-back task have been conducted (Kwon et al., 2002; Vuontela et al., 2003; Schleppen and Jonkman,

**TABLE 1 | Age groups of the study population.**

Age group	N	Gender		Age (months)		% Completed		
		Boys	Girls	M	SD	1-back	2-back	3-back
7	387	193	194	91.32	2.52	98.71	77.78	34.88
8	592	285	307	101.32	3.49	99.83	83.45	41.89
9	606	310	296	113.82	3.37	99.67	86.96	49.01
10	618	297	321	125.35	3.50	99.51	91.59	60.68
11	601	315	286	137.50	3.47	99.67	91.35	66.72
12	476	253	223	149.39	3.39	100.00	92.65	70.59
13	442	233	209	161.38	3.38	99.10	94.57	78.96
Total	3722	1886	1836					

Number, gender, and mean age of participants for each age group and percentage of participants completing each level of the task.

2010). In general, this research has revealed gradual performance improvements during childhood (Vuontela et al., 2003). These changes continue through adolescence at the more demanding levels (i.e., 2-back; e.g., Schleepen and Jonkman, 2010).

The current study aims to provide normative data on n-back task performance among children and adolescents aged 7–13 years, contributing useful information on a task which is becoming increasingly more frequent and for which, to our knowledge, there is no normative data available. Previous studies have used only one or two levels of n-back in which the information had to be updated in memory. They have included only two or three groups of children of different ages for comparison with a rather limited number of children in each age group (e.g., less than 25). The present study used a more fine-grained approach to analyze the age trends in n-back performance through childhood and early adolescence. Thus, a wide range of ages, from 7 to 13 years, and a large sample for each of the ages was considered.

The current study also represents an opportunity to analyze age trends in the task. Given the nature of the process involved at the different levels of n-back, different age-related patterns for each n-back level might be obtained. Since the 2-back or 3-back versions require the involvement of higher executive control processes to a greater extent than the 1-back condition, we expected a more pronounced increase in performance with age in the more complex conditions than in the 1-back version of the task.

To this end, three levels of WM load (1-back, 2-back, and 3-back) were administered as part of a large-scale study assessing different cognitive functions considered essential for adequate academic performance, such as attention, memory, language, and reasoning.

## Materials and Methods

### Participants

A total of 3722 children and adolescents (1886 boys and 1836 girls) aged 7–13 years took part in this study. Only children whose parents consented to their participation in the study were administered the task. No other criteria were applied for exclusion. This was to obtain the largest sample size possible and to represent the full range of children as accurately as possible. The participants made up a sample of school children recruited from 43 schools across several Spanish cities. The number of participants in each

age group, their gender and mean age is shown in **Table 1**. Prior to testing, informed consent to children's participation was obtained from a parent or legal guardian. Approval from the ethical committee of the University of La Laguna (Spain) was obtained previous to the study.

### Materials and Procedure

Children were given a series of cognitive tasks over three sessions as a part of a larger study on the development of cognitive skills involved in academic performance. The n-back task was administered at the beginning of the session. Each session lasted about 60 min.

The n-back task programmed by Robinson and Fuller (2004) in E-prime (Schneider et al., 2002) was used in this study. The version of the task applied here consisted of three levels: 1-back, 2-back, and 3-back. The items to be updated were letters, that is, linguistic material. The following 20 consonants were used as stimuli in all blocks: B, C, D, F, G, H, J, K, L, M, N, P, Q, R, S, T, V, W, Y, Z. Stimuli were presented one by one in the center of the screen (font: Palatino Linotype, size: 30). On the 1-back task, children had to compare the current trial to the previous one; they were asked to press the “yes” key only when the two subsequently presented letters were the same; otherwise they were to press the “no” key. The 2-back and 3-back trials were similar, with the exception that on 2-back trials, children had to compare the currently presented letter to the one presented two trials before, and on 3-back trials, to the letter presented three trials before. The first three trials of each block were always non-target trials where the correct response was “no.” Each stimulus appeared on the screen for 500 ms, followed by a screen that remained blank for another 3000 ms. The children had 3500 ms, from stimulus onset until the beginning of the subsequent trial to press the corresponding response key. Response times as well as the number of hits, correct rejections, misses, false alarms, and non-responses were recorded.

Each level of the task (1-back, 2-back, etc.) was preceded by (1) instructions, (2) examples that included a sequence of six letters with the corresponding correct responses, and (3) a practice block that consisted of 20 trials (30% “yes” trials). And additional practice block was administered if a participant did not reach a correct percentage of 60% on the “yes” trials. There were two test blocks made up of 20 trials each (also with 30% “yes” trials), so that

there were a total of 40 trials per level. In order to prevent fatigue and frustration, the task was discontinued when the children did not reach a correct percentage of 60% on a given test block.

### Data Analyses

Two sets of analyses were conducted. The first was carried out to determine the age trends across the different levels of the n-back task. To this end, separate analyses were performed on the percentage of correctly identified targets (hits), the percentage of false alarms (“yes” responses to non-targets), the signal-detection parameter d-prime ( $d'$ ) that reflects the sensitivity of the participants to discriminate items as previously presented (or not) n trials back, and the mean response times for correctly identified targets.

The second set of analyses was aimed to provide normative data on the different levels of the n-back task and consisted of multiple regression analyses on accuracy that included the following predictors: age (in months), age squared, and gender.

## Results

### Age-Related Data

A mixed ANOVA was conducted at 1- and 2-back for each of the aforementioned variables, including age (7, 8, 9, 10, 11, 12, and 13 years) and gender as the between-group factors. If the more difficult level (3-back) had been included, the total sample would have been reduced by up to 42%. However, by keeping to the 1- and 2-back levels, 11% of participants were excluded. **Table 1** shows the number of participants by age that performed levels 1 and 2 of the n-back task<sup>1</sup>.

### Hits

A 2 (n-back load) × 7 (age) × 2 (gender) ANOVA was conducted on the percentage of hits. **Table 2** shows the ANOVA summary table for all the dependent variables. There was a significant main effect of load, with the percentage of hits in the 1-back condition ( $M = 74.04, SD = 16.8$ ) higher than that observed at 2-back ( $M = 50.55, SD = 25.22$ ). The main effects were also significant for age and gender. The gender differences indicated that girls ( $M = 62.59, SD = 16.82$ ) had a slightly higher percentage of hits than boys ( $M = 61.42, SD = 16.80$ ). There was also a significant interaction between load and age. All other interactions failed to reach the standard significance level ( $p > 0.075$ ).

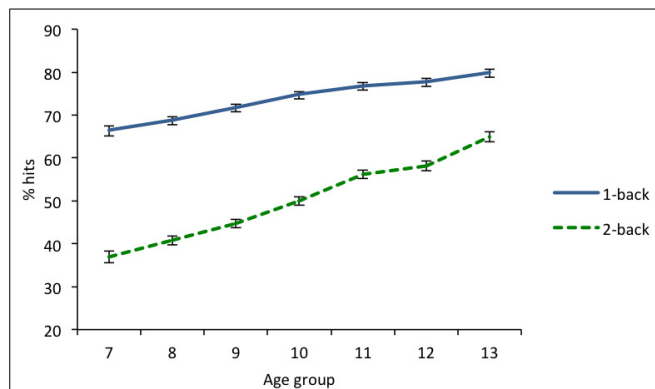
There was an increase in the percentage of hits in relation to age, although the patterns observed for each n-back load were different (see **Figure 1**). This was confirmed by the crucial interaction between the linear trend for age and load. A trend analysis for each n-back load indicated a relatively small but significant linear

<sup>1</sup>A parallel set of analyses including all the levels of n-back load was run. The main effects and interactions were roughly equivalent to those presented when considering only the 1-back and 2-back levels. Importantly, the age linear trend for 3-back was significant on hits:  $F(1,2134) = 103.63, p < 0.001, \eta_p^2 = 0.046$ ; false alarms:  $F(1,2134) = 90.58, p < 0.001, \eta_p^2 = 0.041$ ; d-prime:  $F(1,2134) = 180.96, p < 0.001, \eta_p^2 = 0.078$ ; and response times:  $F(1,2086) = 78.49, p < 0.001, \eta_p^2 = 0.036$ .

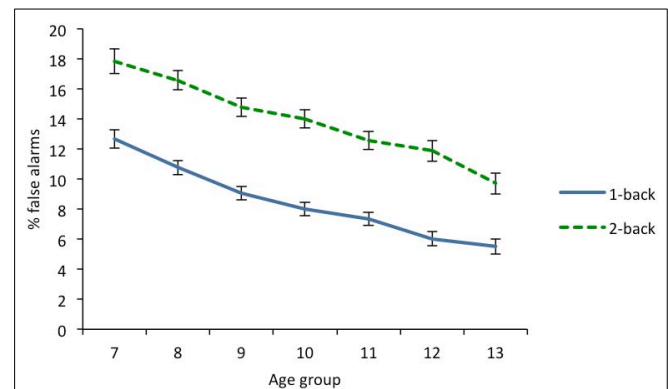
**TABLE 2 | Summary for ANOVAs performed on hits, false alarms, d-prime, and response time.**

Source	Hits				False alarms				D-prime				Response times			
	F	df	P	$\eta_p^2$	F	df	P	$\eta_p^2$	F	df	P	$\eta_p^2$	F	df	P	$\eta_p^2$
N-back load	3080.52	1,3282	0.000	0.48	554.69	1,3282	0.000	0.14	3703.15	1,3282	0.000	0.53	841.12	1,3107	0.000	0.21
Age	84.45	6,3282	0.000	0.13	23.92	6,3282	0.000	0.04	92.48	6,3282	0.000	0.14	84.52	6,3107	0.000	0.14
Gender	3.96	1,3282	0.047	0.01	8.75	1,3282	0.003	0.01	15.74	1,3282	0.000	0.01	57.46	1,3107	0.000	0.02
N-back load × age	20.27	6,3282	0.000	0.04	1.06	6,3282	0.385	0.01	12.00	6,3282	0.000	0.02	0.69	6,3107	0.660	0.01
N-back load × gender	1.24	1,3282	0.265	0.00	2.94	1,3282	0.086	0.01	6.61	1,3282	0.010	0.01	6.38	1,3107	0.012	0.01
N-back load × age × gender	0.41	6,3282	0.872	0.01	1.10	6,3282	0.362	0.01	0.66	6,3282	0.684	0.01	1.49	6,3107	0.176	0.01
Age × gender	1.91	6,3282	0.075	0.01	475.76	1,3289	0.000	0.13	1.63	6,3282	0.134	0.01	1.56	6,3107	0.154	0.01
Age (linear)													419.66			
Age (linear) × n-back load	108.16	1,3289	0.000	0.03					50.92	1,3289	0.000	0.01				
Age (linear) in 1-back	208.90	1,3289	0.000	0.06					273.75	1,3289	0.000	0.08				
Age (linear) in 2-back	415.71	1,3289	0.000	0.11					502.58	1,3289	0.000	0.13				





**FIGURE 1 | Percentage of hits as a function of age in 1-back and 2-back conditions.** Error bars indicate two standard error of the mean.



**FIGURE 2 | Percentage of false alarms as a function of age in 1-back and 2-back conditions.** Error bars indicate two standard error of the mean.

increase in performance with age at the 1-back level and a more pronounced linear increase with age at 2-back.

A posteriori analyses using the Bonferroni method were conducted to determine between which age groups the differences occurred. At 1-back no differences were found between some of the contiguous age groups, that is, 7 vs. 8 ( $p = 1$ ), 8 vs. 9 ( $p = 0.079$ ), and 10 vs. 11 years ( $p = 0.762$ ). From age 11, there were also no differences between the different groups (11, 12, and 13,  $p > 0.076$ ). The remaining comparisons were significant ( $p < 0.001$ ). At 2-back no differences were observed in some contiguous age groups, namely 7 vs. 8 ( $p = 0.487$ ), 8 vs. 9 ( $p = 0.168$ ), and 11 vs. 12 ( $p = 1$ ). The rest of the comparisons between the different age groups showed statistically significant differences ( $p < 0.007$ ).

### False Alarms

A 2 (n-back load)  $\times$  7 (age)  $\times$  2 (gender) ANOVA was conducted on the percentage of false alarms (see **Table 2**). There was a significant effect of load, with the percentage of false alarms in the 1-back condition ( $M = 8.32$ ,  $SD = 10.66$ ) lower than that observed at 2-back ( $M = 13.81$ ,  $SD = 14.56$ ). The main effect of age was also significant. As can be seen in **Figure 2**, there was a linear decrease in the percentage of false alarms with age similar at both n-back levels. A trend analysis revealed that only the linear component was significant, accounting for 99.12% of the variance. A posteriori analyses using the Bonferroni correction revealed differences in recognition between each age group ( $p < 0.05$ ), except at contiguous age groups ( $p > 0.20$ ) and between 10 and 12 year old ( $p = 0.06$ ).

There were also a significant albeit small gender effect. This was caused by boys ( $M = 11.77$ ,  $SD = 10.99$ ) committing more false alarms than girls ( $M = 10.63$ ,  $SD = 11.01$ ). None of the interactions approached significance ( $p > 0.086$ ).

### D-Prime

D-prime was estimated as  $d' = Z_{Hits} - Z_{FalseAlarms}$ . The fourth method proposed by Stanislaw and Todorov (1999) was used to avoid that  $d'$  might be undetermined when the hit or the false-alarm rate was equal to 0 or 1. Specifically, scores equal to 0 were replaced by  $0.5/n$  and scores equal to 1 were replaced

by  $(n - 0.5)/n$ , with  $n$  representing the number of signal and noise trials. A 2 (n-back load)  $\times$  7 (age)  $\times$  2 (gender) ANOVA was conducted on these values, similar to the previous analyses (see **Table 2**).

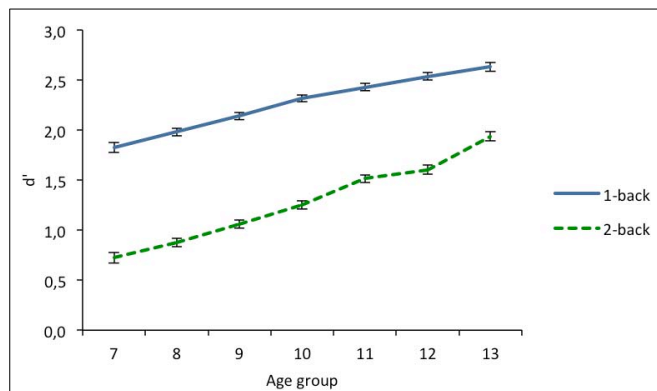
The ANOVA revealed a significant effect of load given that  $d'$  was greater in the 1-back condition ( $M = 2.28$ ,  $SD = 0.89$ ) than at 2-back ( $M = 1.29$ ,  $SD = 0.99$ ). The main effects were also statistically significant for age and gender. In addition, the interactions of age by load and gender by load were also significant. All other interactions failed to reach the standard significance level ( $p = 0.134$ ).

An interaction analysis between the linear trend of age and load revealed that even though  $d'$  increased with age at both levels of load, the age trend was different for each level as shown by the significant interaction. As with the previous analysis on hits, an age-trend analysis for each n-back load indicated an age-related linear increase on  $d'$  at both 1-back and 2-back. Thus, the interaction was again due to the fact that the linear increase was more pronounced at 2-back than at 1-back level (see **Figure 3**). This pattern was also evidenced by *post hoc* analyses with the Bonferroni correction that showed that at 1-back, all the age groups differed ( $p < 0.05$ ), except at contiguous age groups ( $p > 0.06$ ), excluding 9 vs. 10 ( $p = 0.009$ ). At 2-back significant differences were observed between all age groups ( $p < 0.027$ ), except between 7 vs. 8 ( $p = 0.514$ ) and 11 vs. 12 ( $p = 1$ ).

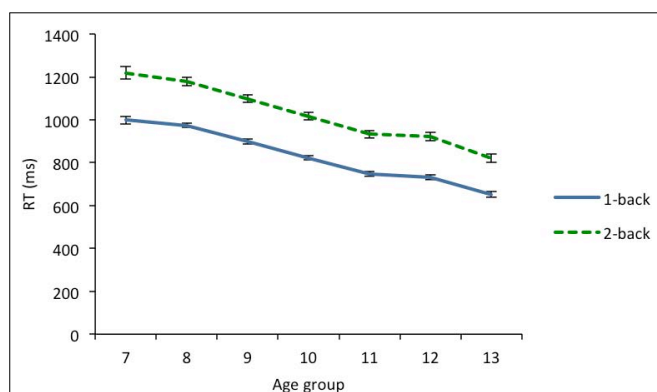
As regards the interaction between n-back levels and gender, a univariate analysis for each level revealed a significant gender difference on  $d'$  at 1-back,  $F(1,3295) = 19.58$ ,  $p < 0.001$ , given that girls ( $M = 2.34$ ;  $SD = 0.85$ ) showed a higher  $d'$  than boys ( $M = 2.19$ ,  $SD = 0.85$ ). However, at 2-back, girls ( $M = 1.31$ ,  $SD = 0.93$ ) and boys ( $M = 1.25$ ,  $SD = 0.93$ ) obtained similar  $d'$ ,  $F(1,3295) = 2.29$ ,  $p = 0.13$ .

### Response Times

Only response times for hits were analyzed. A 5% of participants were excluded from this analysis because they did not have hits at the 2-back level. These participants were not distributed evenly across all age groups: 97 children were excluded from the 7-year-old group, 21 from the 8-year-old group, 17 from the 9-year-old group, 14 from the 10-year-old group, 11 from the 11-year-old



**FIGURE 3 | D-prime scores ( $d'$ ) as a function of age in 1-back and 2-back conditions.** Error bars indicate two standard error of the mean.



**FIGURE 4 | Response time as a function of age in 1-back and 2-back conditions.** Error bars indicate two standard error of the mean.

group, 7 from the 12-year-old group, and 8 from the 13-year-old group.

The results of the mixed ANOVA revealed significant main effects for all variables (see **Table 2**). Load had a significant effect, since response times at 1-back were lower ( $M = 821.66$ ;  $SD = 277.95$ ) than at 2-back ( $M = 1014.32$ ;  $SD = 432.22$ ). An age effect was also found (see **Figure 4**). A trend analysis showed a significant linear component for age that explained 97.7% of the variance. Finally, there were a significant gender effect that was qualified by a significant interaction between load and gender. Although boys ( $M = 887.78$ ,  $SD = 283.82$ ) tended to respond quicker than girls ( $M = 970.94$ ,  $SD = 284.44$ ), this difference was greater at 2-back (106 ms) than at 1-back (73 ms) level. No other interaction was significant ( $p = 0.154$ ). *Post hoc* comparison with the Bonferroni correction revealed significant differences between all age groups ( $p < 0.001$ ), except between 11 and 12 years where no differences were found ( $p = 1$ ).

Given that girls outperformed boys but took longer time to respond, there was the possibility of a speed-accuracy trade-off. To assess this explanation a complementary analysis of the inverse efficiency scores was performed (Townsend and Ashby, 1983). The inverse efficiency score is a composite index of response time and accuracy that takes speed-accuracy trade-offs into account. This score was calculated by dividing the mean correct response

times by the proportion of hits for each participant and for each level of n-back. Then, a 2 (n-back load)  $\times$  7 (age)  $\times$  2 (gender) ANOVA was conducted on the inverse efficiency score. The expected main effects of load,  $F(1,3107) = 724.96$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.19$ , and age,  $F(6,3107) = 51.58$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.09$ , as well the interaction between load and age,  $F(6,3107) = 23.40$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.043$  were significant. More importantly, the main effect of gender reached significance,  $F(1,3107) = 6.90$ ,  $p = 0.009$ ,  $\eta_p^2 = 0.002$ , with girls showing a higher inverse efficiency score (2225) than boys (2030). Thus, the gender effect was again obtained with an index uncontaminated by possible speed-accuracy trade-offs.

## Normative Data

A set of multiple regression analyses on accuracy were carried out. The predictors were age (in months), age squared, and gender. By including age squared it was possible to assess the quadratic effects of age. Age was centered to avoid multicollinearity between predictors before calculating the quadratic term. Gender was coded as 1 for boys and 2 for girls.

The analyses of each level included all participants who carried out that level of the task. To analyze total hits, all participants were included and the hits for each level of load were summed. If participants had not performed the task beyond a certain level (e.g., 2-back), they were given a score equal to 0 on that and higher levels.

Regression models for the percentage of hits at each n-back level as well as for total hits are shown in **Table 3**. Results reveal that age was the strongest predictor, contributing to an increase in performance at all n-back levels as well as for total hits. The quadratic age component also predicted performance, but only at 1-back. This component showed that age-related increases in performance diminished with age. Gender was found to be a predictor only in the 1-back condition and for total hits. Girls obtained a higher number of hits than boys at both levels of the task.

Normative data stratified by age and gender for total hits and those obtained at each n-back level are shown in **Table 4**.

## Discussion

The n-back task is a WM measure that is being used more frequently in research, clinical settings and in training studies. This research aimed to provide normative data collected from a large-scale cross-sectional study on a wide sample of children and adolescents aged 7–13 years. Participants were individually administered a computerized verbal n-back task with three levels of load: 1-back, 2-back, and 3-back. The results revealed clear effects in relation to the task's level of difficulty as well as consistent age patterns in the different variables analyzed.

The results showed an overall improvement in performance with age, which, in general, translated into an increase in the percentage of hits and in the discrimination parameter  $d'$  accompanied by a decrease in response times and in the percentage of false alarms. However, slightly different patterns were observed depending on the dependent variable and the task's level of difficulty.

**TABLE 3 | Multiple linear regression models of percentage of total hits and percentage of hits on each version of the task with gender, age, and age squared as predictors.**

	Variable	B	SE	Standardized $\beta$	t	p	R <sup>2</sup>
Total (n = 3722)	(Constant)	19.05	0.50		37.94	0.000	
	Gender	0.93	0.30	0.05	3.09	0.002	
	Age	0.16	0.01	0.37	24.32	0.000	
	Age <sup>2</sup>	-0.01	0.00	-0.03	-1.68	0.092	0.14
1-back (n = 3722)	(Constant)	66.21	1.14		58.02	0.000	
	Gender	2.68	0.69	0.06	3.91	0.000	
	Age	0.27	0.01	0.28	17.52	0.000	
	Age <sup>2</sup>	-0.01	0.01	-0.05	-3.42	0.001	0.08
2-back (n = 3296)	(Constant)	49.09	1.38		35.59	0.000	
	Gender	0.83	0.82	0.02	1.01	0.317	
	Age	0.39	0.02	0.35	2.93	0.000	
	Age <sup>2</sup>	-0.01	0.01	-0.01	-0.61	0.541	0.12
3-back (n = 2141)	(Constant)	5.54	1.48		34.24	0.000	
	Gender	0.45	0.88	0.01	0.51	0.611	
	Age	0.23	0.02	0.24	1.73	0.000	
	Age <sup>2</sup>	0.00	0.01	-0.01	-0.20	0.844	0.06

Gradual improvement in performance was observed in both hits and the discrimination index,  $d'$ , across the different age groups, which was more pronounced at 2-back than at 1-back. This implies that differences between both levels decrease progressively with age. Thus, while at age 7 the percentage of hits was about 66% in the easiest condition and 37% in the most difficult, at age 13 the difference between both levels decreased from 80% at 1-back to 65% at 2-back. This suggests that the age-related improvement was more pronounced at the 2-back level than at 1-back. The regression analyses showed a linear effect of age at the more complex levels (2-back and 3-back), while at 1-back a quadratic effect was also found. Other authors have reported that, at 1-back, a mature level is reached at 10–12 years, whereas at 2-back performance continues to improve even after adolescence (Brahmbhatt et al., 2010; Schleepen and Jonkman, 2010).

Interestingly, the age-related patterns found in hits were not observed in false alarms. In the case of hits at 1-back, there were only small differences between the older age groups. In contrast, the number of false alarms experienced a parallel decline with age at both n-back levels. This discrepancy implies that the mirror effect, a regularly occurring phenomenon in recognition memory tasks (Glanzer and Adams, 1985), does not occur across the different age levels included in this study. The mirror effect is observed on single “yes-no” recognition tasks when a factor that increases the hit rate decreases the false alarm rate or *vice versa*. For example, the advantage that word concreteness produces on recognition hit-rate is mirrored in the false alarms rate. In this study, in which age was the factor considered, it was found that the relationship between the ability to accurately recognize and reject items in the 1-back condition differed across the age groups.

The reduction of the load effect on hit-rate observed across the age groups was not found for response times. Thus, the difference between the times taken on both levels of difficulty was similar across all age groups. Other studies (e.g., Schleepen and Jonkman, 2010) have found a similar discrepancy between both variables.

The disparity in the pattern between time and accuracy may indicate some independence in the development of factors related to storage and speed. In this regard, Bayliss et al. (2005) have suggested that both aspects can develop independently.

The results described above may have different but not mutually exclusive explanations. First, it is clear that both 1-back and 2-back involve different storage demands. The 1-back level requires one item to be stored in memory whereas 2-back requires two. In addition, both levels entail varying demands of information manipulation. Second, the 2-back condition requires greater executive control demands than the 1-back condition. It has been found that when executive control demands are not excessively high, children aged 11 may perform well (e.g., Best et al., 2009). However, when the conditions are more complex, clear improvements in performance in older age groups are detected, suggesting further executive development (e.g., Gathercole, 1999; Huizinga et al., 2006).

Also related to changes in executive control, Schleepen and Jonkman (2010) have reported that developmental differences are specifically due to an improvement in interference control. In support of this idea, these authors have found significant correlations between the difference in scores obtained on 1-back and 2-back levels and performance in a flanker task used as a measure of interference control. Thus, developmental changes in the ability to manage interference in WM could be a factor contributing to the age-related differences shown in this study.

An additional possible explanation for the age trends previously described is related to the ability to switch the focus of attention. At 1-back the last item presented, which is still held active in the focus of attention, should be compared with the new item; however, at 2-back the penultimate item must be retrieved into the focus of attention in order to compare it with the new one (e.g., McElree, 2001). Studies with older people have found age-related differences on n-back performance, which have been attributed to the increased cost of switching the focus of attention with age (e.g., Verhaeghen and Basak, 2005; Van Gerven et al., 2008).

**TABLE 4 | Normative data for total hits and for hits on each version of the task stratified by gender and age (7, 8, 9, 10, 11, 12, 13).**

		Boys (n = 1886)											Girls (n = 1836)																	
		Age group											Age group																	
		7	8	9	10	11	12	13	7	8	9	10	11	12	13	7	8	9	10	11	12	13								
		(n = 193)	(n = 285)	(n = 310)	(n = 297)	(n = 315)	(n = 253)	(n = 233)	(n = 194)	(n = 307)	(n = 296)	(n = 321)	(n = 286)	(n = 223)	(n = 209)	(n = 193)	(n = 285)	(n = 310)	(n = 297)	(n = 315)	(n = 253)	(n = 233)	(n = 194)	(n = 307)	(n = 296)	(n = 321)	(n = 286)	(n = 223)	(n = 209)	
Total	M	14.45	16.71	17.63	20.23	21.94	22.5	24.82	14.25	15.58	18.70	21.07	23.24	24.70	27.20	14.45	16.71	17.63	20.23	21.94	22.5	24.82	14.25	15.58	18.70	21.07	23.24	24.70	27.20	
	SD	8.49	9.20	9.17	9.44	9.90	9.66	9.85	8.58	8.41	9.29	9.10	9.28	9.21	8.74	8.49	9.20	9.17	9.44	9.90	9.66	9.85	8.58	8.41	9.29	9.10	9.28	9.21	8.74	
	Percentile	5	3	4	4	4	4	4	2	3	4	5	5	7	11	5	3	4	4	4	4	4	2	3	4	5	5	7	11	
		25	7	10	13	14	14	18	8	10	12	13	16	18	23	25	7	10	13	14	14	18	8	10	12	13	16	18	23	
		50	13	15	21	24	24	28	13	15	18	22	25	26	30	50	13	15	21	24	24	28	13	15	18	22	25	26	30	
		75	22	25	28	30	30	32	20	22	26	28	31	31	34	75	22	25	28	30	30	32	20	22	26	28	31	31	34	
		95	28	32	35	35	37	37	30	30	34	35	35	38	39	95	28	32	35	35	37	37	30	30	34	35	35	38	39	
1-back	M	8.05	8.82	9.14	9.72	10.01	10.20	10.28	8.21	8.60	9.54	10.11	10.41	10.73	11.29	8.05	8.82	9.14	9.72	10.01	10.20	10.28	8.21	8.60	9.54	10.11	10.41	10.73	11.29	
	SD	3.04	3.09	2.82	2.96	2.98	2.96	2.96	3.29	3.09	3.12	2.86	2.77	2.55	2.34	3.04	3.09	2.82	2.96	2.98	2.96	2.96	3.29	3.09	3.12	2.86	2.77	2.55	2.34	
	Percentile	5	3	4	4	4	4	4	2	3	4	5	5	6	7	5	3	4	4	4	4	4	2	3	4	5	5	6	7	
		25	6	7	8	8	8	9	6	6	7	8	9	9	10	25	6	7	8	8	8	9	6	6	7	8	9	9	10	
		50	8	9	10	11	11	11	8	9	10	11	11	11	12	50	8	9	10	11	11	11	8	9	10	11	11	11	11	12
		75	10	11	12	12	13	12	11	11	12	12	12	13	13	75	10	11	12	12	13	12	11	11	12	12	12	13	13	13
		95	13	13	14	14	14	14	14	13	14	14	14	14	14	95	13	13	14	14	14	14	14	13	14	14	14	14	14	14
2-back	M	4.10	5.02	5.23	6.29	6.96	7.26	8.16	3.96	4.53	5.74	6.52	7.44	7.93	9.11	4.10	5.02	5.23	6.29	6.96	7.26	8.16	3.96	4.53	5.74	6.52	7.44	7.93	9.11	
	SD	3.64	3.67	3.81	3.78	4.08	3.91	3.99	3.59	3.49	3.76	3.50	3.68	3.77	3.75	3.64	3.67	3.81	3.78	4.08	3.91	3.99	3.59	3.49	3.76	3.50	3.68	3.77	3.75	
	Percentile	5	-	-	-	-	-	-	-	-	-	-	-	-	2	5	-	-	-	-	-	-	-	-	-	-	-	-	2	
		25	2	2	3	4	5	6	-	1	3	4	5	6	7	25	2	2	3	4	5	6	-	1	3	4	5	6	7	
		50	4	5	6	7	8	9	3	5	6	7	8	8	10	50	4	5	6	7	8	9	3	5	6	7	8	8	10	
		75	7	8	9	10	11	11	6	7	9	9	10	11	12	75	7	8	9	10	11	11	6	7	9	9	10	11	12	12
		95	11	11	12	13	13	13	11	10	12	12	13	13	14	95	11	11	12	13	13	13	11	10	12	12	13	13	13	14
3-back	M	2.30	2.87	3.26	4.23	4.97	5.04	6.37	2.07	2.45	3.41	4.44	5.38	6.05	6.80	2.30	2.87	3.26	4.23	4.97	5.04	6.37	2.07	2.45	3.41	4.44	5.38	6.05	6.80	
	SD	3.41	3.81	3.91	4.15	4.27	4.12	4.27	3.46	3.46	3.98	4.24	4.28	4.36	4.11	3.41	3.81	3.91	4.15	4.27	4.12	4.27	3.46	3.46	3.98	4.24	4.28	4.36	4.11	
	Percentile	5	-	-	-	-	-	-	-	-	-	-	-	-	-	5	-	-	-	-	-	-	-	-	-	-	-	-	-	
		25	-	-	-	-	-	2	-	-	-	-	-	-	5	25	-	-	-	-	-	-	-	-	-	-	-	-	5	
		50	-	-	4	6	6	7	-	-	-	5	6	7	8	50	-	-	4	6	6	7	-	-	-	-	-	-	8	
		75	5	6	7	8	8	10	4	5	7	8	9	10	10	75	5	6	7	8	8	10	4	5	7	8	9	10	10	
		95	9	10	11	11	11	12	10	10	10	11	12	12	13	95	9	10	11	11	11	12	10	10	10	11	12	12	13	



Recently, Lendínez et al. (2015) have shown that the ability to accurately retrieve information from outside the focus of attention changes across childhood and adolescence. Therefore, changes in the ability for focus switching could contribute to explain the age trends observed in this study.

Finally, the age patterns described may also reflect the differential contribution of familiarity and recognition mechanisms to the different levels of the n-back task. Responses in the 1-back level may be based on familiarity, whereas at 2-back and at higher levels, recollection is also necessary (e.g., Oberauer, 2005). It might be possible that changes related to the familiarity mechanism are more moderate than those of the recollection process. This pattern is also consistent with the results obtained with episodic tasks. Although changes in both processes have been observed since early childhood, familiarity seems to stabilize by mid-childhood, whereas recollection continues to improve during adolescence (for a review, see Ghetti and Lee, 2013).

To summarize, changes in executive control may be mediated by developmental improvements relating to different mechanisms such as interference resolution, focus switching, and even recollection. Note however, that we have not tested the different possible theoretical explanations of the observed age trends, since our aim was to provide normative data for different age groups. Hence, the investigation of the relative merits of these explanations may be a relevant objective for future studies. In addition, it would also be important to determine the specific relationship between performance at the different n-back levels and other complex WM measures for different age groups.

In this study, boys tended to respond more quickly than girls, although with less accuracy, committing more false alarms and obtaining fewer hits. These slight differences persisted across all ages. Even though the effect size of these differences was not large ( $\eta_p^2$  between 0.001 and 0.018), the differences would suggest that boys exhibit greater impulsiveness (see also Vuontela et al., 2003). The findings are consistent with other studies reporting differences in favor of girls in some WM and STM tasks involving verbal processing (e.g., Lowe et al., 2003; Lynn and Irwing, 2008; however, for no differences see Nagel et al., 2007; Lejbak et al., 2011). Vuontela et al. (2003) have suggested that these gender differences could be due to differences in neurodevelopment. It has also been argued that boys and girls could use different strategies while performing the task (Speck et al., 2000).

The current study provided normative data on children and adolescents aged 7–13 years performing a computerized n-back task. The sample size allowed for accurate estimations of typical performance for the different age groups under study, and to determine at what age children were able to deal with the different levels of the task. In this regard, practically all children performed adequately at 1-back. However, a considerable number of 7-year-olds did not manage to complete the 2-back level or only achieved low levels of performance. Task difficulty increased notably at 3-back, given that most of the 7-year-olds children were not able to complete it. The results of the regression analyses indicated that the 2-back was the level that explained age-related performance best. Future studies should include older age groups in order to assess when asymptotic values are reached for the different levels of difficulty. It is possible that better performance at higher n-back

levels might not be reached until after adolescence, a time when executive functioning seems to reach maturity (Best and Miller, 2010).

In addition to the clear age patterns found, high variability was observed in performance across the different ages. For example, at 1-back, there were participants from all age groups who obtained all possible scores. At 2-back the variability was reduced to some extent given that a number of younger children (i.e., aged 7) were unable to perform the task, and those who were, did not achieve the highest scores. Despite the difficulties of the task for the younger children, there was evidence of 7-year-olds performing the task at a level similar to that of many 13-year-old children. An analogous pattern emerged at 3-back. Although most children younger than 9 were unable to reach this level, nor did many of those who were older, there were participants across all ages who did manage to perform it.

These normative data could be used for the assessment of verbal WM and the updating function. Individual differences in WM capacity have important consequences for children's ability to acquire knowledge and new skills (see Swanson and Alloway, 2012, for a review). The ability to update information has also been shown to have an important role in accounting for individual differences in different areas of academic attainment (e.g., Passolunghi and Pazzaglia, 2004; Swanson and Beebe-Frankenberger, 2004; St Clair-Thompson and Gathercole, 2006; Van der Sluis et al., 2007; Carretti et al., 2009; Cornoldi et al., 2012; Lechuga et al., 2015; Pelegrina et al., 2015). Low scores for a specified age on the n-back task could indicate that this system, involved in numerous complex cognitive abilities, is functioning poorly. Therefore, these normative data could contribute to detect children at risk of academic difficulties and thus support early intervention programs and strategies.

These normative data may also be valuable in the context of WM training programs. There is some evidence that WM capacity can be increased by training, at least in the short-term (e.g., Jaeggi et al., 2014). The effects of WM training may also transfer to academic attainment (for a review, see Titz and Karbach, 2014). The n-back is a task frequently used to train WM and to assess WM capacity in both children and adults (see Melby-Lervag and Hulme, 2012). The task presented in this study could be used to obtain baseline scores and to determine the possible changes after training. Baseline scores and possible variations after the training could be readily interpreted in relation to the corresponding age group.

There are some limitations in the present study that suggest ideas for future research. First, it should be noted that the variables considered in the regression analyses explained a small portion of the variance. This suggests that additional factors are influencing performance on this task as well as individual differences. The n-back task requires multiple executive control processes and these could be a source of performance variability. It has been shown that executive functions component processes may develop at different rates (Huizinga et al., 2006; Best et al., 2009; Xu et al., 2013). Thus, it might be possible that different processes involved in the task contribute to the observed age patterns. Further studies should analyze the specific cognitive processes that are involved in this task.

Furthermore, although the cross-sectional design may be useful to identify age differences, only a longitudinal design permits the capture of intra-individual changes over time. Thus, for instance, it might be relevant to analyze different developmental patterns (e.g., early- or late-maturing children). In addition, a longitudinal design would be informative not only in regards to the individual differences in performance shown in this study, but also in regards to their possible changes over time. Moreover, this design would also permit determining the relationship over time between some cognitive processes and performance on the n-back task.

Finally, it should also be relevant to compare the results obtained with different n-back tasks. The n-back task we administered followed a procedure analogous to that used in other studies. However, given that there is no standard procedure for this task, it should be appropriate to investigate the extent to which these results can be generalized to other versions.

To conclude, the results indicate age-related changes that might be related to different cognitive processes involved in

updating information in WM. The normative data obtained may be useful in interpreting the results of other studies in which the n-back task has been applied under similar conditions. Furthermore, given the type of stimuli used, this information might be valuable even when the task is applied in other languages. The normative data collected in this study may be relevant on a practical level, given that they provide information about performance on a WM task that taps executive processes that are central to cognitive development and learning.

## Acknowledgments

The research reported here was supported by a grant from the Spanish Ministry of Science and Innovation (CSD2008-00048) in the CONSOLIDER-INGENIO 2010 programme and two grants from the Spanish Ministry of Science and Technology: PSI2012-37764 and PSI2012-33625.

## References

- Alloway, T. P., and Alloway, R. G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *J. Exp. Child Psychol.* 106, 20–29. doi: 10.1016/j.jecp.2009.11.003
- Baddeley, A. D. (2007). *Working Memory, Thought and Action*. Oxford: Oxford University Press.
- Baddeley, A. D., and Logie, R. H. (1999). “Working memory: the multiple-component model,” in *Models of Working Memory*, eds A. Miyake and P. Shah (New York: Cambridge University Press), 28–61.
- Bayliss, D., Jarrold, C., Baddeley, A. D., Gunn, D. M., and Leigh, E. (2005). Mapping the developmental constraints on working memory span performance. *Dev. Psychol.* 41, 579–597. doi: 10.1037/0012-1649.41.4.579
- Belachi, C., Carretti, B., and Cornoldi, C. (2010). The role of working memory and updating in Coloured Raven Matrices performance in typically developing children. *J. Cogn. Psychol.* 22, 1010–1020. doi: 10.1080/09541440903184617
- Best, J. R., and Miller, P. H. (2010). A developmental perspective on executive function. *Child Dev.* 81, 1641–1660. doi: 10.1111/j.1467-8624.2010.01499.x
- Best, J. R., Miller, P. H., and Jones, L. L. (2009). Executive functions after age 5: changes and correlates. *Dev. Rev.* 29, 180–200. doi: 10.1016/j.dr.2009.05.002
- Brahmbhatt, S. B., White, D. A., and Barch, D. M. (2010). Developmental differences in sustained and transient activity underlying working memory. *Brain Res.* 1354, 140–151. doi: 10.1016/j.brainres.2010.07.055
- Cain, K. (2006). “Children’s reading comprehension: the role of working memory in normal and impaired development,” in *Working Memory and Education*, ed. S. J. Pickering (Amsterdam: Academic Press), 61–91.
- Carretti, B., Borella, E., Cornoldi, C., and De Beni, R. (2009). Role of working memory in explaining the performance of individuals with specific reading comprehension difficulties: a metaanalysis. *Learn. Individ. Differ.* 19, 246–251. doi: 10.1016/j.lindif.2008.10.002
- Case, R., Kurland, M., and Goldberg, J. (1982). Operational efficiency and the growth of short-term memory. *J. Exp. Child Psychol.* 33, 386–404. doi: 10.1016/0022-0965(82)90054-6
- Colom, R., Abad, F. J., Quiroga, A., Shih, P. C., and Flores-Mendoza, C. (2008). Working memory and intelligence are highly related construct, but why? *Intelligence* 36, 584–606. doi: 10.1016/j.intell.2008.01.002
- Conway, A. R. A., Kane, M. J., and Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends Cogn. Sci.* 7, 547–552. doi: 10.1016/j.tics.2003.10.005
- Cornoldi, C., Drusi, S., Tencati, C., Giofrè, D., and Mirandola, C. (2012). Problem solving and working memory updating difficulties in a group of poor comprehenders. *J. Cogn. Educ. Psychol.* 11, 39–44. doi: 10.1891/1945-8959.11.1.39
- Cowan, N. (1995). *Attention and Memory: An Integrated Framework*. New York: Oxford University Press.
- Daneman, M., and Carpenter, P. A. (1980). Individual differences in working memory and reading. *J. Verbal Learn. Verbal Behav.* 19, 450–466. doi: 10.1016/S0022-5371(80)90312-6
- Deschuyteneer, M., Vandierendonck, A., and Muylaert, I. (2006). Does solution of mental arithmetic problems such as  $2 + 6$  and  $3 \times 8$  rely on the process of “memory updating”? *Exp. Psychol.* 53, 198–208. doi: 10.1027/1618-3169.53.3.198
- Elosúa, M. R., García Madruga, J. A., Gutiérrez, F., Luque, J. L., and Gárate, M. (1997). Un estudio sobre las diferencias evolutivas en memoria operativa: capacidad o eficiencia? *Estudios Psicol.* 58, 15–27. doi: 10.1174/021093997320954827
- Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., De Fries, J. C., and Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychol. Sci.* 17, 172–179. doi: 10.1111/j.1467-9280.2006.01681.x
- Friedman, N. P., Miyake, A., Young, S. E., De Fries, J. C., Corley, R. P., and Hewitt, J. K. (2008). Individual differences in executive functions are almost entirely genetic in origin. *J. Exp. Psychol. Gen.* 137, 201–225. doi: 10.1037/0096-3445.137.2.201
- García-Madruga, J. A., Gárate, M., Elosúa, M. R., Luque, J. L., and Gutiérrez, F. (1997). Comprensión lectora y memoria operativa: un estudio evolutivo. [Reading Comprehension and working memory: a developmental study]. *Cognitiva* 9, 99–132.
- García-Madruga, J. A., Gutiérrez, F., Carriedo, N., Luzón, J. M., and Vila, J. O. (2005). “Working memory and propositional reasoning: searching for new working memory tests,” in *The Shape of Reason. Essays in Honour of Paolo Legrenzi*, eds V. Girotto and P. N. Johnson-Laird (London: Psychology Press), 69–89.
- García-Madruga, J. A., Gutiérrez, F., Carriedo, N., Luzón, J. M., and Vila, J. O. (2007). Checking the role of central executive in propositional reasoning. *Think. Reason.* 13, 370–393.
- Gathercole, S. E. (1999). Cognitive approaches to the development of short-term memory. *Trends Cogn. Sci.* 3, 410–418. doi: 10.1016/S1364-6613(99)01388-1
- Gathercole, S. E., Pickering, S. J., Ambridge, B., and Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Dev. Psychol.* 40, 177–190. doi: 10.1037/0012-1649.40.2.177
- Ghetti, S., and Lee, I. K. (2013). “The development of recollection and familiarity during childhood,” in *The Wiley Handbook on The Development of Children’s Memory*, eds P. J. Bauer and R. Fivush (Oxford: Wiley-Blackwell), 309–335.
- Glanzer, M., and Adams, J. K. (1985). The mirror effect in recognition memory. *Mem. Cogn.* 13, 8–20. doi: 10.3758/BF03198438
- Huizinga, M., Dolan, C. V., and Van der Molen, M. W. (2006). Age-related change in executive function: developmental trends and a latent variable analysis. *Neuropsychologia* 44, 2017–2036. doi: 10.1016/j.neuropsychologia.2006.01.010

- Jaeggi, S. M., Buschkuhl, M., Perrig, W. J., and Meier, B. (2010). The concurrent validity of the N-back task as a working memory measure. *Memory* 18, 394–412. doi: 10.1080/09658211003702171
- Jaeggi, S. M., Buschkuhl, M., Shah, P., and Jonides, J. (2014). The role of individual differences in cognitive training and transfer. *Mem. Cogn.* 42, 464–480. doi: 10.3758/s13421-013-0364-z
- Kane, M., Conway, A., Miura, T., and Colflesh, G. (2007). Working memory, attention control, and the n-back task: a question of construct validity. *J. Exp. Psychol. Learn. Mem. Cogn.* 33, 615–622. doi: 10.1037/0278-7393.33.3.615
- Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *J. Exp. Psychol.* 55, 352–358. doi: 10.1037/h0043688
- Kwon, H., Reiss, A. L., and Menon, V. (2002). Neural basis of protracted developmental changes in visuo-spatial working memory. *Proc. Natl. Acad. Sci. U.S.A.* 99, 13336–13341. doi: 10.1073/pnas.162486399
- Lechuga, M. T., Moreno, V., Pelegrina, S., Gómez-Ariza, C., and Bajo, M. T. (2006). Age differences in memory control: evidence from updating and retrieval-practice tasks. *Acta Psychol.* 123, 279–298. doi: 10.1016/j.actpsy.2006.01.006
- Lechuga, M. T., Pelegrina, S., Pelaez, J. L., Martín-Puga, M. E., and Justicia, M. J. (2015). Working memory updating as a predictor of academic attainment. *Educ. Psychol.* doi: 10.1080/01443410.2014.950193 (advance online publication.)
- Lejbak, L., Crossley, M., and Vrbancic, M. (2011). A male advantage for spatial and object but not verbal working memory using the n-back task. *Brain Cogn.* 76, 191–196. doi: 10.1016/j.bandc.2010.12.002
- Lendínez, C., Pelegrina, S., and Lechuga, M. T. (2015). Age differences in working memory updating. The role of interference, focus switching and substituting information. *Acta Psychol.* 157, 106–113. doi: 10.1016/j.actpsy.2015.02.015
- Lowe, P. A., Mayfield, J. W., and Reynolds, C. R. (2003). Gender differences in memory test performance among children and adolescents. *Arch. Clin. Neuropsychol.* 18, 865–878. doi: 10.1093/arclin/18.8.865
- Luciana, M., Conklin, H., Hooper, C., and Yarger, R. (2005). The development of nonverbal working memory and executive control processes in adolescents. *Child Dev.* 76, 697–712. doi: 10.1111/j.1467-8624.2005.00872.x
- Luna, B., Garver, K. E., Urban, T. A., Lazar, N. A., and Sweeney, J. A. (2004). Maturation of cognitive processes from late childhood to adulthood. *Child Dev.* 75, 1357–1372. doi: 10.1111/j.1467-8624.2004.00745.x
- Lynn, R., and Irwing, P. (2008). Sex differences in mental arithmetic, digit span, and g defined as working memory capacity. *Intelligence* 36, 226–235. doi: 10.1016/j.intell.2007.06.002
- McElree, B. (2001). Working memory and focal attention. *J. Exp. Psychol. Learn. Mem. Cogn.* 27, 817–835. doi: 10.1037/0278-7393.27.3.817
- Melby-Lervag, M., and Hulme, C. (2012). Is working memory training effective? A meta-analytic review. *Dev. Psychol.* 49, 270–291. doi: 10.1037/a0028228
- Nagel, B., Ohannissian, A., and Cummins, K. (2007). Performance dissociation during verbal and spatial working memory tasks. *Percept. Mot. Skills* 105, 243–250. doi: 10.2466/pms.105.5.243-250
- Oberauer, K. (2002). Access to information in working memory: exploring the focus of attention. *J. Exp. Psychol. Learn. Mem. Cogn.* 28, 411–421. doi: 10.1037/0278-7393.28.3.411
- Oberauer, K. (2005). Binding and inhibition in working memory: individual and age differences in short-term recognition. *J. Exp. Psychol. Gen.* 134, 368–387. doi: 10.1037/0096-3445.134.3.368
- Oberauer, K. (2006). Is the focus of attention in working memory expanded through practice? *J. Exp. Psychol. Learn. Mem. Cogn.* 32, 197–214. doi: 10.1037/0278-7393.32.2.197
- Owen, A. M., McMillan, K. M., Laird, A. R., and Bullmore, E. (2005). n-back working memory paradigm: a meta-analysis of normative functional neuroimaging studies. *Hum. Brain Mapp.* 25, 46–59. doi: 10.1002/hbm.20131
- Passolunghi, M. C., and Pazzaglia, F. (2004). Individual differences in memory updating in relation to arithmetic problem solving. *Learn. Individ. Differ.* 14, 219–230. doi: 10.1016/j.lindif.2004.03.001
- Pelegrina, S., Capodici, A., Carretti, B., and Cornoldi, C. (2015). Magnitude representation and working memory updating in children with arithmetic and reading comprehension disabilities. *J. Learn. Disabil.* doi: 10.1177/0022219414527480 [Epub ahead of print].
- Redick, T. S., and Lindsey, D. R. B. (2013). Complex span and n-back measures of working memory: a meta-analysis. *Psychon. Bull. Rev.* 20, 1102–1113. doi: 10.3758/s13423-013-0453-9
- Robinson, B., and Fuller, B. (2004). *N-Back Test [Computer software]*. Available at: <http://step.psy.cmu.edu/scripts-plus/> (accessed June 21, 2012)
- Schleepen, T. M. J., and Jonkman, L. M. (2010). The development of non-spatial working memory capacity during childhood and adolescence and the role of interference control: an n-back task study. *Dev. Neuropsychol.* 35, 37–56. doi: 10.1080/87565640903325733
- Schneider, W., Eschman, A., and Zuccolotto, A. (2002). *E-Prime Users' Guide*. Pittsburgh, PA: Psychology Software Tools, Inc.
- Siegel, L. S. (1994). Working memory and reading: a life-span perspective. *Int. J. Behav. Dev.* 17, 109–124. doi: 10.1177/016502549401700107
- Speck, O., Ernst, T., Braun, J., Koch, C., Miller, E., and Chang, L. (2000). Gender differences in the functional organization of the brain for working memory. *Neuroimage* 11, 2581–2585. doi: 10.1097/00001756-200008030-00046
- Stanislaw, H., and Todorov, N. (1999). Calculation of signal detection theory measures. *Behav. Res. Methods Inst. Comput.* 31, 137–149. doi: 10.3758/BF03207704
- St Clair-Thompson, H. L., and Gathercole, S. E. (2006). Executive functions and achievements in school: shifting, updating, inhibition, and working memory. *Q. J. Exp. Psychol.* 59, 745–759. doi: 10.1080/17470210500162854
- Swanson, H. L., and Alloway, T. P. (2012). “Working memory, learning, and academic achievement,” in *APA Educational Psychology Handbook*, Vol. 1, *Theories, Constructs, and Critical Issues*, eds K. R. Harris, S. Graham, T. C. Urdan, B. McCormick Christine, G. M. Sinatra, and J. Sweller (Washington, DC: American Psychological Association), 327–366.
- Swanson, H. L., and Beebe-Frankenberger, M. (2004). The relationship between working memory and mathematical problem solving in children at risk and not at risk for math disabilities. *J. Educ. Psychol.* 96, 471–491. doi: 10.1037/0022-0663.96.3.471
- Titz, C., and Karbach, J. (2014). Working memory and executive functions: effects of training on academic abilities. *Psychol. Res.* 78, 852–868. doi: 10.1007/s00426-013-0537-1
- Townsend, J. T., and Ashby, F. G. (1983). *The Stochastic Modeling of Elementary Psychological Processes*. Cambridge: Cambridge University Press.
- Turner, M. L., and Engle, R. W. (1989). Is working memory capacity task dependent? *J. Mem. Lang.* 28, 127–154. doi: 10.1016/0749-596X(89)90040-5
- Van der Sluis, S., de Jong, P., and Van der Leij, A. (2007). Executive functioning in children, and its relations with reasoning, reading, and arithmetic. *Intelligence* 35, 427–449. doi: 10.1016/j.intell.2006.09.001
- Van Gerven, P. W. M., Meijer, W. A., Prickaerts, J. H. M., and Van der Veen, F. M. (2008). Aging and focus switching in working memory: excluding the potential role of memory load. *Exp. Aging Res.* 34, 367–378. doi: 10.1080/03610730802274165
- Van Leeuwen, M., Van den Berg, S. M., Hoekstra, R. A., and Boomsma, D. I. (2007). Endophenotypes for intelligence in children and adolescents. *Intelligence* 35, 369–380. doi: 10.1016/j.intell.2006.09.008
- Verhaeghen, P., and Basak, C. (2005). Aging and switching of the focus of attention in working memory: results from a modified n-back task. *Q. J. Exp. Psychol. A* 58, 134–154. doi: 10.1080/02724980443000241
- Vuontela, V., Steenari, M. R., Carlson, S., Koivisto, J., Fjallberg, M., and Aronen, E. T. (2003). Audiospatial and visuospatial working memory in 6-13-year-old school children. *Learn. Mem.* 10, 74–81. doi: 10.1101/lm.53503
- Xu, F., Han, Y., Sabbagh, M. A., Wang, T., Ren, X., and Li, C. (2013). Developmental differences in the structure of executive function in middle childhood and adolescence. *PLoS ONE* 8:e77770. doi: 10.1371/journal.pone.0077770

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Pelegrina, Lechuga, García-Madruga, Elosúa, Macizo, Carreiras, Fuentes and Bajo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.