Working Memory Capacity Mediates the Relationship between Removal and Fluid Intelligence

Krishneil A. Singh, Gilles E. Gignac, Christopher R. Brydges, & Ullrich K. H. Ecker

School of Psychological Science

University of Western Australia

Word Count: 15,505 (excluding title page, abstract, references, tables, and figure captions)

Address for Correspondence:

Kris Singh

School of Psychological Science

University of Western Australia

35 Stirling Highway, Perth, WA 6009, Australia

Email: kris.singh@research.uwa.edu.au

Abstract

A process of active, item-wise removal of information from working memory (WM) has been proposed as the core component process of WM updating. Consequently, we investigated the associations between removal efficiency, WM capacity, and fluid intelligence (gF) in a series of three individual-differences studies via confirmatory factor analysis. In each study, participants completed a novel WM updating task battery designed to measure removal efficiency. In Study 1, participants additionally completed a WM capacity task battery. In Study 2, participants completed a battery of well-established measures of gF in addition to the updating battery. In Study 3, participants completed the updating, WM capacity, and gF task batteries. The results suggested that removal efficiency was related to both WM capacity and gF. Furthermore, based on a mediation analysis, the relationship between removal efficiency and gF was found to be entirely indirect via removal's influence on WM capacity. The results were interpreted to suggest that removal ability may contribute to performance in reasoning tasks effectively through increasing WM capacity, presumably through reducing interference from distracting information.

Keywords: memory updating; working memory capacity; removal; fluid intelligence.

Working Memory Capacity Mediates the Relationship between Removal and Fluid Intelligence

Working memory (WM) is a capacity-limited system responsible for the temporary maintenance and manipulation of a select set of representations for ongoing cognition (Baddeley, 2000; Miyake & Shah, 1999; Oberauer, 2009). Traditionally, research has focused on the storage capacity of WM. However, some have argued that the main purpose of a WM system is to serve higher-level cognition (Oberauer, 2009). Indeed, WM capacity is one of the best known predictors of reading comprehension (Daneman & Merikle, 1996), reasoning ability (Süß, Oberauer, Wittmann, Wilhelm, & Schulze, 2002), and general intelligence (Conway, Kane, & Engle, 2003).

Serving higher-level cognition in an environment that constantly changes, however, poses a conundrum for WM: the system needs to construct representations that are both stable—a purpose of any memory system—and flexible at the same time. In other words, contents of WM need to be protected from forgetting and interference, while the system needs to be able to flexibly replace subsets of information in order to meet varying processing requirements. Recent evidence points to the idea that these dual goals are achieved by switching between maintenance and updating modes of operation (Kessler & Oberauer, 2014; also see Artuso & Palladino, 2011; Ecker, Oberauer, & Lewandowsky, 2014; Murty et al., 2011; Roth, Serences, & Courtney, 2006). The exact processing mechanisms underlying maintenance and updating, however, are still under debate.

Regarding maintenance, there are differing perspectives regarding the nature of WM's capacity limitation. Two of these perspectives include the decay theory and the interference theory. The former proposes that forgetting is exclusively a function of time, because memory traces decay passively (e.g., Baddeley, 2000; Barrouillet, Bernardin, Portrat, Vergauwe, & Camos, 2007; Barrouillet, De Paepe, & Langerock, 2012). This necessitates a process of attentional refreshing

or sub-vocal rehearsal to counteract decay and keep selected items available for ongoing processing (Baddeley, 1986; Camos, 2015; Vergauwe & Langerock, 2017). Rehearsal, however, can only keep a certain number of representations active in the face of decay. This may be considered an intuitive explanation of the capacity limitation of WM. However, it has been largely rejected, at least in the verbal WM domain, based on an accumulation of evidence that the mere passage of time does not cause forgetting (e.g., Berman, Jonides, & Lewis, 2009; Farrell et al., 2016; Lewandowsky & Oberauer, 2015; Lewandowsky, Oberauer, & Brown, 2009; Nairne, 2002; Oberauer, Farrell, Jarrold, & Lewandowsky, 2016; Oberauer & Lewandowsky, 2008, 2013; Souza & Oberauer, 2015).

By contrast, the latter view argues that interference between representations is the primary limiting factor of WM. According to this perspective, the processing of events, as and when they occur, interferes with our ability to retrieve information (Ecker & Lewandowsky, 2012; Lewandowsky, Geiger, & Oberauer, 2008; Kane & Engle, 2000; Lustig, May, & Hasher, 2001; May, Hasher, & Kane, 1999; Oberauer & Kliegl, 2006). Interference-based models thus view forgetting as a result of distracting information creating interference by being encoded into WM. There are both theoretical and empirical reasons to accept this. First, computational models of WM can account for a large body of data based on interference mechanisms without decay (e.g., Farrell & Lewandowsky, 2002; Oberauer & Lin, 2017). Second, studies show that interference influences performance on complex-span tasks[1]. For example, Lewandowsky et al. (2008) conducted a study where participants studied lists of five consonants for serial recall. The interval between

---

[1] The complex-span paradigm is the most prominent experimental paradigm used to measure WM capacity; it involves both maintenance and manipulation aspects, requiring maintenance of a list of memoranda (e.g., a set of letters or words) for immediate serial recall while undertaking a distracting processing task (e.g., mental arithmetic) concurrently (cf. Conway, Kane, Bunting, Hambrick, Wilhelm, & Engle, 2005; Wilhelm, Hildebrandt, & Oberauer, 2013).

memoranda was held constant, but the number of distractor words presented in-between list items was manipulated. Moreover, in one condition, distractors were unique (i.e., consecutive months such as "January", "February", "March"), whereas in another condition distractors were repeated (e.g., "January", "January", "January"). Results showed that a greater number of distractors caused more forgetting in the unique-distractors condition but not the repeated-distractors condition, arguably because with unique distractors, interference increased with the number of distractors. Comparatively, with repeated distractors, participants were faced with interference from only one distractor word, not multiple. In other words, forgetting was a function of the amount of distracting information processed in a given interval of time, rather than a function of the time interval's duration itself. It follows that—to avoid interference-based forgetting in the absence of time-based decay—WM requires a mechanism to remove out-dated or distracting information.

With respect to updating, previous research has identified the substitution of information as the core process involved in WM updating (Ecker, Lewandowsky, Oberauer, & Chee, 2010). The substitution process itself can be subdivided into two component processes: (1) out-dated, irrelevant information needs to be removed, and (2) newer, more relevant information needs to be encoded. Thus, the requirement of a removal mechanism follows not only from WM's capacity limitations, but also from a decomposition of the WM updating process. In other words, the implication of a WM system that has limited capacity and requires representational flexibility is that there needs to be a process that 'clears' WM of information that is or has become irrelevant.

Therefore, it has recently been proposed that an active removal process exists, serving (a) to minimize interference (and thus forgetting) in WM, and (b) to facilitate the updating of information held in WM. In the domain of WM capacity, active item removal has been incorporated successfully into a computational model of WM as a Hebbian anti-learning process,

in order to model the reduction of distractor interference in the complex-span task (Oberauer, Lewandowsky, Farrell, Jarrold, & Greaves, 2012). In the domain of WM updating, Ecker and colleagues (Ecker, Lewandowsky, & Oberauer, 2014; Ecker, Oberauer et al., 2014) have provided experimental evidence for the importance of active item removal in WM updating. They correspondingly introduced a novel WM updating task that serves to measure reliably a person's removal efficiency. A brief discussion of this removal process, and the evidence supporting it, follows.

A Measure of Removal Efficiency

A disadvantage of using traditional updating tasks to measure WM updating is that they confound updating with generic WM processes. Tasks such as the running memory task (Morris & Jones, 1990) and the *n*-back task (Kirchner, 1958) require participants to maintain information in WM, while substituting outdated items with newer ones—therefore, while they do measure WM updating, they also measure generic processes such as encoding, maintenance, and retrieval. In fact, it has been argued that traditional updating tasks measure WM capacity equally as well as complex span tasks (Schmiedek, Hildebrandt, Lovdén, Wilhelm, & Lindenberger, 2009). For example, in the running memory task, participants encode an item list of unpredictable length (e.g., K, A, S, G, F, V), before recalling the last 3 letters presented when prompted. Successful updating would imply that, over time, participants represent this list in WM as "K…KA…KAS…ASG…SGF…GFV" and that they recall "GFV" at the end of the trial. Even though one could argue that individual items of a set need to be repeatedly updated in this task, the task just measures final recall of the set and, thus, conflates updating, maintenance, and retrieval processes. In addition, it is possible that people who successfully recall the final set in

this task do not actively engage in WM updating throughout the trial, but instead base their recall on item recency alone (Bunting, Cowan, & Saults, 2006).

To measure updating more directly, Kessler and Meiran (2008) measured the time it took people to replace items held in WM. Participants encoded a set of three items, and individual items were repeatedly replaced with new ones before a final recall. At each updating step, participants pressed a key when they had completed the updating, providing a measure of updating efficiency. However, because participants in Kessler and Meiran's task did not know what item was going to be replaced, Ecker, Lewandowsky et al. (2014) argued that this measure still confounded the core updating process of removal with the generic operations of attention reorienting and encoding, as all three processes could only begin when the new item was presented.

To disentangle the removal and encoding confound, Ecker et al. (Ecker, Lewandowsky et al., 2014; Ecker, Oberauer et al., 2014) varied Kessler and Meiran's (2008) task in two ways. First, they cued the to-be-updated item. Providing a cue gave participants information on what needed to be removed, thus, facilitating the updating process. Specifically, at each updating step, the cue signalled to participants what item was about to be updated, prior to the presentation of the new to-be-encoded item. It was assumed that people could use the time between the presentation of the cue and the presentation of the new to-be-encoded item (i.e., the cue-target interval, CTI) to remove the out-dated representation from their WM.

Secondly, Ecker et al. varied the CTI, thereby manipulating the time available for item removal: It was assumed that a short CTI (i.e., 200 ms) did not permit any removal because it only allowed sufficient time to focus attention on the to-be-updated item; updating response latencies in this condition were thus assumed to include both the time required to remove the old item and the time required to encode the new item. By contrast, a long CTI (i.e., 1,500 ms) was assumed to

allow for removal of the cued item prior to the presentation of the new item; updating RTs in this condition should thus not include the time taken for removal.

As expected, updating latencies in Ecker et al. (Ecker, Lewandowsky et al., 2014; Ecker, Oberauer et al., 2014) were considerably faster in the long-CTI condition than the short-CTI condition. In addition, it was found that a speed-up of updating observed in previous research (a) when the new item matches the old to-be-replaced item (i.e., occasional item repetition; cf. Ecker et al., 2010) or (b) when the new item is similar to the old to-be-replaced item (i.e., item similarity; cf. Lendínez, Pelegrina, & Lechuga, 2011) was evident in the short-CTI condition but much diminished in the long-CTI condition. These results indicated that representational overlap associated with item repetition/similarity can only facilitate updating to the degree that the to-be-replaced item is still in WM, and thus supported the removal notion (for a more detailed discussion of the removal-efficiency measure's validity, see Ecker, Lewandowsky et al., 2014 and Ecker, Oberauer et al., 2014).

Ecker et al. (Ecker, Lewandowsky et al., 2014; Ecker, Oberauer et al., 2014) proposed two different measures of removal efficiency, based on the difference in updating RTs between the two CTI conditions. One measure was a proportional-gain score, calculated as the updating-RT difference between the two CTI conditions divided by the updating RT of the short-CTI condition (to control for general processing speed differences). The second was a regression-residual score, calculated as the individual residuals obtained from a simple linear regression model predicting the short-CTI RTs from the long-CTI RTs.

Using these measures in an individual differences study, Ecker, Lewandowsky et al. (2014) did not observe a significant association between removal efficiency and WM capacity using both the proportional-gain ($r = -.04$, $p = .67$) and the regression-residual ($r = -.07$, $p = .44$; $N = 167$)

removal scores. On the one hand, this makes sense to the degree that WM updating and general WM functions may be separable (cf. Ecker et al., 2010). However, given the proposed importance of the removal process to support general WM function by reducing distractor interference and thus allowing for more focussed processing of relevant items (Ecker, Oberauer et al., 2014; Oberauer et al., 2012), it would be expected that removal ability is related to WM capacity.

The Present Study

It may be that the failure to observe an association between removal and WM capacity in Ecker, Lewandowsky et al. (2014) was a false negative finding, potentially brought about by their methodological approach of using only a single updating task to represent individual-differences in removal efficiency. It is well-known from the individual-differences literature that it is preferable to utilize a task battery rather than a single task. This serves to reduce the impact of task-specific variance, and additionally allows for the measurement of a construct at the latent-variable level (Lewandowsky, Oberauer, Yang, & Ecker, 2010; Oberauer, 2005; Wittmann, 1988).

There is reason to believe that a person's efficiency at removing outdated items from WM should be related to the maximum number of items that is held in their WM (i.e., their WM capacity; WMC): The more efficiently a removal mechanism can discard irrelevant or out-dated information, the less interference there will be, and the more WM processing can focus on relevant information. Thus, on average, a person who is efficient at removing should effectively have greater WM capacity.

Study 1, therefore, had two aims. First, to re-investigate the link between the proposed removal process and WM capacity using three different variants of the updating task used in Ecker, Lewandowsky et al. (2014) to measure removal ability. It was hypothesised that greater removal efficiency would be associated with greater WM capacity. The second aim of Study 1 was to

replicate Ecker, Lewandowsky et al.'s non-significant finding regarding the relation between removal efficiency and WM capacity using only the letter updating task used in their study.

In addition to relating to WM capacity, a person's removal efficiency may also be related to their ability to solve novel problems using fluid reasoning (i.e., their fluid intelligence; gF). Previous research has indeed claimed that updating is the only executive function to predict gF (Chen & Li, 2007; Friedman et al., 2006); however, these studies used 'traditional' updating tasks that—as discussed earlier—measure mainly generic WM functioning (Ecker et al., 2010; Ecker, Lewandowsky et al., 2014; Ecker, Oberauer et al., 2014; Kessler & Meiran, 2008; Schmiedek et al., 2009). Consequently, the proposed link between updating and gF may simply reflect the well-established association between WM capacity and gF (Colom et al., 2015; Engle, Tuholski, Laughlin, & Conway, 1999; Gignac, 2014; Kane, Hambrick, & Conway, 2005; Little, Lewandowsky, & Craig, 2014; Oberauer, Schulze, Wilhelm, & Süß, 2005; Unsworth, Brewer, & Spillers, 2009; but see Gignac & Watkins, 2015).

Despite this methodological issue, Ecker et al. (2010) already suggested that a person's ability to substitute information in WM—in essence, their removal efficiency—may make an independent contribution toward the execution of gF-type tasks. Indeed, there are theoretical reasons to assume that removal should relate to gF. For example, good removal skills could make reasoning more efficient by allowing disengagement from irrelevant information and facilitating the discarding of unhelpful operators during problem-solving (see Shipstead, Lindsey, Marshall, & Engle, 2014). Shipstead and colleagues (Harrison, Shipstead, & Engle, 2014; Shipstead, Harrison, & Engle, 2016; Shipstead et al., 2014) have suggested that a potential mechanism contributing to performance on reasoning tasks is the ability to disengage from irrelevant

information. Thus, to the extent that disengagement involves the removal of content from active WM, a person who is efficient at removing should also effectively have a higher level of gF.

Arguably, reasoning takes place in WM, hence the same mechanism of interference prevention that we argued allows for more efficient WM functioning could also be argued to support reasoning (i.e., performance on gF tasks): Efficiently removing unsuccessful steps when solving novel problems may help free attentional resources for the generation of additional relevant steps and avoid perseveration. Hence, the aim of Study 2 was to use the same three updating tasks to investigate the potential relation between removal and gF. It was hypothesized that greater removal efficiency would be associated with higher levels of gF.

Finally, the aim of Study 3 was to simultaneously investigate the nature of the relationship between removal, WM capacity, and gF in a mediation analysis. Two hypotheses are presented: (1) removal skills may contribute to greater WM capacity by reducing interference from distracting information, and good WM capacity in turn may facilitate reasoning ability. This theoretical standpoint suggests that WM capacity fully mediates the relationship between removal efficiency and gF; (2) removal skills may contribute directly to both WM capacity and gF (in addition to a potential indirect effect of removal on gF via WM capacity). This would imply only partial mediation, such that there would still be a significant direct effect of removal on gF in a mediation model including WM capacity as the mediator.

In summary, the present article presents three studies. Study 1 re-investigates the relationship between removal efficiency and WM capacity. Study 2 investigates the potential relationship between removal efficiency and gF. Study 3 investigates the nature of the relations between removal efficiency, WM capacity, and gF via a mediation analysis.

Study 1

Method

Participants. A total of 181 undergraduate students from the University of Western Australia participated for partial course credit. Based on various outlier criteria (see below), 19 participants were excluded from the analyses, yielding a final sample of $N = 162$ participants (116 females, 46 males; age: $M = 22.25$ years, $SD = 6.57$; age range: 18 to 59 years). Of these, 148 were native English speakers, 13 regarded themselves as being highly proficient in English despite it being their second language, and only one participant regarded their English as "adequate".

Materials.

Updating tasks. The present study used three updating tasks. All tasks used the paradigm introduced by Ecker et al. (Ecker, Lewandowsky et al., 2014), but tasks differed with respect to the stimuli employed: one task (like Ecker et al.'s original task) used letters, one used digits, and one used words. Specifically, in all three updating tasks, participants were initially presented with a 1-second fixation cross in the centre of the screen. Following this, each trial began with the simultaneous 2-second presentation of three items, presented individually in a single row of three black, rectangular frames. This was then followed by a series of updating steps. On each updating step, one of the items was cued for updating, with either a short (200 ms) or a long (1500 ms) CTI. During the CTI, one of the frames turned bold and red, signalling that the item in that frame was about to be replaced by a new item. Then the new item was presented in the cued frame, and participants pressed a key (space bar) to indicate that they had encoded the new item and had successfully updated their WM set; the maximum response time was 5 s.[2]

---

[2] For a pictorial representation of a trial sequence and an updating step, see Figures 2 and 3 of Ecker, Lewandowsky et al. (2014; pp. 81-82).

The number of updating steps per trial varied between 1 and 21, with a constant 10 % stopping probability after each updating step. This meant that participants were presented with an unpredictable number of updating steps per trial, resulting in an equal incentive to undertake each updating step independent of the elapsed duration of the trial.

The updating phase of each trial was followed by a cued recall of all three items. Recall was prompted by a blue question mark appearing in each frame in random order (in the word task, three blue underscores were instead presented in each frame, cuing participants to type the first three letters of the remembered word, which were unique). Recall prompts were presented until a response was given or the maximum response time of 5 s per frame was reached.

Each participant completed 2 practice trials before completing a total of 12 trials, with an average of 9 updates per trial. There were approximately 108 updating steps in total, or 54 per CTI condition. In the letter task, all items were consonants; the minimum alphabetic distance between to-be-encoded letters was 2, to avoid sequences like X-Y-Z. The digit task used single-digit numbers (i.e., 1 to 9) as stimuli; the minimum numerical distance between to-be-encoded items was 2. The word task used neutral, monosyllabic words, taken from the online version of the MRC Psycholinguistic Database (Coltheart, 1981; Wilson, 1988; http://websites.psychology.uwa.edu.au/school/MRCDatabase/mrc2.html). Selected words were 2 to 5 letters in length, with a Kucera-Francis frequency $f \geq 50$ (least and most frequent words, respectively, were *chain*, $f = 50$, and *are*, $f = 4393$). Each task took approximately 12 minutes.

WM capacity tasks. Participants' WM capacity was measured with a task battery comprising four tasks (Lewandowsky et al., 2010): two complex span tasks—an operation-span task (OS) and a sentence-span task (SS)—as well as a traditional memory updating task (MU), and

a spatial short-term memory task (SSTM). Each task used the default settings of Lewandowsky et al. (2010), and tasks are therefore discussed only briefly here.

The two complex span tasks (OS and SS) required participants to memorize a set of items for serial recall, with set size varying between 4 and 8 items, while concurrently undertaking a secondary processing task of mental arithmetic (OS) or sentence judgement (SS). Specifically, encoding of memoranda alternated with judging the correctness of a mathematical equation in the OS task (e.g., $2 + 3 = 5$), or the veracity of a sentence in the SS task (e.g., *"The moon is made of cheese"*).

The MU task was a standard updating task that—as discussed in the Introduction—measures mainly generic WM functions. Participants encoded a set of 3 to 5 digits presented in individual frames. After encoding, cues for arithmetic operations (e.g., "+2", "-4") were randomly presented in the frames; the cues ranged from +7 to -7. Participants had to apply the arithmetic operation to the digit that was being remembered for the particular frame and remember the result. After between 2 and 6 updating steps, there was a cued recall of all items, which provided the dependent measure of interest.

The SSTM task required participants to memorize a spatial pattern comprising between 2 and 6 dots, which were presented sequentially in a 10×10 grid. After presentation of the dots, the grid was briefly masked and participants reproduced the dot pattern by clicking on grid positions with a mouse. Only the accuracy of the pattern recall (i.e., the positions of the dots relative to each other) was relevant, not memory for the specific dot locations within the grid.

Procedure. The experiment was run with the aid of MatLab and the Psychophysics toolbox (version 2.54; Brainard, 1997). Participants were individually tested in booths, sitting approximately 70cm from a 17 inch thin-film transistor monitor. Participants first completed the

three variants of the modified updating task in a fixed order (i.e., letter task, digit task, word task) to obtain a measure of their removal ability. Next, participants completed the WMC battery in a fixed task order (i.e., OS, SS, MU, SSTM). The entire experiment took about 70 minutes.

Results

Data screening. Due to a coding error, 32 participants completed a shortened version of the word updating task with a reduced maximum of 9 instead of 21 updating steps per trial. However, in an attempt to retain as many participants as possible, only participants with fewer than 40 updating steps in total were discarded ($n = 7$). Participants were also excluded from further analyses if their recall performance in any of the updating tasks ($n = 3$), any of their removal-efficiency scores ($n = 2$), or their performance on any of the WM tasks (including the complex-span secondary tasks; $n = 7$) fell 3 *SD*s above or below the respective grand mean of all participants.

Updating tasks.

Recall accuracy. As expected, recall accuracy was high for each updating task (letters: $M = .95$, $SD = .05$; digits: $M = .98$, $SD = .03$; words: $M = .95$, $SD = .05$).

Updating RTs. Before aggregation, individual responses < 300 ms were discarded, as were RTs more than 3 *SD*s away from the individual participant's mean RT in a given task. Mean updating RTs for all tasks are provided in Figure 1. A 3×2 repeated measures ANOVA on updating RTs, with the factors task (letters, digits, words) and CTI (short, long) yielded a significant main effect of task, $F(2,322) = 112.82$, $MSE = 0.05$, $p < 0.01$, $\eta_p^2 = 0.41$, a significant main effect of CTI, $F(1,161) = 555.42$, $MSE = 0.02$, $p < 0.01$, $\eta_p^2 = 0.78$, as well as a slight but significant interaction, $F(2,322) = 17.69$, $MSE = 0.01$, $p < 0.01$, $\eta_p^2 = 0.10$. The main result of

interest here was that, as expected, the effect of the CTI manipulation found by Ecker, Lewandowsky et al. (2014) was replicated across all three tasks.

Removal-efficiency scores. Three removal-efficiency scores per participant were calculated from the updating response times in the three updating tasks. Removal-efficiency scores were calculated as regression residuals following Ecker, Lewandowsky et al. (2014). Spearman-Brown corrected reliability indices, bootstrapped with $N = 1,000$ runs, were $\rho = .47$ for the letter task, $\rho = .57$ for the digit task, and $\rho = .60$ for the word task. These estimates implied that approximately 50 % to 60 % of the scale variances were reliability variance. We note that these estimates are lower than the reliabilities reported in Ecker, Lewandowsky et al. (2014). However, the present study used latent variables to represent the dimensions of interest. Consequently, as latent variables are devoid of measurement error (Fan, 2003), the reported effects were not considered affected by the relatively low reliabilities associated with the indicator observed scores (see also Friedman & Miyake, 2017; Little, Lindenberger, & Nesselroade, 1999). Table 1 displays descriptive statistics for the removal-efficiency scores; on this score, negative values indicate greater removal efficiency and positive values indicating lower removal efficiency. As expected, the means for the regression-residual removal scores were close to zero. Table 2 presents the inter-correlations of the removal-efficiency scores. All tasks correlated moderately with each other, indicating a common removal factor, but did not correlate perfectly, indicating task-specific variance.

WM capacity battery. Spearman-Brown corrected split-half reliability estimates for the OS, SS, MU, and SSTM tasks were $\rho = .82, .86, .85$, and $.75$, respectively. Table 1 displays descriptive statistics for the four WMC tasks. All values were in the approximate ranges reported in the validation studies of Lewandowsky et al. (2010). Task inter-correlations are presented in

Table 2. Again, all correlations were in the approximate ranges reported in Lewandowsky et al.

(2010), with the unexpected exception of the correlations between the SSTM task and the OS ($r =$

.13, $p = .11$) and SS ($r = .07$, $p = .40$) tasks, which were lower than in previous research.

The relation between removal efficiency and WM capacity. In order to assess the link

between removal efficiency and WM capacity, a series of confirmatory factor analytic models was

tested in AMOS version 22 (Arbuckle, 2013). Model solutions were estimated via maximum

likelihood. Model-fit was evaluated with the following indices: the comparative fit index (CFI;

Bentler, 1990), the Tucker-Lewis index (TLI; Tucker & Lewis, 1973), the root mean square error

of approximation (RMSEA; Browne & Cudeck, 1993; Steiger & Lind, 1980), and the standardized

root mean square residual (SRMR; Bentler, 1995; Jöreskog & Sörbom, 1986). In order to control

for the effect of age on cognitive ability, a simple linear regression model that regressed each

cognitive ability onto age was conducted. The standardized residuals from these regression

analyses were used as observed variables for the purposes of defining the removal and WM

capacity latent variables.

First, two basic measurement models (one for removal, one for WMC) were tested. The

WMC measurement model was initially defined by four observed variables (one for each task—

OS, SS, MU, SSTM). However, the modelling is reported based on the exclusion of the SSTM

indicator, as it was observed to have a trivial standardized loading on the WMC latent variable

(i.e., .19, $p = .03$).[3] Thus, both measurement models were defined by three observed variables

(removal: letters, digits, words; WMC: OS, SS, MU) linked to a single latent factor. The removal

latent variable had significant moderate positive standardized loadings from all three removal

---

[3] The outcome of the modelling in Study 1 was identical when the SSTM indicator was included. However, the exclusion of the SSTM indicator in Study 1 helped achieve greater comparability with Study 3, where the SSTM task was omitted due to testing time constraints.

tasks. The standardized loadings were .51, .66, and .66 for letters, digits, and words, respectively; all ps < .001. This indicated the three tasks measured a common removal factor. Similarly, the three WMC tasks all had significant moderate to substantial positive standardized loadings on the WMC latent variable. The loadings were .89, .78, and .64 for OS, SS, and MU, respectively; all ps < .001. This indicated that these three tasks measured a common WMC factor.

Next, a single-factor (general) model was tested. This model consisted of one latent common abilities factor (*CAF*) defined by the six observed variables and their associated error terms. Model identification was achieved by assigning a variance of 1 to the *CAF* latent variable. The general factor model was found to be associated with poor model fit, $\chi^2(9) = 70.64$, $p < .001$, *CFI* = .742, *TLI* = .570, *RMSEA* = .206, *SRMR* = .138. Thus, the inter-associations between the removal and WMC measures should not be considered representative of a single cognitive process.

Next, a correlated two-factor model was tested, consisting of two latent variables representing removal efficiency and WM capacity (*REM* and *WMC)*, each receiving loadings only from the three associated observed variables, respectively (see Figure 2). For model identification purposes, both latent variables were assigned a variance of 1. This model was associated with a good fit, $\chi^2(8) = 14.04$, $p = .081$, *CFI* = .975, *TLI* = .953, *RMSEA* = .068, *SRMR* = .045. As can be seen in Figure 2, all six observed variables were associated with significant ($p < .001$) and moderate to substantial standardized loadings on their respective factors. Importantly, the inter-latent variable correlation between *REM* and *WMC* was significant, $r = -.23$ (95% CI: -.44 / -.02; *SE* = .101; $p = .020$), indicating that greater removal efficiency was associated with greater WM capacity. Additionally, the chi-squared difference test suggested that the correlated two-factor model was statistically significantly better fitting than the single-factor model, $\Delta\chi^2(1) = 56.60$, $p < .001$.

Unique true score variance: Removal and WM capacity. Comparing model fit estimates between single-factor and correlated two-factor models is a common type of analysis approach; however, it is limited in two ways. First, endorsement of a correlated two-factor model does not necessarily support the plausibility of two specific dimensions (Gignac, 2014; Gignac & Kretzschmar, 2017). Second, correlated two-factor models cannot estimate the unique variance associated with each latent variable (Reise, 2012). Thus, a higher-order model was tested to supplement the correlated-factor analysis.

As shown in Figure 3, the higher-order model consisted of a common abilities factor (*CAF*) defined by two first-order factors (*REM* and *WMC*; note that a higher-order model with two first-order factors was used strictly for the purpose of decomposing specific factor variance from common factor variance, and thus no theoretical conclusions should be drawn regarding the second-order factor). Both first-order factors were associated with a residual (*REM_r* and *WMC_r*), which represents the true score variance unaccounted for by the *CAF* factor. Both first-order factors received loadings from their three associated observed variables. The two second-order loadings and the first-order factor residual coefficients were specified at 1 for identification purposes.

It should be noted that the statistical significance of the residual estimates associated with the first-order factors are affected by sample size. Relative to precedent, the sample size in the present study was reasonable ($N \approx 200$), but not large by confirmatory factor analytic standards and therefore statistical tests had limited power. Thus, the theoretical significance of the first-order residual estimates was additionally assessed using an effect-size criterion based on the coefficient

omega-hierarchical-subscale (henceforth omega-HS or ωhs)[4]. Following Gignac and Kretzschmar (2017), the following omega-HS effect size guidelines were used: relatively small, $< .20$; typical, .20 to .30; and relatively large, $> .30$. Omega-HS calculations were performed with the software tool Omega (Watkins, 2013; see also Brunner, Nagy, & Wilhelm, 2012; Reise, Bonifay, & Haviland, 2013; Zinbarg, Yovel, Revelle, & McDonald, 2006), based on a higher-order model with reflected *REM* observed variables (i.e., each removal score was subtracted from 1 such that higher scores reflected better abilities across all variables in the model; these ranged between -.05 to .05 approximately). Prior to calculating omega-HS, a Schmid-Leiman decomposition was performed (see Gignac, 2007; Schmid & Leiman, 1957). The decomposition allowed for the estimation of shared variance between (1) the *CAF* factor and the observed first-order factors, and (2) the first-order factors and their respective observed variables.

Based on the higher-order model, only the *WMC* first-order factor was associated with a statistically significant residual variance term, $S^2_{WMC} = .68$, $p < .001$; the residual variance term of the *REM* first-order factor just missed the conventional significance criterion, $S^2_{REM} = .16$, $p = .080$. As this may be a power issue, the ωhs estimates were evaluated. The obtained ωhs values were greater than .20 for both *WMC* and *REM*, $\omega hs_{WMC} = .71$; $\omega hs_{REM} = .38$, which suggests that the unique true score variance associated with both WM capacity and removal ability is relatively large (Gignac & Kretzschmar, 2017).

---

[4] Omega-HS is a standardized index ranging from 0 to 1 and can be considered an indicator of specific (unique) latent variable strength. Values close to 0 reflect a weak specific latent variable, and values closer to 1 reflect a strong specific latent variable. Omega-HS can therefore be viewed as a representation of effect size that is unaffected by sample size. Comparatively, tests of statistical significance of first-order residual terms are affected by statistical power, and it should be noted that statistical significance and omega-HS may not necessarily agree with each other (see Gignac & Kretzschmar, 2017).

Replication of Ecker, Lewandowsky et al. (2014). The correlated two-factor model suggested a significant relationship between removal efficiency and WM capacity, based on a battery of updating tasks. In order to further validate the use of a removal task battery to investigate removal efficiency at the latent variable level, we attempted to replicate Ecker, Lewandowsky et al.'s (2014) finding of a non-significant relation between removal efficiency and WM capacity by using only the letter updating task. Following Ecker, Lewandowsky et al., calculations were based on random thirds of the data in order to obtain a stable latent *REM* factor with three observed variables. In addition to the regression-residual score (*REMRR*), we also calculated a proportional gain score (*REMPG*), following Ecker, Lewandowsky et al. The Spearman-Brown corrected reliability index of this latter score, bootstrapped with $N = 1,000$ runs, was $\rho = .45$. The mean *REMPG* score was $M = 0.16$ ($SE = 0.008$; Range $= -0.07 - 0.38$). The mean *REMRR* score was $M = -0.00$ s ($SE = 0.008$; Range $= -0.21$ to 0.26 s).

Correlational Analyses. Similar to Ecker, Lewandowsky et al. (2014), the *REMPG* scores did not correlate significantly with any of the WM capacity tasks (the correlations with the MU, OS, SS, and SSTM tasks were -.12, -.03, -.02, and -.01, all p's > .13). Apart from a significant correlation with the MU task ($r = -.22$, $p = .01$), the *REMRR* scores also did not correlate significantly with the other three WM capacity tasks (the correlations with the OS, SS, and SSTM tasks were -.13, -.09, and -.06, all p's > .10).

Latent variable analysis. To replicate the non-significant inter-latent variable relationship between removal efficiency and WM capacity using just the letter updating task, a latent variable analysis was performed. First, two measurement models were tested for the two removal efficiency scores. Both the *REMPG* and the *REMRR* removal efficiency latent variables were defined by three observed variables derived from random thirds of data obtained from only the letter removal

task. The *REMPG* latent variable had significant moderate standardized loadings from its three observed variables. The loadings were .45, .48, and .49 (all *p*'s < .05). Similarly, the *REMRR* latent variable had significant moderate standardized loadings from its three observed variables. The loadings were .59, .40, and .31 (all *p*'s < .05). These results indicate that both types of removal scores measured latent removal factors. As in Ecker, Lewandowsky et al. (2014), the WM capacity measurement model for this analysis was defined by four observed variables (one for each task—OS, SS, MU, SSTM) linked to a single latent factor. The loadings for this model were reported earlier in the Results section.

Next, two correlated two-factor models were tested in order to estimate the relationship between removal efficiency—as measured by one task—and WM capacity. The first correlated two-factor model estimated the relationship between removal and WM capacity using the *REMPG* scores. This model was associated with a good fit, $\chi^2(13) = 16.74$, *p* = .211, *CFI* = .981, *TLI* = .969, *RMSEA* = .042, *SRMR* = .048. However, the inter-latent variable correlation between *REMPG* and *WMC* was non-significant (*r* = -.08, 95% CI: -.32 / .16, *p* = .54).

The second correlated two-factor model estimated the relationship between removal efficiency and WM capacity using the *REMRR* scores. This model was also associated with a good fit, $\chi^2(13) = 17.93$, *p* = .160, *CFI* = .974, *TLI* = .959, *RMSEA* = .049, *SRMR* = .052. Again, the inter-latent variable correlation between *REMRR* and *WMC* was non-significant (*r* = -.23, 95% CI: -.54 / .00, *p* = .07). Thus, the present study was able to replicate Ecker, Lewandowsky et al.'s (2014) non-significant relationship between removal efficiency and WM capacity using only the letter updating task.

Discussion

The main aim of Study 1 was to investigate the association between removal efficiency and WM capacity. Previous research by Ecker, Lewandowsky et al. (2014) failed to observe a significant association between removal efficiency and WM capacity. While this finding makes sense to the degree that updating and WM functions represent separate cognitive processes, one possible reason Ecker, Lewandowsky et al. (2014) did not find any effect may be due to the fact that they utilised only one removal task to represent individual differences in removal efficiency. Consequently, our goal in the present study was to measure removal efficiency more reliably and validly, by the administration of a battery of removal tasks. Arguably, our goal was achieved based on three outcomes: (1) the three removal tasks showed a pattern of inter-task correlations that suggested substantial shared variance; (2) the three removal tasks also showed substantial task-specific variance; and (3) the removal measurement-model fit suggested that removal efficiency was successfully measured at the latent-variable level.

Based on the notion that a process of active item-wise removal is a core updating process but also serves to remove distractor information in WM tasks, it was hypothesized that removal efficiency would be significantly related to WM capacity. Indeed, it was found that removal efficiency, as measured by our removal task battery, covaried significantly with WM capacity. This finding suggests that a removal process may be important for WM updating, specifically, and WM functioning more generally (see Ecker et al., 2010). Based on previous research and theorizing, the most likely candidate mechanism to explain this relation is that the removal process serves to reduce interference from distracting information in WM, thus allowing for more focussed processing of goal-relevant information (Ecker, Lewandowsky et al., 2014; Oberauer et al., 2012). In other words, individuals with good removal skills will tend to have greater WM capacity, presumably because they can more efficiently minimize interference from distracting information

in WM tasks. Thus, the findings from Study 1 are consistent with computational models of WM that have incorporated an interference-reduction mechanism necessary for a WM system not reliant upon time-based decay (Oberauer et al., 2012). For example, Oberauer et al.'s SOB-CS model of WM proposes that active removal is necessary to remove no-longer-relevant items whenever there is free time in-between distractor operations using Hebbian anti-learning, which undoes the association between a distractor and its position marker (also see Kessler & Meiran, 2008). Replication of the covariation between removal efficiency and WM capacity in Study 3 will be important in order to support the theoretical conclusions drawn here.

A second aim of Study 1 was to replicate Ecker, Lewandowsky et al.'s (2014) non-significant association between removal and WM capacity at the latent-variable level using just one updating task. This replication was achieved: Removal efficiency—as measured by a single letter updating task—failed to covary significantly with WM capacity at the latent-variable level. This finding further supports the use of a task battery to measure removal efficiency at the latent-variable level.

Study 2

Having established some support for the notion that an active removal process is related to WM capacity, we now turn our focus to the relationship between updating and gF. As discussed in the Introduction, updating has been claimed to be the only executive function to predict gF (Chen & Li, 2007; Friedman et al., 2006). However, the 'traditional' updating tasks used in these studies measure updating in addition to generic WM functions such as encoding, maintenance, and retrieval (Ecker et al., 2010; Ecker, Lewandowsky et al., 2014; Ecker, Oberauer et al., 2014; Kessler & Meiran, 2008; Schmiedek et al., 2009). In addition, traditional updating tasks have been shown to measure WM capacity equally as well as complex span tasks (Schmiedek et al., 2009).

It is argued, therefore, that the proposed link between updating and gF may simply reflect the well-established association between WM capacity and gF (Colom et al., 2015; Engle et al., 1999; Gignac, 2014; Kane et al., 2005; Little et al., 2014; Oberauer et al., 2005; Unsworth et al., 2009; but see Gignac & Watkins, 2015).

Despite this, Ecker et al. (2010) have suggested that a person's ability to substitute information (i.e., their removal efficiency) may make an independent contribution to performance on reasoning tasks. Study 2, therefore, was an individual differences study which investigated whether removal efficiency—as measured by the same removal task battery that was used in Study 1—covaried significantly with well-established measures of gF.

Method

Participants. A total of 185 undergraduate students from the University of Western Australia participated for partial course credit. Based on various outlier criteria (see below), 22 participants were excluded from the analyses, yielding a final sample of $N = 163$ participants (121 females, 42 males; age: $M = 21.19$ years, $SD = 5.20$; age range: 18 to 53 years). Of these, 147 were native English speakers, and 16 regarded themselves as being highly proficient in English despite it being their second language.

Materials.

Updating tasks. Study 2 used the same three modified updating tasks that were used in Study 1, in addition to five measures of gF.

Fluid intelligence tasks.

Cattell's Culture Fair Intelligence Test (CFIT; Cattell & Cattell, 1960). The CFIT comprises four paper and pencil subtests performed under timed conditions. Participants have 12.5 minutes to complete all of the subtests, allowing 2.5-4 minutes for each one. Specifically, subtest

1 ("Progressive series completion"; CFIT1) contains 13 items; for each item, participants have to choose a shape to complete a series of progressive shapes (e.g., a series of "growing" bars). Subtest 2 ("Classification"; CFIT2) contains 14 items; for each item, participants are presented with five stimuli and have to pick the two stimuli that are different from the others (e.g., horizontal instead of vertical bars). Subtest 3 ("Matrices" ; CFIT3) contains 13 items; for each item, participants have to select a stimulus out of five options that completes a 2×2 or 3×3 grid of stimuli, satisfying pattern constraints set by the existing stimuli. Subtest 4 ("Conditions"; CFIT4) contains 10 items; for each item, participants are presented with a compound stimulus (e.g., comprising a circle and a square) containing a dot in a particular location (e.g., outside the square but inside the circle), and have to select one out of five similar stimuli that would allow them to place a dot in the same position. Rather than combining the four subtest scores into a single CFIT score, the sum of all correct answers for each subtest provided four CFIT scores and thus four observed gF variables for modelling.

Raven's Advanced Progressive Matrices—Short form (RAPM-S; Arthur & Day, 1994; Raven, 1962). A short paper-and-pencil form of the RAPM-S was used (Arthur & Day, 1994), which consists of 12 items selected from the 36 items in Set II of the full test version[5]. Similar to the "Matrices" subtest of the CFIT, participants are presented with a series of 3×3 grids, each containing a pattern of stimuli with a missing element, and participants are required to select one out of eight options to complete the pattern. While there was no time limit for this test, participants generally took around 10 to 15 minutes to complete the 12 items.

---

[5] The 12 Raven's items used were items 1, 4, 8, 11, 15, 18, 21, 23, 25, 30, 31, and 35 (see Arthur & Day, 1994).

Procedure. Participants first completed the three variants of the modified updating task (in the same fixed order as in Study 1) before they completed the gF tasks in a fixed order (i.e., the four CFIT subtests followed by the RAPM-S). The entire experiment took approximately 70 minutes.

Results

Data screening. Performance-based outliers were excluded. Specifically, participants were excluded from further analyses if their recall performance in any of the updating tasks ($n = 8$), any of their removal-efficiency scores ($n = 6$), or their performance on any of the gF tasks ($n = 8$) fell 3 *SD*s above or below the respective grand mean of all participants.

Updating tasks.

Recall accuracy. As expected, recall accuracy was high for each updating task (letters: $M = .96$, $SD = .04$; digits: $M = .97$, $SD = .03$; words: $M = .94$, $SD = .06$).

Updating RTs. Before aggregation, individual responses < 300 ms were discarded, as were RTs more than 3 *SD*s away from the individual participant's mean RT in a given task. Mean updating RTs for all tasks are provided in Figure 4. A 3×2 repeated measures ANOVA on updating RTs, with the factors task (letters, digits, words) and CTI (short, long) yielded a significant main effect of task, $F(2,324) = 163.56$, $MSE = 0.04$, $p < 0.01$, $\eta_p^2 = 0.50$, a significant main effect of CTI, $F(1,162) = 586.50$, $MSE = 0.02$, $p < 0.01$, $\eta_p^2 = 0.78$, as well as a slight but significant interaction, $F(2,324) = 10.94$, $MSE = 0.004$, $p < 0.01$, $\eta_p^2 = 0.06$. As in the first experiment, the effect of the CTI manipulation found by Ecker, Lewandowsky et al. (2014) was replicated across all three tasks.

Removal-efficiency scores. As in Study 1, three removal-efficiency scores per participant were calculated from the updating response times in the three updating tasks. Spearman-Brown

corrected reliability indices, bootstrapped with $N = 1,000$ runs, were $\rho = .43$ for the letter task, $\rho = .48$ for the digit task, and $\rho = .55$ for the word task. It is noted that these estimates are lower than the reliabilities reported by Ecker, Lewandowsky et al. (2014). However, similar to Study 1, the present study used latent variables to represent dimensions of interest. The reported effects, therefore, were not considered affected by the relatively low reliabilities associated with the indicator observed scores (see also Friedman & Miyake, 2017; Little et al., 1999).Table 3 displays descriptive statistics for the removal-efficiency scores. Table 4 shows the inter-correlations of the removal scores. Similar to Study 1, all tasks correlated moderately with each other indicating a common removal factor, but did not correlate perfectly indicating task specific variance.

Fluid intelligence tasks. Cronbach's alpha reliability coefficients for the gF tasks were as follows: $\alpha_{CFIT1} = .26$, $\alpha_{CFIT2} = .54$, $\alpha_{CFIT3} = .35$, $\alpha_{CFIT4} = .28$, and $\alpha_{RAPM-S} = .65$. As detailed below, we dealt with the low CFIT subscale reliabilities by combining the subtests together to form two CFIT composite scores (rather than four). Table 3 displays descriptive statistics for the five gF tasks. Task inter-correlations are presented in Table 4. As expected, all gF tasks correlated moderately with each other, except the forth CFIT subtest, which correlated with no other gF task.

The relation between removal efficiency and fluid intelligence. A series of confirmatory factor analytic models was tested to assess the link between removal efficiency and gF. Similar to Study 1, a simple linear regression model that regressed each cognitive ability onto age was performed, in order to control for the effect of age. The standardized residuals from these regression analyses were used as observed variables for the purposes of defining the removal and gF latent variables.

First, two basic measurement models (one for removal, one for gF) were tested. The gF measurement model was defined by three observed variables. Specifically, the four CFIT subtests

were combined to create two of the three observed variables such that the average of the residuals from the CFIT1 and CFIT2 subtests was used as one observed variable, and the average of the residuals from the CFIT3 and CFIT4 subtests was used as another observed variable. The age-corrected residuals from the RAPM-S formed the third observed variable for the gF measurement model. This was done in order to (1) reduce the impact of processing speed from the four CFIT tests on the gF latent variable, (2) improve the reliability of the model indicators, and (3) provide a balanced modelling approach where both measurement models were defined by three observed variables linked to a single latent factor (removal: letters, digits, words; gF: CFIT1&2, CFIT3&4, RAPM-S). The inter-correlations between the gF indicators that were used in the modelling and the gF and removal tasks are provided in Table 4. The removal latent variable had significant moderate positive standardized loadings from all three removal tasks. The weights were .55, .71, and .70 for letters, digits, and words, respectively; all $ps < .001$. This indicated the three tasks measured a common removal factor. Similarly, the three gF indicators all had significant moderate to substantial positive standardized loadings on the gF latent variable. The weights were .68, .45, and .67 for CFIT1&2, CFIT3&4, and RAPM-S, respectively; all $ps < .001$. This indicated that the five tasks measured a common gF factor.

Next, a single-factor (general) model was tested, consisting of one latent common abilities factor (*CAF*) and the six observed variables and their associated error terms. Model identification was achieved by assigning a variance of 1 to *CAF*. The single-factor model was associated with a poor level of model fit, $\chi^2(9) = 65.37$, $p < .001$, *CFI* = .612, *TLI* = .353, *RMSEA* = .197, *SRMR* = .136. Thus, the inter-associations between the removal and gF measures should not be considered representative of a single cognitive process.

Next, a correlated two-factor model was tested, consisting of two latent variables representing removal efficiency and fluid intelligence (*REM* and *gF*), each receiving loadings from their respective observed variables (see Figure 5). For model identification purposes, both latent variables were assigned a variance of 1. This model was associated with a good fit, $\chi^2(8) = 12.99$, $p = .112$, *CFI* = .966, *TLI* = .935, *RMSEA* = .062, *SRMR* = .064. As can be seen in Figure 5, all observed variables were associated with significant ($p < .001$) and moderate to substantial standardized loadings on their respective factors. The inter-latent variable correlation between *REM* and *gF* was significant, $r = -.30$ (95% CI: -.55 / .01; *SE* = .103, $p = .003$), indicating that greater removal efficiency was associated with greater fluid intelligence. Additionally, the chi-squared difference test suggested that the correlated two-factor model was statistically significantly better fitting than the single-factor model, $\Delta\chi^2(1) = 52.38$, $p < .001$.

Unique true score variance: Removal and fluid intelligence. Similar to Study 1, model fit evaluation was supplemented with higher-order model analysis. The higher-order model consisted of a common abilities factor (*CAF*) defined by two first-order factors (*REM* and *gF*). Both first-order factors were associated with a residual (*REM_r* and *gF_r*). Both first-order factors received loadings from their respective observed variables (see Figure 6). Regression weights associated with the *CAF* factor and first-order residuals were fixed at 1. Based on the higher-order model, only the *REM* first-order factor was associated with a statistically significant residual variance term, $S^2_{REM} = .22$, $p = .012$; the residual variance term of the *gF* first-order factor just missed the conventional significance criterion, $S^2_{gF} = .10$, $p = .061$. Again, we calculated and assessed ω*hs* estimates following the procedure used in Study 1. The obtained ω*hs* values for both

gF and *REM* were greater than .20, $\omega hs_{gF} = .38$; $\omega hs_{REM} = .53$, which suggests that the unique true score variance associated with gF and removal was relatively large (Gignac & Kretzschmar, 2017).

Discussion

The aim of Study 2 was to investigate the potential association between removal efficiency and gF. Compared to Study 1, the three removal tasks showed a similar pattern of inter-task correlations, suggesting substantial shared variance in addition to substantial task-specific variance. Moreover, the measurement model suggested that we were again able to successfully measure removal efficiency at the latent-variable level.

It was hypothesized, in line with suggestions by Ecker et al. (2010), that a person's ability to remove information from WM may be a predictor of gF. The findings of Study 2 seem to support this hypothesis. While this appears to be consistent with previous findings suggesting a link between WM updating and gF, our finding is novel in that the present study isolated the core process of updating, viz. item-specific removal, rather than relying on a compound measure of WM updating and other generic WM functions (Chen & Li, 2007; Friedman et al., 2006).

The results from Study 2 thus suggest that a removal process may contribute to performance in reasoning tasks. Specifically, an individual's ability to remove irrelevant information from WM may assist them when solving novel problems that require (a) identification of a pattern, rule, or set of rules, (b) focus on only some aspects of stimuli and not others, or (c) running through a number of operational steps towards a solution. Individuals who are more efficient at removing information that is not conducive to a solution may be better at solving these problems, as they will be able to disengage more efficiently from rules, patterns, task aspects, or operational steps that are not relevant to the solution, allowing them to focus on information that is task-relevant (see Harrison et al., 2014; Shipstead et al., 2014).

Study 3

Studies 1 and 2 provided two key pieces of empirical evidence. Respectively, the findings suggested that: (1) greater removal efficiency is associated with greater WM capacity; and (2) greater removal efficiency is associated with greater fluid reasoning. However, the pattern of inter-relationships between these three constructs remains unclear. In particular, it remains to be determined whether the relationship between removal and fluid intelligence is mediated, partially or fully, by WM capacity. Engle et al. (1999) proposed that WM capacity and gF reflect the ability to keep representations in a highly active state in the face of interference. Removal has been successfully modelled in computational models of WM as an interference reduction mechanism allowing for more focussed processing of goal-relevant items (Oberauer et al., 2012). Additionally, Fukuda, Vogel, Mayr, and Awh (2010) found that the relationship between WM capacity and gF was mediated by the number of representations that can be simultaneously maintained in WM. Thus, based on previous empirical research and theory (e.g., Carpenter, Just, & Shell, 1990; Engle et al., 1999; Fukuda et al., 2010), it may be hypothesised that good removal skills contribute to greater WM capacity by reducing interference from distracting information, and good WM capacity in turn may facilitate reasoning ability. This theoretical standpoint suggests that WM capacity should fully mediate the relationship between removal efficiency and gF reported in Study 2. The above hypothesis can be tested within the context of a mediation analysis paradigm (e.g., Chen & Li, 2007; MacKinnon, Fairchild, & Fritz, 2007). Thus, the aim of Study 3 was primarily to investigate the nature of the relationship between removal ability, WM capacity, and gF, focusing in particular on the question of whether the relationship between removal efficiency and gF is mediated by WM capacity, or whether removal efficiency indeed relates directly to gF.

Method

Participants. A total of 243 undergraduate students from the University of Western Australia participated for partial course credit. Based on various outlier criteria (see below), 39 participants were excluded from the analyses, yielding a final sample of $N = 204$ participants (129 females, 75 males; age: $M = 20.77$ years, $SD = 5.99$; age range: 17 to 57 years). Of these, 169 were native English speakers, 33 regarded themselves as being highly proficient in English despite it being their second language, and only 2 participants regarded their English as "adequate."

Materials.

Updating tasks. Study 3 used the same three modified updating tasks that were used in Studies 1 and 2.

WM capacity tasks. Working memory capacity measures consisted of three of the four tasks that were used in Study 1. The SSTM task—which was the least significant indicator in the WM capacity measurement model in Study 1—was excluded from the present study due to pragmatic testing time constraints.

Fluid intelligence tasks. Fluid intelligence measures consisted of the same five tasks that were used in Study 2.

Procedure. Participants first completed the three variants of the modified updating task in the same fixed order as Studies 1 and 2 (i.e., letters, digits, words) to obtain a measure of their removal ability. Next, they completed the WM capacity battery in the same fixed order as in Study 1 (i.e., MU, OS, SS). Finally, participants completed the five gF tasks in the same fixed order as Study 2 (i.e., four CFIT tests followed by the RAPM-s). The entire experiment took approximately 90 minutes.

Results

   Data Screening. Participants were excluded if they had missing data on the removal tasks ($n = 4$) and WM capacity tasks ($n = 3$). Next, performance-based outliers were excluded. Specifically, participants were excluded from further analyses if their recall performance in any of the updating tasks ($n = 11$), any of their removal-efficiency scores ($n = 8$), their performance on any of the WM tasks (including the complex-span secondary tasks; $n = 7$) or gF tasks ($n = 6$) fell 3 *SD*s above or below the respective grand mean of all participants.

   Updating tasks.

   Recall accuracy. Similar to the first two studies, recall accuracy was high for each updating task (letters: $M = .95$, $SD = .05$; digits: $M = .97$, $SD = .03$; words: $M = .94$, $SD = .06$).

   Updating RTs. Before aggregation, individual responses < 300 ms were discarded, as were RTs more than 3 *SD*s away from the individual participant's mean RT in a given task. Mean updating RTs for all tasks are provided in Figure 7. A 3×2 repeated measures ANOVA on updating RTs, with the factors task (letters, digits, words) and CTI (short, long) yielded a significant main effect of task, $F(2,406) = 166.80$, $MSE = .048$, $p < .001$, $\eta_p^2 = 0.45$, a significant main effect of CTI, $F(1,203) = 595.60$, $MSE = .019$, $p < .001$, $\eta_p^2 = 0.75$, and a significant interaction, $F(2,406) = 11.20$, $MSE = .006$, $p < .001$, $\eta_p^2 = 0.05$. As in the first two studies, the effect of the CTI manipulation found by Ecker, Lewandowsky et al. (2014) was replicated across all three tasks.

   Removal-efficiency scores. As in Studies 1 and 2, three removal-efficiency scores per participant were calculated from the updating response times in the three updating tasks. Spearman-Brown corrected reliability indices, bootstrapped with $N = 1,000$ runs, were $\rho = .46$ for the letter task, $\rho = .64$ for the digit task, and $\rho = .60$ for the word task. Again, we noted that these

estimates are lower than the reliabilities reported by Ecker, Lewandowsky et al. (2014). However, similar to Studies 1 and 2, the present study used latent variables to represent the dimensions of interest and the reported effects are, therefore, considered not affected by the relatively low reliabilities associated with the indicator observed scores (Friedman & Miyake, 2017; Little et al., 1999). Table 5 displays descriptive statistics for the three removal-efficiency scores. Negative values indicate greater removal efficiency and positive values indicate lower removal efficiency. As expected, the means for the regression-residual removal scores were close to zero. Table 6 displays the inter-correlations of the removal scores. Similar to the first two studies, all tasks correlated moderately with each other indicating a common removal factor, but did not correlate perfectly indicating task specific variance.

WM capacity battery. Spearman-Brown corrected split-half reliability estimates for the OS, SS, and MU tasks were $\rho = .82$, .86, and .85, respectively. Table 5 displays descriptive statistics for the three WMC tasks. All values were in the approximate ranges reported in the validation studies of Lewandowsky et al. (2010). Task inter-correlations are provided in Table 6. The three WMC tasks had large, positive correlations with each other, which were in the approximate ranges reported by Lewandowsky et al. (2010).

Fluid intelligence tasks. Cronbach's alpha coefficients for the fluid-intelligence tasks were as follows: $\alpha_{CFIT1} = .45$, $\alpha_{CFIT2} = .51$, $\alpha_{CFIT3} = .29$, $\alpha_{CFIT4} = .54$, and $\alpha_{RAPM-S} = .63$. As in Study 2, the low CFIT subscale reliabilities were counteracted by combining the subtests together to form two CFIT composite scores. Table 5 displays descriptive statistics for the five gF tasks. Task inter-correlations are provided in Table 6. As expected, all gF tasks had moderate, positive relationships with each other. Unlike Study 2, the CFIT4 subtest had positive associations with the other three

CFIT subtests. Additionally, most of the gF tasks had positive correlations with the WMC tasks. Table 6 also shows that the gF tasks did not significantly correlate with the removal tasks.

The relation between removal efficiency, WM capacity, and fluid intelligence. In order to examine the relationship between removal efficiency, WM capacity, and gF, a series of confirmatory factor analytic models was tested. Similar to Studies 1 and 2, a simple linear regression model that regressed each cognitive ability onto age was initially performed in order to control for the effect of age on cognitive ability. The standardized residuals from these regression analyses were used as observed variables for the purposes of defining the removal, WM capacity, and gF latent variables.

Measurement models. First, three basic measurement models (one for removal, one for WM capacity, one for gF) were tested. Following the same procedure as Study 2, the gF measurement model combined the four CFIT tasks into two observed variables. Thus, all three latent variables were defined by three observed variables (*REM*: letters, digits, words; *WMC*: OS, SS, MU; *gF*: CFIT1&2, CFIT3&4, RAM-S). The inter-correlations between the gF indicators used in the modelling and the gF, removal, and WM capacity tasks are provided in Table 6. The *REM* latent variable had significant, moderate, positive standardized loadings from all three removal tasks. The weights were .55, .54, and .67 for letters, digits, and words, respectively; all ps < .001. This indicated that the three removal tasks measured a common removal factor. The three WMC tasks all had significant substantial positive standardized loadings on the WMC latent variable. The loadings were .85, .82, and .66 for OS, SS, and MU, respectively; all ps < .001. This indicated that these three tasks measured a common WM capacity factor. Similarly, the *gF* latent variable had significant moderate to substantial positive standardized loadings from all three observed

variables. The weights were .77, .60, and .56 for CFIT1&2, CFIT3&4, and RAPM-S, respectively; all ps < .001. This indicated that these five tasks measured a common gF factor.

Correlated 3-factor model. In order to estimate the relationships for the mediation analysis, a correlated three-factor model was tested. As shown in Figure 8, the model consisted of three latent variables (*REM*, *WMC*, and *gF*) that received loadings from each of their respective observed variables. Model identification was achieved by assigning a variance of 1 to each of the three latent variables. The correlated three-factor model was associated with a good fit, $\chi^2(24) = 38.43$, $p = .031$, *CFI* = .966, *TLI* = .950, *RMSEA* = .054, *SRMR* = .051. All nine observed variables were associated with significant ($p < .001$) and moderate to substantial standardized loadings on their respective factors. Additionally, the inter-latent variable correlations between *REM* and *WMC* ($r = -.30$; 95% CI: -.51 / -.09; *SE* = .092; $p = .001$), *REM* and *gF* ($r = -.21$; 95% CI: -.41 / .02; *SE* = .104; $p = .044$), and *WMC* and *gF* ($r = .49$; 95% CI: .31 / .65; *SE* = .076; $p < .001$) were all significant. We were, therefore, able to replicate the significant relationships that removal efficiency had with WM capacity and gF in Studies 1 and 2, respectively.

Mediation analysis. To test whether the effect of removal ability on gF was mediated by WM capacity, a latent-variable mediation analysis was performed. As shown in Figure 9, the model consisted of removal (*REM*) as a predictor variable, WM capacity (*WMC*) as the mediating (i.e., intervening) variable, and fluid intelligence (*gF*) as the dependent variable. The model, therefore, estimated a direct regression path from *REM* to *gF* to represent the direct effect of removal efficiency on gF. Additionally, the model estimated an indirect regression path from *REM* to *gF* via *WMC*. A full mediation effect was considered consistent with the observation of a reduction of the total effect between *REM* and *gF* to a non-significant direct effect. In contrast, a partial mediation effect was considered consistent with the observation of a statistically significant

reduction in the direct effect, in comparison to the total effect between *REM* and *gF*, but not to non-significance.

In order to facilitate interpretation of the results, the regression-residual removal efficiency indices were reflected. Thus, in the current investigation, positive values indicated better ability across all variables. Model solutions were estimated via maximum likelihood. However, bootstrapping was also used to calculate *p*-values and bias-corrected 95% confidence intervals for all point-estimates (bootstrap samples = 1,000).

As can be seen in Figure 9, *REM* was associated with a statistically significant standardized indirect effect onto *gF* via WMC, β = .14, *p* = <.001 (95% CI: .06 / .28). Furthermore, *REM* was not associated with a statistically significant direct effect onto *gF*, β = .07, *p* = .547 (95% CI: -.16 / .26). Thus, as the indirect effect was statistically significant, it necessarily implies that the total effect of *REM* onto *gF* was reduced in a statistically significant manner. These results are consistent with full mediation. Finally, the squared multiple correlation for *gF* was estimated as R^2 = .24. Thus, approximately 24 % of the true score variance in gF was accounted for by individual differences in WM capacity and removal efficiency.

Alternative mediation model 1. We acknowledge that there may be alternative models regarding the nature of the relationships between the investigated constructs. Thus, for thoroughness, two alternative mediation models were tested. As shown in Figure 10, the first alternative model consisted of fluid intelligence (*gF*) as the primary predictor variable, removal (*REM*) as the mediating/intervening variable, and WM capacity (*WMC*) as the dependent variable. Model solutions were estimated via maximum likelihood, and bootstrapping was used to calculate the *p*-values and bias-corrected 95 % confidence intervals for all point-estimates (bootstrap samples = 1,000).

As can be seen in Figure 10, *gF* was not associated with a statistically significant standardized indirect effect onto *WMC* via *REM*, β = .04, *p* = .057 (95% CI: -.00 / .14). Furthermore, *gF* was not associated with a statistically significant standardized direct effect onto *REM*, β = .21, *p* = .072 (95% CI: -.01 / .41). However, the standardized direct effect of *REM* onto *WMC* was statistically significant, β = .21, *p* = .045 (95% CI: .00 / .41). As expected, *gF* was associated with a statistically significant direct effect onto *WMC*, β = .45, *p* = .002 (95% CI: .25 / .63). The results from the first alternative model, where *REM* was specified to mediate the relationship between *gF* and *WMC*, are consistent with partial mediation. Finally, the squared multiple correlation for *WMC* was estimated as R^2 = .28. Thus, approximately 28 % of the true score variance in WM capacity was accounted for by individual differences in gF and removal efficiency.[6]

Alternative mediation model 2. The second alternative mediation model that was tested was a model where WM capacity predicts both gF and removal. As shown in Figure 11, this model consisted of WM capacity (*WMC*) as the primary predictor variable, fluid intelligence (*gF*) as the mediating/intervening variable, and removal (*REM*) as the dependent variable. Again, model solutions were estimated via maximum likelihood, and bootstrapping was used to calculate *p*-values and bias-corrected 95 % confidence intervals for all point-estimates (bootstrap samples = 1,000).

---

[6] We additionally tested a structural model without a mediating variable whereby *gF* predicted both *REM* and *WMC*. In this model, *gF* had statistically significant direct effects onto both *REM* (β = .25, *p* =.031; 95% CI: .03 - .44) and *WMC* (β = .50, *p* = .002; 95% CI: .33 - .65). Additionally, this model was associated with good fit statistics, $\chi^2(25)$ = 43.18, *p* = .013, *CFI* = .958, *TLI* = .939, *RMSEA* = .060, *SRMR* = .067. We compared AIC and BIC estimates for this model (AIC = 83.18; BIC = 149.54) and the proposed mediation model (AIC = 80.43; BIC = 150.11). The absolute difference between the AIC estimates was > 2, while the difference in BIC was negligible. Taken together, these results provide positive evidence that the proposed mediation model is a better fitting model (e.g., see Raftery, 1995).

As can be seen in Figure 11, *WMC* was not associated with a statistically significant standardized indirect effect onto *REM* via *gF*, $\beta = .04$, $p = .474$ (95% CI: -.10 / .17). Furthermore, *gF* was not associated with a statistically significant direct effect onto *REM*, $\beta = .08$, $p = .538$ (95% CI: -.19 / .31). However, *WMC* was associated with a statistically significant direct effect onto *REM*, $\beta = .26$, $p = .038$ (95% CI: .01 / .52). The results from the second alternative model, where *gF* mediates the relationship between *WMC* and *REM*, are consistent with partial mediation. The squared multiple correlation for *REM* was estimated as $R^2 = .10$. Thus, approximately 10 % of the true score variance in removal efficiency was accounted for by individual differences in WM capacity and gF.

Comparison of mediation models. As can be seen in Figures 9, 10, and 11, the number and nature of the estimated parameters associated with each of the three mediation models were the same. That is, the models were equivalent with respect to model fit testing. Correspondingly, all three mediation models were associated with the same level of model close-fit: $\chi^2(24) = 38.43$, $p = .031$, *CFI* = .966, *TLI* = .950, *RMSEA* = .054, *SRMR* = .051. Thus, despite the theoretical distinctions between the models, the models were considered empirically equivalent.

Discussion

The aim of Study 3 was to investigate simultaneously the nature of the associations between removal efficiency, WM capacity, and gF via a mediation analysis. Two rival hypotheses were entertained. One the one hand, good removal skills may contribute to greater WM capacity by reducing interference from distracting information, and good WM capacity in turn may facilitate reasoning ability. This theoretical standpoint suggests that WM capacity fully mediates the relationship between removal efficiency and gF. On the other hand, removal skills may contribute directly to both WM capacity and gF (in addition to a potential indirect effect of removal

on gF via WM capacity). This would imply only partial mediation, such that there would still be a significant direct effect of removal on gF in a mediation model including WM capacity as the mediator.

In conjunction with Studies 1 and 2, the results from Study 3 allow for two conclusions. The first relates to the relationship between removal and WM capacity. Ecker et al. (Ecker, Lewandowsky et al., 2014; Ecker, Oberauer et al., 2014) suggested that the removal process is important not only for WM updating, but also WM maintenance processing. However, Ecker, Lewandowsky et al. (2014) did not find a significant association between removal and WM capacity. In contrast, Study 1 of this investigation did, in fact, find a relationship between removal and WM capacity, suggesting that the lack of an effect found by Ecker, Lewandowsky et al. (2014) may have been a false-negative, potentially brought about by their methodological approach of using only a single updating task to represent individual-differences in removal efficiency. Study 3 replicated the latter finding and confirmed that there is a genuine relationship between an individual's ability to remove information from WM and their WM capacity. More specifically, the more efficient a person is at removing outdated/irrelevant information, the greater their effective WM capacity. Furthermore, the results from Studies 1 and 3 suggest that removing irrelevant items during updating is a similar process to removing distractors during complex-span tasks. The findings, thus, support Ecker et al.'s (Ecker, Lewandowsky et al., 2014; Ecker, Oberauer et al., 2014) theoretical notion that removal may be a key element for both WM updating and WM maintenance. In addition, the results from Studies 1 and 3 are consistent with computational models of WM incorporating an interference-reduction mechanism not reliant upon decay, viz. removal (Oberauer et al., 2012).

The second conclusion that can be made from the results of Studies 2 and 3 relates to the relationship between removal and gF. Shipstead and colleagues (Harrison et al., 2014; Shipstead et al., 2016; Shipstead et al., 2014) suggested that the ability to disengage from irrelevant information may be a potential mechanism contributing to performance on reasoning tasks. Study 2 supported this idea by demonstrating a relationship between an individual's ability to remove information from WM and their gF. However, Study 2 was not designed to ascertain whether this relation would hold up when a measure of WM capacity was entered into the modelling as a mediator. The results from our modelling in Study 3 fell in line with the hypothesis that WM capacity fully mediates the relationship between removal efficiency and gF: There was an indirect effect of removal efficiency on gF through WM capacity, but no direct effect of removal onto gF. This implies that removal efficiency may influence reasoning ability, but only by its influence on WM capacity.

We acknowledge that there may be alternative models regarding the nature of the relationship between removal, WM capacity, and gF. Consequently, for thoroughness, we tested two alternative mediation models in Study 3. The first alternative model proposed that gF predicts both removal and WM capacity, and that removal mediates the relationship between gF and WM capacity. This reflects the assumption that individuals with higher levels of fluid intelligence are more efficient at removing irrelevant information from WM, which in turn results in greater WM capacity. The results suggested that removal partially mediates the relationship between gF and WM capacity. There was no indirect effect of gF onto WM capacity through removal.

The second alternative mediation model proposed that WM capacity predicts both gF and removal, and that gF mediates the relationship between WM capacity and removal. This reflects the assumption that individuals with greater WM capacity have higher levels of fluid intelligence

and, therefore, possess greater executive abilities. The results suggested that gF partially mediates the relationship between WM capacity and removal. There was no indirect effect of WM capacity onto removal through gF.

Although the three mediation models tested in this investigation were associated with different theoretical implications, they were associated with the same model fit statistics and, therefore, considered indistinguishable empirically. The problem of equivalent models pervades differential psychology (Raykov & Marcoulides, 2001). While we acknowledge that there may be various explanations regarding the nature of the relationships between removal, WM capacity, and gF, our preference is for a model that includes WM as a mediator of the total effect between removal efficiency and gF. Our preference is influenced principally by empirical and theoretical considerations (e.g., Ecker et al., 2010; Ecker, Lewandowsky et al., 2014; Ecker, Oberauer et al., 2014; Oberauer, 2009; Oberauer et al., 2012), which we discuss further in the General Discussion. Consequently, we propose that the removal process be characterized as an executive updating process that may facilitate the execution of WM capacity tasks.

General Discussion

Ecker et al. (2010; Ecker, Lewandowsky et al., 2014; Ecker, Oberauer et al., 2014) have argued that an active, item-wise removal process is a core WM updating process. The present article presents three studies investigating how the efficiency of this removal process relates to WM capacity and gF. Studies 1 and 2 showed that the ability to remove information from WM—measured at the latent-variable level through a novel battery of updating tasks—relates to a person's WM capacity and their fluid intelligence, respectively. Study 3 showed that WM capacity fully mediates the relationship between removal and gF, and that removal efficiency does not directly contribute to performance on reasoning tasks. A detailed discussion of the findings

follows. Specifically, the first three sub-sections of the general discussion relate to the three studies in this paper, respectively. This is followed by a brief discussion of alternative interpretations of the processing mechanisms associated with our removal task, before finally addressing some limitations and directions for future research.

Removal and WM Capacity

Oberauer et al. (2012) argued that removal is not only important for information substitution during WM updating, but that removal is also important for general maintenance of information in WM to the extent that it can prevent interference through the removal of distracting information. Thus, we hypothesized that removal ability should predict general WM functioning, and in particular, WM capacity. Our results were in line with this prediction. We therefore argue that active item-wise removal serves to minimize interference during WM tasks, such that efficient removal skills allow maintenance processing to focus on relevant information, hence effectively increasing WM's capacity. In other words, removal efficiency predicted WM capacity because good performance in WM tasks—such as complex-span tasks—requires the removal of irrelevant information.

An alternative view why our removal measure might relate to WM capacity is that all tasks involved require attentional control in order to focus processing on task-relevant items (Kane, Bleckley, Conway, & Engle, 2001; McCabe, Roediger, McDaniel, Balota, & Hambrick, 2010; Unsworth & Spillers, 2010). An individual with poor attentional control will struggle with updating in the short-CTI condition, where two items compete for representation, but may cope well in the long-CTI condition, where the prolonged CTI may facilitate focusing on the target item. Hence, such a person would seem to show low removal efficiency. In WM tasks, they may generally be more easily distracted and thus show lower WMC scores.

The current study cannot differentiate between these two possibilities. However, there are two more general reasons not to embrace the alternative attentional-control view: (1) this notion cannot explain what happens to irrelevant items if there is no decay; (2) neither the decay nor the attentional-control view can explain a pattern observed in Experiment 3 of Ecker, Oberauer et al. (2014), where with two frames being updated, the updating-RT benefit associated with occasional item repetition disappeared with a long CTI—but only in the left not the right updated frame. Ecker et al. argued that people scan a list from left to right during updating and remove only one item during the long CTI even when two are to-be-updated, to avoid costs associated with frame switches and processing-mode switches (see Kessler & Oberauer, 2014, 2015). Thus, the benefit of item repetition remained intact in the right to-be-updated frame, which was not affected by removal, while it disappeared in the left to-be-updated frame because the memorized item had been removed (and thus a repeated item was encoded just like a new item, as there was no representational overlap anymore with an existing item in WM). Assuming that focussing on (or selectively rehearsing) relevant items "pushes" irrelevant items out of the active WM system (or lets them decay) cannot explain how one irrelevant item can stay in WM while another one does not.

Studies 1 and 3 provide empirical evidence demonstrating that efficient removal significantly predicts good performance in WM capacity tasks that require maintaining relevant WM representations in an active state. These findings are consistent with Oberauer et al.'s (2012) SOB-CS computational model of WM, in which outdated representations are removed when there is free time in-between distractor operations using Hebbian anti-learning to allow for maintenance processing of relevant information. We therefore argue that removal is an executive process that serves WM maintenance processing—along with other generic processes such as encoding,

maintenance, and retrieval (see Colom, Abad, Quiroga, Shih, & Flores-Mendoza, 2008; Martínez et al., 2011)—in addition to WM updating. With regards to WM updating, our findings are also in line with work by Kessler and Oberauer (2014, 2015) investigating the costs associated with switching between maintenance and updating modes of WM. While Kessler and Oberauer did not explicitly model the removal process, they speculated that item removal could run in parallel with the switch from maintenance to updating mode.

Removal and Fluid Intelligence

Regarding the relationship between WM and gF, Oberauer, Süß, Wilhelm, and Wittmann (2008) argued that WM and gF share common processing demands, in particular the requirement to build and maintain flexible bindings. In WM, remembering a memory set such as a list of items in serial order requires individuals to form relational representations between each item and its list position (or between items). Similarly, reasoning tasks require individuals to form spontaneous arbitrary relationships between patterns and rules. If there is an inherent capacity limit to the number of relational representations (i.e., bindings) that can be formed at any one time, the relation between WM capacity and gF may be due to this shared capacity limit relating to the ability to form new relational bindings (Oberauer et al., 2008).

From this assumption, it is only a small step to propose that WM and gF also share a common *unbinding* mechanism. WM contents frequently need to be updated, requiring not only the new binding of items to positions, but also the unbinding—or removal—of old items (Oberauer et al., 2012). Likewise, in gF tasks, relations between patterns and rules often need to be established in order to allow the formulation of predictions and testing of assumptions; however, when these efforts are not fruitful during the task, the relations need to be unbound or removed (also see Harrison et al., 2014; Shipstead et al., 2014).

Previous proposals in the literature that updating has a strong relationship to gF (e.g., Chen & Li, 2007; Martínez et al., 2011), and that updating is the only executive function to predict gF (Friedman et al., 2006), have relied on traditional, unspecific updating tasks. In this context, it has been argued that both WM capacity and updating relate to gF because of the common requirement for attentional control, serving to retain relevant information in the face of concurrent processing demands and distraction, as well as substitute information when it becomes irrelevant (Engle et al., 1999; Miyake et al., 2000). By contrast, we have argued that the traditional updating tasks measure mainly WM capacity; thus, the relation between memory updating and gF reported in the literature is more likely to actually reflect the established link between WM capacity and gF. From our point of view, it is the efficient removal of outdated information that lies at the core of memory updating; any direct relation between WM updating and gF would thus involve the efficiency of this removal process. While Study 2 seemed to support this notion, Study 3 demonstrated no direct relation between removal and gF. We thus conclude that gF does not directly benefit from updating abilities over and above the benefits conveyed through more efficient WM processing.

Removal, WMC, and Fluid Intelligence

WM and gF have been widely found to be strongly related constructs (e.g., median $r = .72$, $N = 3,100$; Kane et al., 2005). In fact, Martínez et al. (2011) found that WM capacity was hardly distinguishable from WM updating and gF at the latent-variable level (also see Schmiedek et al., 2009). They argued that individual differences in gF, therefore, can be largely accounted for by generic short-term storage processes, such as encoding, maintenance, and retrieval. Indeed, WM and intelligence share common capacity limitations because reasoning tasks use WM as a workspace (e.g., a certain number of items or relationships between items may need to be kept

active during the reasoning process to come to a solution; Colom et al., 2008; Colom, Chuderski, & Santarnecchi, 2016; Halford, Cowan, & Andrews, 2007).

Short-term storage is not the only explanation for the WM-gF relationship. Kane and Engle (2002) argued more specifically that executive attention—that is, attentional control to focus on goal-relevant information in the face of distraction—is the important underlying cause of the correlation between WMC and gF. In addition, the relationship between WM and gF has also been explained by individual differences in the ability to cope with novelty and complexity (e.g., Salthouse, 2014) as well as the ability to form new structural representations (e.g., Oberauer et al., 2008).

We had an alternative hypothesis: we speculated that removal efficiency may relate to both WMC and gF through a common underlying function, viz. the unbinding of irrelevant/outdated WM content and the associated reduction of interference. Like others (e.g., Ackerman, Beier, & Boyle, 2005; Kane et al., 2005), this would assume that the two constructs of WMC and gF are clearly separable—for example, Shipstead and colleagues (2016) recently proposed that performance on WM tasks primarily calls upon maintenance processes, whereas performance on gF tasks primarily call upon disengagement processes (also see Harrison et al., 2014; Shipstead et al., 2014). However, we assumed that it may be the prevention of interference through an active, item-wise removal process that may contribute to performance in both WM and gF tasks. To illustrate, the ability to manage interference is important for performance in WM tasks, such as complex span tasks that involve distractors, but also for gF tasks like the RAPM, because good performance requires individuals to be able to maintain rule combinations relevant to solving a problem while efficiently removing distractions from all other irrelevant combinations. Results from Studies 1 and 2 seemed to support this notion. However, Study 3 clearly showed that there

was no direct effect of removal on gF. Thus, it seems more appropriate to characterize removal as an executive process that exclusively serves WM and does not have direct influence on gF. Nonetheless, we argue that removal can play a role in determining a person's effective WM capacity, which in turn facilitates performance on gF tasks.

An indirect effect of removal on gF via WM capacity seems particularly plausible as reasoning items on gF tests increase in difficulty. Simple or early items do not place a heavy demand on WM, such that removal may not be required to maintain efficient processing. The demand placed on the WM system increases as reasoning problems become more complex, as more complex problems will require the maintenance of multiple rule combinations that require greater WM capacity (cf. Carpenter et al., 1990; Harrison et al., 2014; Little et al., 2014; Salthouse, 1993; Unsworth & Engle, 2005; Wiley, Jarosz, Cushen, & Colflesh, 2011). Thus, the importance of removal may increase with difficulty. Future research could investigate this novel hypothesis.

Alternative Interpretations Regarding the Processing Involved in the Removal Task

We assume that the presently obtained results provide additional evidence for an active, item-wise removal process that is an executive process serving WM maintenance and updating. However, it is acknowledged that there may be alternative explanations regarding the processing involved in the removal tasks. One alternative explanation is in respect to cue utilization. While providing a cue during updating disentangles the removal, encoding, and attention reorientation confound, it can be argued that the results of Ecker, Lewandowsky et al. (2014) and the present set of studies may also be interpreted in terms of individual differences in the general capacity for cue utilization. This hypothesis would assume that participants may differ primarily in their ability to use cues to anticipate task demands and prepare for updating rather than their efficiency to perform the actual removal operation itself (see, e.g., Brouwers, Wiggins, Helton, O'Hare, &

Griffin, 2016; Xiang et al., 2016). While we acknowledge that individuals may differ in the effectiveness of their cue utilization, we argue against the notion that this cue-utilization hypothesis can explain present findings. Assuming that only the long-CTI condition would allow time to utilize the cue, poor cue utilization would mean delayed removal onset and potentially incomplete removal during the CTI, thus prolonging updating RTs in the long-CTI condition. Hence, a person with poor cue utilization may ironically seem to have high removal efficiency (i.e., a small RT difference between short and long CTI conditions). In the present context, this would imply that participants with greater WM capacity have poorer cue utilization skills. We thus argue that the removal notion offers a superior interpretation of the present findings, as well as the findings of Ecker et al. (Ecker, Lewandowsky et al., 2014; Ecker, Oberauer et al., 2014).

Similarly, we acknowledge that rather than our proposed process of active item-wise removal, there may instead be an inhibitory process serving to suppress outdated or irrelevant information. Indeed, Friedman and Miyake (2004) identified a type of inhibition that bears some resemblance to our proposed removal process, namely 'resistance to distractor interference'. This refers to the ability to resist or resolve interference from information in the external environment that is irrelevant to the task at hand. However, we note that it has proven difficult to establish reliable inhibition measures at the latent-variable level (both in our lab and others; e.g., Huizinga, Dolan, & van der Molen, 2006; Rey-Mermet, Gade, & Oberauer, 2017; Singh, Gignac, Brydges, Rey-Mermet, & Ecker, 2018; van der Sluis, de Jong, & van der Leij, 2007; see also Friedman & Miyake, 2017), suggesting that there may be many different types of inhibition (Engle & Kane, 2003; Friedman & Miyake, 2004). Thus, while removal of information from WM could be seen as a type of inhibition, from our perspective a label loses its usefulness if it describes many different things, and we thus argue that the proposed "removal" term is the preferable term.

Limitations and Future Research

      Although the quality and breadth of measures included in this investigation may be regarded as a strength, we acknowledge that it is a limitation that we did not include a measure of pure processing speed. While the use of proportional gain and regression-residual measures mitigate against an influence of processing speed, it is impossible to rule out entirely a role for individual differences in processing speed. Such a possible confound is potentially important, as individual differences in processing speed are known to be related to both WM and fluid intelligence (Conway, Cowan, Bunting, Therriault, & Minkoff, 2002; Redick, Unsworth, Kelly, & Engle, 2012). Consequently, we encourage future research to investigate the extent to which the proposed removal process relates to WM capacity, gF, and processing speed.

      Correspondingly, we note that the measurement of the proposed removal process is a relatively new undertaking, and further developments and improvements can be expected. In particular, in this investigation, the removal tasks used alphanumeric stimuli, which may produce domain-specific effects. In order to derive a more general removal latent factor, future investigations of removal may consider including figural or spatial tasks, in order to filter out domain- or stimulus-specific variance. Likewise, with respect to the relationship between removal and WM capacity, future studies may also consider investigating this relationship using a different set of WM tasks to allow for generalization. Finally, the comparatively small correlation estimated between WM capacity and gF in Study 3 was likely a result of range restriction in a rather homogenous sample of participants. Future studies could investigate the relationship between removal, WM capacity, and gF in a more diverse sample.

Conclusion

To conclude, the present work provides further evidence from three studies that the removal of information from WM may play a role in both the maintenance and updating of WM. The current work is the first to investigate the nature of the relations between removal efficiency, WM capacity, and fluid intelligence. The modelling results suggest that WM capacity fully mediates the relationship between removal efficiency and gF. It is, therefore, proposed that active, item-wise removal is an executive updating process that serves the WM system directly, with no direct impact on fluid reasoning.

Author Note

References

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin*, *131*, 30-60.

Arbuckle, J. L. (2013). *IBM SPSS Amos 22 User's Guide*. Crawfordville: Amos Development Corporation.

Arthur, W., & Day, D. V. (1994). Development of a short form for the Raven Advanced Progressive Matrices Test. *Educational and Psychological Measurement*, *54*, 394-403.

Artuso, C., & Palladino, P. (2011). Content-context binding in verbal working memory updating: On-line and off-line effects. *Acta Psychologica, 136,* 363-369.

Baddeley, A. D. (1986). *Working memory*. Oxford: Clarendon Press.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*, 417-423.

Barrouillet, P., Bernardin, S., Portrat, S., Vergauwe, E., & Camos, V. (2007). Time and cognitive load in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 570-585.

Barrouillet, P., De Paepe, A., & Langerock, N. (2012). Time causes forgetting from working memory. *Psychonomic Bulletin & Review, 19,* 87-92.

Berman, M. G, Jonides, J., & Lewis, R. L. (2009). In search of decay in verbal short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 317-333.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107,* 238-246.

Bentler, P. M. (1995). *EQS structural equations program manual.* Encino, CA: Multivariate Software.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10,* 433-436.

Brouwers, S., Wiggins, M. W., Helton, W., O'Hare, D., & Griffin, B. (2016). Cue utilization and
      cognitive load in novel task performance. *Frontiers in Psychology, 7,* 1-12.

Browne, M. W. & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen &
      J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA:
      Sage.

Brunner, M., Nagy, G., & Wilhelm, O. (2012). A tutorial on hierarchically structured constructs.
      Journal of Personality, 80, 796-846.

Bunting, M., Cowan, N., & Saults, J. S. (2006). How does running memory span work? *Quarterly
      Journal of Experimental Psychology, 59*, 1691-1700.

Camos, V. (2015). Storing verbal information in working memory. *Current Directions in
      Psychological Science*, *24*, 440-445.

Carpenter, P.A., Just, M.A., & Shell, P. (1990). What one intelligence test measures: A theoretical
      account of the processing in the Raven Progressive Matrices Test. *Psychological Review,
      97*, 404-431.

Cattell, R. B., & Cattell, A. K. S. (1960). *Culture Fair Intelligence Test, Scale 2*. Champaign, IL:
      Institute for Personality and Ability Testing.

Chen, T., & Li, D. (2007). The roles of working memory updating and processing speed in
      mediating age-related differences in fluid intelligence. *Aging, Neuropsychology, &
      Cognition, 14,* 631-646.

Colom, R., Abad, F. J., Quiroga, M. A., Shih, P. C., & Flores-Mendoza, C. (2008). Working
      memory and intelligence are highly related constructs, but why? *Intelligence, 36,* 584-606.

Colom, R., Chuderski, A. & Santarnecchi, E. (2016). Bridge over troubled water: commenting on Kovacs and Conway's process overlap theory. *Psychological Inquiry, 27*, 181-189.

Colom, R., Privado, J., García, L. F., Estrada, E., Cuevas, L., & Shih, P. C. (2015). Fluid intelligence and working memory capacity: Is the time for working on intelligence problems relevant for explaining their large relationship? *Personality and Individual Differences, 79,* 75-80.

Coltheart, M. (1981). The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology, 33A*, 497-505.

Conway, A. R. A., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence, 30*, 163-183.

Conway, A. R. A., Kane, M. J., Bunting, M. F, Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12*, 769-786.

Conway, A. R. A., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences, 7*, 547-552.

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review, 3,* 422-433.

Ecker, U. K. H., & Lewandowsky, S. (2012). Computational constraints in cognitive theories of forgetting. *Frontiers in Psychology, 3,* 400.

Ecker, U. K. H., Lewandowsky, S., & Oberauer, K. (2014). Removal of information from working memory: A specific updating process. *Journal of Memory and Language, 74*, 77-90.

Ecker, U. K. H., Lewandowsky, S., Oberauer, K., & Chee, A. (2010). The components of working memory updating: An experimental decomposition and individual differences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 170-189.

Ecker, U. K. H., Oberauer, K., & Lewandowsky, S. (2014). Working memory updating involves item-specific removal. *Journal of Memory and Language, 74,* 1-15.

Engle, R. W., & Kane, M. J. (2003). Executive attention, working memory capacity, and a two-factor theory of cognitive control. *Psychology of Learning and Motivation*, *44*, 145-199.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. *Journal of Experimental Psychology: General*, *128*, 309-331.

Fan, X. (2003). Two approaches for correcting correlation attenuation caused by measurement error: Implications for research practice. *Educational and Psychological Measurement, 63*(6), 915-930.

Farrell, S., & Lewandowsky, S. (2002). An endogenous distributed model of ordering in serial recall. *Psychonomic Bulletin & Review, 9,* 59-79.

Farrell, S., Oberauer, K., Greaves, M., Pasiecznik, K., Lewandowsky, S., & Jarrold, C. (2016). A test of interference versus decay in working memory: Varying distraction within lists in a complex span task. *Journal of Memory and Language, 90,* 66-87.

Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: A latent-variable analysis. *Journal of Experimental Psychology: General*, *133*, 101-135.

Friedman, N. P., & Miyake, A. (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex, 86,* 186-204.

Friedman, N. P., Miyake, A., Corley, R. P., Young, S., DeFries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science, 17*, 172-179.

Fukuda, K., Vogel, E., Mayr, U., & Awh, E. (2010). Quantity not quality: the relationship between fluid intelligence and working memory capacity. *Psychonomic Bulletin & Review, 17*, 673-679.

Gignac, G. E. (2007). Multi-factor modeling in individual differences research: Some recommendations and suggestions. *Personality and Individual Differences, 42,* 37-48.

Gignac, G. E. (2014). Fluid intelligence shares closer to 60% of its variance with working memory capacity and is a better indicator of general intelligence. *Intelligence, 47,* 122-133.

Gignac, G. E., & Kretzschmar, A. (2017). Evaluating dimensional distinctness with correlated-factor models: Limitations and suggestions. *Intelligence*, *62*, 138-147.

Gignac, G. E. & Watkins, M. W. (2015). There may be nothing special about the association between working memory capacity and fluid intelligence. *Intelligence, 52,* 18-23.

Halford, G. S., Cowan, N., & Andrews, G. (2007). Separating cognitive capacity from knowledge: A new hypothesis. *Trends in Cognitive Sciences*, *11*, 236-242.

Harrison, T. L., Shipstead, Z., & Engle, R. W. (2014). Why is working memory capacity related to matrix reasoning tasks? *Memory & Cognition, 43*, 389-396.

Huizinga, M., Dolan, C. V., & van der Molen, M. W. (2006). Age-related change in executive function: Developmental trends and a latent variables analysis. *Neuropsychologia, 44,* 2017-2036.

Jöreskog, K. G., & Sörbom, D. (1986). LISREL VI: Analysis of linear structural relationships by maximum likelihood and least squares methods. Mooresville, IN: Scientific Software.

Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General, 130*, 169-183.

Kane, M. J., & Engle, R. W. (2000). Working-memory capacity, proactive interference, and divided attention: limits on long-term memory retrieval. *Journal of Experimental Psychology, Learning, Memory, and Cognition, 26*, 336-358.

Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin & Review*, *9*, 637-671.

Kane, M. J., Hambrick, D. Z., & Conway, A. R. A. (2005). Working memory capacity and fluid intelligence are strongly related constructs: Comment on Ackerman, Beier, and Boyle. *Psychological Bulletin, 131*, 66-71.

Kessler, Y., & Meiran, N. (2008). Two dissociable updating processes in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 1339-1348.

Kessler, Y., & Oberauer, K. (2014). Working memory updating latency reflects the cost of switching between maintenance and updating modes of operation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40,* 738-754.

Kessler, Y., & Oberauer, K. (2015). Forward scanning in verbal working memory updating. *Psychonomic Bulletin & Review*, *22*, 1770-1776.

Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology, 55*, 352-358.

Lendínez, C., Pelegrina, S., & Lechuga, T. (2011). The distance effect in numerical memory-updating tasks. *Memory & Cognition, 39*, 675-685.

Lewandowsky, S., Geiger, S. M., & Oberauer, K. (2008). Interference-based forgetting in verbal short-term memory. *Journal of Memory and Language, 59,* 200-222.

Lewandowsky, S., & Oberauer, K. (2015). Rehearsal in serial recall: An unworkable solution to the non-existent problem of decay. *Psychological Review, 122,* 674-699.

Lewandowsky, S., Oberauer, K., & Brown, G. D. (2009). No temporal decay in verbal short-term memory. *Trends in Cognitive Sciences*, *13*, 120-126.

Lewandowsky, S., Oberauer, K., Yang, L.-X., & Ecker, U. K. H. (2010). A working memory test battery for MATLAB. *Behavior Research Methods, 42*, 571-585.

Little, D. R., Lewandowsky, S., & Craig, S. (2014). Working memory capacity and fluid abilities: The more difficult the item, the more more is better. *Frontiers in Psychology, 5,* 239.

Little, T. D., Lindenberger, U., & Nesselroade, J. R. (1999). On selecting indicators for multivariate measurement and modelling with latent variables: When "good" indicators are bad and "bad" indicators are good. *Psychological Methods, 4,* 192-211.

Lustig, C., May, C. P., & Hasher, L. (2001). Working memory span and the role of proactive interference. *Journal of Experimental Psychology: General, 130*, 199-207.

MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. *Annual Review of Psychology, 58*, 593-614.

Martínez, K., Burgaleta, M., Román, F. J., Escorial, S., Shih, P. C., Quiroga, M. Á., & Colom, R. (2011). Can fluid intelligence be reduced to 'simple' short-term storage? *Intelligence, 39*, 473-480.

May, C. P., Hasher, L., & Kane, M. J. (1999). The role of interference in memory span. *Memory & Cognition., 27*, 759-767.

McCabe, D. P., Roediger, H. L., McDaniel, M. A., Balota, D. A., & Hambrick, D. Z. (2010). The relationship between working memory capacity and executive functioning: Evidence for a common executive attention construct. *Neuropsychology, 24*, 222-243.

Miyake, A., Friedman, N. P., Emerson, M.J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex ''frontal lobe'' tasks: A latent variable analysis. *Cognitive Psychology, 41*, 49-100.

Miyake, A., & Shah, P. (1999). *Models of working memory: Mechanisms of active maintenance and executive control*. Cambridge University Press.

Morris, N., & Jones, D. M. (1990). Memory updating and working memory: The role of the central executive. *British Journal of Psychology, 81,* 111-121.

Murty, V. P., Sambataro, F., Radulescu, E., Altamura, M., Iudicello, J., Zoltick, B., Weinberger, D. R., Goldberg, T. E., & Mattay, V. S. (2011). Selective updating of working memory content modulates meso-cortico-striatal activity. *NeuroImage, 57,* 1264-1272.

Nairne, J. S. (2002). Remembering over the short-term: The case against the standard model. *Annual Review of Psychology, 53,* 53-81.

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

Oberauer, K. (2009). Design for a working memory. *Psychology of Learning and Motivation, 51,* 45-100.

Oberauer, K., Farrell, S., Jarrold, C., & Lewandowsky, S. (2016). What limits working memory capacity? *Psychological Bulletin, 142,* 758-799.

Oberauer, K., & Kliegl, R. (2006). A formal model of capacity limits in working memory. *Journal of Memory and Language, 55*, 601-626.

Oberauer, K., & Lewandowsky, S. (2008). Forgetting in immediate serial recall: Decay, temporal distinctiveness, or interference? *Psychological Review, 115,* 544-576.

Oberauer, K., & Lewandowsky, S. (2013). Evidence against decay in verbal working memory. *Journal of Experimental Psychology: General, 142,* 380-411.

Oberauer, K., Lewandowsky, S., Farrell, S., Jarrold, C., & Greaves, M. (2012). Modeling working memory: An interference model of complex span. *Psychonomic Bulletin & Review, 19,* 779-819.

Oberauer, K., & Lin, H. Y. (2017). An interference model of visual working memory. *Psychological Review, 124,* 21-59.

Oberauer, K., Schulze, R., Wilhelm, O., & Süß, H. M. (2005). Working memory and intelligence-Their correlation and their relation: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin, 131*, 61-65.

Oberauer, K., Süß, H. M., Wilhelm, O., & Wittmann, W. W. (2008). Which working memory functions predict intelligence? *Intelligence*, *36*, 641-652.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology, 25,* 111-163.

Raven, J. C. (1962). *Advanced Raven's progressive matrices.* Melbourne: Australian Council for Educational Research.

Raykov, T., & Marcoulides, G. A. (2001). Can there be infinitely many models equivalent to a given covariance structure model? *Structural Equation Modeling*, *8*(1), 142-149.

Redick, T. S., Unsworth, N., Kelly, A. J., & Engle, R. W. (2012). Faster, smarter? Working memory capacity and perceptual speed in relation to fluid intelligence. *Journal of Cognitive Psychology*, *24*(7), 844-854.

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioural Research, 47*, 667-696.

Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment, 95*, 129-140.

Rey-Mermet, A., Gade, M., & Oberauer, K. (2017). Should we stop thinking about inhibition? Searching for individual and age differences in inhibition ability. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*

Roth, J. K., Serences, J. T., & Courtney, S. M. (2006). Neural system for controlling the contents of object working memory in humans. *Cerebral Cortex, 16,* 1595-1603.

Salthouse, T. A. (1993). Influence of working memory on adult age differences in matrix reasoning. *British Journal of Psychology, 84,* 171-199.

Salthouse, T. A. (2014). The relations between running memory and fluid intelligence. *Intelligence, 43,* 1-7.

Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika, 22*, 53-61.

Schmiedek, F., Hildebrandt, A., Lovdén, M., Wilhelm, O., & Lindenberger, U. (2009). Complex span versus updating tasks of working memory: The gap is not that deep. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 1089-1096.

Shipstead, Z., Harrison, T. L., & Engle, R. W. (2016). Working memory capacity and fluid intelligence: maintenance and disengagement. *Perspectives on Psychological Science, 11*, 771-799.

Shipstead, Z., Lindsey, D. R. B., Marshall, R. L., & Engle, R. W. (2014). The mechanisms of working memory capacity: Primary memory, secondary memory, and attention control. *Journal of Memory and Language, 72,* 116-141.

Singh, K. A., Gignac, G. E., Brydges, C. R., Rey-Mermet, A., & Ecker, U. K. H. (2018). Interference control in working memory. Retrieved from: osf.io/h8cgs. doi: 10.17605/OSF.IO/H8CGS.

Souza, A. S., & Oberauer, K. (2015). Time-based forgetting in visual working memory reflects temporal distinctiveness, not decay. *Psychonomic Bulletin & Review, 22,* 156-162.

Steiger, J. H. & Lind, J. C. (1980, June). *Statistically based tests for the number of common factors.* Paper presented at the Psychometric Society Annual Meeting, Iowa City, IA.

Süß, H.-M., Oberauer, K., Wittmann, W. W., Wilhelm, O., & Schulze, R. (2002). Working-memory capacity explains reasoning ability—and a little bit more. *Intelligence, 30,* 261-288.

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38,* 1-10.

Unsworth, N., Brewer, G. A., & Spillers, G. J. (2009). There's more to the working memory capacity-fluid intelligence relationship than just secondary memory. *Psychonomic Bulletin & Review, 16*, 931-937.

Unsworth, N., & Engle, R. W. (2005). Working memory capacity and fluid abilities: Examining the correlation between operation span and Raven. *Intelligence, 33,* 67-81.

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

van der Sluis, S., de Jong, P. F., van der Leij, A. (2007). Executive functioning in children, and its relations with reasoning, reading, and arithmetic. *Intelligence, 35,* 427-449.

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.

Watkins, M. W. (2013). Omega [Computer software]. Phoenix, AZ: Ed & Psych Associates.

Wiley, J., Jarosz, A. F., Cushen, P. J., & Colflesh, G. J. H. (2011). New rule use drives the relation between working memory capacity and Raven's Advanced Progressive Matrices. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 37,* 256-263.

Wilhelm, O., Hildebrandt, A. H., & Oberauer, K. (2013). What is working memory capacity, and how can we measure it? *Frontiers in Psychology, 4*, 1-22.

Wilson, M. (1988). MRC Psycholinguistic Database: Machine-usable dictionary, version 2. *Behavior Research Methods, Instruments, & Computers*, *20*, 6-10.

Wittmann, W. W. (1988). Multivariate reliability theory: Principles of symmetry and successful validation strategies. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of Multivariate Experimental Psychology* (pp. 505-560). New York: Plenum.

Xiang, L., Zhang, B., Wang, B., Jiang, J., Zhang, F., & Hu, Z. (2016). The effect of aging on the dynamics of reactive and proactive cognitive control of response interference. *Frontiers in Psychology, 7,* 1640.

Zinbarg, R. E., Yovel, I., Revelle, W., & McDonald, R. P. (2006). Estimating generalizability to a

    latent variable common to all of a scale's indicators: A comparison of estimators for ωh.

    Applied Psychological Measurement, 30, 121-144.

Table 1.

Descriptive Statistics of Performance on Removal and Working Memory Tasks in Study 1.

Task	Mean	*SD*	Range	Skewness	Kurtosis
RemL	0.00	0.13	-0.32 - 0.39	0.41	0.07
RemD	-0.01	0.11	-0.28 – 0.23	0.19	-0.62
RemW	0.01	0.16	-0.53 – 0.45	-0.17	0.33
OS	0.69	0.12	0.33 - 0.96	-0.55	0.05
OS*pt*	0.91	0.06	0.65 - 1.00	-1.26	1.75
SS	0.68	0.13	0.29 – 0.98	-0.52	-0.10
SS*pt*	0.91	0.05	0.66 – 1.00	-1.24	2.31
MU	0.70	0.15	0.28 – 1.00	-0.40	-0.26
SSTM	0.83	0.05	0.66 – 0.95	-0.49	0.33

Note. Removal tasks: RemL = letter updating; RemD = digit updating; RemW = word updating. Working memory capacity tasks: OS = operation span; SS = sentence span; *pt* denotes secondary processing task; MU = memory updating; SSTM = spatial short-term memory. The *M*'s for the letter, digit, and word updating tasks were based on regression-residual scores, calculated as the individual residuals obtained from a simple linear regression model predicting the short-CTI RTs from the long-CTI RTs. *SD* = standard deviation. $N = 162$.

Table 2.

Task Inter-correlations Between Removal Efficiency Scores and WMC Tasks for Study 1.

_____

Task	1	2	3	4	5	6	7
1. RemL	-						
2. RemD	.34**	-					
3. RemW	.34**	.44**	-				
4. OS	-.17*	-.10	-.14	-			
5. SS	-.01	-.07	-.10	.69**	-		
6. MU	-.10	-.17*	-.25*	.57**	.49**	-	
7. SSTM	.09	-.04	-.08	.13	.07	.29**	-

_____

Note. Removal tasks: RemL = letter updating; RemD = digit updating; RemW = word updating. Working memory capacity tasks: OS = operation span; SS = sentence span; MU = memory updating; SSTM = spatial short-term memory. $N = 162$. **$p < .001$; *$p < .05$.

Table 3.

Descriptive Statistics of Performance on Removal and Fluid Intelligence Tasks in Study 2.

Task	Mean	*SD*	Range	Skewness	Kurtosis
RemL	0.01	0.12	-0.28 - 0.30	0.05	-0.35
RemD	0.01	0.11	-0.22 – 0.27	0.39	-0.29
RemW	0.01	0.14	-0.38 – 0.45	0.21	0.27
CFIT1	8.15	1.51	4.00 – 11.00	-0.14	-0.16
CFIT2	7.79	2.12	2.00 – 13.00	-0.28	-0.47
CFIT3	7.19	1.46	3.00 – 11.00	-0.09	0.23
CFIT4	6.24	1.30	2.00 – 9.00	-0.40	0.01
RAPM-S	9.08	2.11	4.00 – 12.00	-0.46	-0.53

Note. Removal tasks: RemL = letter updating; RemD = digit updating; RemW = word updating. Fluid intelligence tasks: CFIT = Culture Fair Intelligence Test; numbers (e.g., 1-4) denote subtest number; RAPM-S = Raven's Advanced Progressive Matrices – Short form. The *M*'s for the letter, digit, and word updating tasks were based on regression-residual scores, calculated as the individual residuals obtained from a simple linear regression model predicting the short-CTI RTs from the long-CTI RTs. *SD* = standard deviation. *N* = 163.

Table 4.

Task Inter-correlations Between Removal Efficiency Scores and Fluid Intelligence Tasks and Measures Included in the Factor Models in Study 2.

Measure	1	2	3	4	5	6	7	8	9	10
1. RemL	-									
2. RemD	.40**	-								
3. RemW	.39**	.50**	-							
4. CFIT1	.09	.02	-.05	-						
5. CFIT2	.05	-.02	-.01	.17*	-					
6. CFIT3	-.09	-.08	.03	.20*	.34**	-				
7. CFIT4	.19*	.02	-.03	.07	.09	.08	-			
8. RAPM-S	-.10	-.24*	-.26*	.31**	.36**	.31**	.13	-		
9. CFIT1&2	-.08	.06	.05	.76**	.76**	.35**	.10	.46**	-	
10. CFIT3&4	-.07	.04	.00	.18*	.29**	.74**	.74**	.30**	.31**	-

Note. Removal tasks: RemL = letter updating; RemD = digit updating; RemW = word updating. Fluid intelligence tasks: CFIT = Culture Fair Intelligence Test; numbers (e.g., 1-4) denote subtest number; RAPM-S = Raven's Advanced Progressive Matrices – Short form. CFIT1&2 = average of the age-corrected regression residuals from subtests 1 and 2 of the CFIT; CFIT3&4 = average of the age-corrected regression residuals from subtests 3 and 4 of the CFIT. $N = 163$. $**p < .001$; $*p < .05$.

Table 5.

Descriptive Statistics of Performance on Removal, Working Memory Capacity, and Fluid Intelligence Tasks in Study 3.

Task	Mean	*SD*	Range	Skewness	Kurtosis
RemL	-0.00	0.13	-0.36 – 0.36	0.05	-0.11
RemD	-0.00	0.14	-0.32 – 0.53	0.62	0.68
RemW	0.00	0.15	-0.39 – 0.46	0.22	-0.19
OS	0.71	0.11	0.41 – 0.95	-0.45	-0.29
OS*pt*	0.91	0.07	0.60 – 1.00	-1.99	5.42
SS	0.69	0.13	0.32 – 0.99	-0.42	0.05
SS*pt*	0.91	0.05	0.70 – 0.99	-1.47	2.71
MU	0.68	0.16	0.23 – 1.00	-0.52	-0.55
CFIT1	0.60	0.14	0.15 – 0.92	-0.51	0.40
CFIT2	0.51	0.15	0.07 – 0.86	-0.13	-0.54
CFIT3	0.55	0.12	0.23 – 0.77	-0.44	-0.34
CFIT4	0.57	0.17	0.10 – 1.00	-0.28	0.05
RAPM-S	0.67	0.19	0.17 – 1.00	-0.88	-0.60

Note. Removal tasks: RemL = letter updating; RemD = digit updating; RemW = word updating. Working memory capacity tasks: OS = operation span; SS = sentence span; *pt* denotes secondary processing task; MU = memory updating. Fluid intelligence tasks: CFIT = Culture Fair Intelligence Test; numbers (e.g., 1-4) denote subtest number; RAPM-S = Raven's Advanced Progressive Matrices – Short form. The *M*'s for the letter, digit, and word updating tasks were based on regression-residual scores, calculated as the individual residuals obtained from a simple linear regression model predicting the short-CTI RTs from the long-CTI RTs. *SD* = standard deviation. $N = 204$.

Table 6.

Inter-correlations Between Removal Efficiency Scores, Working Memory Capacity, and Fluid Intelligence Tasks and Measures Included in the Factor Models in Study 3.

Measure	1	2	3	4	5	6	7	8	9	10	11	12	13
1. RemL	-												
2. RemD	.30**	-											
3. RemW	.37**	.36**	-										
4. OS	-.11	-.14	-.11	-									
5. SS	-.12	-.17*	-.11	.69**	-								
6. MU	-.20**`	-.30**`	-.18*	.56**	.54**	-							
7. CFIT1	-.12	-.21**	-.21**	.31**	.25**	.36**	-						
8. CFIT2	.03	-.06	-.01	.20**	.11	.18	.27**	-					
9. CFIT3	-.02	-.10	.02	.22**	.25**	.22**	.38**	.35**	-				
10. CFIT4	-.11	-.08	-.05	.17*	.26**	.24**	.16*	.30**	.25**	-			
11. RAPM-S	.11	-.06	-.03	.24**	.15*	.21**	.30**	.39**	.26**	.28**	-		
12. CFIT1&2	.07	17*	.14*	.31**	.23**	.29**	.79**	.79**	.44**	.30**	.43**	-	
13. CFIT3&4	.08	.12	.02	.24**	.32**	.29**	.33**	.40**	.79**	.79**	.34**	.47**	-

Note. Removal tasks: RemL = letter updating; RemD = digit updating; RemW = word updating. Working memory capacity tasks:

OS = operation span; SS = sentence span; MU = memory updating. Fluid intelligence tasks: CFIT = Culture Fair Intelligence Test; numbers (e.g., 1-4) denote subtest number; RAPM-S = Raven's Advanced Progressive Matrices – Short form. CFIT1&2 = average of the age-corrected regression residuals from subtests 1 and 2 of the CFIT; CFIT3&4 = average of the age-corrected regression residuals from subtests 3 and 4 of the CFIT. $N = 204$. **$p < .001$; *$p < .05$.
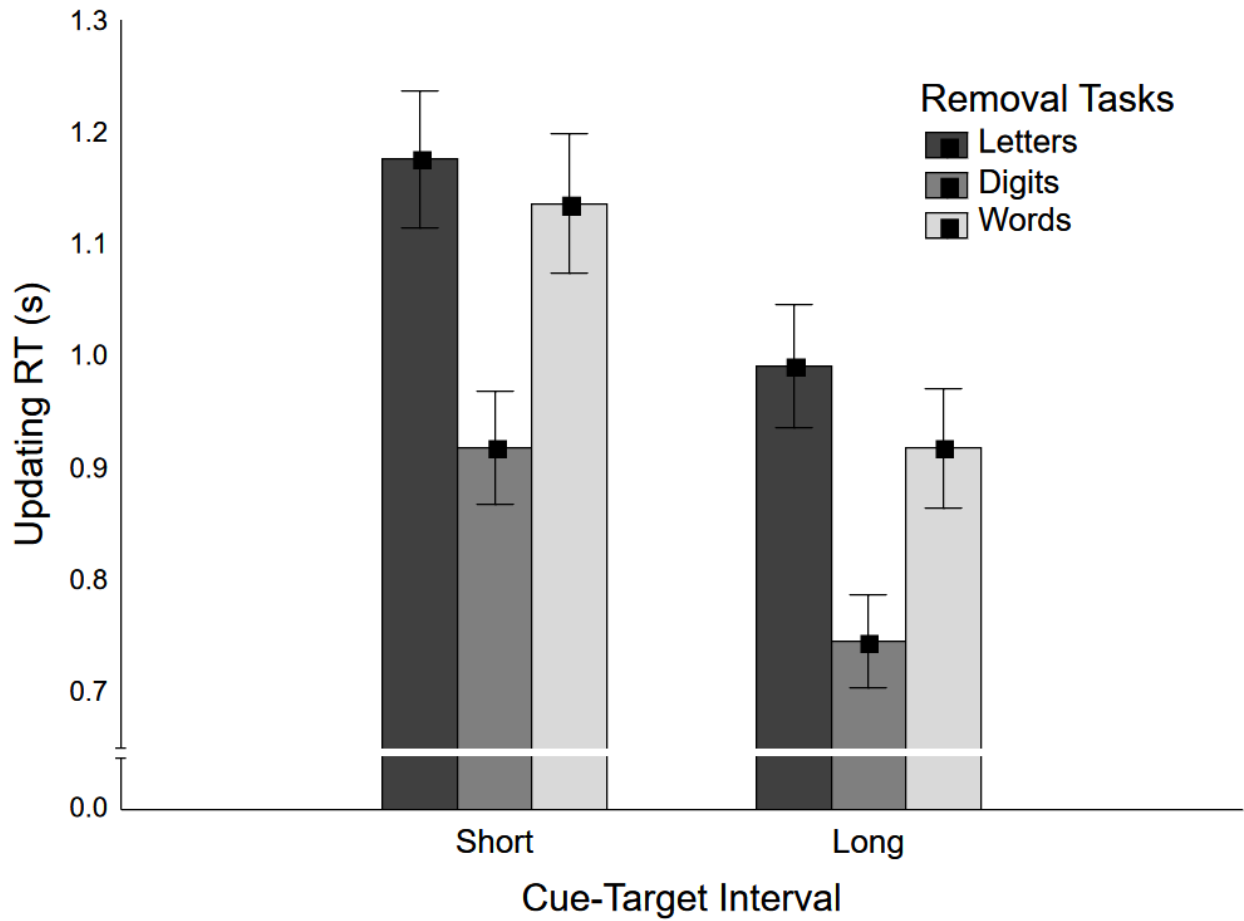
Figure 1.
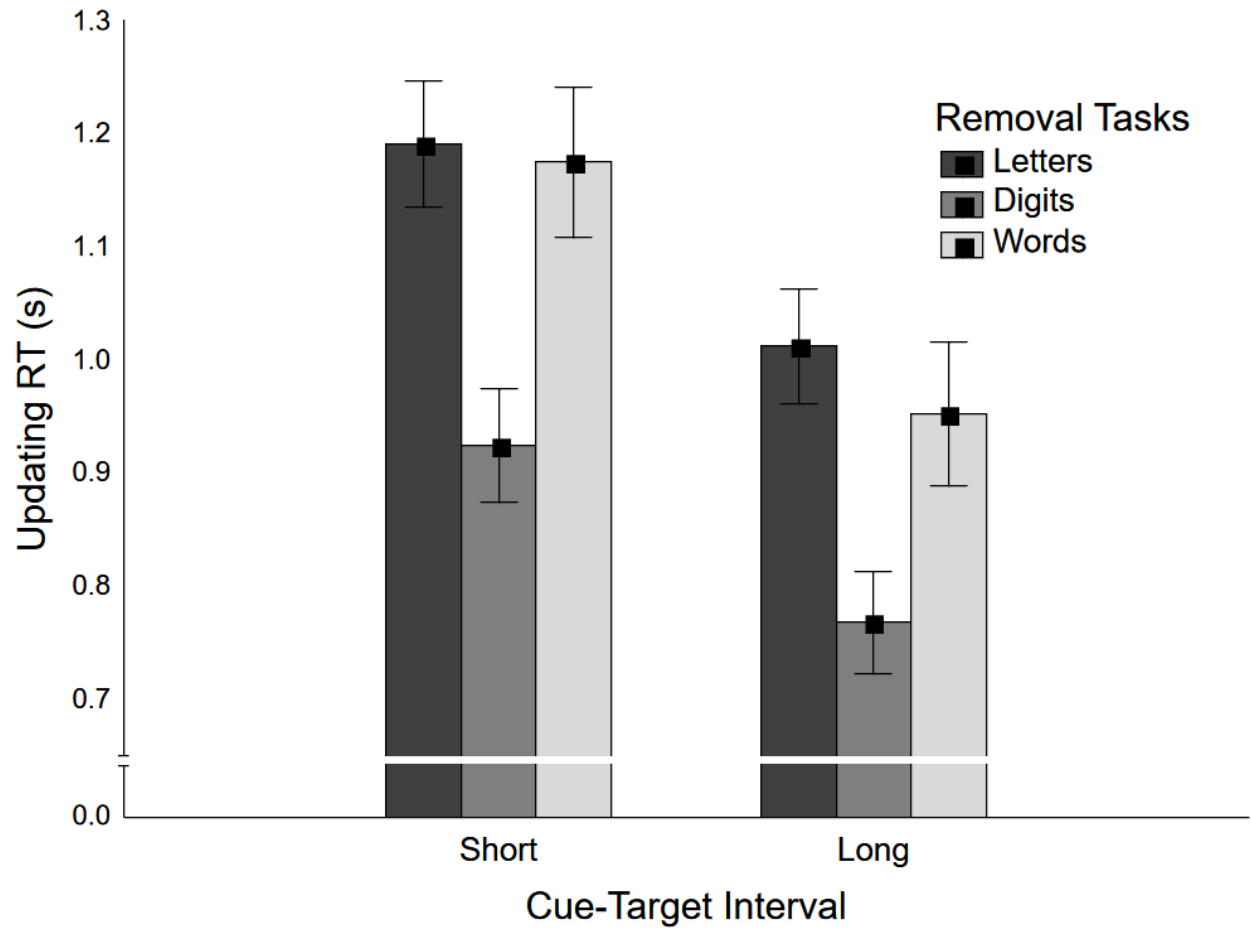
Figure 1. Updating response times for the three removal tasks per CTI condition for Study 1. Short CTI = 200ms; long CTI = 1500ms. Vertical bars denote within-subject standard errors of the mean.
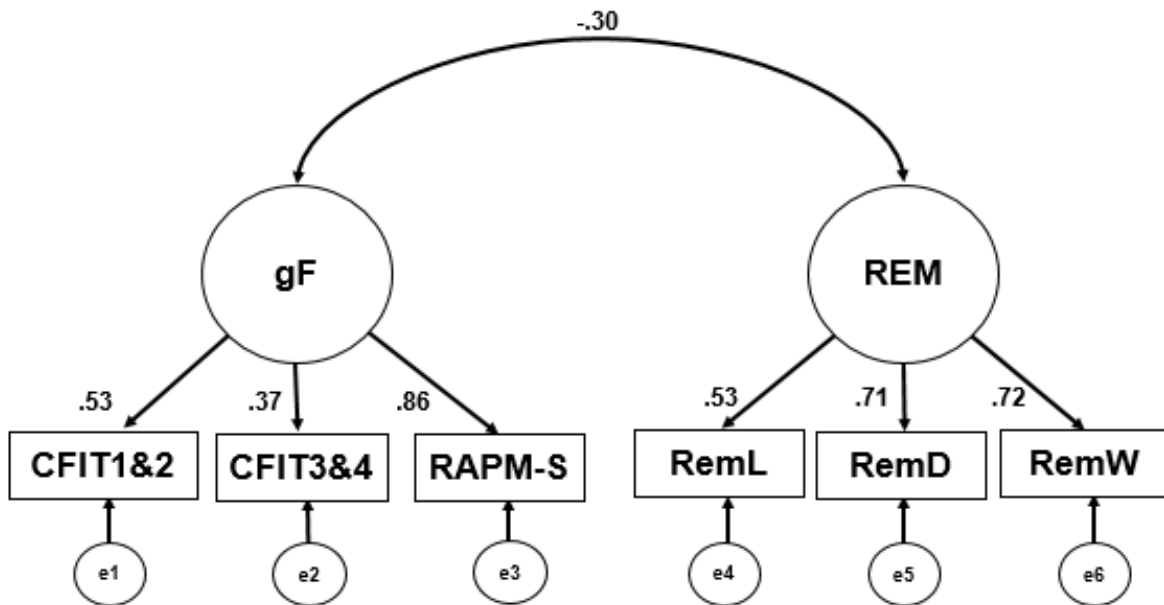
Figure 2.
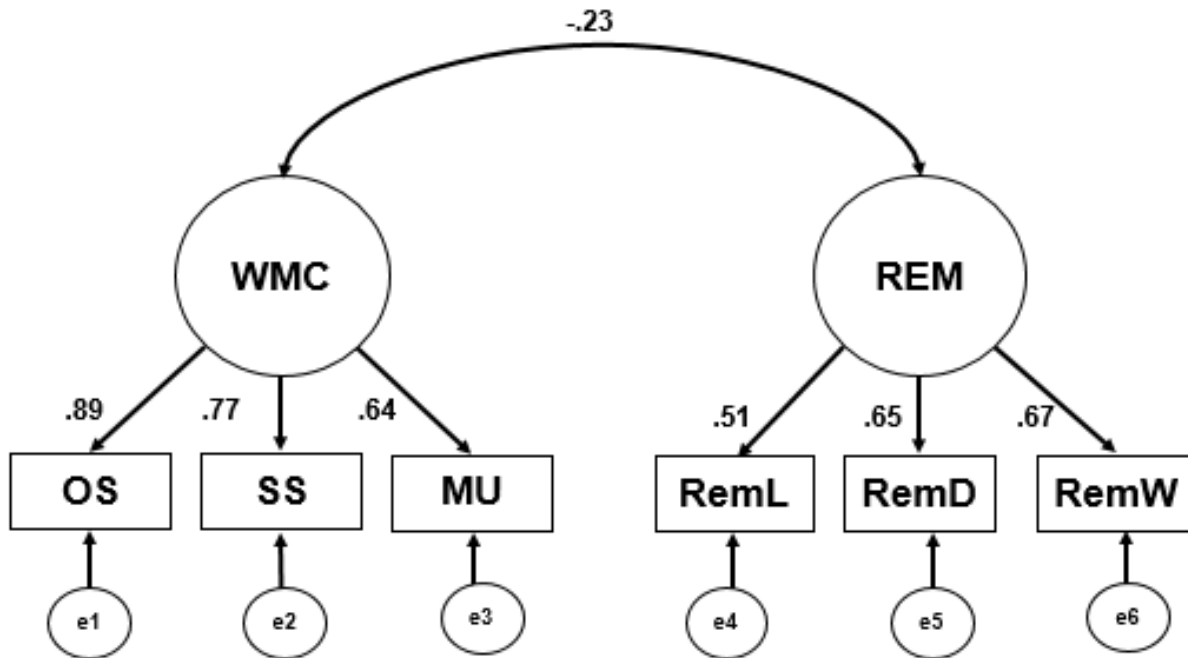
Figure 2. Graphical representation of the correlated two-factor model with standardized regression weights for Study 1, showing a significant relationship between WM capacity (WMC) and removal efficiency (REM). Observed variables related to WMC reflect mean performance in WM tasks that were calculated as regression residuals after controlling for age: OS (operation span), SS (sentence span), and MU (memory updating). Observed variables related to REM are the age-controlled regression-residual-based removal efficiency indices from the letter updating task (RemL), the digit updating task (RemD), and the word updating task (RemW). e1-e6 = error variables.
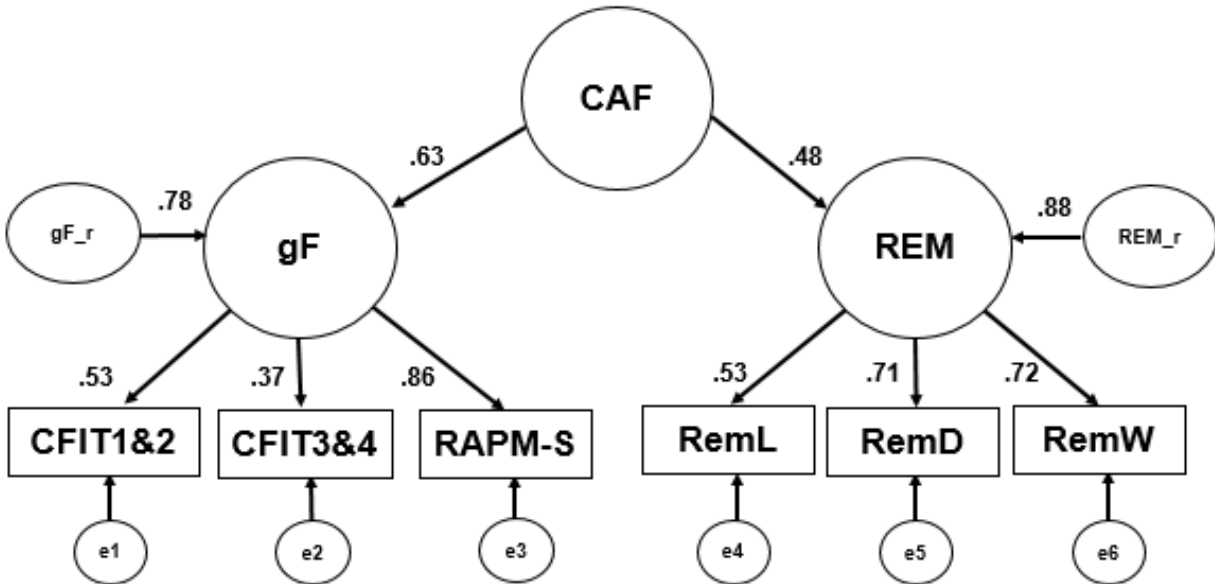
Figure 3.
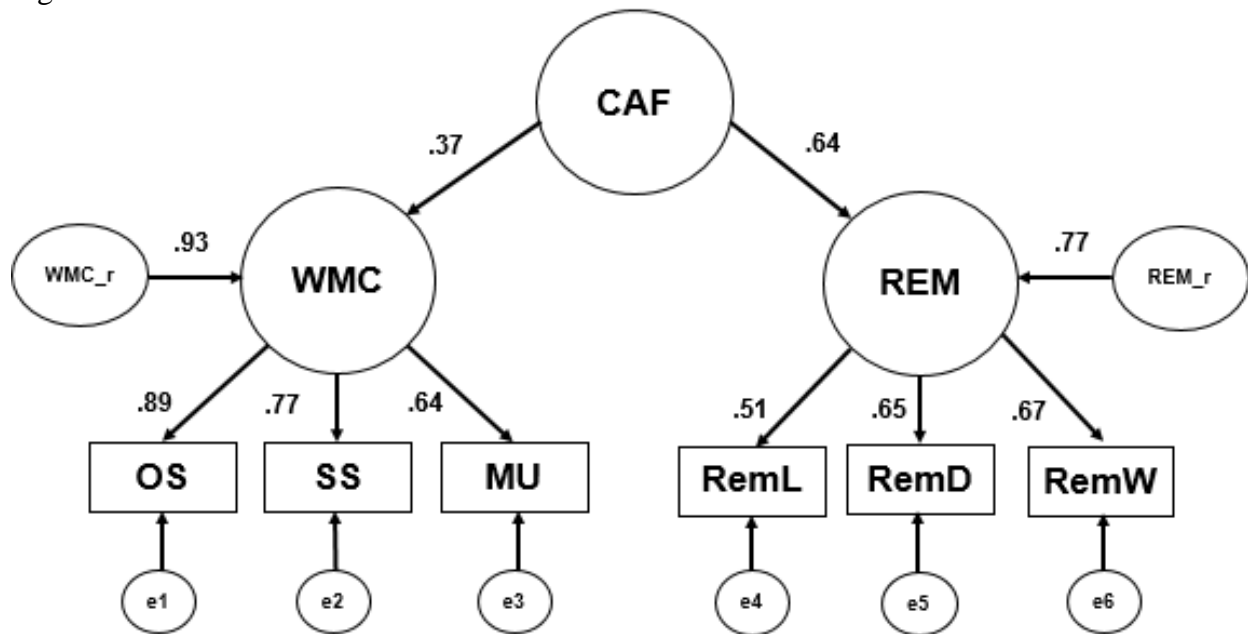
Figure 3. Graphical representation of the higher-order model with standardized regression weights for Study 1, showing the unique variance associated with each of the first-order latent factors, WM capacity (WMC) and removal efficiency (REM). For acronym spellings, consult notes in Tables 1-7. e1-e6 = error variables; CAF = common abilities factor.

Figure 4.

Figure 4. Updating response times for the three removal tasks per CTI condition for Study 2. Short CTI = 200ms; long CTI = 1500ms. Vertical bars denote within-subject standard errors of the mean.

Figure 5.

Figure 5. Graphical representation of the correlated two-factor model with standardized regression weights for Study 2, showing a significant relationship between fluid intelligence (gF) and removal efficiency (REM). Observed variables related to gF reflect mean performance in gF tasks that were calculated as regression residuals after controlling for age; four subtests of Cattell's (Cattell & Cattell, 1960) Culture Fair Intelligence Test combined into two observed variables (CFIT1&2 and CFIT3&4), and the short-form of the Raven's Advanced Progressive Matrices (RAPM-S; Arthur & Day, 1994). Observed variables related to REM are the age-controlled regression-residual-based removal efficiency indices from the letter updating task (RemL), the digit updating task (RemD), and the word updating task (RemW). e1-e6 = error variables.

Figure 6.

Figure 6. Graphical representation of the higher-order model with standardized regression weights for Study 2, showing the unique variance associated with each of the first-order latent factors, fluid intelligence (gF) and removal efficiency (REM). For acronym spellings, consult notes in Tables 1-7. e1-e6 = error variables; CAF = common abilities factor.
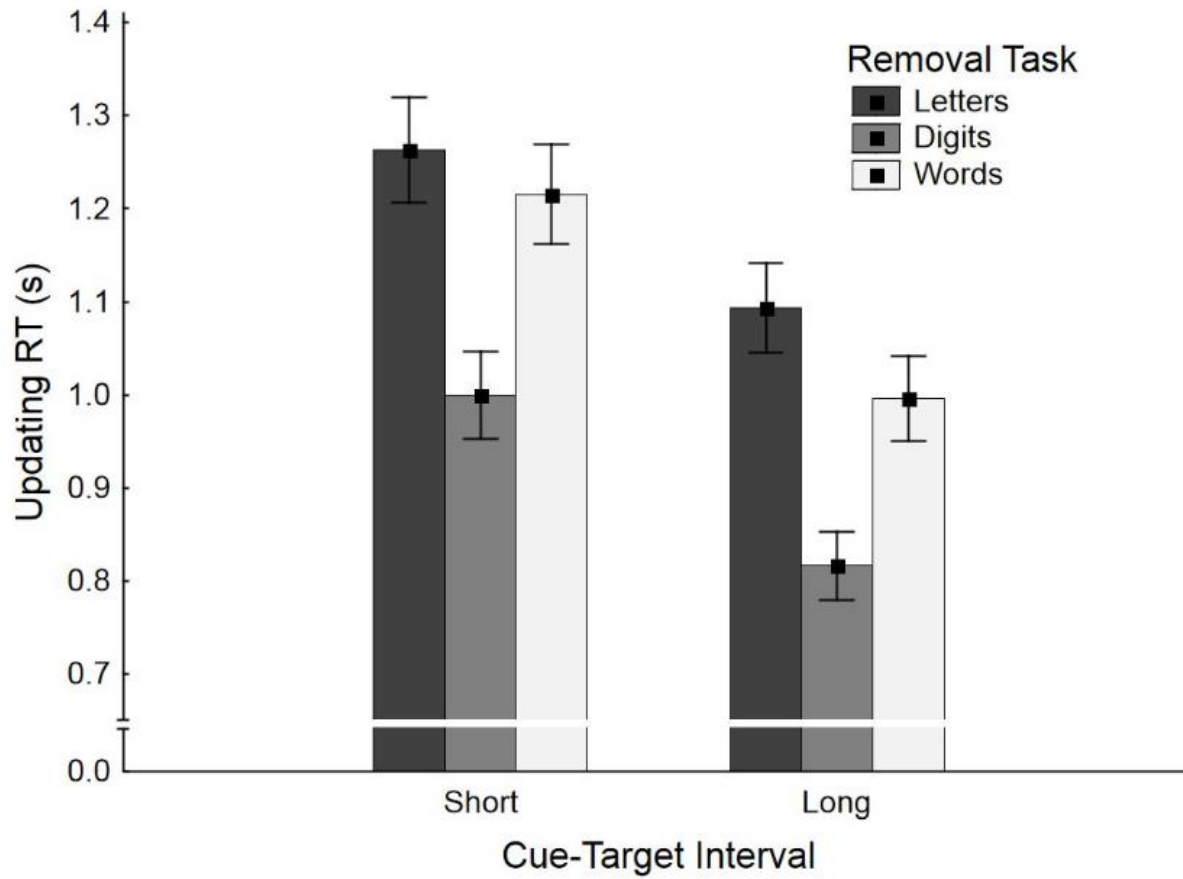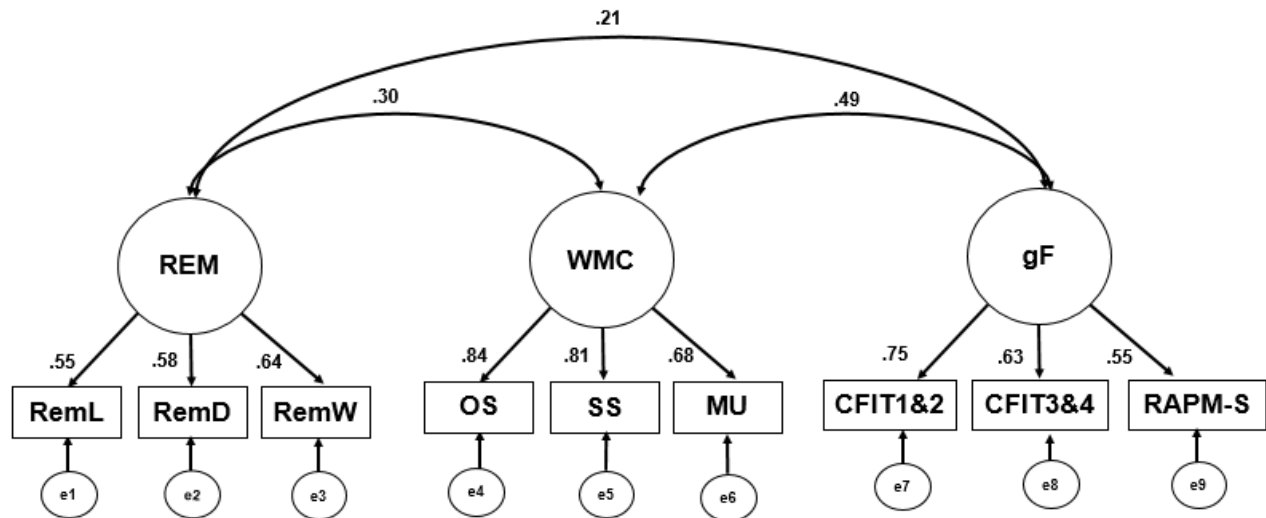
Figure 7.

Figure 7. Updating response times for the three removal tasks per CTI condition for Study 3. Short CTI = 200ms; long CTI = 1500ms. Vertical bars denote within-subject standard errors of the mean.

Figure 8.

Figure 8. Graphical representation of the correlated three-factor model with standardized regression weights for Study 3, showing significant inter-latent variable relationships between removal efficiency (REM) and WM capacity (WMC), removal efficiency (REM) and fluid intelligence (gF), and WM capacity (WMC) and fluid intelligence (gF). Observed variables related to REM are the age-controlled regression-residual-based removal efficiency indices from the letter (RemL), digit (RemD), and word (RemW) updating tasks. Observed variables related to WMC reflect mean performance in WM tasks that were calculated as regression residuals after controlling for age: OS (operation span), SS (sentence span), and MU (memory updating). Observed variables related to gF reflect mean performance in gF tasks that were calculated as regression residuals after controlling for age; four subtests of Cattell's (Cattell & Cattell, 1960) Culture Fair Intelligence test combined into two observed variables (CFIT1&2 and CFIT3&4), and the short-form of the Raven's Advanced Progressive Matrices (RAPM-S; Arthur & Day, 1994). e1-e9 = error variables.
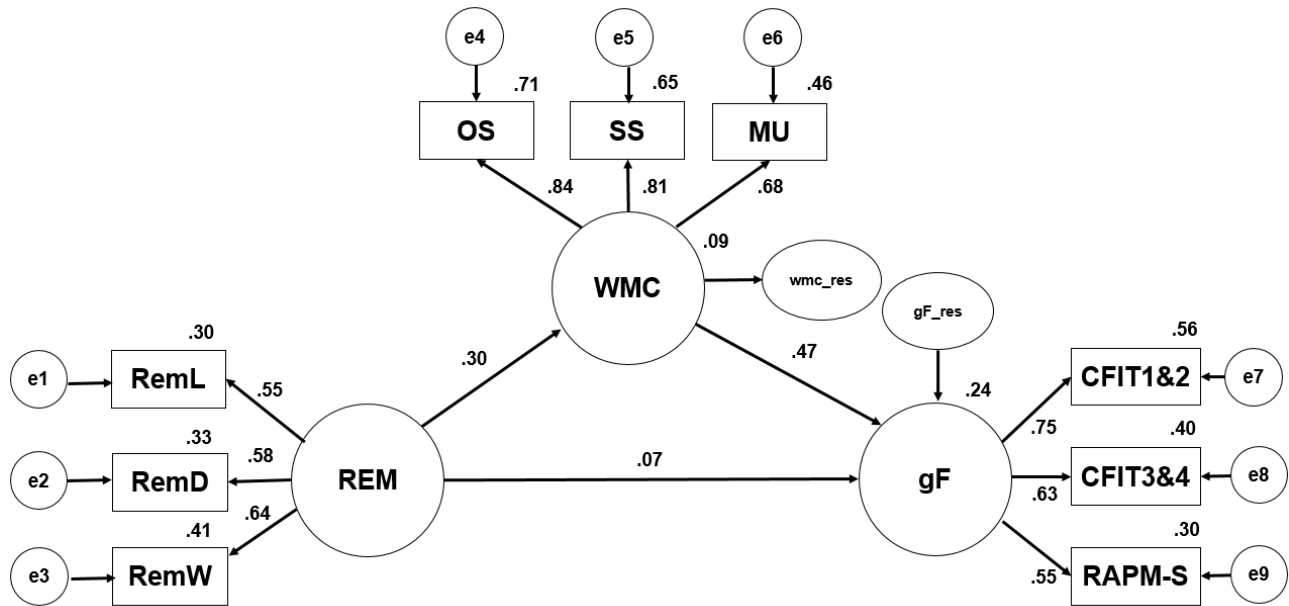
Figure 9.

Figure 9. Graphical representation of the key mediation model with standardized regression weights for Study 3. Removal efficiency (REM) is the predictor variable and fluid intelligence (gF) is the dependent variable, with WM capacity (WMC) as the mediating variable. wmc_res = residual error variance associated with the WMC latent variable; gF_res = residual error variance associated with the gF latent variable.
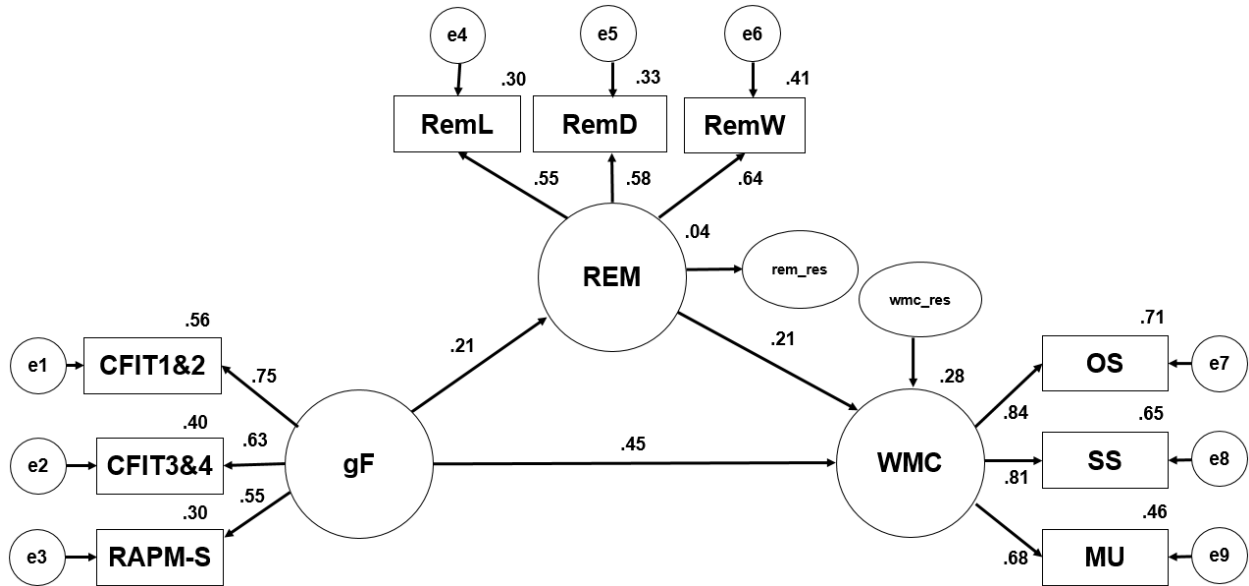
Figure 10.

Figure 10. Graphical representation of the first alternative mediation model with standardized regression weights for Study 3. Fluid intelligence (gF) is the predictor variable and WM capacity (WMC) is the dependent variable, with removal efficiency (REM) as the mediating variable. rem_res = residual error variance associated with the REM latent variable; wmc_res = residual error variance associated with the WMC latent variable.
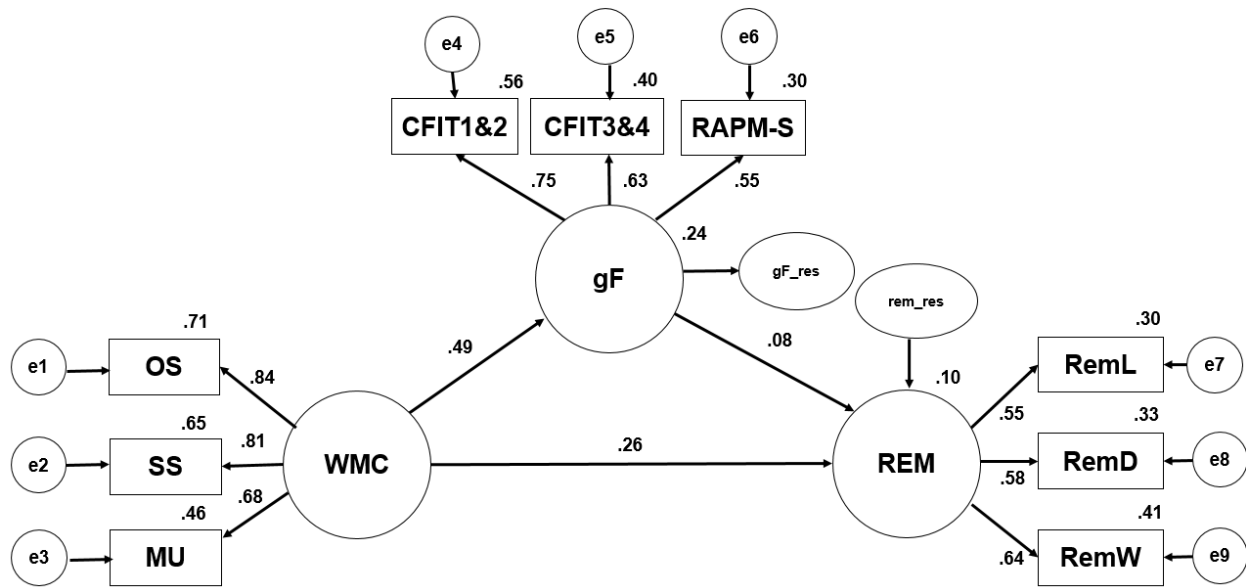
Figure 11.

Figure 11. Graphical representation of the second alternative mediation model with standardized regression weights for Study 3. WM capacity (WMC) is the predictor variable and removal efficiency (REM) is the dependent variable, with fluid intelligence (gF) as the mediating variable. gF_res = residual error variance associated with the gF latent variable; rem_res = residual error variance associated with the REM latent variable.