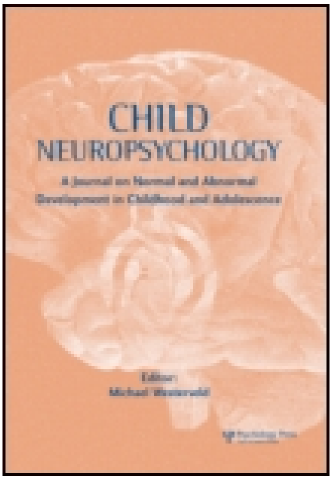

Child Neuropsychology: A Journal on Normal and Abnormal Development in Childhood and Adolescence
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/ncny20

Recall initiation strategies must be controlled in training studies that use immediate free recall tasks to measure the components of working memory capacity across time

Bradley S. Gibson[a], Dawn M. Gondoli[a], Ann C. Johnson[a] & Matthew K. Robison[a]

[a] Department of Psychology, University of Notre Dame, Notre Dame, Indiana, USA
Published online: 12 Aug 2013.

Child Neuropsychology, 2014
Vol. 20, No. 5, 539–556, http://dx.doi.org/10.1080/09297049.2013.826185

Routledge
Taylor & Francis Group

Recall initiation strategies must be controlled in training studies that use immediate free recall tasks to measure the components of working memory capacity across time

Bradley S. Gibson, Dawn M. Gondoli, Ann C. Johnson, and Matthew K. Robison

Department of Psychology, University of Notre Dame, Notre Dame, Indiana, USA

There has been great interest in using working memory (WM) training regimens as an alternative treatment for ADHD, but it has recently been concluded that existing training regimens may not be optimally designed because they target the primary memory component but not the secondary component of WM capacity. This conclusion requires the ability to accurately measure changes in primary and secondary memory abilities over time. The immediate free recall task has been used in previous studies to measure these changes; however, one concern with these tasks is that the recall order required on training exercises may influence the recall strategy used during free recall, which may in turn influence the relative number of items recalled from primary and secondary memory. To address this issue, previous training studies have explicitly controlled recall strategy before and after training. However, the necessity of controlling for recall strategies has not been explicitly tested. The present study investigated the effects of forward-serial-order training on free recall performance under conditions in which recall strategy was not controlled using a sample of adolescents with ADHD. Unlike when recall order was controlled, the main findings showed selective improvement of the secondary memory component (as opposed to the primary memory component) when recall order was uncontrolled. This finding advances our understanding of WM training by highlighting the importance of controlling for recall strategies when free recall tasks are used to measure changes in the primary and secondary components of WM across time.

Keywords: ADHD; Working memory training; Dual-component model of working memory.

A debate has emerged over the past decade concerning the extent to which working memory (WM) training regimens can ameliorate the putative WM deficiencies and related behavioral symptoms associated with attention deficit/hyperactivity disorder (ADHD; see, e.g., Diamond & Lee, 2011; Klingberg, 2010; Melby-Lervåg & Hulme, 2012; Shipstead, Hicks, & Engle, 2012; Shipstead, Redick, & Engle, 2012, for recent reviews). In consideration of the key factors involved in this debate, Gibson, Gondoli, Johnson, Steeger, and Morrissey (2012) have recently argued that the full potential of WM training has not yet been adequately tested, primarily because existing span-based regimens do not target the

Address correspondence to Bradley S. Gibson, Department of Psychology, 118 Haggar Hall, University of Notre Dame, Notre Dame, IN 46556, USA. E-mail: bgibson@nd.edu

appropriate theoretical mechanisms likely to be involved in ADHD. For instance, according to the dual-component model (Unsworth & Engle, 2007a, 2007b; Unsworth & Spillers, 2010), WM capacity is thought to reflect contributions from both an attention component that serves to actively maintain a limited amount of goal-relevant information in immediate or "primary memory" (PM) and a memory component that serves to retrieve goal-relevant information from episodic or "secondary memory" (SM) back into PM, once that information has been lost from PM. Within this theoretical framework, Gibson, Gondoli, Flies, Dobrzenski, and Unsworth (2010) showed that middle-school students with ADHD were primarily deficient in the SM component, but not the PM component, relative to age-matched individuals without ADHD.

To address whether WM-training regimens target the deficient SM component, Gibson, Gondoli, Johnson, Steeger, Dobrzenski, and Morrissey (2011) investigated whether the PM, SM, or both components of WM capacity could be enhanced by one well-known and widely used adaptive WM-training regimen known as "Cogmed-RM™," which contains a mixture of both verbal and spatial simple span exercises. Because spatial simple span tasks may engage the components of WM capacity more than verbal simple span tasks (Kane et al., 2004; Miyake, Friedman, Rettinger, Shah, & Hegarty, 2001; Oberauer, 2005; Shah & Miyake, 1996), the exercises were divided into two separate training conditions—a verbal training condition and a spatial training condition—to examine whether spatial training might engage the SM component more than verbal training. The main findings showed that the training regimens selectively improved the number of items recalled from PM ($d = 0.52$), but not the number of items recalled from SM ($d = 0.15$). Furthermore, the same pattern was observed across both the verbal and spatial training conditions. Because the component of WM capacity that was observed to be improved by the training regimen (PM) was not the same component that was observed to be deficient in ADHD (SM), Gibson et al. concluded that this training regimen may not be optimally designed, which may explain why this regimen has not been established to be effective in treating this disorder (Melby-Lervåg & Hulme, 2012; Shipstead, Hicks, & Engle, 2012).

The conclusion that adaptive WM-training regimens may selectively target specific components of WM capacity requires the ability to accurately measure changes in PM and SM over time. The study reported by Gibson et al. (2011) used verbal and spatial immediate free recall (IFR) tasks to measure the PM and SM components of WM capacity, based largely on Unsworth and Engle's (2007a) conclusion that reliable and valid measures of these two components can be derived from performance on these tasks (see also, Craik & Birtwistle, 1971; Unsworth, Spillers, & Brewer, 2010; Watkins, 1974). According to Unsworth and Engle (2007a), IFR tasks may be better suited for assessing recall from PM and SM than complex (e.g., operation span) or simple span (e.g., digit span) tasks because IFR tasks can provide separate measures of each component, whereas complex and simple span tasks typically provide a single measure that may reflect contributions from both components. For instance, Tulving and Colotla (1970) developed a method that can be applied to free recall that estimates the number of items recalled from PM and SM (Gibson et al., 2010, 2011, 2013; Gibson, Kronenberger, et al., 2012; Unsworth & Engle, 2007a; Unsworth et al., 2010).

According to Tulving and Colotla (1970), estimates of the number of items that can be recalled from PM and SM must take into consideration both input and output interference. Specifically, the greater the amount of interference preceding the recall of an item, the more likely the item will be recalled from SM as opposed to PM. Using this method, an item is considered to be recalled from PM when seven or fewer other items have been presented (input interference) or recalled (output interference) between the presentation of the item in question and its recall; whereas an item is considered to be recalled from SM when eight or more items have been presented or recalled between the presentation of the item in question and its recall. Notice that the assignment of an item is typically determined by the item's serial position and its recall order. For instance, the last item in the list will be recalled from PM if seven or fewer items are recalled before it, but this item will be recalled from SM if more than seven items are recalled before it. The one exception is when the presentation of an item is followed by the presentation of seven or more other items. In this case, the recall assignment of the item will be determined solely by its serial position; that is, the item will be recalled from SM regardless of the order in which it is recalled.

One concern with using Tulving and Colotla's (1970) method to estimate the PM and SM components of WM capacity is that the relative magnitude of these estimates can depend on an individual's "recall initiation strategy" (Gibson et al., 2010; Unsworth, Brewer, & Spillers, 2011; see also, Gibson et al., 2013, for a discussion of other concerns). For instance, Unsworth et al. identified three distinct recall initiation strategies in a sample of 150 undergraduate students who completed a verbal IFR task. In this sample, approximately 45% of the participants recalled mainly recency items (i.e., the last few items in the list), approximately 20% of the participants recalled mainly primacy items (i.e., the first few items in the list), and approximately 35% of the participants recalled equal proportions of both recency and primacy items.

Likewise, using a similar verbal IFR task in their study of middle-school students with and without ADHD, Gibson et al. (2010) found that approximately 54% of the participants recalled mainly recency items, approximately 27% of the participants recalled mainly primacy items, and approximately 19% of the participants recalled equal proportions of both recency and primacy items (similar percentages were also found when a spatial IFR task was used). Furthermore, the use of different recall initiation strategies did not differ between the two diagnostic groups.

Perhaps more importantly, Unsworth et al. (2011) found that individuals who favored a recency strategy tended to recall more items from PM than from SM, whereas individuals who favored a primacy strategy tended to recall more items from SM than from PM, even though the two groups did not differ in overall WM capacity (as measured by the operation span, reading span, and symmetry span tasks). Thus, differences in the order of recall can produce a trade-off in the number of items recalled from PM and SM, without reflecting a difference in WM capacity per se.

Understanding the linkage between these different recall initiation strategies and the extent to which items are recalled from PM or SM is potentially important in the context of WM training because the recall initiation strategy used after several weeks of training may differ from the recall initiation strategy used before training. Indeed, existing span-based training regimens such as Cogmed-RM™ typically use exercises that require strict forward serial-order recall, which may increase the likelihood that participants will use a primacy strategy on IFR tasks performed after training. If this shift toward the use of a primacy strategy on IFR tasks performed after training had been preceded by the use of a recency strategy on IFR tasks performed before training, then there would likely be a net increase in the number of items recalled from SM (as well as a potential decrease in the number of items recalled from PM). However, the observed increase in the SM outcome measure need not reflect an actual improvement in the ability to encode and retrieve information from SM. Rather such an increase may simply reflect a change in the recall initiation strategy without any coincident change in WM capacity.

In order to disentangle potential changes in the ability to recall information from PM and SM from potential changes in recall initiation strategy, Gibson et al. (2011) chose to control recall initiation strategy by instructing participants to use a recency strategy during both pretraining and posttraining administrations of the IFR tasks; in addition, compliance was monitored throughout the duration of the tasks, and reminders were provided as needed (see also, Gibson, Kronenberger, et al., 2012, 2013). The use of a recency strategy was chosen over the use of a primacy strategy because the dual-component model typically assumes that participants are using a recency recall initiation strategy during free recall (Gibson et al., 2010; Unsworth & Engle, 2007a).

However, although Gibson et al. (2011) controlled the recall initiation strategy, the rationale for doing so was not explicitly discussed in their study. Perhaps more importantly, the necessity of experimentally controlling for the recall initiation strategy used on IFR tasks administered before and after training was based on an untested assumption. Thus, it is currently unknown whether the recall initiation strategy used on IFR tasks performed after training would actually differ from the recall initiation strategy used on IFR tasks performed before training, if these strategies had been allowed to vary. In addition, if training does cause a change in the use of recall initiation strategies, it is also currently unknown whether such a change also affects the relative pattern of enhancement that is observed in the PM and SM components of WM capacity across time.

To address these issues, the present study was designed to be identical to the study conducted by Gibson et al. (2011), with the sole exception being that middle-school students with ADHD were now free to use different recall initiation strategies across time. These participants were randomly assigned to either a verbal or spatial training condition. If extended training with span-based exercises requiring forward serial-order recall increases the likelihood that participants will adopt a primacy recall initiation strategy on IFR tasks performed after training when the recall initiation strategy is uncontrolled, then the beneficial effects of the training regimen may now be confined to the SM component. Thus, the primary question of interest in the present study is not whether the pattern of enhancement observed in a training condition is greater than the pattern of enhancement observed in a control condition. Rather, the primary question of interest in the present study is whether the pattern of enhancement observed under conditions in which the recall initiation strategy is uncontrolled will be opposite to the known pattern that has been observed when the recall initiation strategy is controlled in previous studies (Gibson et al., 2011). These findings would be important because they would show that the same WM training regimens can produce different patterns of enhancement depending on IFR task instructions, and they would explicitly highlight the importance of controlling for recall initiation strategies in WM-training studies that use IFR tasks to measure potential changes in the PM and SM components of WM capacity across time.

METHOD

Recruitment of Participants

Following Gibson et al. (2011), adolescents with ADHD were recruited from two middle schools (Grades 6 to 8) in a Midwestern public school district. Initial contact letters briefly describing the study were provided to school officials who then mailed the letters to the parents of all adolescents in the district who were known to have a diagnosis of ADHD based on medical information provided to the school by the parents. Approximately 6% of the total population was contacted. A total of 31 families responded and were scheduled for an initial screening. The protocol for the present study was approved by the institutional review board at the University of Notre Dame.

Initial Screening

Following Gibson et al. (2011), the primary caregiver and adolescent began this 2-hour session by providing informed consent and assent, respectively, in a laboratory located in the Psychology Department at the University of Notre Dame. The primary caregiver provided information about the adolescent's treatment plan during this session. The type, time, and dosage of all medications were documented using a questionnaire. The primary caregiver then participated in a structured interview using the Computerized Diagnostic Interview Schedule for Children Version 4 (C-DISC-IV; Shaffer, Fisher, Lucas, Dulcan, & Schwab-Stone, 2000) to verify the presence of the number, age of onset, associated impairment, and cross-situational pervasiveness of the *Diagnostic and Statistical Manual of Mental Disorders*, fourth edition (*DSM-IV*; American Psychiatric Association, 1994) ADHD symptoms. Because there were no specific hypotheses regarding the subtypes of ADHD, no attempt was made to differentiate between ADHD-PI and ADHD-C. In addition, a variety of other, potentially comorbid, psychiatric conditions (such as anxiety, depression, and oppositional defiant disorder) were also examined. A diagnosis of ADHD was required for inclusion in the study. One of the 31 respondents failed to meet this diagnostic criterion.

The Wechsler Abbreviated Scale of Intelligence (WASI) was administered to each of the adolescents to obtain a general measure of cognitive functioning. Note that the initial phone contact with the primary caregiver revealed that 27 of the 31 adolescents were being treated with medication at the time of the initial screening. All stimulant ADHD medication and atomoxetine was withheld for at least 24 hours prior to the WASI; however, non-ADHD medication was not withheld. A full-scale IQ greater or equal to 70 was required for inclusion in the study. Two of the 31 respondents failed to meet this intellectual criterion.

Pretraining Assessment

This assessment was identical to the pretraining assessment used by Gibson et al. (2011). Immediately before the intervention, the adolescents and their primary caregivers participated in a 2-hour laboratory session at the same location as the initial screening. Primary caregivers were instructed to follow their adolescent's normal medication treatment plan before arriving to the pretraining session based on the information they provided at the initial screening session. The type, time, and dosage of that day's medication were also documented at the pretraining session using the same questionnaire that was used at the initial screening session.

During each session, the adolescent performed a verbal and spatial version of the IFR task. Both of these tasks were identical to the IFR tasks used by Gibson et al. (2011) and both were administered by a research assistant who was blind to training condition. In the verbal IFR task, participants were presented with 15 lists of 12 unique words that were randomly combined. The words were printed in 20-point font, and all words appeared white against the black background of a standard CRT monitor.[1] Each word was presented consecutively for one second in the middle of the computer screen. Following the presentation of a single list, question marks appeared in the center of the screen prompting a response by the participant. At this point, participants were instructed to verbally recall as many words as possible in any order. Participants reported their answers into a microphone that was connected to a digital recorder. A response was scored correct if it matched one of the list items, if it was a plural version of a singular list item ("boards" instead of "board" or vice versa), or if it was a past-tense version of a present-tense list item ("shot" instead of "shoot" or vice versa). Participants were given 30 seconds to recall the word lists, and they were required to wait the full 30 seconds before proceeding to the next trial. This mandatory 30-second recall period ensured that participants could not prematurely terminate their recall in the event that they found the delay imposed by the recall period aversive (Sonuga-Barke, 2003). Three practice trials using letter stimuli (instead of words) preceded the experimental trials. The word lists were presented in the same random order to all subjects.

In the spatial IFR task, participants were presented with 15 lists of 12 different locations that were marked by white squares. Squares appeared at any one of 15 x 12 or 180 unique screen locations. Each location was cued only once across the 15 different lists. Locations were cued by temporarily changing the color of the square from white to red. Each of the 12 locations in a list was cued in consecutive order for one second. At the conclusion of the list, participants were prompted to use the computer mouse to recall as many locations as possible by clicking on the relevant locations. As in the verbal IFR task, participants were told that they could recall the locations in any order. For each response, the computer recorded the location of the mouse click, the order of the mouse click, the time of each mouse click, and the number of correct recall responses. A response was scored correct if it matched one of the list items. As in the verbal task, participants were given 30 seconds to recall the spatial lists, and they were required to wait the full 30 seconds before proceeding to the next trial. In order to make the task manageable, only 60 of the possible 180 squares appeared at any one time. These 60 squares were selected randomly from the set of 180 possible locations and they remained visible for five consecutive trials (5 trials x 12 cued locations = 60 locations). At the conclusion of the fifth trial in each set, a new set of 60 locations was randomly selected from the 180 possible locations without replacement. This sequence of five trials repeated three times for a total of 15 trials. The three sets of 60 randomly selected squares were determined separately for each participant. As in the verbal task, three practice trials preceded the experimental trials. Primary caregivers and adolescents were paid $20.00 each for participating in the pretraining session.

[1] One potential limitation in the present study, as well as in our previous training study (Gibson et al., 2011), is that participants were not screened for reading problems before completing the verbal IFR task. However, there is no reason to believe that the presence versus absence of a reading disability should be associated with a change in recall initiation strategy. In addition, the likelihood of recruiting a participant with a reading disability should be similar across the present study and our previous training study because the participants were the same age, had the same diagnosis and IQ scores and were recruited from the same three middle schools. Finally, the presence versus absence of a reading disability should not affect performance on the spatial IFR task.

Outcome Measures

Three primary outcome measures were calculated from the IFR tasks. First, "probability of first recall" was used to assess recall initiation strategies. Probability of first recall refers to the number of times an item presented at each serial position was reported first divided by the total number of trials. Participants who use a recency recall initiation strategy tend to begin recall by reporting items from the end of the list first whereas participants who use a primacy recall initiation strategy tend to begin recall by reporting items from the beginning of the list first (Gibson et al., 2010; Unsworth et al., 2011). In addition to allowing probability of first recall to vary in a continuous fashion, we also used this measure to classify participants into one of three groups for descriptive purposes. In particular, following Unsworth et al., we also averaged probability of first recall across the last three serial positions (Positions 10, 11, and 12) and the first three serial positions (Positions 1, 2, and 3) and then took the difference between these two averages (Last average – First average). Participants were classified as primarily using a recency recall initiation strategy when the difference was 0.10 or greater whereas participants were classified as primarily using a primacy recall initiation strategy when the difference was −0.10 or less. Participants were classified as using a combination of both recall initiation strategies when the difference was in between −0.10 and 0.10.

The second outcome measure used Tulving and Colotla's (1970) method to provide estimates of the number of items that can be recalled from PM and SM. An item was considered to be recalled from PM when there were seven or fewer items intervening between that item's presentation and its recall. In contrast, an item was considered to be recalled from SM when there were more than seven items intervening between that item's presentation and its recall.

Third, probability correct as a function of serial position was used to reinforce the conclusions based on Tulving and Colotla's (1970) method. If training targets only the PM component, then improvements should be confined to the recency portion of the curve; conversely, if training targets only the SM component, then improvements should be confined to the prerecency portion of the curve. If training targets both components, then improvements should extend across the entire range of serial positions.

Working Memory Training

The verbal and spatial WM training conditions were identical to those used by Gibson et al. (2011). The 28 adolescents who met the diagnostic and intellectual inclusion criteria were randomly assigned to one of the two WM-training groups. A programmer employed by Cogmed generated the random allocation sequence and then assigned participants to one of the two conditions in a consecutive fashion according to their ID number. Thirteen adolescents were randomly assigned to the verbal WM-training group and 15 adolescents were randomly assigned to the spatial WM-training group. Participants were informed that they would be randomly assigned to one of two training conditions, but the nature of these two training conditions was not described further. Both the participants and the research assistants assessing outcomes were blind to training condition. Note that the primary caregiver was asked to maintain their adolescent's normal treatment plan during the training phase of the study based on information provided at the pretraining session.

Each of the two training conditions was composed of a total of six exercises: Five exercises were unique to each training condition and one exercise was common to both training conditions. Nine of the 11 exercises were extracted directly from the standard Cogmed-RM intervention; the two remaining exercises were modified versions of the standard exercises.

The five unique exercises in the verbal training group were Corrector, Decoder, Input Module with Lid (forward version), Stabilizer (standard version), and Stabilizer (random version). These five simple span exercises involved remembering lists of letters or digits. Note that in the two Stabilizer exercises, dots of light appeared either in an ordered sequence (standard version) or in a random sequence (random version) along with each of the to-be-remembered letters. At the completion of the list, one of the letters from the list was presented and participants had to click on the dot that had been associated with the letter. Thus, some of the verbal tasks may have contained a spatial component. All the exercises required participants to remember the lists in the forward serial order.

The five unique exercises in the spatial training group were Asteroids, Data Room, Rotating Dots, Visual Data Link (stationary version), and Visual Data Link (rotating version). These five simple span exercises involved remembering the positions of dots in two- or three-dimensional grids that translated, rotated, or remained stable. All the exercises required participants to remember the lists in the forward serial order.

The one common exercise (Space Whack) was included to help maintain blindness to training condition in the event that two participants happened to discuss their training experience with each other (though this did not occur). In addition, the common exercise was also included to break the monotony of the five unique exercises as this exercise resembled the popular "whack-a-mole" arcade game (albeit involving space aliens that emerged from craters) and it was judged to be the most "game-like" by the researchers.

For each of the exercises, the length of the list was automatically adjusted, on a trial-by-trial basis, to match the WM span of the participant on that particular exercise. Both WM-training groups completed their exercises at home on a personal computer that was connected to the Internet. Participants were allowed to complete the exercises in any order. All recall responses were reported by clicking a cursor on the display using a computer mouse. Positive feedback was provided verbally by the program after most successful trials; in addition, participants also received "energy" on all successful trials that they could use in a video game involving racing robots at the end of each day of training. Participants completed a total of 120 trials per day (20 trials x 6 exercises) before they were allowed to progress to the next day of training. Performance on the training exercises was automatically uploaded to a secure Web site that was monitored for compliance by the authors. All the authors were certified by Cogmed to administer the intervention, and they made weekly phone calls to the adolescents to provide feedback and to answer questions about the training in accordance with this certification.

Both WM-training groups were instructed to complete a maximum of 25 days of training, and a minimum of 20 days was required for inclusion in the study. Four of the 28 adolescents (14.89%) withdrew from the experiment before they started the training ($n = 2$ in each of the two training conditions) but after they had been randomly assigned. In addition, 2 of the 28 adolescents failed to complete at least 20 days of training (both were in the verbal training condition).

Posttraining Assessment

Within one week of completing the intervention, the primary caregiver and adolescent returned to the laboratory for a 2-hour laboratory session at the same location as the pretraining session. During scheduling of the posttraining session, the primary caregiver was reminded to medicate their adolescent at the same time and with the same dosage based on the information they provided at the pretraining session. The posttraining session was also scheduled at the same time of day as the pretraining session. The same primary outcome measures that were administered at the pretraining session were re-administered at the posttraining session, and all participants were compensated at the same rate. All participants completed the study during the spring semester.

RESULTS

Table 1 describes the characteristics of the 22 adolescents who met the initial diagnostic, intellectual, and WM-training inclusion criteria. As expected, the adolescents in the two training conditions did not differ in full-scale IQ, $t(20) = -0.42$, $p > .65$, verbal IQ, $t(20) = -0.13$, $p > .85$, performance IQ, $t(20) = -0.46$, $p > .65$, or age, $t(20) = -0.81$, $p > .40$.

Analysis of Working Memory Training

Figure 1 depicts average span length as a function of training duration in each of the two training conditions. The effect of training duration on average span length was evaluated using a two-way, mixed analysis of variance (ANOVA) with training duration (Day 1 to Day 20) as the sole within-subjects factor and with training condition (verbal vs. spatial) as the sole between-subjects factor. The *p* values associated with the effects of training duration were Greenhouse-Geisser corrected to offset the statistical consequences

Table 1 Intelligence Scores and Ages of the Participants Listed as Function of Training Condition (Standard Error).

	Verbal training	Spatial training
Full-Scale IQ	102.10 (4.31)	104.50 (3.94)
Verbal IQ	103.50 (5.21)	104.42 (4.75)
Performance IQ	100.70 (4.12)	103.17 (3.76)
Age in years	12.00 (0.38)	12.33 (0.34)

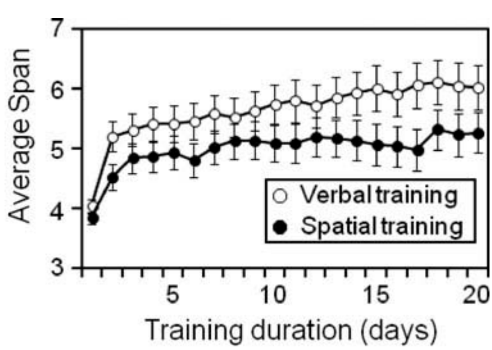

Figure 1 Average training spans depicted as a function of training duration and training condition. Error bars represent standard errors.

of violating the assumption of sphericity. As expected, average span length increased sharply over the first few days in both training groups and then increased more gradually over the remainder of the training period, as indicated by a significant main effect of training duration, $F(19, 380) = 22.58$, $p < .0001$, $\eta_p^2 = .53$. Although participants in the verbal training condition appeared to achieve greater average spans over time than participants in the spatial training condition, neither the main effect of training condition nor the training Duration x Training condition interaction achieved significance: $F(1, 20) = 2.85$, $p > .10$, $\eta_p^2 = .12$, and $F(19, 380) = 1.56$, $p > .15$, $\eta_p^2 = .07$, respectively.

Probability of First Recall

The percentage of participants that were classified in each of the three recall initiation strategy groups based on Last average – First average are listed in Table 2 as a function of time, IFR task modality, and training condition. Examination of the descriptive information provided in Table 2 revealed that a greater percentage of participants used the recency strategy on the IFR tasks performed during the pretraining assessment ($M = 60\%$) than on the IFR tasks performed during the posttraining assessment ($M = 27\%$). Conversely, a greater percentage of participants used the primacy strategy on the IFR tasks performed during the posttraining assessment ($M = 39\%$) than on the IFR tasks performed during the pretraining assessment ($M = 12\%$). Approximately the same percentage of participants used a combination of the two strategies on the IFR tasks performed during the pretraining assessment ($M = 29\%$) and on the IFR tasks performed during the posttraining assessments ($M = 34\%$).

In order to examine these apparent changes more closely, a four-way, mixed-model ANOVA was conducted on probability of first recall with time (pretraining vs. posttraining), serial position (Positions 1 to 12), and task modality (verbal IFR vs. spatial IFR) as the three within-subjects factors and with training condition (verbal vs. spatial) as the sole between-subjects factor. The *p* values associated with the effects of serial position were Greenhouse-Geisser corrected to offset the statistical consequences of violating the assumption of sphericity. The probability of first recall is shown in Figure 2 as a function of time, serial position, and IFR task modality in each of the verbal (left panels) and spatial (right panels) training conditions.

Of greatest importance, there was a significant Time x Serial position x Modality x Training condition interaction, $F(11, 220) = 5.05$, $p < .001$, $\eta_p^2 = .20$, indicating that the likelihood of initiating recall with the first list item (and therefore of using a primacy recall

Table 2 Percentage of Participants in Each of the Three Strategy Groups Listed as Function of Time, Immediate Free Recall (IFR) Task Modality, and Training Condition.

	Pre-verbal IFR	Post-verbal IFR	Pre-spatial IFR	Post-spatial IFR
Verbal training ($n = 10$)				
Recency	50%	20%	80%	30%
Primacy	20%	60%	10%	30%
Both	30%	20%	10%	40%
Spatial training ($n = 12$)				
Recency	58%	50%	50%	8%
Primacy	8%	16%	8%	50%
Both	33%	33%	42%	42%

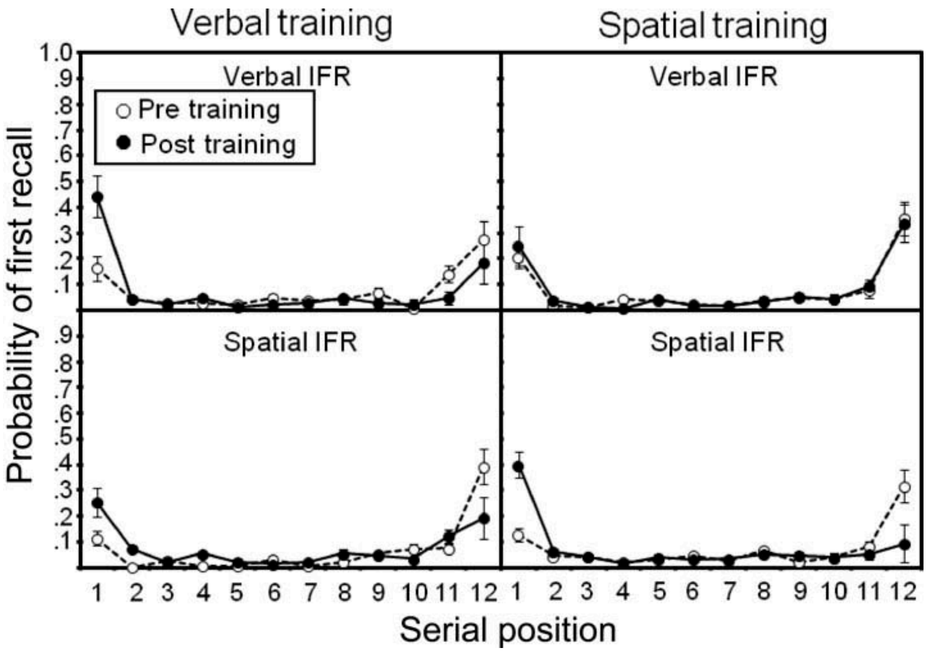

Figure 2 Probability of first recall depicted as a function of time, serial position, and task modality in each of the verbal and spatial training conditions. Error bars represent standard errors.

initiation strategy) increased substantially after WM training, especially when the modality of the IFR task matched the modality of the training condition. The nature of this four-way interaction was explored further by analyzing the effects of time, serial position, and task modality within each of the two training conditions separately.

With respect to the verbal training condition, the three-way interaction between time, serial position, and task modality was found to be significant, $F(11, 99) = 2.99$, $p < .05$, $\eta_p^2 = .25$. As can be seen in the left panel of Figure 2, the probability of initiating recall with the first list item was 0.28 higher on the verbal IFR task after verbal WM training, $t(9) = 3.24$, $p < .01$, whereas this probability was only 0.14 higher on the spatial IFR task after verbal WM training, though this latter increase also attained significance, $t(9) = 2.77$, $p < .025$. In addition, when the probability of initiating recall with the first list item increased significantly after verbal WM training, the probability of initiating recall with the last or the second-to-last list item (and therefore of using a recency recall initiation strategy) often decreased after verbal WM training. For instance, the probabilities of initiating recall with the last and the second-to-last list items were both 0.09 lower on the verbal IFR task after verbal WM training, $t(9) = 1.25$, $p > .20$, and $t(9) = 2.43$, $p < .05$, respectively. Likewise, the probability of initiating recall with the last list item was 0.20 lower on the spatial IFR task after verbal WM training, $t(9) = 3.06$, $p < .02$.

With respect to the spatial training condition, the three-way interaction between time, serial position, and task modality was also found to be significant, $F(11, 121) = 5.89$, $p < .001$, $\eta_p^2 = .35$. As can be seen in the right panel of Figure 2, the probability of initiating recall with the first list item was 0.27 higher on the spatial IFR task after spatial WM training, $t(11) = 3.63$, $p < .005$, whereas this probability was only 0.04 higher on the verbal

IFR task after spatial WM training, and this latter increase did not approach significance, $t(11) = 1.12$, $p > .25$. Conversely, the probability of initiating recall with the last list item was 0.22 lower on the spatial IFR task after spatial WM training, $t(11) = 3.70$, $p < .005$. Altogether, these findings are consistent with the hypothesis that intensive WM training involving forward serial-order recall can increase the likelihood that participants will use a primacy recall initiation strategy on IFR tasks when these strategies are left uncontrolled. In addition, the present findings also provided some evidence that the observed change in recall initiation strategy was stronger when the modality of the IFR task matched the modality of the training exercises.

Number of Items Recalled from PM and SM

The number of items recalled from PM and SM is shown in Figure 3 as a function of time and IFR task modality in each of the verbal (left panels) and spatial (right panels) training conditions. A four-way, mixed-model ANOVA was conducted on the number of items recalled with time (pretraining vs. posttraining), WM component type (PM vs. SM), and task modality (verbal IFR vs. spatial IFR) as the three within-subjects factors and with training condition (verbal vs. spatial) as the sole between-subjects factor.

If the number of items recalled from SM is related to the use of a primacy recall initiation strategy (Unsworth et al., 2011), then there should be a significant increase in the number items recalled from SM in the present study. As can be seen in Figure 3, there was selective enhancement of the SM component after training when recall order was not

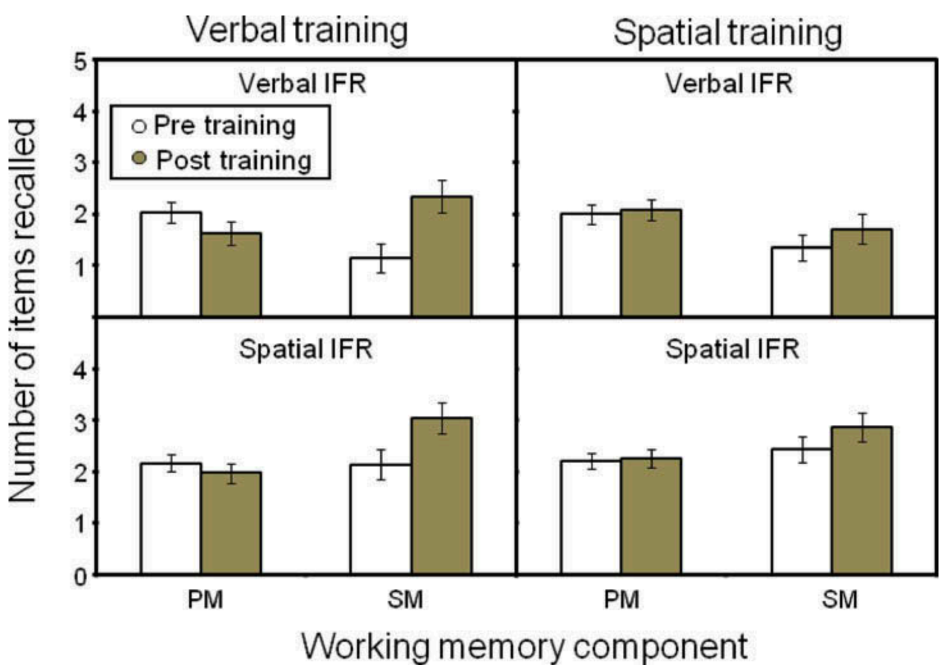

Figure 3 Estimates of the number of items recalled using Tulving and Colotla's (1970) method. These estimates are depicted as a function of time, working memory component, and task modality in each of the verbal and spatial training conditions. Error bars represent standard errors.

controlled. In addition, this selective enhancement of the SM component was also numerically larger after training when the modality of the IFR task matched the modality of the training condition. For instance, participants recalled 1.20 more items from SM on the verbal IFR task after verbal WM training whereas they only recalled 0.894 more items from SM on the spatial IFR task after verbal WM training. Conversely, participants recalled 0.44 more items from SM on the spatial IFR task after spatial WM training whereas they only recalled 0.36 more items from SM on the verbal IFR task after spatial WM training. Despite this pattern, the four-way interaction did not approach significance, $F(1, 20) < 1$. However, there was a significant three-way interaction between time, WM-component type, and training condition, $F(1, 20) = 5.44$, $p < .05$, $\eta_p^2 = .21$, that was consistent with the current expectations.

The nature of this three-way interaction was explored further by analyzing the effects of time and training condition on each of the PM and SM conditions separately. With respect to the SM component, there was a significant main effect of time, $F(1, 20) = 31.30$, $p < .0001$, $\eta_p^2 = .61$. In addition, there was also a significant Time x Training condition interaction, $F(1, 20) = 6.33$, $p < .025$, $\eta_p^2 = .24$, indicating that the increase in the number of items recalled after training was significantly larger in the verbal training condition ($M = 1.04$ more items) than in the spatial training condition ($M = 0.40$ more items), though both changes were found to be significant, $t(9) = 5.24$, $p < .002$, and $t(11) = 2.38$, $p < .05$, respectively. In contrast, with respect to the PM component, neither the main effect of time, $F(1, 20) = 1.25$, $p > .25$, $\eta_p^2 = .06$, nor the Time x Training condition, $F(1, 20) = 3.07$, $p > .10$, $\eta_p^2 = .13$, attained significance.

Probability Correct

Probability correct is shown in Figure 4 as a function of time, serial position, and IFR task modality in each of the verbal (left panels) and spatial (right panels) training conditions. A four-way, mixed-model ANOVA was conducted on probability correct with time (pretraining vs. posttraining), serial position (Positions 1 to 12), and task modality (verbal IFR vs. spatial IFR) as the three within-subjects factors and with training condition (verbal vs. spatial) as the sole between-subjects factor. The p values associated with the effects of serial position were Greenhouse-Geisser corrected to offset the statistical consequences of violating the assumption of sphericity.

If probability correct as a function of serial position is related to the use of a primacy recall initiation strategy (Unsworth et al., 2011), then there should be a significant increase in the proportion of prerecency items that are correctly recalled in the present study. As can be seen in Figure 3, there was selective enhancement of the prerecency items after training when recall order was not controlled. Although this selective enhancement of the prerecency items appeared to be slightly larger when the modality of the IFR task matched the modality of training, the four-way interaction between time, serial position, IFR task modality, and training conditions did not approach significance, $F < 1$. However, there was a significant three-way interaction between time, serial position, and training condition, $F(11, 220) = 3.64$, $p < .02$, $\eta_p^2 = .15$, that was consistent with the current expectations.

The nature of this three-way interaction was explored further by analyzing the effects of time and training condition on each of the recency and prerecency items separately. Following Unsworth et al. (2011), the analysis of recency items focused on the last three items and the analysis of prerecency items focused on the first three items. With respect

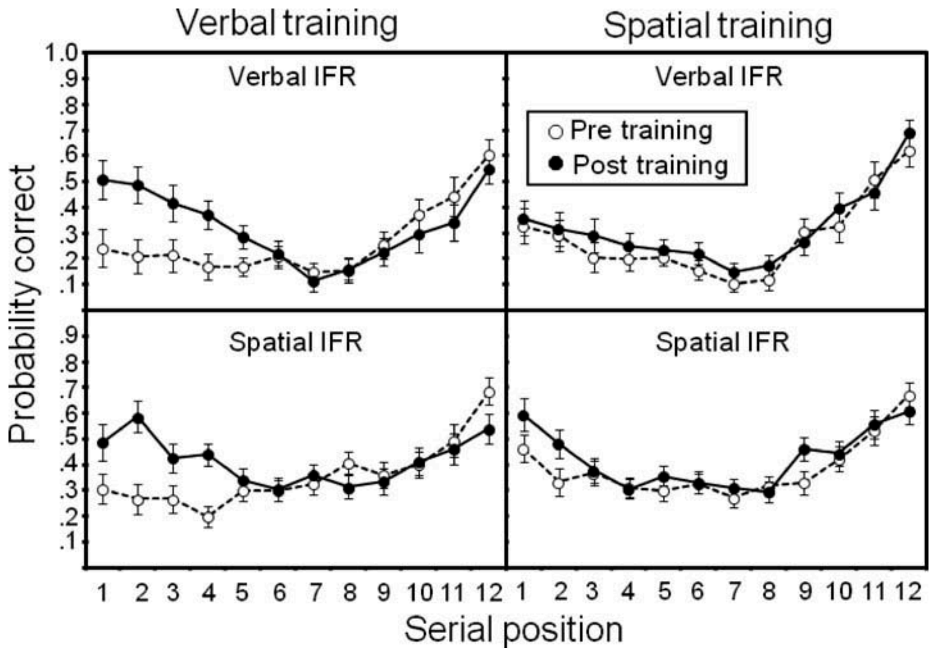

Figure 4 Probability correct depicted as a function of time, serial position, and task modality in each of the verbal and spatial training conditions. Error bars represent standard errors.

to the prerecency items, there was a significant main effect of time, $F(1, 20) = 31.47$, $p < .0001$, $\eta_p^2 = .61$. In addition, there was also a significant Time x Training condition interaction, $F(1, 20) = 8.82$, $p < .01$, $\eta_p^2 = .31$, indicating that the increase in probability correct after training was significantly larger in the verbal training condition ($M = 0.23$ more items), $t(9) = 5.66$, $p < .001$, than in the spatial training condition ($M = 0.07$ more items), $t(11) = 2.00$, $p = .07$. In contrast, with respect to the recency items, neither the main effect of time, $F(1, 20) = 0.77$, $p > .35$, $\eta_p^2 = .04$, nor the Time x Training condition, $F(1, 20) = 1.63$, $p > .20$, $\eta_p^2 = .08$, attained significance. Overall, the findings obtained when probability correct was analyzed corroborated the findings obtained when the number of items recalled from PM and SM was analyzed.

GENERAL DISCUSSION

The conclusion that adaptive WM-training regimens may selectively target the specific components of WM capacity requires the ability to accurately measure changes in PM and SM over time. Accordingly, the present study was conducted to highlight the importance of controlling for various recall initiation strategies that may influence the measurement of these components. Two important findings were obtained in the present study. First, as expected, examination of the probability of first recall responses suggested that participants' use of the primacy recall initiation strategy generally increased after training when no explicit recall instructions were given. This was indicated by a significant increase in the probability that recall was initiated from the beginning of the list after training, and this increase was attributed to the fact that participants had trained with simple span tasks that required forward serial-order recall. Second, examination of the number of items recalled from PM and SM, as well as serial position effects, suggested that the SM component was selectively enhanced when recall order was not controlled and use of the primacy recall initiation strategy increased after training. This finding stands in contrast to the findings reported in our previous study that concluded that the PM component was selectively enhanced when middle-school students with ADHD were instructed to maintain a recency recall initiation strategy both before and after training (Gibson et al., 2011).

The present findings are therefore important because they showed that identical training regimens and identical outcome measures can produce two quite different patterns of change in the PM and SM components of WM capacity simply by modifying the IFR task instructions. Obviously, these two disparate sets of findings would have been difficult to reconcile had they been reported without explicitly emphasizing the rather subtle difference between controlling versus not controlling the order of recall used on IFR tasks before and after training. Hence, we believe it is important to highlight how potentially subtle changes in task instructions can lead to radically different, and potentially misleading, conclusions about the effects of WM training on the PM and SM components of WM capacity.

Perhaps more importantly, we also contend that only one of the two observed outcomes can actually be interpreted to reflect an increase in WM capacity. Specifically, we contend that training-induced changes in WM capacity can be accurately inferred only when participants use the same recall initiation strategy before and after training (as they did in Gibson et al., 2011), but not when they use different recall initiation strategies before and after training (as they did in the present study). This is because the observed changes may simply reflect a change in the order in which items are recalled rather than a change in capacity per se. Indeed, although Unsworth et al. (2011) found a trade-off in the number of items recalled from PM and SM as a function of whether participants used the primacy or recency recall initiation strategy, both groups were found to have the same overall WM capacity (as measured by the operation span, reading span, and symmetry span tasks).

In the present study, we did not find a strict trade-off in the number of items recalled from PM and SM after training in the uncontrolled recall condition. Rather, we found a significant increase in the number of items recalled from SM after training without a coincident decrease in the number of items recalled from PM, leading to an overall significant main effect time in the uncontrolled recall condition. In other words, the observed decrease in number of items recalled from PM appeared to be smaller than expected in the uncontrolled recall condition. This smaller-than-expected decrease in the number of items recalled from PM likely reflects the selective enhancement of the PM component that is known to follow from these training regimens (Gibson et al., 2011; Gibson, Kronenberger, et al., 2012), and that can be observed more clearly when participants are instructed to recall items from the end of the list first (as in the controlled recall condition). However, it is also worth considering whether this unexpected imbalance may have arisen because there was a greater-than-expected increase in the number of items recalled from SM after training rather than a smaller-than-expected decrease in the number of items recalled from PM.

For instance, although Unsworth et al. (2011) found no difference in overall WM capacity between participants who used the primacy and recency recall initiation strategies, the two groups were found to differ in other abilities that are not directly related to WM capacity (Unsworth, 2009): Namely, participants who used the primacy recall initiation strategy group tended to be more resistant to proactive interference than participants who used the recency recall initiation strategy. Thus, it is possible that extended training with simple span tasks requiring forward serial-order recall did more than simply change the order in which participants recalled the items. In particular, this training may have also made participants more resistant to proactive interference which in turn may have enabled them to increase the number of items they could recall from SM when they used a primacy recall initiation strategy. Such a notion is consistent with the idea that increases in certain cognitive abilities can enable more effective memory strategies (Dunlosky & Kane, 2007).

In order to investigate this possibility, we calculated the number of items from previous lists that were recalled on current lists—termed "previous-list intrusions" (PLIs). A three-way, mixed-model ANOVA was conducted on these PLIs with time and IFR task modality as the two within-groups factors and training condition as the sole between-group factor. However, neither the main effect of time nor any of the interactions involving time approached significance in this study (all $ps > .20$ or more). Thus, the present findings provided no evidence that there was a greater-than-expected increase in the number of items recalled from SM after training arising from greater resistance to proactive interference. Consequently, we conclude that the unexpected imbalance is due to a smaller-than-expected decrease in the number of items recalled from PM arising from the selective enhancement of the PM component (Gibson et al., 2011; Gibson, Kronenberger, et al., 2012).

We have thus far attributed the observed increase in the use of the primacy recall initiation strategy to extended exposure to training tasks that required forward serial-order recall. However, others have reported intraindividual changes in the relative use of different recall initiation strategies over shorter durations of time and without any intervening training. For instance, Unsworth et al. (2011) found that participants who tended to use the recency recall initiation strategy on the majority of trials actually tended to use the primacy recall initiation strategy on the first few trials of a verbal IFR task—a finding that had also been reported in several earlier studies (Dallett, 1963; Huang, 1986; Murdock, 1974). According to Unsworth et al., this strategy shift occurs as proactive interference builds over the first few trials, and it gets more difficult to differentiate early-list items from previous-list items. However, although there have been frequent reports of a primacy-to-recency strategy shift over time, there have been no previous reports of the opposite recency-to-primacy strategy shift, as was observed in the present study. Thus, it is unlikely that this recency-to-primacy strategy shift is simply due to repeated exposure to the task. That said, the main conclusion of the present study would still hold even if the observed increase in the use of the primacy recall initiation strategy was due to test-retest effects and not training: Namely, training-induced changes in WM capacity can be properly inferred only when participants use the same recall initiation strategy before and after training. Hence, it is necessary to highlight the importance of controlling for recall initiation strategies in WM-training studies that use IFR tasks to measure potential changes in the PM and SM components of WM capacity across time.

Original manuscript received May 9, 2013
Revised manuscript accepted July 14, 2013
First published online August 12, 2013

REFERENCES

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: American Psychiatric Association.

Craik, F. I. M., & Birtwistle, J. (1971). Proactive inhibition in free recall. *Journal of Experimental Psychology*, *91*, 120–123.

Dallett, K. M. (1963). Practice effects in free and ordered recall. *Journal of Experimental Psychology*, *66*, 65–71.

Diamond, A., & Lee, K. (2011). Interventions shown to aid executive function development in children 4 to 12 years old. *Science*, *333*, 959–964.

Dunlosky, J., & Kane, M. J. (2007). The contributions of strategy use to working memory span: A comparison of strategy-assessment methods. *Quarterly Journal of Experimental Psychology*, *60*, 1227–1245.

Gibson, B. S., Gondoli, D. M., Flies, A. C., Dobrzenski, B. A., & Unsworth, N. (2010). Application of the dual-component model of working memory to ADHD. *Child Neuropsychology*, *16*, 60–79.

Gibson, B. S., Gondoli, D. M., Johnson, A. C., Steeger, C. M., Dobrzenski, B. A., & Morrissey, R. A. (2011). Component analysis of verbal versus spatial working memory training in adolescents with ADHD: A randomized, controlled trial. *Child Neuropsychology*, *17*, 546–563.

Gibson, B. S., Gondoli, D. M., Johnson, A. C., Steeger, C. M., & Morrissey, R. M. (2012). The future promise of Cogmed working memory training. *Journal of Applied Research in Memory and Cognition*, *1*, 214–216.

Gibson, B. S., Gondoli, D. M., Kronenberger, W. G., Johnson, A. C., Steeger, C. M., & Morrissey, R. M. (2013). Exploration of an adaptive training regimen that can target the secondary memory component of working memory capacity. *Memory & Cognition*. Advance online publication. doi: 10.3758/s13421-013-0295–8

Gibson, B. S., Kronenberger, W. G., Gondoli, D. M., Johnson, A. C., Morrissey, R. M., & Steeger, C. M. (2012). Component analysis of simple span vs. complex span adaptive working memory exercises: A randomized, controlled trial. *Journal of Applied Research in Memory and Cognition*, *1*, 179–184.

Huang, I. (1986). Transitory changes of primacy and recency in successive single-trial free recall. *The Journal of General Psychology*, *113*, 5–21.

Kane, M. J., Hambrick, D. Z., Tulholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General*, *133*, 189–217.

Klingberg, T. (2010). Training and plasticity of working memory. *Trends in Cognition Sciences*, *14*, 317–324.

Melby-Lervåg, M., & Hulme, C. (2012). Is working memory training effective?: A meta-analytic review. *Developmental Psychology*. Advance online publication. doi: 10.1037/a0028228

Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related?: A latent-variable analysis. *Journal of Experimental Psychology: General*, *130*, 621–640.

Murdock, B. B. (1974). *Human memory: Theory and data*. Potomac, MD: Lawrence Erlbaum.

Oberauer, K. (2005). The measurement of working memory capacity. In O. Wilhelm & R. W. Engle (Eds.), *Understanding and measuring intelligence* (pp. 393–408). Thousand Oaks, CA: Sage.

Schaffer, D., Fisher, P., Lucas, C. P., Dulcan, M. K., & Schwab-Stone, M. E. (2000). NIMH Diagnostic Interview Schedule for Children Version-IV (DISC-IV): Description, differences from previous versions and reliability of some common diagnoses. *Journal of the American Academy of Child and Adolescent Psychiatry*, *39*, 28–38.

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology General*, *125*, 4–27.

Shipstead, Z., Hicks, K. L., & Engle, R. W. (2012). Cogmed working memory training: Does the evidence support the claims? *Journal of Applied Research in Memory and Cognition*, *1*, 185–193.

Shipstead, Z., Redick, T. S., & Engle, R. W. (2012). Is working memory training effective? *Psychological Bulletin*, *138*, 628–654.

Sonuga-Barke, E. J. S. (2003). The dual pathway model of AD/HD: An elaboration of neuro-developmental characteristics. *Neuroscience and Biobehavioral Reviews*, *27*, 593–604.

Tulving, E., & Colotla, V. A. (1970). Free recall of trilingual lists. *Cognitive Psychology*, *1*, 86–98.

Unsworth, N. (2009). Variation in working memory capacity, fluid intelligence, and episodic recall: A latent variable examination of differences in the dynamics of free recall. *Memory & Cognition*, *37*, 837–849.

Unsworth, N., Brewer, G. A., & Spillers, G. J. (2011). Inter- and intra-individual variation in immediate free recall: An examination of serial position functions and recall initiation strategies. *Memory*, *19*, 67–82.

Unsworth, N., & Engle, R. W. (2007a). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, *114*, 104–132.

Unsworth, N., & Engle, R. W. (2007b). On the division of short-term and working memory: An examination of simple and complex span and their relation to higher order abilities. *Psychological Bulletin*, *133*, 1038–1066.

Unsworth, N., & Spillers, G. J. (2010). Working memory capacity: Attentional control, secondary memory, or both?: A direct test of the dual-component model. *Journal of Memory and Language*, *62*, 392–406.

Unsworth, N., Spillers, G. J., & Brewer, G. A. (2010). The contributions of primary and secondary memory to working memory capacity: An individual differences analysis of immediate free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 240–247.

Watkins, M. J. (1974). Concept and measurement of primary memory. *Psychological Bulletin*, *81*, 695–711.