Decoding the content of working memory in school-aged children

Nora Turoman[1*], Prosper Agbesi Fiave[1], Clélia Zahnd[1], Megan T. deBettencourt[2] and Evie Vergauwe[1]

[1] Faculty of Psychology and Educational Sciences, University of Geneva, Geneva, Switzerland

[2] Department of Psychology, University of Chicago, IL, USA

[*] Correspondence should be addressed to Nora Turoman, University of Geneva, 40 Boulevard Pont d'Arve, 1205 Geneva, Switzerland. Email: nora.turoman@unige.ch

Abstract

Developmental improvements in working memory (WM) maintenance predict many real-world outcomes, including educational attainment. It is thus critical to understand which WM mechanisms support these observed behavioral improvements, and how WM maintenance strategies might change through development. One challenge is that specific WM neural mechanisms cannot easily be measured behaviorally, especially in a child population. However, new multivariate decoding techniques have been designed, primarily in adult populations, that can sensitively decode the contents of working memory. The goal of this study was to deploy multivariate decoding techniques to decode the contents of WM in children. We created a simple computerized WM game for children, in which children maintained different categories of information (visual, spatial or verbal). We collected electroencephalography (EEG) data from 20 children (7-12-year-olds) while they played the game. Using Multivariate Pattern Analysis (MVPA) on children's EEG signals, we reliably decoded the category of the maintained information during the sensory and maintenance period. In a set of exploratory reliability and validity analyses, we examined the robustness of these results when trained on less data, and how these patterns generalized within individuals throughout the testing session. Furthermore, these results matched theory-based predictions of WM across individuals and across ages. As the first study of its kind, our proof-of-concept provides a direct and age-appropriate potential alternative to exclusively behavioral WM maintenance measures in children. Our study demonstrates the utility of MVPA to directly measure and track the spontaneously-generated representational content of children's WM.

Keywords: short-term memory, multivariate pattern analyses, child development

Decoding the content of working memory in school-aged children

Working memory (WM), the brain's limited-capacity system which temporarily maintains information that is no longer physically present (Baddeley & Hitch, 1974; Cowan, 1998; Miller, 1956), has been recognized as the primary determinant of cognitive development in children (for review see Cowan, 2016), and a key predictor of scholastic skills and academic achievement (e.g., Alloway & Alloway, 2010; Bayliss et al., 2003; Bull et al., 2008). It is known that WM performance improves with age (e.g., Gathercole, 1999; Gathercole et al., 2004; Salthouse, 1994), and the emergence of spontaneously used maintenance mechanisms in WM has been proposed as an underlying cause of such improvements (e.g., Camos & Barrouillet, 2011; Gathercole & Adams, 1994; Geier et al., 2009; Magimairaj & Montgomery, 2013; Shimi & Scerif, 2017). However, the way in which some of these maintenance mechanisms are typically measured has made it difficult to build accurate theories of their developmental trajectory. Specifically, children's WM maintenance mechanisms are usually assessed using behavioral tasks. Yet, the spontaneous use of maintenance mechanisms that occur covertly is difficult to assess behaviorally without introducing specific task manipulations. One concern is that these manipulations can bias whether the to-be-measured maintenance mechanisms can be detected, and how they appear to operate. Such methodological issues can weaken the derivation chain from hypothesis building based on extant theory, through their testing with a given set of methods, to yielding results that are used to build new theory (Meehl, 1990; Scheel et al., 2021). Since science is based on the constant continuation of this cycle, if one element is weak, this puts the inferences that can be drawn from the data at risk on a grand scale. Given the clear educational importance of understanding WM development, methods must be developed which accurately assess how children spontaneously maintain memoranda in WM, i.e., that can uncover and track children's maintained memoranda without the use of potentially interfering task manipulations.

Issues in measuring children's WM maintenance

There are several different strategies that can be used to maintain different types of information in WM for short periods of time (e.g., Bjorklund et al., 2008; Kail & Hagen, 1977). Some of these mechanisms can be measured relatively simply. For example, rehearsal, which involves subvocal repetition of the information to be remembered (Baddeley, 1986; Conrad, 1964) can be tracked by observing lip movements associated with the maintained information (e.g., Elliott et al., 2021; Flavell et al., 1966). Accordingly, there is ample evidence that rehearsal strategies are spontaneously employed without being instructed, from age 7 onwards (e.g., Baddeley et al., 1998; Ferguson et al., 2002; Flavell et al., 1966; Hitch et al., 1991). In another example, children may organize memoranda by a common category during a brief WM delay (Bower, 1970; Mandler, 2002). This organization strategy has typically been measured in children by presenting memoranda in visual forms (e.g., flashcards) that could be spatially grouped by categories (for procedure see e.g., Salatas & Flavell, 1976). Interestingly, only older children (around age 10 onwards) seem to use this strategy spontaneously (Bjorklund & de Marchena, 1984; Hasselhorn, 1992; Schleepen & Jonkman, 2012). However, even 4-year-olds (Sodian et al., 1986), 7-year-olds (Lange & Jackson, 1974), and 9-year-olds (Corsale & Ornstein, 1980) are shown to organize memoranda when task instructions are modified to emphasize the usefulness of the underlying category information. Thus, specific task instructions may alter the WM maintenance strategies and influence the measurement of WM maintenance mechanisms. This example foreshadows the complexity of measuring mechanisms that are not directly visible.

Other covert WM mechanisms cannot be easily inferred without employing specific task settings. One such covert mechanism is refreshing, which involves briefly reactivating to-be-remembered information by focusing limited internal attentional resources onto the representation (e.g., Barrouillet et al., 2004; Camos et al., 2018). Though it is not the only covert WM maintenance mechanism in existence, it has received renewed research interest (Káldi & Babarczy, 2021; Lintz & Johnson, 2021; Oberauer & Souza, 2020; Vergauwe et al., 2021; Vergauwe & Langerock, 2022), and is a fitting example mechanism to demonstrate both the necessity to manipulate task parameters to measure it, and how such manipulations can interfere with its measurement. One popular way to measure refreshing is by varying the attentional demands of a secondary processing task (Barrouillet et al., 2009; Bayliss et al., 2003; Bertrand & Camos, 2015; Camos & Barrouillet, 2011; Conlin et al., 2005; Oftinger & Camos, 2015, 2017, 2018; Tam et al., 2010). In such dual-task setups, refreshing would be indexed by declines in WM performance in the maintenance task as a function of the attentional demands of the processing task (e.g., Barrouillet et al., 2009; Bayliss et al., 2003; Camos & Barrouillet, 2011; Tam et al., 2010). A key question has been when refreshing emerges during development, with multiple studies observing an impact of the attentionally demanding processing task (interpreted as the reliance on refreshing) starting at age 7 and becoming progressively greater with age (Barrouillet et al., 2009; Gaillard, 2011; Portrat et al., 2009). However, as Vergauwe and colleagues (2021) have noted, neither the emergence of spontaneous refreshing at age 7, nor the increases in refreshing efficiency from then onto adolescence are unequivocally supported in the literature. First, the detrimental effects of attentional demands on memory performance which are characteristic for refreshing have been observed even at 4–6 years of age (Bertrand & Camos, 2015; Tam et al., 2010). Second, memory performance decreases as a result of attentionally demanding concurrent tasks have not been found to differ much between 6-year-olds and 8-year-olds (e.g., Conlin et al., 2005; Oftinger & Camos, 2015, 2017, 2018). Though dual-task paradigms have yielded mixed results on the developmental trajectory of refreshing, these paradigms have still all detected refreshing in children. Removing the secondary task, however, seems to make refreshing undetectable. Namely, in a paradigm without a dual task design, and using a different outcome measure than the above studies (based on Vergauwe & Langerock, 2017; see also Vergauwe & Langerock, 2022), Vergauwe and colleagues (2021) did not find evidence for spontaneous refreshing in children older than

7 years of age (see also Valentini & Vergauwe, 2023). However, the removal of the processing task could have had other consequences, for example rendering the paradigm less challenging for children and easier to understand. The results presented here demonstrate that specific paradigm design choices, for example the outcome measure and the inclusion of a secondary task, can influence whether or not refreshing can be detected in children. The contrast between paradigms with and without secondary tasks has the potential to upend common conceptions of WM development and refreshing as a maintenance mechanism. Combined with the results regarding the organization strategy, on a larger scale, these outcomes (as well as e.g., Brady et al., 2021) highlight the need for deeper considerations of methodology before testing theoretical assumptions in the WM field.

Moving forward: Designing better methods for assessing WM maintenance in children

Most studies on children's WM mechanisms have used behavioral measures exclusively, meaning that detecting maintenance mechanisms was reliant either on the observable production of said mechanisms (for overt mechanisms), or inferences based on results under specific task settings (for covert mechanisms). Though such approaches have been instrumental to understanding WM development, we have seen that they are limited, in that it may be challenging to disentangle whether failure to detect an effect in children is due to a) a true absence of an effect, b) the children not being motivated to do a difficult task, or c) the children not understanding the instructions of the task. Therefore, it would be powerful to develop a measure that is independent from task demands, which would allow children to behave, and process information spontaneously.

A potential solution lies in departing from behavioral-only measures. Assessing maintenance mechanisms using neural measures would remove the need for introducing secondary tasks or elaborate instructions, thus simplifying and removing sources of bias from behavioral tasks, and allow children to maintain memoranda in whichever way was most natural to them. In particular, Multivariate Pattern

Analysis (MVPA) can index different perceptual or cognitive states, by training classifiers to distinguish between different patterns of activity that are distributed across the brain (Haxby et al., 2001; Norman et al., 2006). Multivariate analyses of electroencephalography (EEG) data, has successfully decoded the contents of attended information in WM (LaRocque et al., 2013), the amount of WM load (Adam et al., 2020), individual differences in WM load and attentional focus (Karch et al., 2015), and attentional processes involved in the transfer of information into long-term memory (DeBettencourt et al., 2021), all in adults. It has less frequently been applied to neural data collected from children, though there have been some notable exceptions (e.g., Mares et al., 2020 [face processing in typical development]; Petit et al., 2020 [language processing in typical development and autism]; Lui et al., 2021 [word reading skills with Chinese characters]). To our knowledge, however, this method has yet to be used to elucidate WM processes in children, let alone WM *maintenance* processes in children.

The current study

In the current study, we aimed to develop a measure that can directly assess the spontaneously-generated content of children's WM during maintenance and track its changes over time. To do so, we posited that, instead of changing behavioral task parameters (i.e., without relying on difficult concurrent processing tasks, or hard-to-understand task instructions), we could leverage measures of brain activity during maintenance. We tested this hypothesis in a proof-of-concept study, combining a simple, child-friendly, computerized working memory task with advanced analyses of electroencephalographic (EEG) measures. Our study design allowed us to probe the spontaneously-generated representational content of children's working memory during maintenance, and throughout the course of an entire trial from encoding to response. The ability to decode the contents of WM from children was the key aim of our study, but we also went on to assess the validity and reliability of our measure through a series of exploratory analyses.

Methods

Participants

A total of 25 children (8 female, mean age = 9 years 4 months, *SD* age = 1 year 3 months, Range: 6 years 0 months – 12 years 9 months) were recruited for the present study, through personal contacts, word of mouth, and the participant database of the Working Memory, Cognition and Development lab. Of this, five participants were excluded for failing to finish the task, failing to reach chance-level behavioral accuracy (50% of the total accuracy), or excessive noise in their EEG data (that could not be cleaned from the data such at least 200 trials remain in the dataset, e.g., due to excessive movement throughout the experimental task). Thus, the final sample included 20 children (7 female, mean age = 9 years and 7 months, *SD* age = 1 year and 5months, Range: 7 years 0 months – 12 years and 2 months).

Participants were tested at the EEG lab of the Faculty of Psychology and Educational Sciences of the University of Geneva and were offered a 20 Swiss franc voucher from a popular media store. All research procedures were approved by the University of Geneva Ethical Commission (approval code: CUREG_2021-05-49). Informed consent was obtained from parents/caregivers and verbal assent was obtained from children before participating in the study.

Stimuli

Stimuli belonged to one of three categories: Visual (an image of a robot), Spatial (an image of a rocket ship in a given spatial location on a circular grid), or Verbal (an image of a French-sounding non-word). These categories were selected to reflect the most commonly used stimulus categories in WM research (visual features, spatial location, verbal stimulus), mapping onto domain-specific dissociations of working memory resources as proposed in the popular multi-component model (Baddeley & Hitch, 197; Baddeley, 1986; Baddeley & Logie, 1999). There were 16 stimuli per category. For the spatial category, there were 16 positions that a rocket could occupy on a circular display; based on the memory items from Ricker and Vergauwe (2020). For the visual category, there were 16 images of robots (generated by typing the following numbers into the 'generate' query space at Robohash.org: 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, and saved as a .png file). Finally, for the verbal category, there were 16 images of random nonwords generated by WinWordGen 1.0 (Duyck et al., 2004). These nonwords were legal bigrams with 5 letters per word drawn from Lexique.org, base language French, selected if a native French speaker confirmed it did not remind them of a real French word. All stimuli were presented centrally on a black background subtending 5 degrees of visual angle, to minimize eye movements.

Task procedure

The experimental task involved a single-item delayed-recognition task (Figure 1), based on the Phase 1 task of Experiment 1 in LaRocque et al. (2013). Importantly, participants were only told to remember the information – there were no instructions as to how to maintain the stimuli. Each trial began with an inter-trial interval (ITI, average duration 1000ms), during which participants were presented with a white fixation dot centrally on the screen. To reduce anticipation effects, the ITIs were randomly jittered between 800ms to 1200ms in steps of 50ms. Next, during the Sensory period, the to-be-memorized item was displayed for 1000ms. Stimulus category was randomly determined on each trial. Afterwards, during the maintenance (Delay) period (2000ms), a central white fixation cross appeared on the screen. Then, a probe stimulus of the same category was displayed for 1000ms. Finally, during the Response period, a white question mark was displayed centrally (until the participant responded or 2000ms, whichever was shorter), indicating to the participants to respond. If the probe stimulus matched the to-be-memorized stimulus, participants were to press the 'k' key on the keyboard in front of them. If the probe stimulus did not match the previously encoded target stimulus, participants were supposed to press the 'd' key on the keyboard. The probe stimulus matched the to-be-memorized stimulus 50% of the time. The 'k' and 'd' keys were marked with green circular stickers. Prior to the experiment, participants were told to respond when they saw the question mark on the screen. There were no instructions on which hand to use to provide responses. To minimize blinking artefacts during the Sensory (encoding) and Delay (retention) periods, participants were encouraged to withhold blinking during these times, and we turned on the light in the testing booth for participants that had trouble withholding blinking.

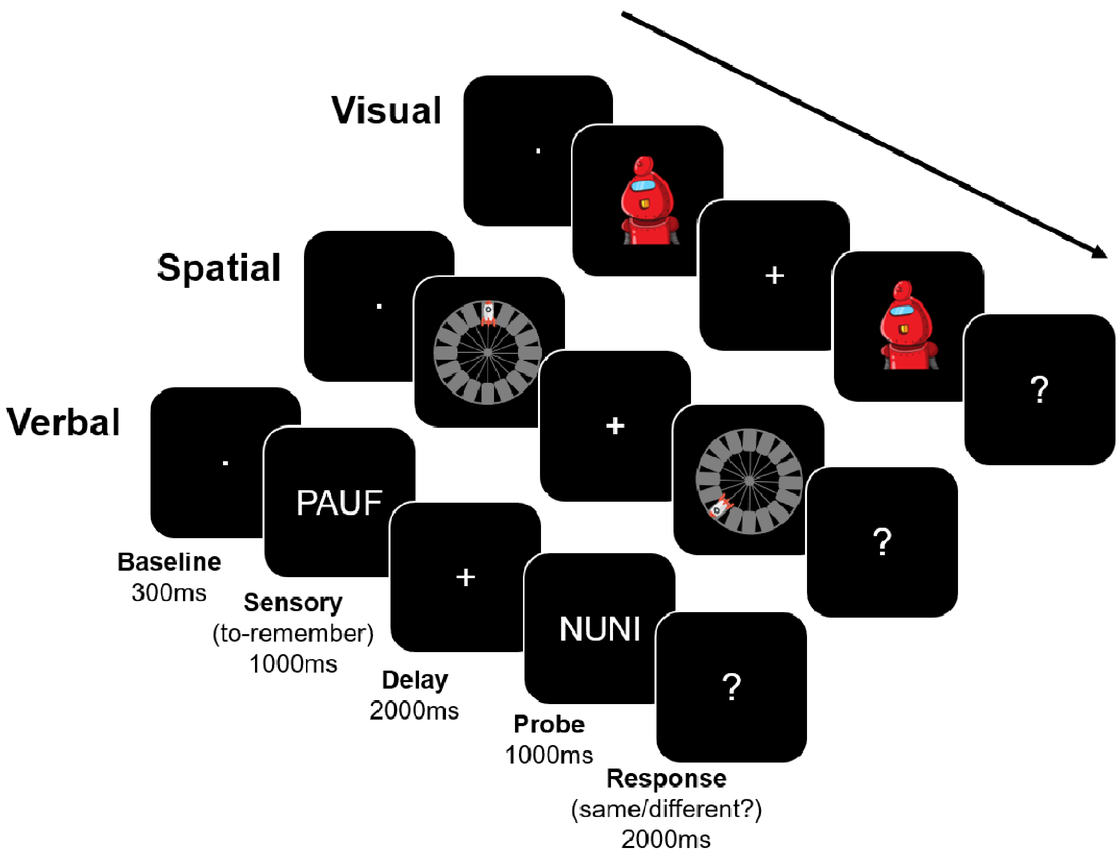


Figure 1. Task schematic. One trial per category (visual, spatial, verbal) is depicted. The stimuli for the visual category were images of robots, the stimuli for the spatial category were rockets in a particular location on a platform, and the stimuli for the verbal category were nonwords. The Baseline period was the last 300ms of the ITI, which lasted 1000ms on average. The stimuli were presented during the Sensory period (1000 ms), followed by a blank Delay period (2000 ms). Then, a probe image appeared (1000 ms) that was either the same image (50% of the time) or a different image from the same category (50% of the time). Finally, during the Response period (2000 ms) a question mark appeared, and participants could respond. Categories were randomly intermixed across trials over the duration of the experiment. Note that the stimuli are shown much larger in the figure, for clarity, than they appeared in the experiment. EEG data

were epoched so as to contain the Baseline, Sensory, and Delay periods (i.e., -300ms to 3000ms relative to Sensory period onset).

Participants completed 128 trials per category, resulting in a total of 384 experimental trials across categories. To help increase children's motivation, the task contained a background story (helping astronauts find their way home from an alien space base) and was presented in a game-like fashion. Participants could take self-timed breaks after every 48 trials. During these breaks, the participants' total number of correct responses out of possible correct responses was shown on the screen, alongside the number of 'blocks' left. Before starting the paradigm, participants completed several slower practice trials. In total, each session took a maximum of 2h, with approximately 45 minutes of data collection.

The experimental paradigm was programmed using the Psychopy Builder Standalone version 2020.2.5 (Peirce et al., 2019), and presented on a 24" LCD monitor (60Hz refresh rate) in a sound-attenuated, shielded booth. A BioSemi ActiveTwo amplifier (BioSemi Inc., Amsterdam, The Netherlands) was used to record EEG data from a 64-electrode BioSemi gel headcap (10/20 electrode layout). All sites were referenced online to electrode Cz, and re-referenced offline to the average reference. To record eye movements and blinks, additional electrodes were placed at the outer canthi of both eyes (for the horizontal electrooculogram; HEOG) and above and below the right eye (for the vertical electrooculogram; VEOG). Electrode impedances were adjusted to below 5 kΩ prior to the start of the experiment. Data were digitized at 2048 Hz.

Behavioral data analyses

Although behavioral data analyses were not the focus of the present study, we calculated accuracy (percentage of correct responses) over the entire task. We excluded the data of participants with accuracy that was lower than a level that would be obtained by pure chance (50%) from further EEG analyses. To estimate whether our participants could successfully complete the task, we derived the average accuracy score across all participants at three different points: before any exclusions, after exclusions based on behavioral criteria, and after exclusions based on EEG-related criteria (i.e., the data retaining at least 200 trials after cleaning).

EEG preprocessing

For EEG data preprocessing, we used the Matlab-based (Natick, Massachusetts: The MathWorks Inc) EEGLAB software (v.2022.0, Delorme & Makeig, 2004). We first down-sampled the data to 500Hz, removed the DC offset, and applied a bandpass filter of 1Hz – 40Hz (12 dB/octave roll-off computed forward and backward to eliminate phase shift). Then, we epoched the data from -300ms to 3000ms relative to the onset of the stimulus for each trial, such that each epoch contained the Baseline period (-300ms to 0ms, during the ITI), the Sensory period (0ms – 1000ms), and the Delay period (1000ms – 3000ms; see Figure 1). A semi-automatic artefact rejection procedure was used to remove artefacts (transient noise, movement, skin potentials, etc.), which consisted of applying an automatic artefact rejection criterion of ±150µV for EEG artefacts (adapted to children's EEG, see e.g., Melinder et al., 2010; Shimi et al., 2015) along with visual inspection. Next, to remove the influence of blinks, we conducted independent component analysis (ICA) using the ICLabel package (Pion-Tonachini et al., 2019) in EEGLAB. We detected those components that contained eye movements and blinks with visual inspection and removed only these components from the data. We discarded any electrodes contaminated by artefacts, based on visual inspection (maximum 13% of the electrode montage) and interpolated the missing data using 3-dimentional splines (Perrin et al., 1987). Our EEG analyses only included participants with over 200 trials (67% of the total number of trials) remaining after the cleaning procedure.

EEG classification

The main goal of this study was to examine the EEG multivariate representations for children performing a working memory task. We approached this goal in three ways: 1) We calculated the average classification accuracy during the Baseline, Sensory, and Delay periods separately (time-average classification). 2) We examined timepoint-by-timepoint classification over the entire trial (time-wise classification), and 3) We examined how classification performance trained at a specific moment in time generalized over other timepoints (temporal generalization classification). This three-pronged analysis allowed us to probe whether we can detect differences in the spontaneously-generated representational content of children's working memory. First, we tested whether we could reliably decode WM content during delay periods, as well as other periods during the trial. Second, we tested the classification accuracy throughout the course of an experimental trial. Third, we further tracked the representational content over time by showing how long the same representational structure can be detected.

Multivariate classification was performed within each subject by employing the MVPA-Light toolbox (Treder, 2020) with the linear discriminant analysis (LDA) classifier. We used each participant's preprocessed single-trial voltage amplitudes as input features for the classifier. Both correct and incorrect response trials were included in the analysis (as in LaRocque et al., 2013), in order to maximize the number of trials per category, assuming that incorrect responses were made because the details of a given stimulus, and not its category membership, were wrongly remembered. First, for time-average classification, we calculated the average voltage amplitudes across each of the respective time-windows: -300ms – 0ms for the Baseline period, 0ms – 1000ms for the Sensory period, and 1000ms – 3000ms for the Delay period. We averaged across the samples (trials), with 5 samples for each average and demeaned the data across trials. Second, for time-wise and temporal generalization classification, we divided each trial into smaller 50-ms and 20-ms time-bins and calculated the mean voltage amplitudes for each bin (as in Adam et al., 2020).

For the time-average and time-wise classification analyses, we performed 100 iterations of the classification analyses for each time-bin (as in Adam et al., 2020). On each iteration, 2/3 of the trials were randomly assigned to a training set and 1/3 of the trials to a held-out test set. The classifier performance was determined by averaging the classification accuracy across the 100 iterations. At each time-bin, we also performed a null classification where, at each iteration, the labels that were associated with trials were randomly shuffled 1000 times. To ensure the same number of trials per category in the training and test set, condition labels with fewer trials were up-sampled. We ensured that the condition labels were balanced during classification using a stratification procedure, natively implemented in the MVPA-light toolbox. The outputs of the classification performance were accuracy (i.e. the proportion of correctly predicted class labels) and a set of confusion matrices (i.e., tables of proportions correctly predicted and incorrectly predicted class labels). Significance was statistically assessed against theoretical chance (33%) via Bonferroni-corrected one-sided *t*-tests at each time-bin (assuming no meaningful values that are below chance). As an additional assessment, we also compared decoding accuracy to empirical chance, which was obtained through a classification analysis on 1000 random shuffles of data labels (reported in the Supplementary Materials). Statistical assessment against both theoretical and empirical chance involved performing subject-wise permutation (see also Fahrenfort et al., 2018), with 1000 iterations per subject, at an alpha level of 0.05, implemented via the "*mv_statistics*" function of the MVPA-light toolbox.

To analyze temporal generalizability, for each time point, we trained the classifier on recorded brain activity for the given time point and tested it on brain activity recorded at all other time points. Significance was statistically assessed against theoretical chance (33%) via Bonferroni-corrected one-sided *t*-tests at each time-bin.

Open Practices Statement

The behavioral data, exclusion criteria, preprocessing, analysis scripts, and study materials (stimuli and task design) are available at https://osf.io/jeh67/?view_only=2a9c2379a1514ce996b446cf1b0690b3. Due to space limitations on OSF, the raw EEG data are stored in a Zenodo repository: [to be shared after manuscript acceptance, per our ethics board requirements], and the preprocessed EEG data are stored in another Zenodo repository: [to be shared after manuscript acceptance, per our ethics board requirements].

Results

Behavioral results

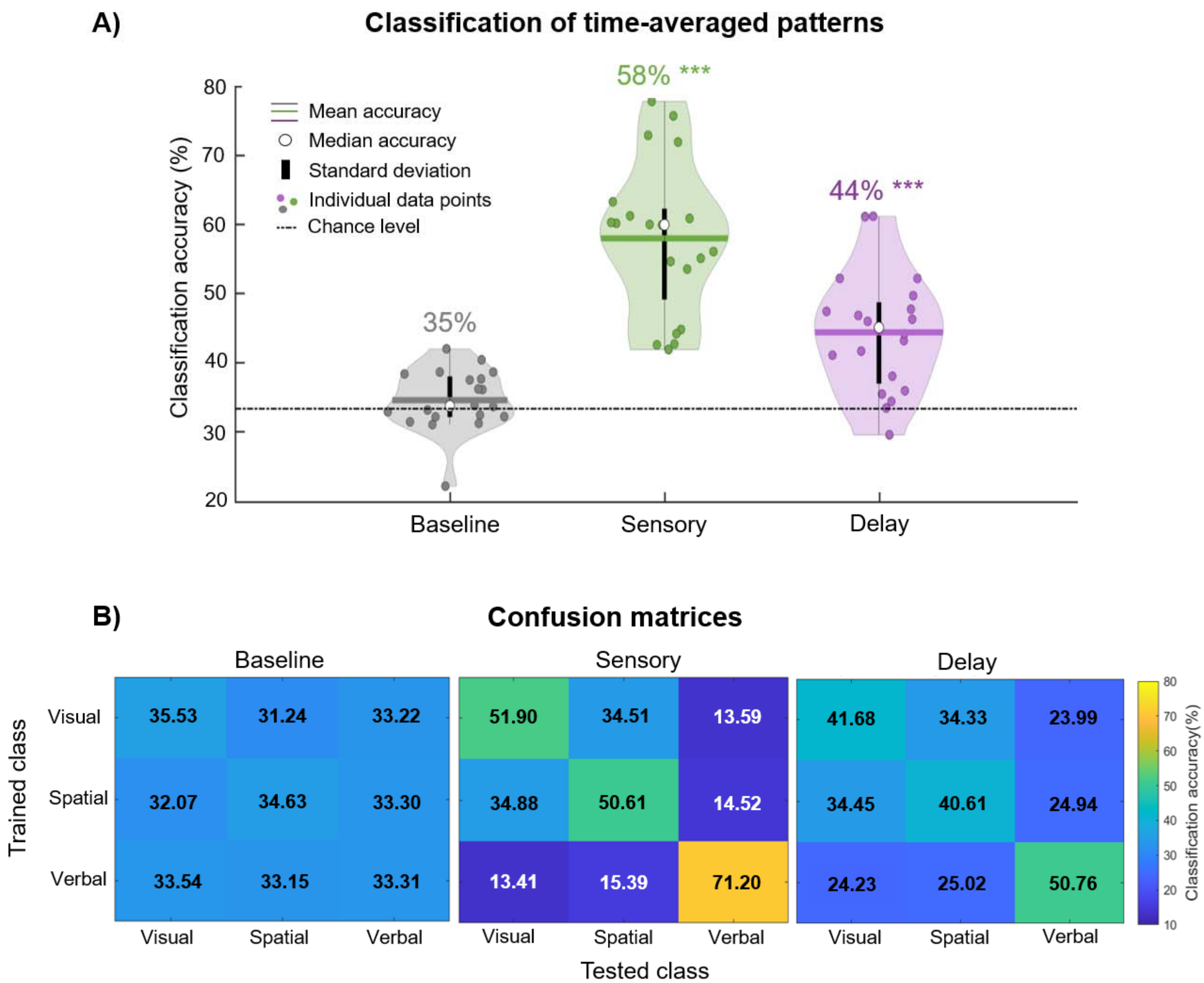
In a first analysis step, we verified whether participants successfully performed the task by checking their average behavioral accuracy results. The total sample had an average accuracy score of 87% (SD = 34%). After applying our behavioral exclusion criteria, the average accuracy score was 89% (SD = 32%). Finally, after also applying our EEG exclusion criteria, the average accuracy score was 90% (SD = 30%). This high accuracy demonstrates that our participants successfully performed the WM task.

EEG classification

The key goal of this study was to use multivariate pattern analysis to decode EEG data collected from children during a WM game. We were interested in whether multivariate EEG patterns differed when children maintained different categories of information (visual, spatial, and verbal). First, we examined the time-average EEG voltage patterns during the Baseline, Sensory, and Delay periods (Figure 2A). As expected, we found that decoding accuracy was above theoretical chance (33%) in the Sensory period (mean = 58.0%, median = 60.0%, SD = 11.1%, p = 2.77e^{-09}) and the Delay period (mean = 44.4%, median = 45.1%, SD = 8.6%, p = 7.38e^{-06}), but not in the Baseline period (mean = 34.6%, median = 33.8%, SD = 4.4%, p = 0.11). This demonstrates reliable decoding of children's WM contents during the time when children maintained information (Delay), and when they viewed information (Sensory), but also verifies that we were unable to decode during the moments prior to the stimulus presentation when they were preparing for the onset of the next trial (Baseline).

Category confusability

We next examined the classification performance of specific categories, and which categories were most confusable, using the confusion matrices from the time-average classification (Figure 2B). First, during the Baseline period, all categories at test were confusable with each other, since all of the values were near chance (33%). Next, during the Sensory period, we observed that verbal information was most accurately decoded (71%) and least confusable with either visual or spatial information. Whereas visual (52%) and spatial (51%) information were more confusable with each other (35%). The confusability patterns in the Sensory period were reechoed during the Delay period, though with less pronounced distinctions between the categories, consistent with slightly lower overall decoding accuracy at Delay than at Sensory. These results show that each category was successfully decoded, but also these differences between categories are consistent with theoretical distinctions between WM processes.



Baseline	Visual	Spatial	Verbal
Visual	35.53	31.24	33.22
Spatial	32.07	34.63	33.30
Verbal	33.54	33.15	33.31

Sensory	Visual	Spatial	Verbal
Visual	51.90	34.51	13.59
Spatial	34.88	50.61	14.52
Verbal	13.41	15.39	71.20

Delay	Visual	Spatial	Verbal
Visual	41.68	34.33	23.99
Spatial	34.45	40.61	24.94
Verbal	24.23	25.02	50.76

Figure 2. A) Classification accuracy of time-averaged patterns during the Baseline (gray), Sensory (green), and Delay (purple) periods. The mean is illustrated as a horizontal line, the median accuracy is depicted as a white circle, and standard error is depicted as a vertical black line. Individual data points are depicted as dots, and theoretical chance is represented by a dashed and dotted horizontal black line. The numerical average accuracy score is displayed above each violin plot. Accuracy scores that are significantly above chance are marked with *** if they are above the *p*<0.001 threshold. B) Confusion matrices for the three categories of information (visual, spatial, and verbal) during the Baseline, Sensory, and Delay periods. Training class labels are on the y-axis, and testing class labels on the x-axis. The tables show the proportion of trials where a category was confused with any of the other three categories.

Moment-by-moment EEG classification

To examine more fine-grained patterns of classification over time, we trained and tested multivariate classifiers on data from smaller time steps across the duration of a trial (Figure 3A). We subdivided each trial into smaller time bins (50ms) and averaged the EEG voltage patterns within each bin (the results of the 20ms time-bin averages are shown in the Supplementary Materials). The average decoding accuracy over the course of of a trial (without the Baseline period, i.e. 0 to 3000ms) was 48.7% (median = 49.2%, SD = 9.4%), which was significantly above chance (33%, $p = 3.36e^{-07}$). Specifically, we observed significant above-chance decoding throughout nearly all of the Sensory period, and until almost 1500ms into the Delay period. All of these results were significant according to a Bonferroni-corrected *t*-test with 1000 permutations (corrected *p*-value was $7.69e^{-04}$). Consistent with our expectations, classification accuracy was not reliably above chance during the Baseline period. These results show that differences in multivariate representations of information categories are readily detectable when memoranda are directly observable, but also, and this is what we aimed to test here, when they are no longer physically present, during their maintenance in WM.

Temporal generalization of EEG patterns

Finally, we examined how children's WM representations unfolded across time by examining how patterns generalized across different moments (Figure 3B). That is, we trained a multivariate classifier at a particular time point, and tested it on all other timepoints. We observed similar multivariate representations across time, particularly from the Sensory to the Delay periods of the trials: Representations that were detected during the early Sensory period could be traced with high accuracy over the entire Sensory period, and importantly, for a portion of the Delay period. This is reflected in reliable decoding at distant moments of time, far from the diagonal of the matrix. In sum, our results suggest not only that differences in observed *and* maintained representational content can be detected, but that the same representations that were formed when stimuli were observed persisted in WM for a portion of the maintenance period, even when the stimuli were absent.

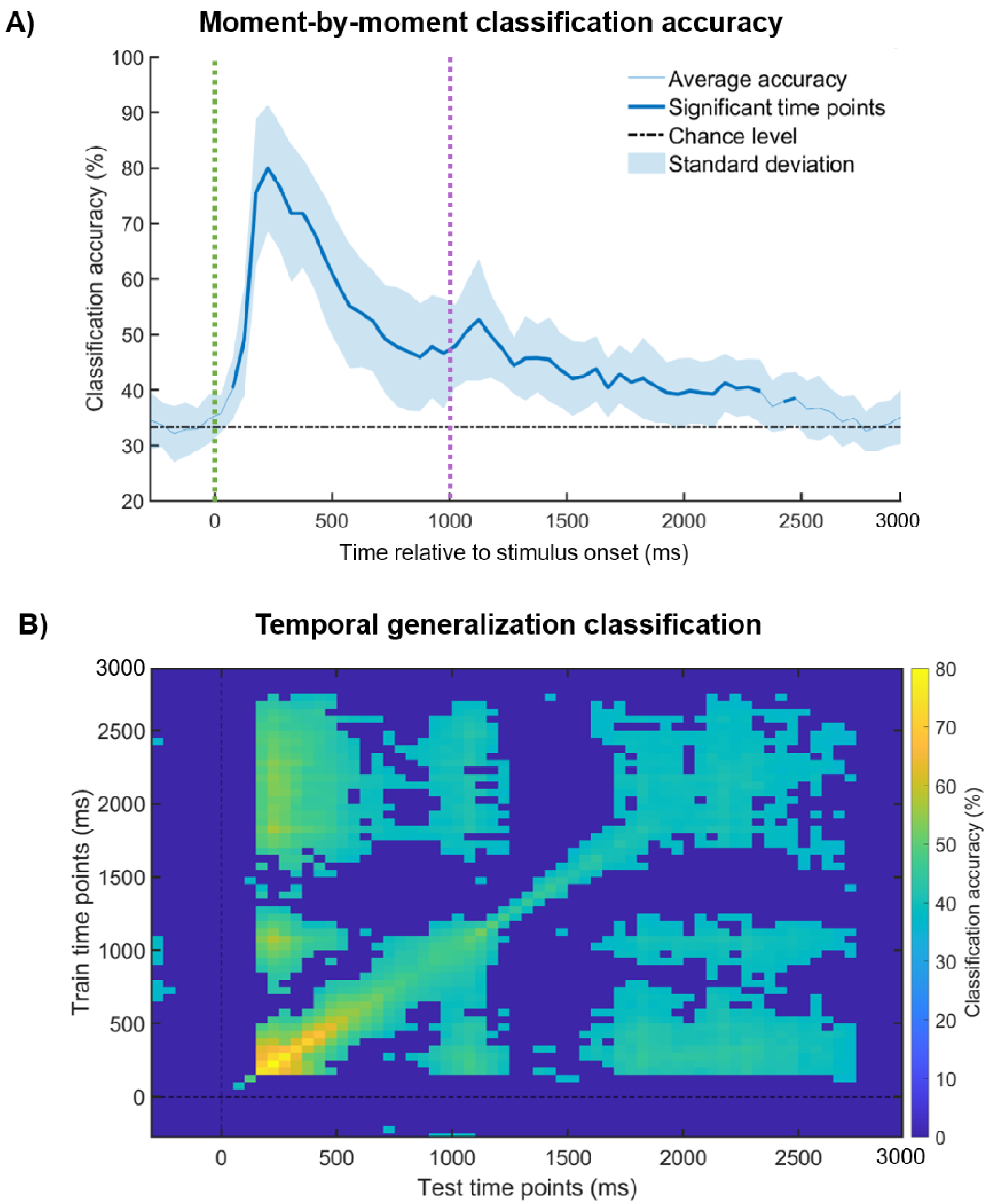


Figure 3. A) Moment-by-moment classification of the category of information. The average across participants is depicted in a blue line (time points that were significant after Bonferroni correction are represented by a thicker blue line). The standard deviation is depicted in the shaded blue area. Theoretical chance (33%) is depicted in a black dashed horizontal line, and the onsets of the Sensory (0ms) and Delay periods (1000ms) relative to stimulus onset are depicted in dashed green and purple vertical lines, respectively. B) Temporal generalization of classification, where a classifier was trained on one time point and tested on all other time points. Average classification values are overlaid by a significance mask: dark blue fields indicate the time points at which classification did not survive the Bonferroni correction for multiple comparisons, while all other colors correspond to above-chance classification generalization.

Further exploratory analyses of reliability and validity

To further validate our method and investigate its robustness, we conducted additional exploratory analyses to assess the reliability and validity of the method. First, we split our EEG data in half and conducted several analyses to establish the robustness of EEG decoding within and across these halves. Second, we compared the results of the main time-average decoding to an expected pattern based on our given task design (i.e., no above-chance decoding at Baseline, high above-chance decoding at Sensory, and lower but still above-chance decoding at Delay). Third, we compared the results of the main confusion matrix results to a pattern that follows classic WM theory (i.e., that verbal stimuli would be the least confusable with other categories, while visual and spatial stimuli would be more confusable with each other, though still distinct). Fourth, we investigated the relationship between decoding accuracy in the Sensory and Delay periods per individual, to capture the consistency of our measures across participants. Finally, we examined decoding accuracy as a function of age. To avoid reiterating the main analysis results, we will only present the results of the split-half analyses, individual difference analyses, and analyses of decoding accuracy by age.

Split-half analyses

In a first set of exploratory split-half analyses, we examined whether, with half as much data, we could still reliably decode the content of children's WM. For each participant, we split the data into an early half (blocks 1 – 4 in the testing session) and a late half (blocks 5 – 8), and applied the same classification procedures to each half separately. Even when training/testing only within the early half, classification was robust during almost the whole Sensory period, and within the first 500ms of the Delay period but not during the Baseline period (Figure 4A). The results for the late-half showed that we reliably classified data during the Sensory period and the first 500ms of the Delay period, and not during the Baseline period (Figure 4B).

We also investigated the consistency or stability of the representational structures in WM throughout the entire session, by conducting cross-decoding analyses between the early and late halves. That is, the data from the early half of the session were used for training the classifier and the data from the late half of the session were used for testing the classifier (see e.g., LaRocque et al., 2013; Lewis-Peacock et al., 2012). We observed reliable cross-decoding during the most of the Sensory period and the first 500ms of the Delay period (Figure 4C). This suggests that representational structures detected in the early half of the testing session were comparable to those detected in the late half of the testing session.

In sum, these split-half classification results show that our method detects robust representations, which are furthermore stable across the testing session, lending support to the overall reliability and validity of the method.

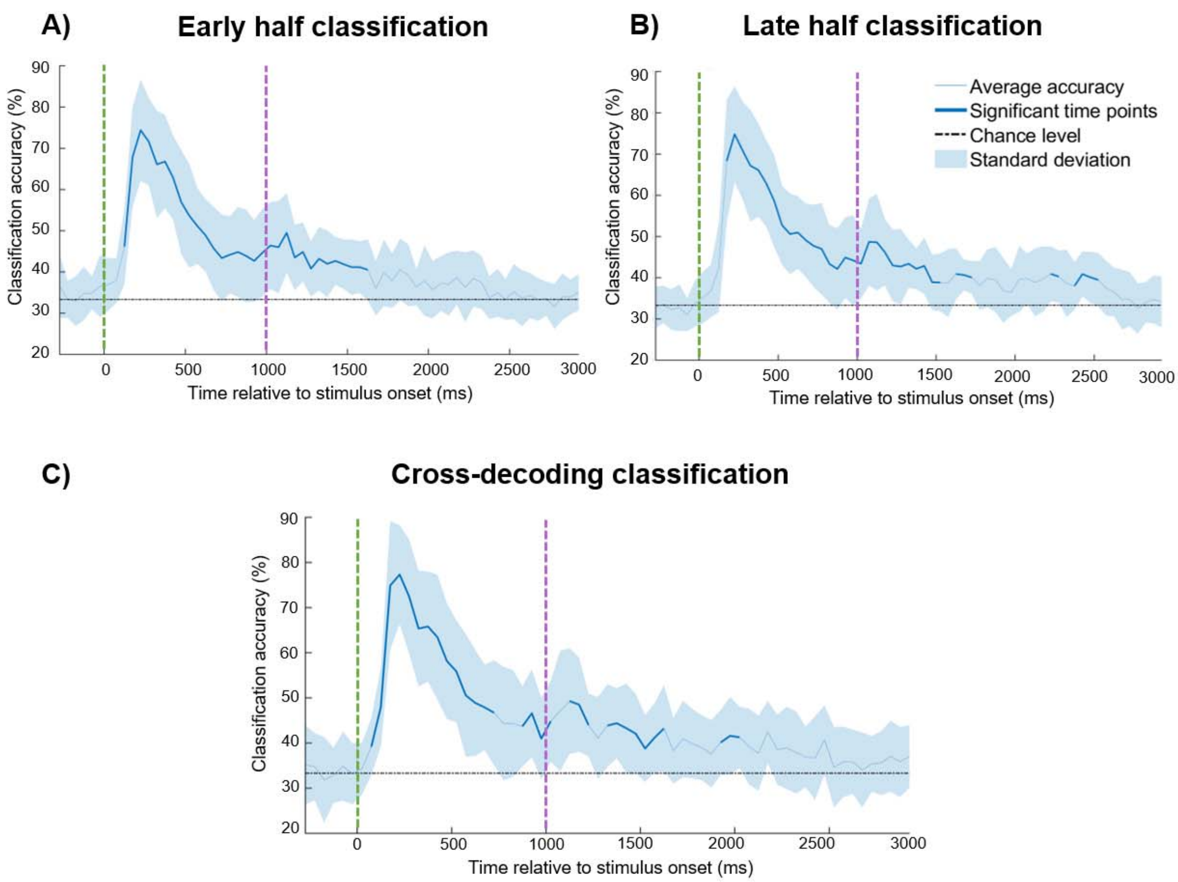


Figure 4. Classification performance on EEG data split into early ('blocks' 1-4) and late ('blocks' 5-8) halves. A) Classification results in the early half of the testing session, B) Classification results in the late half of the testing session and C) Cross-decoding results between the early and late halves (training the classifier on early half data, testing the classifier on late half data). Across all panels, average accuracy is depicted in a blue line (time points that were significant after Bonferroni correction are represented by a thicker blue line). The standard deviation is depicted in the shaded blue area. Theoretical chance (33%) is depicted in a black dashed line.

Individual differences in multivariate decoding

When it comes to individual classification performance, we had two main assumptions with regards to the consistency of our measures across participants. First, we assumed that classification performance when observing the stimulus to be remembered (i.e., during the Sensory period) should be higher than classification performance when maintaining said stimulus in memory in its absence (i.e., during the Delay). Second, we assumed that individuals with higher classification performance at Sensory than others will also have higher classification performance at Delay than others. To test these assumptions, we first compared average classification accuracy at Sensory and Delay using a paired-sample right-tailed *t*-test. Second, we quantified the strength of the relationship between classification accuracy at Sensory and Delay via Pearson correlation.

The results of these analyses revealed two main patterns (Figure 5). First, classification accuracy was higher during the Sensory period than in the Delay period across participants ($t(19) = 9.87$, $p = 3.23e\text{-}09$), by an average of 13.5%. Second, classification performance was highly correlated across individuals ($r(18) = .84$, $p < 0.001$). That is, most participants with high decoding accuracy in the Sensory period also tended to have high decoding accuracy during the Delay period. This suggests that our results were driven by patterns present across the whole sample, and that our method is consistent across individuals since decoding accuracy across individuals was consistent across the Sensory and Delay periods.

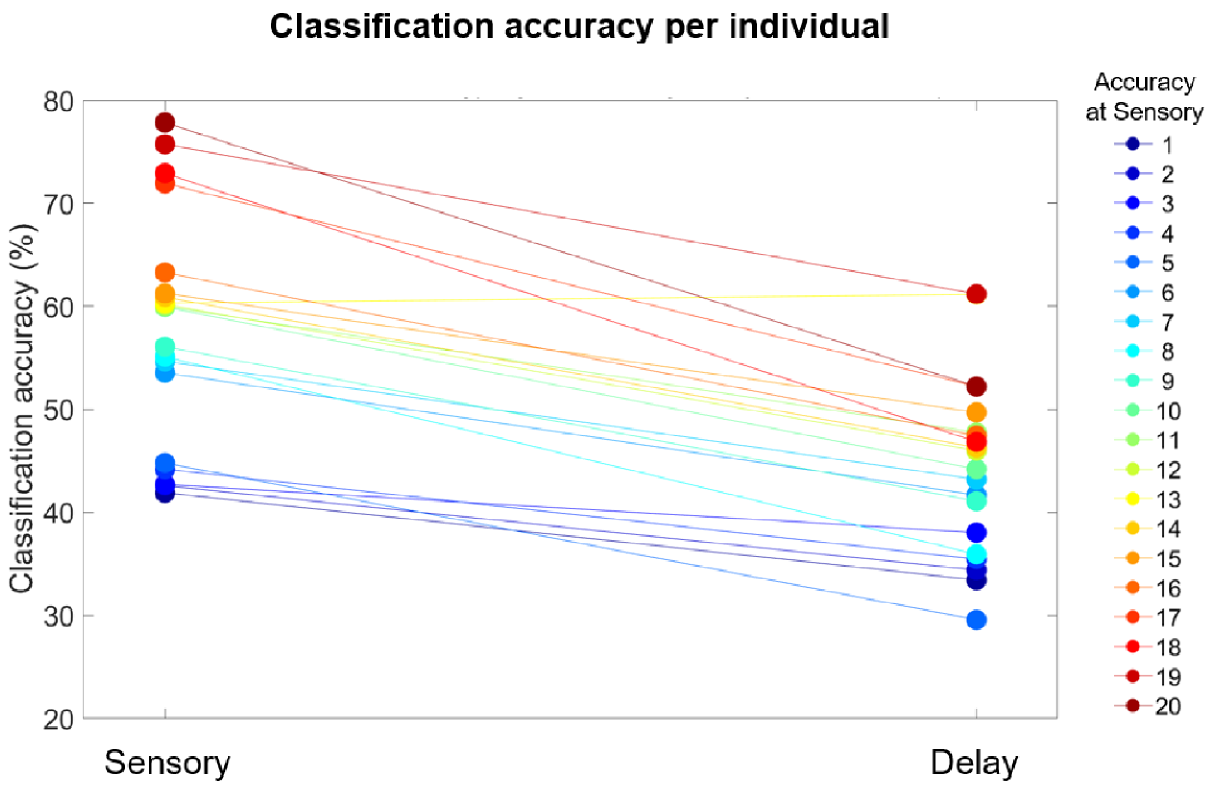


Figure 5. Individual differences in classification accuracy. Each dot depicts the average decoding accuracy from a participant during the Sensory and Delay periods, and dots from the same individual are connected by lines. The colors are organized according to the decoding accuracy during the Sensory period, from lowest accuracy (blue) to highest accuracy (red).

Decoding across age

Another concern could be that our results were driven by older participants, and not the entire range of participants. It is well-established that children's WM performance tends to improve with age (e.g., Gathercole, 1999; Gathercole et al., 2004; Salthouse, 1994), and thus our above-chance decoding results could have stemmed from older 'high performers' with higher decoding accuracy. Thus, in a final exploratory analysis, we investigated whether classification varied as a function of age. We applied a linear regression between participants' age and their average classification accuracy for the Baseline, Sensory, and Delay periods separately, using the "fitlm" and "anova" functions in Matlab.

In Figure 6, decoding accuracy is shown by age. One can immediately see that successful decoding during Sensory and Delay periods was present at all ages, from the youngest to the oldest participant. Our above-chance decoding thus does not seem to be driven by high-performing older children. Moreover, regression analyses on decoding accuracy as a function of age did not reveal any significant relationship between age and decoding accuracy in any period of the trials: Baseline (R^2 = 0.03, p = 0.46), Sensory (R^2 = 0.02, p = 0.57), and Delay (Delay: R^2 = 0.03, p = 0.51). As Figure 6 indeed shows, in each of the periods of interest, decoding accuracy was stable across the entire age range. This further supports that our results were driven by patterns present across the whole sample, regardless of age, and lends evidence to the consistency of our method across individuals.

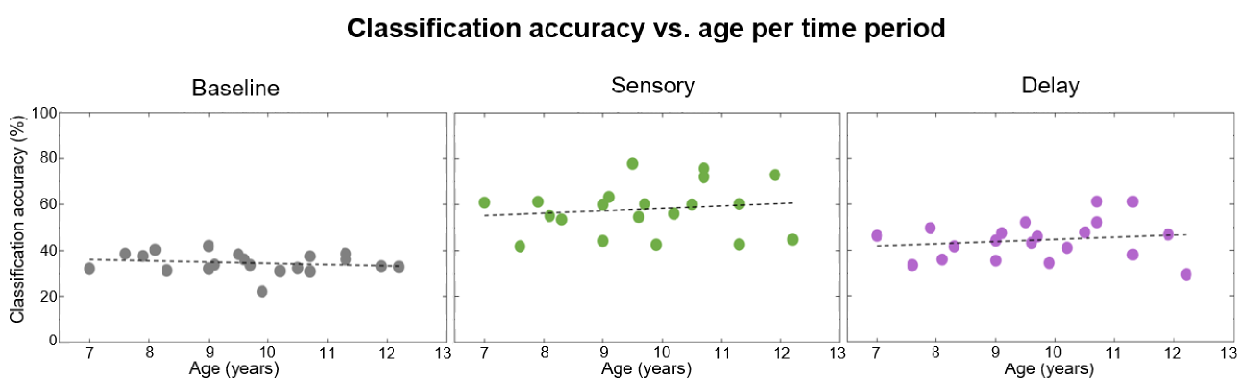


Figure 6. Decoding accuracy as a function of age. For each period (Baseline, Sensory, and Delay) we examined whether there was a relationship between classification accuracy and age. The age of the participant is depicted along the x-axis, and the classification accuracy is along the y-axis. Each participant is represented by one dot in each plot. The black dashed regression line on each panel shows the relationship between age and classification accuracy, which was not reliable for any of the three periods.

Discussion

In the present study, we aimed to develop a direct neural measure of spontaneously-generated WM representational content in children. We developed a game-like computerized task with no maintenance instructions. We collected EEG data from children while they performed this task, and analyzed this data with multivariate pattern analysis methods. We reliably decoded the category of information during a WM maintenance period. Furthermore, we examined the moment-by-moment classification performance during the maintenance period and the consistency of the representations from sensory to delay periods. In addition, we conducted exploratory analyses of the reliability and validity of our results.

Decoding the content of children's WM

Although this is the first study to use MVPA on EEG signals to investigate children's WM maintained memoranda, the results are comparable with the existing literature in adults. Namely, our delay period decoding accuracy between visual, spatial, and verbal categories in children's EEG (44%) was comparable to decoding accuracy between visual, phonological, and semantic categories in young adults' EEG (45.3%) in the Phase 1 task of LaRocque and colleagues (2013). This is especially impressive given that our sample consisted of children, a population whose WM performance is known to be lower than that of young adults (e.g., Gathercole, 1999; Gathercole et al., 2004; Salthouse, 1994), and whose EEG signals are not perfectly comparable to those of adults, due to physiological differences, higher spectral power in children than adults, etc. (e.g., Barriga-Paulino et al., 2011; Scerif et al., 2006). It is equally noteworthy that such results were obtained from an MVPA pipeline based on adult applications (e.g., the mini-block approach by Adam et al., 2020). Thus, even before delving into reliability and validity, these results already suggest that our method is appropriate for decoding children's WM content.

Inspired by the methodological issues in measuring children's maintenance mechanisms, and the resultant theoretical confusion surrounding children's ability to apply certain maintenance mechanisms spontaneously (e.g., in the case of refreshing and organization), we aimed to develop a direct method for tapping into the content of children's WM. Exploring whether it is possible to infer what children are maintaining was the first step towards investigations of whether they apply given maintenance mechanisms spontaneously. As such, the results of this proof-of-concept study cannot in and of themselves answer whether, or at which age, children can spontaneously refresh information. However, the method we propose can be used as a base for answering this and other open questions in the field of WM development. For one, our method is sensitive and appropriate for investigating the nature of neural representations in children's WM. MVPA has already been successful at characterizing memory representations at different WM states in adults (e.g., Christophel et al., 2018; Lewis-Peacock et al., 2012, 2015; Rose et al., 2016). Now that we have demonstrated that children's WM representational content can also be decoded, we can manipulate the original task to investigate, for example, the preferred code that children use to represent given information (as Lewis-Peacock et al., 2015 did in adults). To tap into the maintenance mechanism that children use to maintain information, the current paradigm can help serve as a starting point onto which extensions can be introduced carefully. To avoid running into the same pitfalls as prior behavioral-only work, these extensions need not necessarily be interleaved secondary tasks (Barrouillet et al., 2009; Bayliss et al., 2003; Bertrand & Camos, 2015; Camos & Barrouillet, 2011; Conlin et al., 2005; Oftinger & Camos, 2015, 2017, 2018; Tam et al., 2010) or potentially unclear maintenance instructions (Bjorklund & de Marchena, 1984; Hasselhorn, 1992; Salatas & Flavell, 1976; Schleepen & Jonkman, 2012). Instead, one could have children maintain lists of items from different categories, and investigate when they think of a given category during retention (for a similar approach using fMRI in adults see Vergauwe et al., In prep; for a similar approach using EEG in adults see Jeanneret et al., 2022). We believe the current paradigm can be expanded in several ways that would enable both a deeper exploration of children's WM content, and the investigation of children's WM maintenance mechanisms.

Exploring the reliability and validity of our approach

Given the topic of this special issue, although our paradigm was not optimized for such analyses, and although such analyses are not typically conducted (or at least not always reported) in the adult literature, we carried out exploratory analyses of reliability and validity to assess the use of our method as part of a derivation chain in the field of WM. Within such constraints, these exploratory analyses nonetheless added further support to the main results.

Splitting the data into an early and late half showed two main lines of support for the consistency of the decoding results over time. First, the pattern of decoding accuracy in both the early and late halves was comparable to that of the overall pattern in the main analyses. Second, training the classifier on the early half data and testing it on the late half data showed that representational structures detected in the early half were similar to those in the late half. Both of these results suggest that, despite potential fatigue over the course of an experimental session, decoding is nonetheless reliable, and that WM representations remain similar over the course of a testing session.

Given our task design, we expected decoding accuracy across the different stages of WM to follow a given pattern. At Baseline, we expected decoding not to rise above chance levels, as this period presumably mainly contained noise. Likewise, there was no blocking of trials by category, such that anticipatory category-related activity would be detected during this period. Next, the Sensory period involved the perception, recognition, and encoding of the information to be remembered. Since this information was physically present on the screen during this time, we expected decoding accuracy to be the highest during the Sensory period. Finally, since the Delay period involved the maintenance of information in WM in the *absence* of the to-be-remembered stimuli on the screen, we still expected above-chance decoding accuracy, though lower than in the Sensory period. These patterns were borne out by time-average results, showing that our decoding procedure correctly responded to 1) noise at Baseline, 2) differences between observed categories of stimuli at Sensory, and 3) differences between representations of maintained categories of stimuli at Delay.

Based on classic WM theoretical accounts whereby WM is split into separate visuo-spatial and verbal domains (Baddeley & Hitch, 1974; Baddeley & Logie, 1999), if our method truly captured WM content, the results would reveal verbal information to be distinct from visual and spatial information, while the latter two would be more confusable. To assess this, we examined patterns in the confusion matrices generated as part of the time-average analyses. We specifically observed that information labelled as verbal at training was the least confusable with information labelled as visual or as spatial at test, and that information labelled as visual and spatial at training was relatively confusable with each other at test. This is in line with classic theories of domain-specificity for visuo-spatial and verbal information (Baddeley & Hitch, 1974; Baddeley & Logie, 1999), and suggests that our method measured actual maintained WM content.

The last two analyses confirmed that our results were not driven by a subset of participants, but by consistent values across the entire sample. Namely, we observed that the majority of participants had slightly higher decoding at Sensory than at Delay, and that those individuals that had high decoding accuracy at Sensory also had high decoding accuracy at Delay. Further, the age of the participants did not influence decoding, as classification accuracies were evenly distributed across participants regardless of age.

Taken together, these results suggest that our method is reliable across time within a testing session, across individuals. Further, they suggest that our method measured what we set out to measure, that is, differences in the content of children's WM.

Limitations

An inherent limitation of our study was that its design and resource allocation were optimized for the main analyses, and aim to provide a proof of concept of decoding children's WM content, but not for our analyses of reliability and validity. For instance, our task was designed to be easy to complete successfully across the 7-12 age range. In addition, our sample size (n=20) was appropriate for a proof-of-concept study, but relatively small for regression analyses (20 participants). Now that our proof of concept appears to have been successful, future studies should include retests and replications in larger samples.

Another potential issue with our design lies in its lack of masks, i.e., irrelevant stimuli that would overwrite the contents of sensory memory, and prevent such sensory representations from driving the decoding results. We intentionally refrained from adding any additional information other than the to-be-remembered items and their probes, both to keep the task as simple as possible, and to prevent any potential biasing effects resulting from masks (see "attractive bias" in Lorenc et al., 2021). A point that is rarely acknowledged in the use of masks is that they can function as external distractors (i.e., irrelevant stimuli that should not be attended; reviewed in Rademaker et al., 2015, p. 1-2). Since children are known to be more susceptible to visual distraction than adults (e.g., Plude et al., 1994; Trick & Enns, 1998), we did not wish to risk replacing sensory memory effects with potential distractor effects. This point notwithstanding, our current design does not let us directly assess whether the representational content at maintenance consists merely of a persisting sensory trace. However, that would be unlikely, since the perceptual and informational content of a stimulus only seem to persist for a maximum of 500ms after stimulus offset (Irwin & Yeomans, 1986; Massaro & Loftus, 1996; Sperling, 1960), and we found significant decoding for much longer (around 1500ms into the Delay period), consistent with comparisons of sensory memory and WM such as Cappiello & Zhang, 2016). This, together with our exploratory analysis results, suggest that our method does truly capture maintained representational content in WM.

Conclusion

The current study demonstrates that children's WM contents can be decoded using EEG MVPA together with a simple behavioral paradigm, in a manner that is promising in terms of reliability and validity. Though the main insights the study provides are *what* children are maintaining rather than *how*, the framework we developed as part of this study can serve as a base for investigations of maintenance mechanisms, or questions related to representational content. As such, this study provides a much-needed stepping stone for strengthening the derivation chain in the field of WM development.

Acknowledgments

We would like to thank all the families that participated in the experimental sessions. This work was completed with support from the Swiss National Science Foundation to Evie Vergauwe (Grant PCEFP1_181141), from the Jacobs Foundation to Nora Turoman [Grant 2021-1417-00], and from the NIH to Megan deBettencourt [5F32MH115597 and K99 MH128893].

References

Adam, K. C. S., Vogel, E. K., & Awh, E. (2020). Multivariate analysis reveals a generalizable human electrophysiological signature of working memory load. *Psychophysiology*, *57*(12). https://doi.org/10.1111/psyp.13691

Alloway, T. P., & Alloway, R. G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *Journal of Experimental Child Psychology*, *106*(1), 20–29. https://doi.org/10.1016/j.jecp.2009.11.003

Baddeley, A. D. (1986). Working memory. *Applied Cognitive Psychology*, *2*(2), Article 2. https://doi.org/10.1002/acp.2350020209

Baddeley, A. D., & Hitch, G. (1974). Working Memory. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (Vol. 8, pp. 47–89). Academic Press. https://doi.org/10.1016/S0079-7421(08)60452-1

Baddeley, A. D., & Logie, R. H. (1999). Working Memory: The Multiple-Component Model. In A. Miyake & P. Shah (Eds.), *Models of Working Memory* (1st ed., pp. 28–61). Cambridge University Press. https://doi.org/10.1017/CBO9781139174909.005

Baddeley, A., Gathercole, S. E., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, *105*(1), Article 1. https://doi.org/10.1037/0033-295X.105.1.158

Barriga-Paulino, C. I., Flores, A. B., & Gómez, C. M. (2011). Developmental Changes in the EEG Rhythms of Children and Young Adults: Analyzed by Means of Correlational, Brain Topography and Principal Component Analysis. *Journal of Psychophysiology*, *25*(3), 143–158. https://doi.org/10.1027/0269-8803/a000052

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time Constraints and Resource Sharing in Adults' Working Memory Spans. *Journal of Experimental Psychology: General*, *133*(1), Article 1. https://doi.org/10.1037/0096-3445.133.1.83

Barrouillet, P., Gavens, N., Vergauwe, E., Gaillard, V., & Camos, V. (2009). Working memory span development: A time-based resource-sharing model account. *Developmental Psychology*, *45*(2), 477–490. https://doi.org/10.1037/a0014615

Bayliss, D. M., Jarrold, C., Gunn, D. M., & Baddeley, A. D. (2003). The complexities of complex span: Explaining individual differences in working memory in children and adults. *Journal of Experimental Psychology: General*, *132*(1), Article 1. https://doi.org/10.1037/0096-3445.132.1.71

Bertrand, R., & Camos, V. (2015). The role of attention in preschoolers' working memory. *Cognitive Development*, *33*, 14–27. https://doi.org/10.1016/j.cogdev.2014.10.002

Bjorklund, D. F., & de Marchena, M. R. (1984). Developmental shifts in the basis of organization in memory: The role of associative versus categorical relatedness in children's free recall. *Child Development*, 952–962.

Bjorklund, D. F., Dukes, C., & Brown, R. D. (2008). The development of memory strategies. In *The development of memory in infancy and childhood* (pp. 157–188). Psychology Press.

Bower, G. H. (1970). Organizational factors in memory. *Cognitive Psychology*, *1*(1), 18–46.

Brady, T. F., Robinson, M. M., Williams, J. R., & Wixted, J. (2021). *Measuring memory is harder than you think: How to avoid problematic measurement practices in memory research* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/qd75k

Bull, R., Espy, K. A., & Wiebe, S. A. (2008). Short-Term Memory, Working Memory, and Executive Functioning in Preschoolers: Longitudinal Predictors of Mathematical Achievement at Age 7 Years. *Developmental Neuropsychology*, *33*(3), 205–228. https://doi.org/10.1080/87565640801982312

Camos, V., & Barrouillet, P. (2011). Developmental change in working memory strategies: From passive maintenance to active refreshing. *Developmental Psychology*, *47*(3), Article 3. https://doi.org/10.1037/a0023193

Camos, V., Johnson, M., Loaiza, V., Portrat, S., Souza, A., & Vergauwe, E. (2018). What is attentional refreshing in working memory?: What is attentional refreshing? *Annals of the New York Academy of Sciences*, *1424*(1), 19–32. https://doi.org/10.1111/nyas.13616

Cappiello, M., & Zhang, W. (2016). A dual-trace model for visual sensory memory. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(11), 1903–1922. https://doi.org/10.1037/xhp0000274

Christophel, T. B., Iamshchinina, P., Yan, C., Allefeld, C., & Haynes, J.-D. (2018). Cortical specialization for attended versus unattended working memory. *Nature Neuroscience*, *21*(4), Article 4. https://doi.org/10.1038/s41593-018-0094-4

Conlin, J. A., Gathercole, S. E., & Adams, J. W. (2005). Children's working memory: Investigating performance limitations in complex span tasks. *Journal of Experimental Child Psychology*, *90*(4), Article 4. https://doi.org/10.1016/j.jecp.2004.12.001

Conrad, R. (1964). Acousitic confusions in immediate memory. *British Journal of Psychology*, *55*(1), 75–84. https://doi.org/10.1111/j.2044-8295.1964.tb00899.x

Corsale, K., & Ornstein, P. A. (1980). Developmental changes in children's use of semantic information in recall. *Journal of Experimental Child Psychology*, *30*(2), 231–245.

Cowan, N. (1998). *Attention and memory: An integrated framework*. Oxford University Press.

Cowan, N. (2016). Working Memory Maturation: Can We Get at the Essence of Cognitive Growth? *Perspectives on Psychological Science*, *11*(2), 239–264. https://doi.org/10.1177/1745691615621279

deBettencourt, M. T., Williams, S. D., Vogel, E. K., & Awh, E. (2021). Sustained Attention and Spatial Attention Distinctly Influence Long-term Memory Encoding. *Journal of Cognitive Neuroscience*, 1–17. https://doi.org/10.1162/jocn_a_01748

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Elliott, E. M., Morey, C. C., AuBuchon, A. M., Cowan, N., Jarrold, C., Adams, E. J., Attwood, M., Bayram, B., Beeler-Duden, S., Blakstvedt, T. Y., Büttner, G., Castelain, T., Cave, S., Crepaldi, D., Fredriksen, E., Glass, B. A., Graves, A. J., Guitard, D., Hoehl, S., … Voracek, M. (2021). Multilab Direct Replication of Flavell, Beach, and Chinsky (1966): Spontaneous Verbal Rehearsal in a Memory Task as a Function of Age. *Advances in Methods and Practices in Psychological Science*, *4*(2), 251524592110181. https://doi.org/10.1177/25152459211018187

Fahrenfort, J. J., van Driel, J., van Gaal, S., & Olivers, C. N. L. (2018). From ERPs to MVPA Using the Amsterdam Decoding and Modeling Toolbox (ADAM). *Frontiers in Neuroscience*, *12*, 368. https://doi.org/10.3389/fnins.2018.00368

Ferguson, A. N., Bowey, J. A., & Tilley, A. (2002). The Association between Auditory Memory Span and Speech Rate in Children from Kindergarten to Sixth Grade. *Journal of Experimental Child Psychology*, *81*(2), 141–156. https://doi.org/10.1006/jecp.2001.2646

Flavell, J. H., Beach, D. R., & Chinsky, J. M. (1966). Spontaneous Verbal Rehearsal in a Memory Task as a Function of Age. *Child Development*, *37*(2), Article 2. https://doi.org/10.2307/1126804

Gaillard, V. (2011). Developmental differences in working memory: Where do they come from? *Journal of Experimental Child Psychology*, *110*, 469–479.

Gathercole, S. E. (1999). Cognitive approaches to the development of short-term memory. *Trends in Cognitive Sciences*, *3*(11), 410–419. https://doi.org/10.1016/S1364-6613(99)01388-1

Gathercole, S. E., & Adams, A.-M. (1994). Children� s phonological working memory: Contributions of long-term knowledge and rehearsal. *Journal of Memory and Language*, *33*(5), 672–688.

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The Structure of Working Memory From 4 to 15 Years of Age. *Developmental Psychology*, *40*(2), Article 2. https://doi.org/10.1037/0012-1649.40.2.177

Geier, C. F., Garver, K., Terwilliger, R., & Luna, B. (2009). Development of Working Memory Maintenance. *Journal of Neurophysiology*, *101*(1), 84–99. https://doi.org/10.1152/jn.90562.2008

Hasselhorn, M. (1992). Task dependency and the role of category typicality and metamemory in the development of an organizational strategy. *Child Development*, *63*(1), 202–214.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex. *Science*, *293*(5539), Article 5539. https://doi.org/10.1126/science.1063736

Hitch, G. J., Halliday, M. S., Schaafstal, A. M., & Heffernan, T. M. (1991). Speech, "inner speech," and the development of short-term memory: Effects of picture-labeling on recall. *Journal of Experimental Child Psychology*, *51*(2), 220–234. https://doi.org/10.1016/0022-0965(91)90033-O

Irwin, D. E., & Yeomans, J. M. (1986). Sensory registration and informational persistence. *Journal of Experimental Psychology: Human Perception and Performance*, *12*(3), 343.

Jeanneret, S., Bartsch, L., Turoman, N., Fiave, P.A., & Vergauwe, E., (2022, November). Hit refresh: on decoding the neural signatures of attentional refreshing in working memory with EEG. Poster to be presented at the 63th Annual meeting of the Psychonomic Society. Boston, Massachusetts, USA. Kail, R., & Hagen, W. (1977). *Perspectives on the development of memory and cognition*. Erlbaum.

Káldi, T., & Babarczy, A. (2021). Linguistic focus guides attention during the encoding and refreshing of Working Memory content. *Journal of Memory and Language*, *116*, 104187. https://doi.org/10.1016/j.jml.2020.104187

Karch, J. D., Sander, M. C., von Oertzen, T., Brandmaier, A. M., & Werkle-Bergner, M. (2015). Using within-subject pattern classification to understand lifespan age differences in oscillatory mechanisms of working memory selection and maintenance. *NeuroImage*, *118*, 538–552. https://doi.org/10.1016/j.neuroimage.2015.04.038

Lange, G., & Jackson, P. (1974). Personal organization in children's free recall. *Child Development*, 1060–1067.

LaRocque, J. J., Lewis-Peacock, J. A., Drysdale, A. T., Oberauer, K., & Postle, B. R. (2013). Decoding Attended Information in Short-term Memory: An EEG Study. *Journal of Cognitive Neuroscience*, *25*(1), 127–142. https://doi.org/10.1162/jocn_a_00305

Lewis-Peacock, J. A., Drysdale, A. T., Oberauer, K., & Postle, B. R. (2012). Neural Evidence for a Distinction between Short-term Memory and the Focus of Attention. *Journal of Cognitive Neuroscience*, *24*(1), Article 1. https://doi.org/10.1162/jocn_a_00140

Lewis-Peacock, J. A., Drysdale, A. T., & Postle, B. R. (2015). Neural Evidence for the Flexible Control of Mental Representations. *Cerebral Cortex*, *25*(10), Article 10. https://doi.org/10.1093/cercor/bhu130

Lintz, E. N., & Johnson, M. R. (2021). Refreshing and removing items in working memory: Different approaches to equivalent processes? *Cognition*, *211*, 104655. https://doi.org/10.1016/j.cognition.2021.104655

Lorenc, E. S., Mallett, R., & Lewis-Peacock, J. A. (2021). Distraction in Visual Working Memory: Resistance is Not Futile. *Trends in Cognitive Sciences*, *25*(3), Article 3. https://doi.org/10.1016/j.tics.2020.12.004

Lui, K. F. H., Lo, J. C. M., Maurer, U., Ho, C. S. -H., & McBride, C. (2021). Electroencephalography decoding of Chinese characters in primary school children and its prediction for word reading performance and development. *Developmental Science*, *24*(3). https://doi.org/10.1111/desc.13060

Magimairaj, B. M., & Montgomery, J. W. (2013). Examining the Relative Contribution of Memory Updating, Attention Focus Switching, and Sustained Attention to Children's Verbal Working Memory Span. *Child Development Research*, *2013*, 1–12. https://doi.org/10.1155/2013/763808

Mandler, G. (2002). Organisation: What levels of processing are levels of. *Memory*, *10*(5–6), 333–338. https://doi.org/10.1080/09658210244000153

Mares, I., Ewing, L., Farran, E. K., Smith, F. W., & Smith, M. L. (2020). Developmental changes in the processing of faces as revealed by EEG decoding. *NeuroImage*, *211*, 116660. https://doi.org/10.1016/j.neuroimage.2020.116660

Massaro, D., & Loftus, G. (1996). Sensory storage: Icons and echoes. *Handbook of Perception and Cognition*, *10*, 68–101.

Meehl, P. E. (1990). *Why Summaries of Research on Psychological Theories are Often Uninterpretable*. *66*, 195–244.

Melinder, A., Gredeback, G., Westerlund, A., & Nelson, C. A. (2010). *Brain activation during upright and inverted encoding of own- and other-age faces: ERP evidence for an own-age bias*. 11.

Miller, G. (1956). Human memory and the storage of information. *IRE Transactions on Information Theory*, *2*(3), 129–137.

Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: Multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, *10*(9), Article 9. https://doi.org/10.1016/j.tics.2006.07.005

Oberauer, K., & Souza, A. S. (2020). How fast can people refresh and rehearse information in working memory? *Memory & Cognition*, *48*(8), 1442–1459. https://doi.org/10.3758/s13421-020-01062-0

Oftinger, A.-L., & Camos, V. (2015). Maintenance Mechanisms in Children's Verbal Working Memory. *Journal of Educational and Developmental Psychology*, *6*, 16. https://doi.org/10.5539/jedp.v6n1p16

Oftinger, A.-L., & Camos, V. (2017). Phonological Similarity Effect in Children's Working Memory: Do Maintenance Mechanisms Matter? *Journal of Child Psychology*, *1*(1), Article 1. https://www.pulsus.com/abstract/phonological-similarity-effect-in-childrens-working-memory-do-maintenance-mechanisms-matter-3908.html

Oftinger, A.-L., & Camos, V. (2018). Developmental improvement in strategies to maintain verbal information in working memory. *International Journal of Behavioral Development*, *42*(2), Article 2. https://doi.org/10.1177/0165025416679741

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Petit, S., Badcock, N. A., Grootswagers, T., & Woolgar, A. (2020). Unconstrained multivariate EEG decoding can help detect lexical-semantic processing in individual children. *Scientific Reports*, *10*(1), 10849. https://doi.org/10.1038/s41598-020-67407-6

Pion-Tonachini, L., Kreutz-Delgado, K., & Makeig, S. (2019). ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. *NeuroImage*, *198*, 181–197. https://doi.org/10.1016/j.neuroimage.2019.05.026

Plude, D. J., Enns, J. T., & Brodeur, D. (1994). The development of selective attention: A life-span overview. *Acta Psychologica*, *86*(2–3), 227–272. https://doi.org/10.1016/0001-6918(94)90004-3

Portrat, S., Camos, V., & Barrouillet, P. (2009). Working memory in children: A time-constrained functioning similar to adults. *Journal of Experimental Child Psychology*, *102*(3), Article 3. https://doi.org/10.1016/j.jecp.2008.05.005

Rademaker, R. L., Bloem, I. M., De Weerd, P., & Sack, A. T. (2015). The Impact of Interference on Short-Term Memory for Visual Orientation. *Journal of Vision*, *41*(6), 1650–1665. https://doi.org/10.1167/13.9.804

Ricker, T. J., & Vergauwe, E. (2020). Consistent Failure to Produce a Cognitive Load Effect in Visual Working Memory Using a Standard Dual-Task Procedure. *Journal of Cognition*, *3*(1), 15. https://doi.org/10.5334/joc.108

Rose, N. S., LaRocque, J. J., Riggall, A. C., Gosseries, O., Starrett, M. J., Meyering, E. E., & Postle, B. R. (2016). Reactivation of Latent Working Memories with Transcranial Magnetic Stimulation. *Science (New York, N.Y.)*, *354*(6316), Article 6316. https://doi.org/10.1126/science.aah7011

Salatas, H., & Flavell, J. H. (1976). Behavioral and metamnemonic indicators of strategic behaviors under" remember" instructions in first grade. *Child Development*, 81–89.

Salthouse, T. A. (1994). The Aging of Working Memory. *Neuropsychology*, *8*(4), 535–543.

Scerif, G., Kotsoni, E., & Casey, B. (2006). Functional neuroimaging of early cognitive development. *Handbook of Functional Neuroimaging of Cognition*, *351*.

Scheel, A. M., Tiokhin, L., Isager, P. M., & Lakens, D. (2021). Why Hypothesis Testers Should Spend Less Time Testing Hypotheses. *Perspectives on Psychological Science*, *16*(4), 12.

Schleepen, T. M. J., & Jonkman, L. M. (2012). Children's use of semantic organizational strategies is mediated by working memory capacity. *Cognitive Development*, *27*(3), 255–269. https://doi.org/10.1016/j.cogdev.2012.03.003

Shimi, A., Nobre, A. C., & Scerif, G. (2015). ERP markers of target selection discriminate children with high vs. Low working memory capacity. *Frontiers in Systems Neuroscience*, *9*. https://doi.org/10.3389/fnsys.2015.00153

Shimi, A., & Scerif, G. (2017). Towards an integrative model of visual short-term memory maintenance: Evidence from the effects of attentional control, load, decay, and their interactions in childhood. *Cognition*, *169*, 61–83. https://doi.org/10.1016/j.cognition.2017.08.005

Sodian, B., Schneider, W., & Perlmutter, M. (1986). Recall, clustering, and metamemory in young children. *Journal of Experimental Child Psychology*, *41*(3), 395–410. https://doi.org/10.1016/0022-0965(86)90001-9

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied*, *74*(11), 1–29.

Tam, H., Jarrold, C., Baddeley, A. D., & Sabatos-DeVito, M. (2010). The development of memory maintenance: Children's use of phonological rehearsal and attentional refreshment in working memory tasks. *Journal of Experimental Child Psychology*, *107*(3), Article 3. https://doi.org/10.1016/j.jecp.2010.05.006

Treder, M. S. (2020). MVPA-Light: A Classification and Regression Toolbox for Multi-Dimensional Data. *Frontiers in Neuroscience*, *14*, 289. https://doi.org/10.3389/fnins.2020.00289

Trick, L. M., & Enns, J. T. (1998). Lifespan changes in attention: The visual search task. *Cognitive Development*, *13*(3), 369–386. https://doi.org/10.1016/S0885-2014(98)90016-8

Valentini, B., & Vergauwe, E. (2023). Stuck on the Last: The Last-Presented Benefit as an Index of Attentional Refreshing in Adolescents. *Journal of Intelligence*, *11*(1), 4. https://doi.org/10.3390/jintelligence11010004

Vergauwe, E., Besch, V., Latrèche, C., & Langerock, N. (2021). The use of attention to maintain information in working memory: A developmental investigation of spontaneous refreshing in school-aged children. *Developmental Science*. https://doi.org/10.1111/desc.13104

Vergauwe, E., Gonzalez-Garcia, C., Wisniewski, D., Langerock, N., & Brass, M. (In prep). *Vergauwe, E., Gonzalez-Garcia, C., Wisniewski, D. Langerock, N., & Brass, M. (in prep). Revealing the dynamic coding of working memory content over time in the face of distraction.*

Vergauwe, E., & Langerock, N. (2017). Attentional refreshing of information in working memory: Increased immediate accessibility of just-refreshed representations. *Journal of Memory and Language*, *96*, 23–35. https://doi.org/10.1016/j.jml.2017.05.001

Vergauwe, E., & Langerock, N. (2022). A (further) test of spontaneous serial refreshing in verbal and spatial working memory. *Attention, Perception, & Psychophysics*. https://doi.org/10.3758/s13414-022-02624-x