Limits of Near Transfer: Content- and Operation-Specific Effects of Working Memory Training

Samsad Afrin Himi[1,4], Matthias Stadler[2], Claudia C. von Bastian[3], Markus Bühner[4], and Sven Hilbert[5]

[1]Department of Psychology, Jagannath University, Bangladesh

[2]Empirical Pedagogy and Educational Psychology, Department of Psychology, Ludwig-Maximilians-Universität, Germany

[3]Department of Psychology, University of Sheffield, UK

[4]Psychological Methods and Assessment, Department of Psychology, Ludwig-Maximilians-Universität, Germany

[5]Educational Science and Sport Science, Faculty of Psychology, University of Regensburg, Regensburg

Author Note

The part of the data of this study was presented at the 51st Conference of German Psychological Association (DGP), Frankfurt am Main, Germany, 2018. We thank Natan Napiórkowski and Clemens Jilg for programming the working memory training task. We also thank Laura Scharl and Miriam Schirmer for their assistance with collecting the data from University of Regensburg. We also thank Dr. Felix Naumann and Dr. Felix Schönbrodt for providing constructive suggestions during the study. The data and analysis scripts are available on the Open Science Framework (https://osf.io/jwr9h/?view_only=21e95876f9a5423da3ab8917d1e3451d).

Correspondence concerning this article should be addressed to Samsad A. Himi, Department of Psychology, Jagannath University, Dhaka 1100, Bangladesh; and Psychological Methods and Assessment, Department of Psychology, Ludwig-Maximilians-Universität München, Leopoldstraße 13, 80802 Munich, Germany. E-mail address: samsad@psy.jnu.ac.bd

Abstract

Working memory (WM) training typically leads to large performance gains in the practiced tasks, but transfer of these gains to other contexts is elusive. One possible explanation for the inconsistent findings of past research is that transfer may only occur when cognitive strategies acquired during training can also be applied in the transfer tasks. Therefore, we systematically varied the content domains and WM operations assessed by training and transfer tasks and, thereby, the extent to which similar cognitive strategies could be applied. We randomly assigned 171 young adults to 1 of 8 experimental groups who trained 1 of 2 working memory operations (storage and processing or relational integration) with materials from 1 of 4 content domains (verbal, numerical, figural-icon, or figural-pattern), to an active or to a passive control group. Before and after 12 sessions of adaptive training within 2-3 weeks, performance was assessed in all eight WM tasks. Bayesian generalized-mixed effects models revealed improved performance in the trained tasks compared to the active control group. However, these improvements did not generalize to tasks measuring the same WM operation with different materials. Moreover, the comparison of the training groups with an active and a passive control group showed considerable differences, thus highlighting the importance of distinguishing between active and passive control. Overall, the findings revealed no evidence for transfer between tasks assumed to afford the same strategies. Therefore, the adoption of specific cognitive strategies alone is unlikely to be responsible for transfer of WM training gains between tasks.

Keywords: working memory training, transfer effect, storage and processing, relational integration

Limits of Near Transfer: Content- and Operation-Specific Effects of Working Memory Training

Working memory (WM), the ability to maintain limited information in the face of interference, is a strong predictor of other higher cognitive functions and real-world behaviors (Barrett et al., 2004). Based on the assumption that the correlations between WM and other cognitive abilities reflect functional overlap, it has been hypothesized that training WM can improve not only WM performance but also complex cognition more broadly (e.g., Jaeggi et al., 2008; Klingberg et al., 2002; for a review see Schwaighofer et al., 2015). However, a growing body of studies in children (e.g., Ang et al., 2015), young adults (e.g., Clark et al., 2017; Foster et al., 2017; Harrison et al., 2013; Minear et al., 2016; Redick et al., 2013), and older adults (e.g., von Bastian et al., 2013) shows that the typically large gains in the trained WM tasks rarely generalize to untrained tasks measuring different yet related constructs (i.e., far transfer), and sometimes not even to untrained WM tasks assessing other WM content domains or operations (near transfer; e.g., De Simoni & von Bastian, 2018; Guye & von Bastian, 2017; Soveri et al., 2017; Sprenger et al., 2013). It is essential to note that the WM training literature may suffer from publication bias – the positive transfer effects found in the previous studies could be overestimated (e.g., Simons et al., 2016).

However, one possible explanation for the limited transfer of improved performance in the trained tasks to other contexts is that trainees acquire strategies during training that are difficult to apply in other contexts. Hence, transfer might occur only if training and transfer tasks afford the same strategies (see also Gathercole et al., 2019). However, the extent to which transfer of training is task-specific (i.e., material-dependent) or process-specific (i.e., material-independent) is yet unclear. For example, Ericsson et al. (1980) showed that training with numbers did not improve the recall of letters, but Hilbert et al. (2014) found transfer between mirror-reversed letters and mirror-reversed numbers. The present study investigated the extent to which transfer is material-dependent by systematically varying the WM content domains and operations assessed by the training and transfer tasks.

The Facet Model of Working Memory

The present study design is based on the facet model of WM (Oberauer et al., 2003) positing that WM capacity comprises two facets: content domains and operations. Each facet is further split into three categories. The content facet contains verbal, numerical, and figural-spatial WM; the operation facet distinguishes the three WM functions storage and processing, relational integration, and supervision. Storage and processing reflects the maintenance of briefly presented new information over a short period of time when simultaneously processing information. Relational integration refers to the ability to build new relations between information elements and integrate them into structures. Supervision corresponds to the shifting factor in Miyake et al.'s (2000) model of executive functions. It involves the activation of relevant goal representations, suppression of distraction, and switching between task sets. Whereas storage and processing and relational integration are consistently found to be highly correlated due to taxing attention control (i.e., ignoring irrelevant information; Himi et al., 2019), supervision is often only weakly related to the other two operations (e.g., Bühner et al., 2005; Hilbert et al., 2017, Oberauer et al., 2003; Oberauer et al., 2008; von Bastian & Oberauer, 2013). Therefore, following the rationale of functional overlap, transfer is most likely to occur between storage and processing and relational integration and, thus, the present study focuses on these two operations.

Strategy Use in Working Memory Training

Similar to the general body of WM training literature, past training studies based on the facet model of WM have also reported mixed evidence for near and far transfer after storage and processing and relational integration training (Lange & Süß, 2015; von Bastian et al., 2013; von Bastian & Oberauer, 2013). For example, in von Bastian and Oberauer's

(2013) study, large gains observed in the trained tasks transferred to distant measures of reasoning but not to untrained measures assessing the same WM function. Similarly, Hilbert and colleagues (2017) observed transfer from verbal to numerical material and vice versa (within WM operations) but not between verbal/numerical and figural material. Von Bastian and Oberauer (2013) as well as Hilbert et al. (2017) speculated that trainees might have acquired task-specific strategies during training that were not applicable to the structurally different near transfer tasks. More precisely, Hilbert et al. assumed that this could be attributed to verbalization strategies acquired in the verbal and numerical tasks, which are not useful for solving the figural tasks. Indeed, there is evidence for content domain-specific differences in strategy use. Specifically, it has been shown that training can yield a significant increase in verbal strategy use when performing untrained verbal tasks (Dunning & Holmes, 2014). Likewise, visual strategies have been found to be effective in enhancing performance in visuospatial WM tasks (Stieff et al., 2020). Furthermore, serial order encoding of verbal or spatial information in WM is not domain-general (Ginsburg et al., 2017, Experiments 1, 2, & 3; Zimmermann et al., 2016). More recently, Forsberg et al. (2020) showed that instructing WM trainees with visualization strategies can considerably boost WM training performance in the absence of transfer to structurally different transfer tasks (see also Laine et al., 2018).

Conversely, if training and transfer tasks afford similar strategies, transfer may be observed. For example, Comblain (1994) showed that training a rehearsal strategy with verbal materials (e.g., pictures of a noun) can improve performance in numerical (digit) and verbal (letter) memory span task performance. Hence, transfer may not be driven by functional overlap but overlap in cognitive strategies and routines that can be applied (see also Gathercole et al., 2019; Norris et al., 2019). Critically, this implies that transfer observed does not reflect increased WM capacity (i.e., the overall amount of information that someone can access and process in the present moment) but instead enhances WM efficiency by making it easier to handle the cognitive load using the pre-existing WM capacity (von Bastian & Oberauer, 2014; von Bastian et al., 2022). This efficiency may well be due to the application of the cognitive strategy used in a given context, meaning how information is processed internally. However, not all cognitive strategies (grouping, sentence generation, mental imagery, or rote repetition) are effective to the same extent (Bailey et al., 2014; Dunlosky & Kane, 2007), and people might not use the same strategy at all times (Morrison et al., 2016). Moreover, Hilbert et al. (2015) showed that individuals relying on verbal processing strategies can be distinguished from those relying on visual strategies in WM tasks. Evidence from the neuroimaging results suggests that the changes occur following training in the middle frontal gyrus (Olesen et al., 2004), which is also responsible for verbal cognitive strategy (Hilbert et al., 2015), corroborating the presumed relation between cognitive strategies and training-related effects.

Gaining a better understanding of how strategy acquisition impacts patterns of transfer requires a task-analytic training procedure that isolates the cognitive processes that are trained and improved (see also Gathercole et al., 2019; Taatgen et al., 2013). So far, few studies have directly ascertained transfer of training with one particular material to other WM tasks with different materials (e.g., Fellman et al., 2020; Linares et al., 2019). For example, Hilbert et al. (2017) systematically varied the stimulus content domains using verbal, numerical, and figural materials for training the two WM functions storage and processing and relational integration. The results indicated transfer between verbal and numerical but not figural materials in tasks measuring the same WM function. Hilbert et al. suggested that processing letters and numbers equally affords silent rehearsal, a strategy frequently used to remember stimuli in WM tasks. However, silent rehearsal is not applicable to all types of visually presented stimuli, as they first need to be recoded into verbal information (see also Hilbert et al., 2015). To further test this proposition, the present study set out to replicate and extend Hilbert et al.'s (2017) study by including a newly developed figural WM task that comprises stimuli that can be easily coded verbally and thus should facilitate silent rehearsal.

The Current Study

The main goal of the current investigation was to determine why WM training leads to material-specific transfer but not to transfer effects on structurally different tasks. We hypothesized that transfer of training occurs if training and transfer tasks involve the same cognitive strategies. For this purpose, we developed a process-based task-specific training regimen and examined the role of strategy by contrasting performance in tasks affording a verbal strategy to tasks that do not. Specifically, we expected that using a verbal strategy would be associated with transfer to tasks comprising easily verbalizable stimuli such as words, numbers, and icons but not to tasks with stimuli that are difficult to verbalize such as figural patterns. Building on Hilbert and colleagues' (2017) study design, we selected training tasks measuring the two WM operations storage and processing and relational integration and extended the content domains by distinguishing between figural materials that allow for applying a verbal strategy (icons) and figural materials that are difficult to verbalize (patterns) and, therefore, are unlikely to evoke the same cognitive strategy. Moreover, this study incorporated both an active and a passive control group, and an adaptive training paradigm to minimize the methodological limitations of Hilbert et al.'s (2017) study.

In addition to testing for transfer between content domains, we also investigated transfer between training the two WM operations storage and processing and relational integration. Notably, however, neither Hilbert et al. (2017) nor von Bastian and Oberauer (2013) observed such effects. The present study provides a further test of this hypothesis. This also allowed for exploring how the newly developed figural-icon tasks related to other, established WM tasks. Moreover, much of the inconsistencies of findings from previous training studies has been attributed to methodological differences such as the type of control group (Shipstead et al., 2012; von Bastian & Oberauer, 2014; but see Au et al., 2020). Previous studies demonstrating transfer effects are often criticized for including only passive control groups (Dougherty et al., 2016). Specifically, whereas a passive control group minimizes confounding training-related improvements with the effects of multiple testing, an active control group additionally controls for non-specific effects of training such as expectancy effects. To estimate the effect of the type of control group, we included a passive as well as an active control group. Finally, the present study uses a robust methodology and well-defined pretest/posttest study design by including both active control and passive control groups, theory-based task selection, and random assignment of the participants to groups.

Methods

The study protocol was preregistered during the process of data collection but before data had been analyzed (https://osf.io/jwr9h/?view_only=21e95876f9a5423da3ab8917d1e3451d). The preregistration was late due to a miscommunication within the lab yet contains all original hypotheses.

Participants

Participants were 171 university students (74.9% female, all others male) recruited from the Ludwig Maximilians-University of Munich (n = 146) and the University of Regensburg (n = 25), Germany. According to an a priori power analysis, our original plan was to collect data from 200 participants, as stated in our preregistration. However, the COVID-19 pandemic forced us to stop data collection early. The median participant age was 22.0 years (1st quartile: 19.0 years; 3rd quartile: 25.0 years). About half of the participants (59.6%) were psychology students. Before training, participants were randomly assigned to 1 of 10 possible groups: storage and processing verbal, numerical, figural-icon, or figural-pattern, or relational integration verbal, numerical, figural-pattern, or figural-icon, or active control, or passive control. Initially, 222 students participated in the pre-test, but 48 did not proceed with the study due to problems installing the training software. Three other participants did not complete the training sessions. The final sample of 171 participants completed all sessions.

All participants provided written informed consent before the pre-test. APA Ethical Principles for human research were followed, and anonymity and confidentiality were maintained. The study was double-blinded, that is, neither the participants nor the experimenter were informed about the group assignment. Participants were informed that they would be assessed in different activities concerning their cognitive functioning. All participants reported normal or corrected-to-normal vision and no neurological or diabetic problems. Participants were compensated with either €35 or course credit. In addition, in order to motivate participants to practice the training task seriously at home, a monetary reward for good performance was promised.

Procedure

Participants were tested in groups of up to five people in a university laboratory. Cognitive assessment was conducted in two sessions (pre- and post-test) on separate days within approximately three weeks, and each session lasted about 1.5 hours (including a 5-min break). All tasks were administered in the same order across participants to minimize subject-by-treatment interactions. During the pre-test session, first all storage and processing tasks were administered (verbal, numerical, figural-icon, figural-pattern), followed by the four relational integration tasks (verbal, numerical, figural-icon, figural-pattern). In the post-test session, these tasks were administered in reverse order to minimize sequence effects. Between the pre- and post-test sessions, participants of the WM training groups and the active control group completed their respective training intervention at home for 20 min on each of 12 days within 2 to 3 weeks. The passive group did not receive any training. All tasks were administered in German.

Pre- and Post-Test Assessment

Participants completed the WM tasks (adapted from Oberauer et al., 2003; von Bastian & Oberauer, 2013), written in Python 2.7 (see https://www.python.org). A standard computer keyboard registered manual responses. We used the same WM tasks for pre- and post-tests. All tasks were used in previous work except for the storage and processing and relational integration figural-icon tasks (see Figure 1). The icons were pronounceable as words with one or two syllables in German. In the pre-post test session participants were asked which strategy they used to perform each task.

Storage and Processing

Participants completed four storage and processing tasks assessing the maintenance of briefly presented information in the face of distraction. These tasks were similar to the Brown-Peterson paradigm (Brown, 1958) with verbal, numerical, figural-icon, and figural-pattern materials, respectively. Participants were first presented a sequence of 3 to 7 words, 4 to 8 numbers, 3 to 7 icons, or 2 to 4 patterns that they were asked to memorize. Following these sequences, participants completed a processing task comprising materials of the same content domain. In the verbal version, the processing task was to categorize words as a city or a country. In the numerical version, numbers were to be classified as odd or even. In both figural tasks, participants had to decide whether arrows pointed upward or downward. After 5 s, participants were asked to recall the memoranda in the same order as they were initially presented. Each of the tasks comprised 15 test trials and 2 practice trials, taking about 12 to 15 min to complete. The mean proportion of correctly recalled elements in each trial (i.e., partial-credit score; cf. Conway et al., 2005) was considered as dependent measure.

Relational Integration

The relational integration tasks required participants to detect a critical relation by integrating single information elements. As for the storage and processing tasks, participants completed four relational integration tasks with verbal, numerical, figural-icon, and figural-pattern materials. For the verbal version, nine words in a 3 × 3 matrix were displayed and one word randomly changed every 2000 ms. Participants were asked to respond when three rhyming words were shown either in a row, column, or diagonal within the matrix. Participants completed 111 test trials and 12 practice trials. The numerical version presented nine three-digit numbers in a 3 × 3 matrix in which one of the numbers was randomly replaced every 2000 ms. Participants had to respond when three identical last digits appeared either in a row, column, or diagonal. The task comprised 112 test trials and14 practice trials. For the figural-icon version, nine sets of icons, with each set consisting of three icons, were presented in a 3 × 3 matrix. One of the sets of icons was randomly replaced every 2000 ms. Participants had to respond when three identical middle icons appeared either in a row, column, or diagonal. Participants completed 112 test trials and 14 practice trials. In the figural-pattern version, 20 black dots were presented in a 10 × 10 matrix, with two black dots changing their location every 2000 ms. Participants were asked to respond when four black dots formed a square. Participants completed 115 test trials and 14 practice trials. The tasks took about 6 min. each. The dependent variable was the discriminability index (d'), reflecting the sensitivity of target detection. It is computed by relating the hit rate and false alarm rate ($d' = z$ (hit rate) $- z$ (false alarm rate)), where z indicates standardized score. In order to correct perfect hit and false alarm rates for the discriminability index we used 1/2N (N = number of false alarms) instead of false alarm rate of 0, and 1- 1/2N (N = number of targets) instead of hit rate of 1.

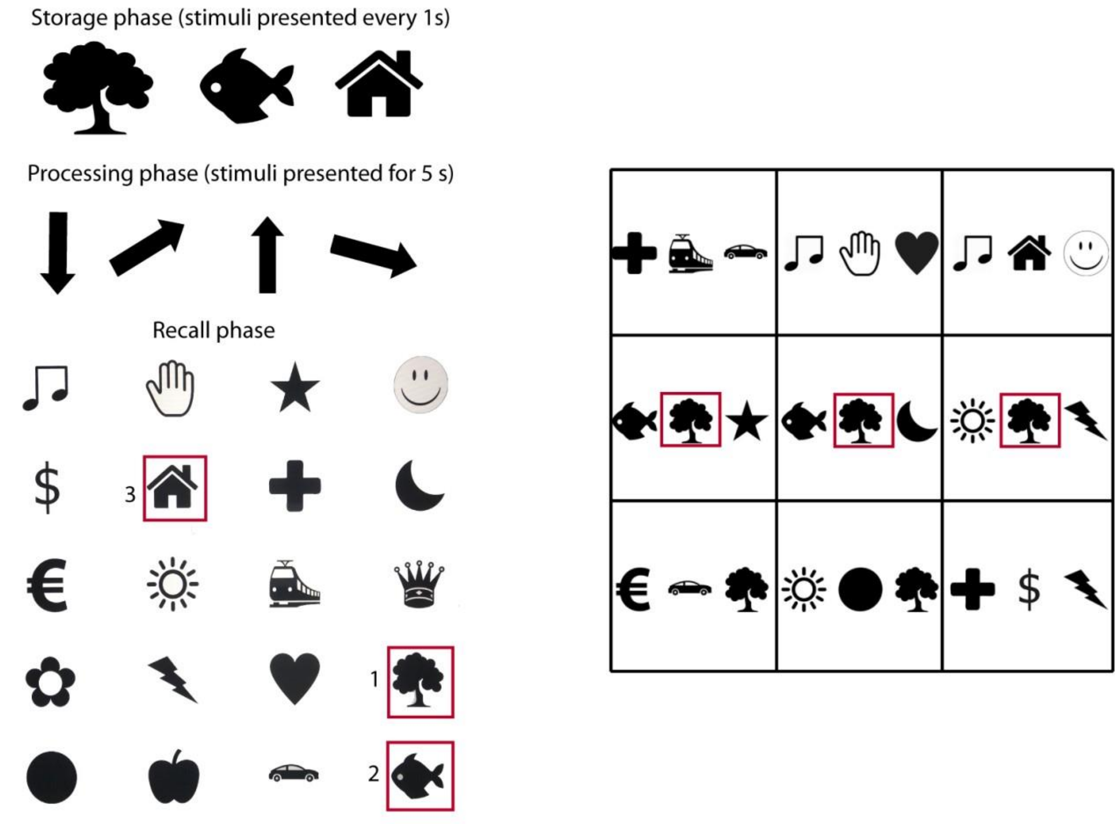
Cognitive Strategies Questionnaires

A cognitive strategies questionnaire was presented at both pre- and post-test. Participants made a forced-choice of whether they primarily used visualization or verbalization in doing storage and processing and relational integration tasks. They were also required to indicate which cognitive strategies they used to complete each of the storage and processing and relational integration tasks: verbalization, visualization, both, a different strategy, or no strategy at all (for details see Himi et al., 2022).[1]

[1] The detailed analysis of cognitive strategies questionnaire was not included in the present study because of the manifold analyses and the reduced sample size.

Figure 1

Figural-Icon WM Tasks Used in the Pre-/Post-Test and Training Sessions.



(a) Storage and processing figural-icon task (b) Relational integration figural-icon task

Note. (a) An example item for the storage and processing figural-icon task in which participants had to remember the icons while judging whether the arrow pointed upwards and downwards. (b) An example item for the relational integration figural-icon task in which participants had to respond when three identical middle icons appeared either in a row, column, or diagonal. The highlighted red boxes are for illustration purposes only.

Working Memory Training

Participants in each WM training group practiced one of the WM operations (storage and processing or relational integration) from one content domain for 12 days, with each session lasting 20 min. Thus, the total training dose was 4 hours spread over 2 to 3 weeks (with a mean of 15.85 days). We developed the WM training tasks based on the tasks of Oberauer et al. (2003) and von Bastian and Oberauer (2013). Whereas the materials were the same as in the pre-/post-test tasks described above, the specific stimuli of the training tasks were different from the stimuli at pre- and post-test to minimize recognition effects. The self-administered training was conducted at home via a Python 2.7 (https://www.python.org) based on a freely accessible online platform hosted by the Leibniz-Rechenzentrum der Bayerischen Akademie der Wissenschaften (LRZ; English: Leibniz Supercomputing Center of the Bavarian Academy of Sciences and Humanities). The training program could be installed on Windows and Mac OS and, thus, was reasonably platform-independent, and could be used at home.

Data produced during training were automatically uploaded and saved on a remote LRZ server and exported as comma-separated raw data (.csv) files. Every time a participant started a training session, the data and settings (i.e., screen resolution, operating system, and time of access) were updated to verify the accuracy of the data with a Hash function. Performance-based feedback was given after each session. The scoring procedures were identical to the ones used for pre- and post-test WM tasks.

Adaptive Algorithm

We used a procedure similar to the adaptive training algorithm used by von Bastian and Oberauer (2013). In the first session, an individual benchmark was established based on the initial performance of the participant (first 40% of trials). Performance was continuously checked after each 40% of trials. If a participant outperformed the benchmark after 40% of trials, difficulty was increased; otherwise, it remained at the same level and the participant had three retries to exceed the benchmark. If they still did not succeed, task difficulty was reset to the initial difficulty level, and a new benchmark was set (this deviates from von Bastian & Oberauer, 2013). The benchmark was restricted to fall between 75% to 95% accuracy to avoid individual benchmarks being too low or too high. Each session started with the same level of difficulty that a participant has attained in the previous session.

In the storage and processing task, difficulty could be titrated either by increasing the processing time duration (5s, 10s, and 15s) or by increasing the number of memoranda (from 3 to 10 for verbal, 4 to 11 for numerical, 3 to 10 for figural-icon, and 2 to 5 for figural-pattern) one at a time. Difficulty was adjusted to individual performance by alternating between these two parameters. Specifically, if a participant outperformed the benchmark with an already increased processing duration, the number of elements was increased and vice versa. If participants' accuracy was below 75% or above 95%, task difficulty was adjusted according to the most recent change. For example, if the most recent increase in difficulty was an increase in the number of memoranda, then the processing duration was increased next. If, after this change, accuracy dropped below 75%, the most recent change was reversed – in this example, the algorithm would reverse the processing duration to its previous level.

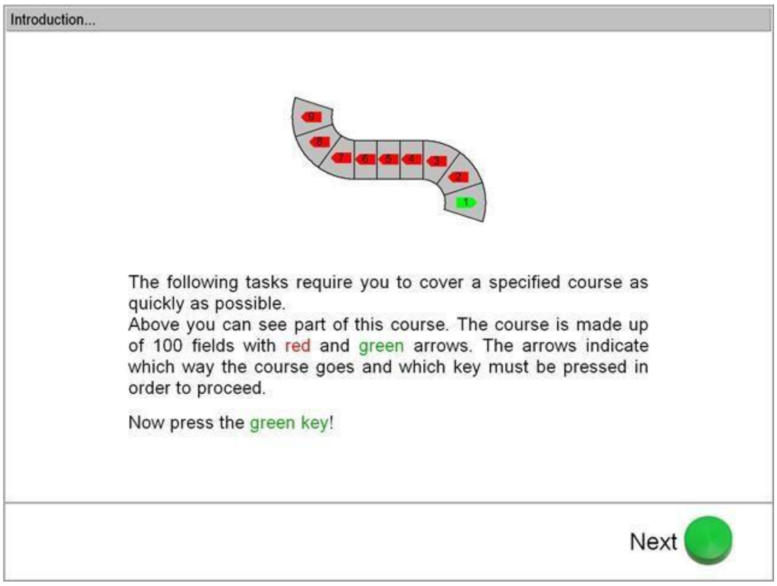
In the relational integration tasks, the difficulty level was adjusted across trials either by decreasing the time interval between changing elements (2.0 s, 1.5 s, 1.0 s, and 0.75 s) or by increasing the number of changing elements (e.g., from 1 to 3 for verbal, numerical, and figural-icon; from 2 to 5 for figural-pattern) one at a time. As for the storage and processing tasks, these two parameters were alternated to adjust the task difficulty. The difficulty level of the training tasks was established by following Hilbert et al. (2017).

Active Control Group Training

Like the WM training groups, the active control group completed 12 sessions (20 min. each). During these sessions, participants completed the Objektiver Leistungsmotivations-Test (OLMT; English: objective achievement motivation test; Schmidt-Atzert, 2004), in which they had to follow a specified route as quickly as possible by pressing the left and right 'shift' button (Figure 2). Each route contained up to 100 fields with red and green arrows. The arrows indicated the direction of the route and, thus, the key that had to be pressed in order to proceed. Green arrows pointed to the right direction, and red arrows pointed to the left. This test is assumed to have only minor WM demands, as evidenced by a non-significant correlation with WM (see test manual; Schmidt-Alzert, 2004). The OLMT comprised three subtests targeting different aspects of motivation. First, for targeting task-related effort, participants had to make as many moves as possible within 10 s. Second, to practice motivation arising from setting personal goals, participants were asked to set and achieve movement targets. Finally, to trigger motivation arising from competition, participants had to outperform a virtual opponent. Each subtest was made up of 10 identical runs that lasted for 10 s. The length of sequence covered by pressing the buttons in the last three runs of the first subtest was the dependent variable. Participants received performance feedback after each run.

Figure 2

Screenshot of the English Version of OLMT Training Platform.



Note. In the OLMT task, participants had to follow a specified route as quickly as possible by pressing the left and right 'shift' button. Green arrows pointed to the right direction, and red arrows pointed to the left.

Statistical Analyses

All analyses were conducted using the statistical programming language R (R Development Core Team, 2015). First, data were preprocessed with the "tidyverse" package

(Wickham, 2021). Second, we confirmed the structure of the facet model of WM using the 'blavaan' package (Merkel et al., 2020) to perform a Bayesian latent-variable analysis on the pre-test and post-test scores. Third, for each WM task, we applied Bayesian generalized linear mixed-effects models using the "blme" package (Dorie, 2015) to assess performance improvement after training. Fourth, the "brms" package (Bürkner et al., 2021) was used to compute Bayes factors (BF) with flat prior settings (normal 0, 1). The 'hypothesis' function of this package computes an evidence ratio (i.e., the BF in favor of H_0) for each hypothesis. BFs range on a continuous scale from 0 to ∞, with a BF of 1 reflecting no evidence. BFs below 1 represent evidence for the null hypothesis, and BFs above 1 indicate evidence in support of the alternative hypothesis (for conventions for the interpretation of the size of BFs, see Wetzels & Wagenmakers, 2012). All figures were created using the "ggplot2" package (Wickham, 2009).

Bayesian Confirmatory Factor Analysis

Bayesian approaches have been shown to provide more accurate estimates of effects than frequentist approaches of latent-variable analysis (see Kruschke, 2013). We used the default priors to fit a two-factor structure representing storage and processing and relational integration. We deemed coefficient estimates credible if zero did not fall within the densest 95% of the distribution, which refers to the Highest Density Interval (HDI).

Bayesian Generalized Linear Mixed-Effects Models

This analysis framework allows for both fixed effects (i.e., experimental conditions or predictors) and random effects (i.e., individuals in experimental conditions) parameters (see Hilbert et al., 2019). Fixed effects describe the relation between the criterion and predictor variables, whereas random effects explain the variability in sampling. We used the default prior of the covariance matrix (inverse Wishart). The random-intercept term was allowed to vary across the subjects.

To evaluate the training gains and transfer effects, we specified a fixed effect associated with a dichotomous growth variable representing linear growth over time that served as a single predictor and was additionally included in the models as an interaction term with the group variables. The group variables were dummy-coded, with the active control group serving as the reference group. This means that each of the eight WM training groups and the passive group were coded 1 for each participant in the respective groups and 0 for everyone else. Participants with the value 0 in all group variables, thus, belonged to the active control group. Accordingly, the dichotomous growth variable models the difference between pre- and post-test for the control group. The difference in gain between pre- and post-test in each WM training group, compared to the active control group, are reflected by the regression weight of the interaction between the corresponding group dummy variable and the growth variable. Finally, the fixed intercept parameter represents the baseline mean in the active control group. The main effects of the group variable were not included in the model to keep the models sparse and because there was no reason to assume group differences at pre-test, due to the random assignment. In addition, we ran a second analysis, in which the passive control group was considered the reference group. All other parameters were remained the same.

Missing data

All participants completed 12 training sessions. However, due to technical issues that led to server downtimes, some training data were not saved online. One day's data was missing for 12 participants, two days for seven participants, and three days for five participants. Consequently, we treated them as missing values and excluded them from the analyses while calculating the mean training performance. If participants completed more than 12 training sessions, these additional sessions were also discarded from the analyses.

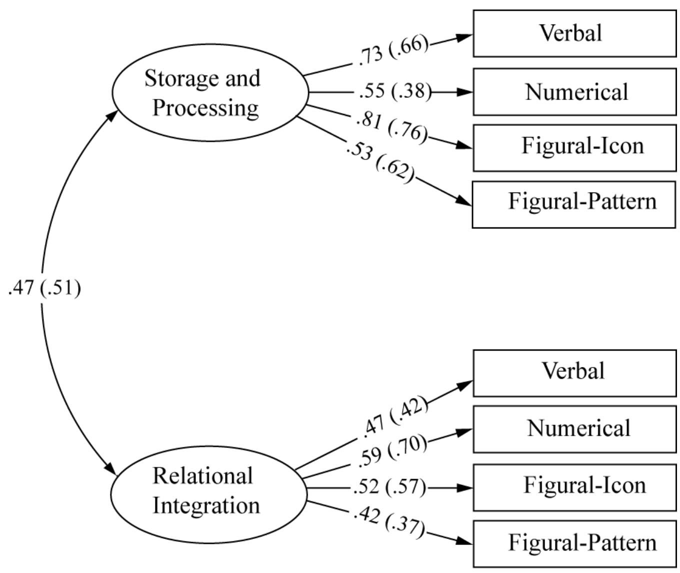
Results

The Facets of WM

The correlated two-factor model (Figure 3) for pre- and post-test scores was tested to see whether storage and processing and relational integration were two separable but correlated factors and how the new tasks related to the latent constructs. Both models were evaluated by Bayesian root mean square error of approximation (BRMSEA). A lower BRMSEA indicates a better model fit (Hu & Bentler, 1999). The BRMSEA values were .080 [95% credible interval (.026 − .135)], and .029 [95% credible interval (.00 − .088)] for the model of pre-test and post-test scores, respectively. Factor loadings (see Table 1) of all the indicators onto their respective latent variables were moderate to high (storage and processing: $\lambda = .38$ to $\lambda = .81$, relational integration: $\lambda = .42$ to $\lambda = .70$), and significantly different from zero. Critically, the new figural-icon tasks loaded significantly on their respective operational WM factors. Storage and processing was correlated with relational integration (pre-test: $r = .47$; post-test: $r = .51$).

Figure 3

Measurement Model Representing the Facet Model of WM.



Note. The values in the parentheses represent the post-test measure. All parameters were statistically significant.

Table 1

Factor Loadings on Latent Variables (Unstandardized Loadings).

Variables	Pre-test Unstandardized Loading	HDI Lower	HDI Upper	Post-test Unstandardized Loading	HDI Lower	HDI Upper
SP						
SPV	1.00			1.00		
SPN	0.28	0.19	0.37	0.18	0.09	0.28
SPFI	0. 67	0.51	0.89	1.45	0.96	2.09
SPFP	0.82	0.56	1.14	0.73	0.52	1.02
RI						
RIV	1.00			1.0		
RIN	1.46	0.77	2.77	1.7	0.87	3.58
RIFI	1.33	0.64	2.70	0.57	0.17	1.37
RIFP	0.60	0.25	1.17	1.26	0.58	2.73
	Pre-test			Post-test		
Covariance						
SP ~~ RI	0.02	0.00	0.02	0.01	0.00	0.02

Note. Highest density interval (HDI) is between 2.5% and 97.5%. SPV = storage and processing verbal; SPN = storage and processing numerical; SPFI = storage and processing figural-icon; SPFP = storage and processing figural-pattern; RIV = relational integration verbal; RIN = relational integration numerical; RIFI = relational integration figural-icon; RIFP = relational integration figural-pattern; SP = storage and processing; RI = relational integration.

Training Performance

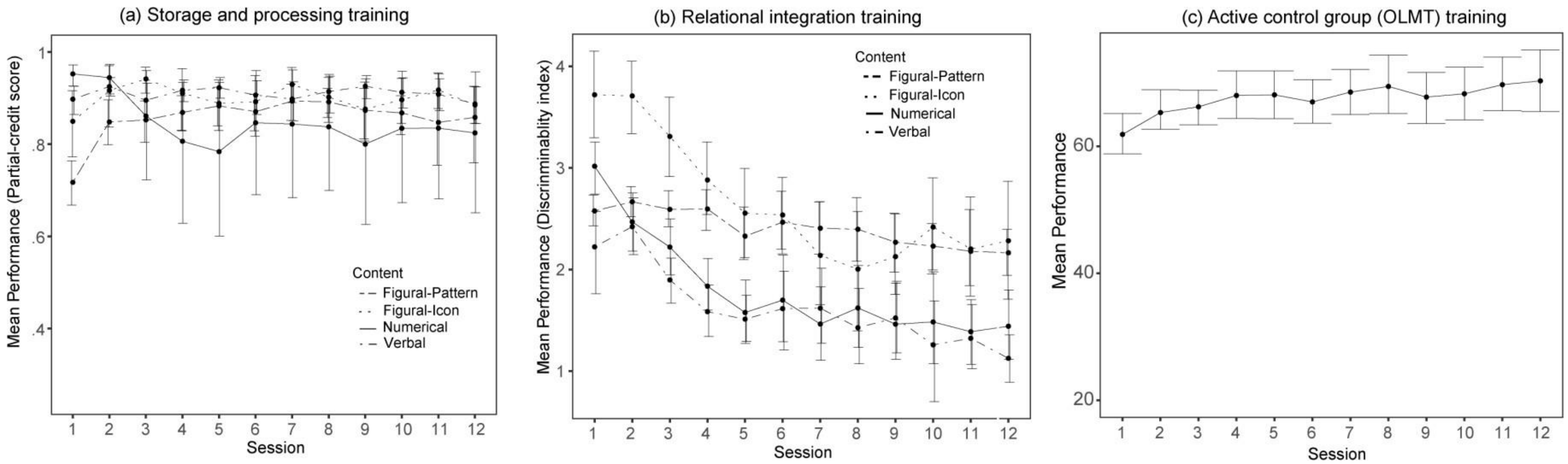
Figure 4 depicts the results of the mean performance achieved by the storage and processing, relational integration, and active control groups over the training period. For the storage and processing groups (Figure 4a), performance somewhat improved for the figural-pattern group, whereas the verbal and figural-icon groups showed relatively consistent performance from the first to the last session. The numerical group showed a decrease in the first five sessions, then started to improve from the sixth session onward and showed relatively consistent performance but did not recover to reach the level from the first session.

All four relational integration groups (Figure 4b) showed a drop in performance with practice. The slopes fluctuated in their steepness, with a less pronounced mean performance decrease for the figural-icon group relative to the other three groups. Different to the relational integration groups, the active control group (Figure 4c) showed a relatively steady increase in performance with training due to practice effect. A closer inspection of the data showed that performance increased across training days if accuracy scores instead of the discriminability index were used as performance measure of the relational integration tasks (see Figure A1). Thus, participants generally did improve in their general task performance (i.e., they identified the critical constellations correctly) but they did not equally improve in reducing their false alarms (i.e., they were not able to detect the critical constellations).

To statistically assess whether performance changed with training, we evaluated training performance across the 12 training sessions with Bayesian linear mixed-effects models, using sessions as fixed-effect and participants as random-effect. The first training session acted as reference category. The results revealed that the training sessions had a significant effect on training performance for six groups: storage and processing figural-pattern, $\chi^2(11) = 48.81$, $p < .001$; relational integration verbal, $\chi^2(11) = 71.70$, $p < .001$; relational integration numerical, $\chi^2(11) = 106.18$, $p < .001$; relational integration figural-icon, $\chi^2(11) = 97.36$, $p < .001$; relational integration figural-pattern, $\chi^2(11) = 36.23$, $p < .001$; active control, $\chi^2(11) = 56.12$, $p < .001$. The regression parameters for the effect of sessions were negative for all relational integration groups but positive for the storage processing figural-pattern and the active control groups.

Figure 4

Training Performance During Twelve Training Sessions.



Note. The values for the storage and processing tasks represent the partial credit scores, whereas it is discriminability index scores for the relational integration tasks. Error bars represent 95% confidence intervals. (a) Storage and Processing Training; (b) Relational Integration Training; (c) Active Control (OLMT) Training.

Training Gains and Transfer Effects

Descriptive statistics for pre- and post-test performance are presented in Table 2. First, we estimated baseline group differences with analysis of variance (ANOVA) using pre-test scores as dependent variables. There was no evidence for baseline differences between the groups in any task (all *p*s $\geq .293$, see Table A1). Next, to examine the training gains from pre- to post-test, we compared the performance improvement in the corresponding WM tasks for all training groups with the active and the passive control groups. In addition, the near/nearest transfer effects to structurally similar (within the same WM operation) and structurally dissimilar WM tasks (between WM operations) were examined.

Mean pretest-posttest scores in WM performance are illustrated in Figures 5 and 6. Overall, each of the storage and processing and relational integration training groups significantly improved from pretest to posttest performance in their respective tasks. The only exception is the storage and processing numerical task in which all the groups performed equally well in the pre- and post-test. Interesting, we found significant near transfer effects on untrained and structurally dissimilar WM tasks in some cases. For example, the relational integration training groups showed improvement in the verbal storage and processing task. However, passive group showed stable performance in the storage and processing figural-pattern, figural-icon, and the relational integration figural-pattern tasks.

Bayesian Multi-level Generalized Linear-Model

Given the relatively moderate group sizes, we estimated WM training effects also with a Bayesian multi-level generalized linear model (see Tables 3 and 4). Because this model is based on a binomial distribution (indicating two outcomes 'yes' and 'no'), we used item-based accuracy (i.e., the correct/wrong responses of each item) of the storage and processing and relational integration tasks as dependent variable for estimating this model.

Active control group. The active control group showed improvement from pre- to post-test in the storage and processing verbal and figural-pattern tasks and the relational integration verbal, numerical, and figural-pattern tasks. The passive control group improved overall from pretest to posttest as well but showed a mean decrease in storage and processing verbal material measures relative to the active control group (b = -0.50, p = .023, CI = -0.91 – -0.08; BF_{10} = 3.13). In addition, all training groups improved in the training tasks, with these effects being supported by at least substantial evidence, $BF_{10} \geq 4.55$. The only exception is the relational integration figural-pattern group for which the effect was non-significant, p = .068, and the evidence was ambiguous, BF_{10} = 1/1.40 (see Tables 3 − 4). Importantly, we found no solid evidence for the near/nearest transfer on structurally similar but untrained WM tasks. One exception to this pattern was that the relational integration figural-icon group showed a mean increase in the relational integration verbal task compared to the active control group, b = 0.32, p = .043, CI = 0.10 – 0.70. Hence, this was the only comparison among all WM tasks supporting our hypothesis that easily verbalizable materials like the figural-icon task might show transfer to other tasks with verbal materials. However, the BF indicated that the evidence was highly ambiguous, BF_{10} = 1.18.

Furthermore, regarding possible transfer to structurally different WM tasks, the results indicated no transfer effect between the two WM operations (i.e., storage and processing and relational integration), with one exception. The group training with the storage and processing figural-pattern task showed higher gains in performance in the relational integration figural-icon task, compared to the active control group, b = 0.73, p = .012, CI = 0.16 – 1.30. However, again, the evidence was ambiguous only, BF_{10} = 2.56.

Passive control group. We conducted further analysis with the passive control group as the reference group (see Tables A2 – A3). The results notably differed from those comparing the WM training groups to the active control group. All trained groups outperformed the passive control group in all tasks, except the figural-pattern and the figural-icon relational integration groups in their respective tasks. Along with the training gains, the results further indicated transfer effects within and between WM operations. Interestingly, training with the figural-icon storage and processing task went along with improvement in verbal task and vice versa, compared to the passive control group, $b = 0.49$, $p = .021$, and $b = 0.51$, $p = .032$, respectively. However, the evidence was ambiguous in both instances, $BF_{10} = 1.27$ and $BF_{10} = 1.39$, respectively.

Table 2

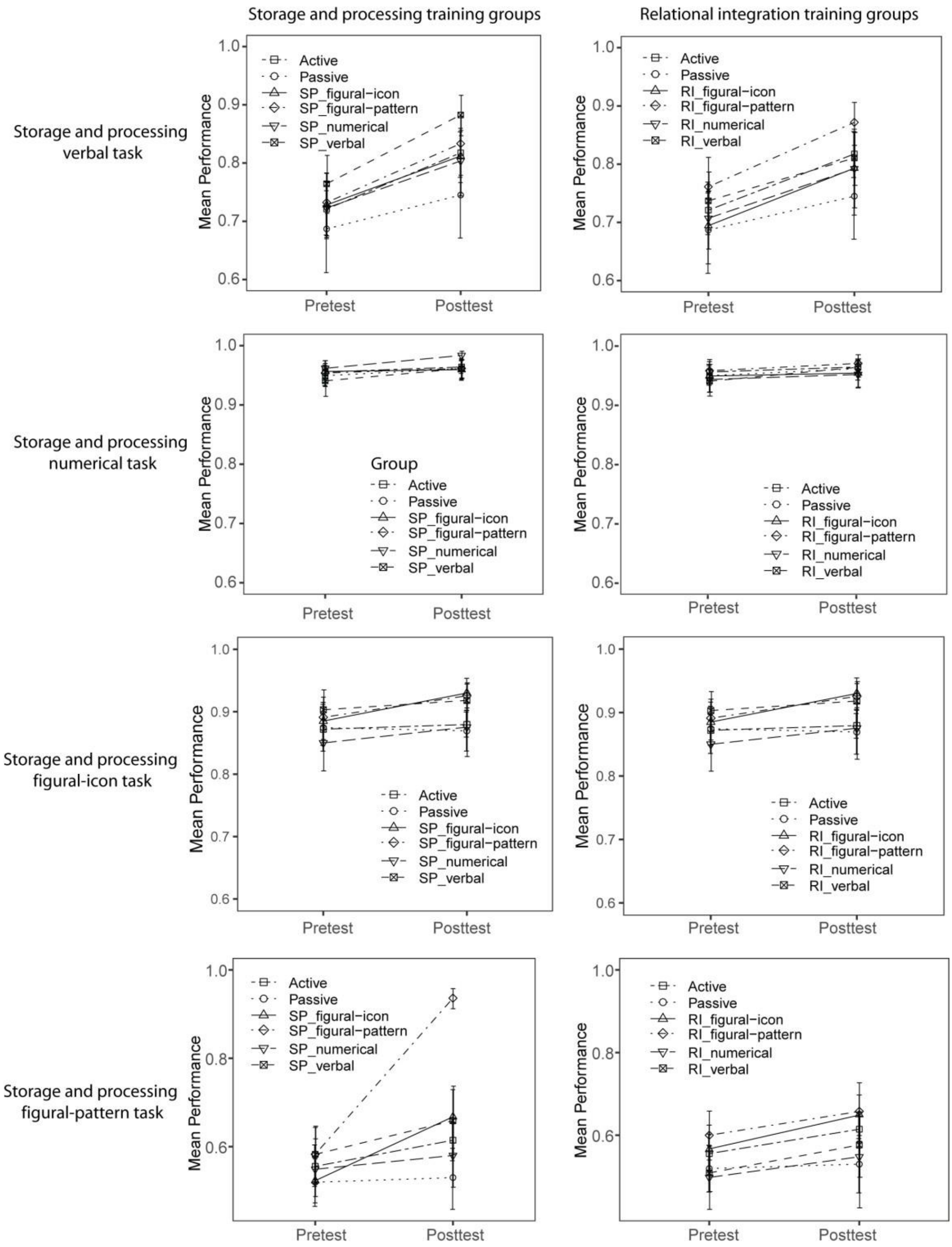
Difference between Pre- and Post-Test in the Working Memory Tasks as a Function of Training Groups.

		Storage and processing training groups				Relational integration training groups				Active Control	Passive
		Verbal	Numerical	Figural -pattern	Figural -icon	Verbal	Numerical	Figural -pattern	Figural -icon		
Group Size (excluding dropouts)		16	17	18	20	17	17	14	16	20	16
Drop out		6	7	4	5	5	7	9	6	2	0
Test											
Storage and Processing											
Verbal	Pre-test	0.76 (0.11)	0.72 (0.10)	0.73 (0.13)	0.73 (0.13)	0.74 (0.11)	0.76 (0.11)	0.71 (0.10)	0.69 (0.13)	0.72 (0.11)	0.69 (0.15)
	Post-test	0.88 (0.07)	0.80 (0.12)	0.80 (0.11)	0.81 (0.11)	0.81 (0.10)	0.87 (0.08)	0.79 (0.12)	0.79 (0.16)	0.82 (0.09)	0.75 (0.14)
Numerical	Pre-test	0.94 (0.05)	0.96 (0.03)	0.95 (0.06)	0.96 (0.04)	0.94 (0.04)	0.96 (0.04)	0.94 (0.05)	0.95 (0.06)	0.96 (0.03)	0.95 (0.03)
	Post-test	0.96 (0.03)	0.98 (0.02)	0.96 (0.03)	0.96 (0.04)	0.96 (0.03)	0.97 (0.35)	0.95 (0.04)	0.94 (0.06)	0.96 (0.03)	0.96 (0.04)
Figural-pattern	Pre-test	0.58 (0.14)	0.55 (0.14)	0.61 (0.13)	0.52 (0.12)	0.51 (0.11)	0.60 (0.14)	0.50 (0.16)	0.57 (0.12)	0.56 (0.11)	0.52 (0.12)
	Post-test	0.66 (0.18)	0.58 (0.16)	0.94 (0.04)	0.67 (0.14)	0.58 (0.16)	0.66 (0.15)	0.55 (0.23)	0.65 (0.10)	0.61 (0.10)	0.53 (0.14)
Figural-icon	Pre-test	0.90 (0.07)	0.85 (0.87)	0.89 (0.08)	0.89 (0.78)	0.89 (0.05)	0.89 (0.06)	0.88 (0.07)	0.88 (0.08)	0.87 (0.06)	0.87 (0.08)
	Post-test	0.92 (0.07)	0.88 (0.09)	0.93 (0.04)	0.93 (0.06)	0.88 (0.10)	0.91 (0.06)	0.89 (0.10)	0.91 (0.06)	0.88 (0.05)	0.87 (0.07)
Relational Integration											
Verbal	Pre-test	2.56 (0.54)	2.29 (0.55)	2.12 (.62)	2.50 (0.78)	2.08 (0.74)	2.42 (.76)	2.15 (0.51)	2.34 (0.70)	2.26 (0.42)	2.28 (0.75)
	Post-test	2.55 (0.55)	2.61 (0.59)	2.55 (.67)	2.87 (0.63)	3.02 (0.78)	2.85 (.74)	2.73 (0.68)	2.96 (0.51)	2.61 (0.51)	2.54 (0.67)
Numerical	Pre-test	2.56 (0.57)	2.70 (0.56)	2.74 (.66)	2.82 (0.70)	2.48 (1.19)	2.98 (.71)	2.75 (0.50)	2.84 (0.60)	2.80 (0.54)	2.32 (0.88)
	Post-test	3.00 (0.52)	2.84 (0.63)	2.93 (.55)	2.98 (0.43)	3.08 (0.70)	3.53 (.58)	2.97 (0.50)	3.32 (0.61)	3.04 (0.60)	2.89 (0.63)
Figural-pattern	Pre-test	2.24 (0.37)	2.25 (0.61)	2.19 (.27)	2.20 (0.41)	2.22 (0.44)	2.36 (.52)	2.17 (0.35)	2.45 (0.29)	2.17 (0.42)	2.36 (0.34)
	Post-test	2.61 (0.45)	2.52 (0.43)	2.55 (.42)	2.45 (0.42)	2.47 (0.41)	2.55 (.43)	2.78 (0.36)	2.62 (0.32)	2.45 (0.40)	2.42 (0.39)
Figural-icon	Pre-test	3.83 (0.81)	3.56 (0.65)	3.41 (.79)	3.66 (0.62)	3.68 (0.66)	3.93 (.64)	3.60 (0.84)	3.64 (0.83)	3.89 (0.69)	3.60 (.95)
	Post-test	4.14 (0.51)	3.94 (0.56)	4.07 (.63)	4.11 (0.55)	4.13 (0.44)	4.15 (.50)	4.12 (0.52)	4.26 (0.34)	3.96 (0.79)	4.05 (0.65)

Note. Means and standard deviation (in parenthesis) of each task performance for the pre-test or post-test are presented. The values for the storage and processing tasks represent the partial credit scores, whereas it is discriminability index scores for the relational integration tasks.

Figure 5

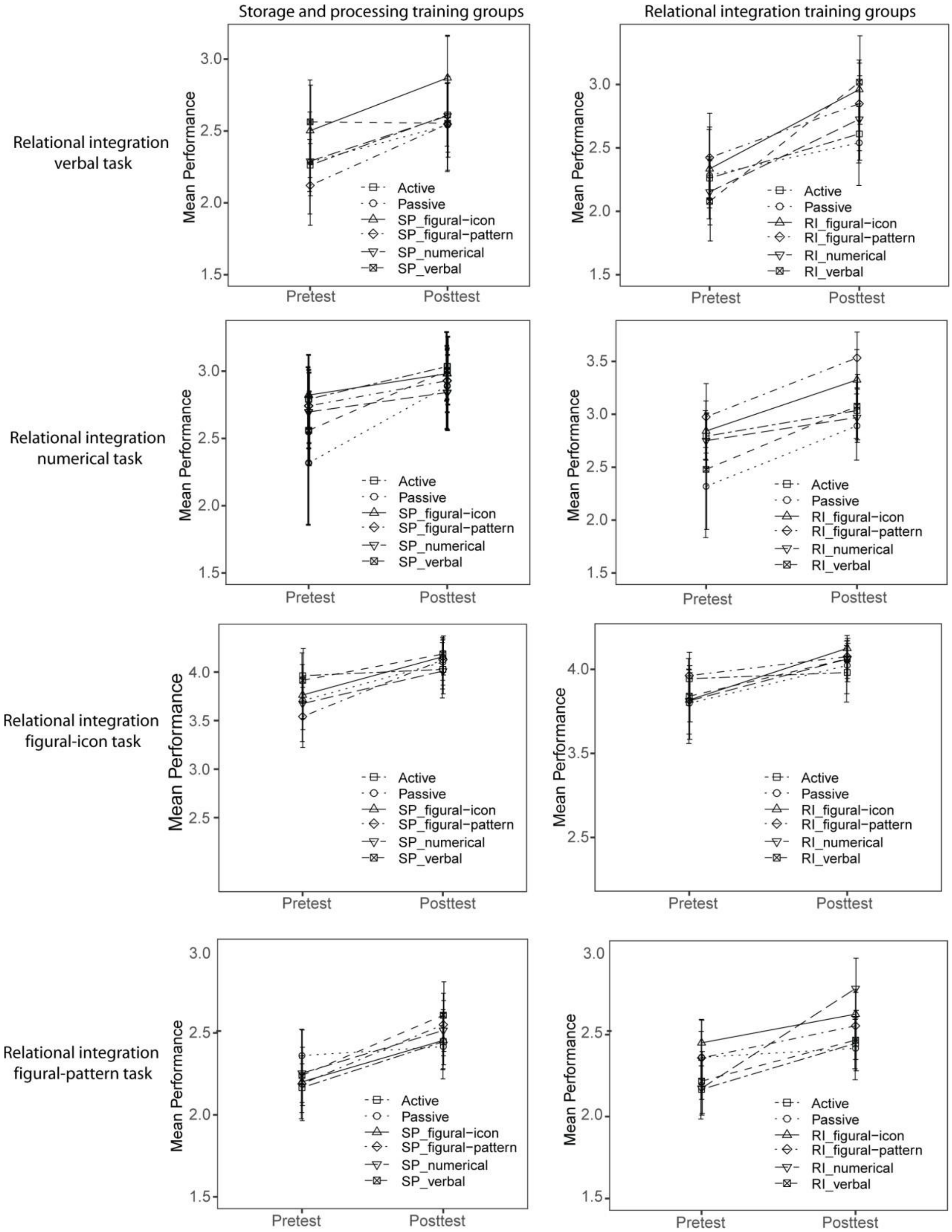
Pre- and Post-Test Performance in Trained and Untrained Storage and Processing Tasks as a Function of Groups.



Note. Error bars represent 95% confidence intervals. RI = relational integration; SP = storage and processing.

Figure 6

Pre- and Post-Test Performance in Trained and Untrained Relational Integration Tasks as a Function of Groups.



Note. Error bars represent 95% confidence intervals. RI = relational integration; SP = storage and processing.

Table 3

Parameters Estimates from Bayesian Generalized Linear Mixed-Effects Models for Storage and Processing Tasks.

Tasks/Covariates	Estimate	*SE*	*z*	*p*	*CI* low	*CI* high	BF_{10}
SP Verbal Task							
Intercept	**-0.54**	**.05**	**-9.95**	**.002**	**-0.65**	**-0.44**	
Active	**0.63**	**.15**	**4.32**	**.001**	**0.34**	**0.92**	
Active*SPV	**0.57**	**.22**	**2.64**	**.008**	**0.15**	**0.99**	**10.00**
Active*SPN	-0.18	.21	-0.86	.389	-0.59	0.23	1/3.65
Active*SPFP	0.15	.21	0.74	.460	-0.25	0.56	1/3.63
Active*SPFI	0.01	.20	-0.01	.993	-0.40	0.39	1/5.15
Active*RIV	-0.08	.21	-0.40	.689	-0.49	0.33	1/4.81
Active*RIN	0.20	.21	0.96	.336	-0.21	0.62	1/2.95
Active*RIFP	-0.05	.22	-0.21	.834	-0.48	0.39	1/4.66
Active*RIFI	-.012	.21	-0.55	.582	-0.53	0.30	1/4.36
Active*Passive	**-0.50**	**.21**	**-2.31**	**.023**	**-0.91**	**-0.08**	**3.13**
SP Numerical task							
Intercept	**0.68**	**.07**	**9.76**	**<.001**	**0.54**	**0.82**	
Active	0.16	.16	1.01	.31	-0.16	0.48	
Active*SPV	0.26	.24	1.06	.29	-0.22	0.74	1/2.74
Active*SPN	**0.85**	**.26**	**3.27**	**< .001**	**0.34**	**1.36**	**50.00**
Active*SPFP	0.15	.24	0.65	.51	-0.31	0.62	1/3.95
Active*SPFI	0.06	.23	0.28	.78	-0.39	0.52	1/4.52
Active*RIV	0.38	.24	1.57	.12	-0.10	0.86	1/1.40
Active*RIN	0.22	.25	0.90	.37	-0.26	0.70	1/3.18
Active*RIFP	-0.17	.25	0.66	.51	-0.66	0.33	1/3.00
Active*RIFI	-0.04	.24	0.18	.86	-0.52	0.43	1/4.24
Active*Passive	-0.07	.24	0.31	.76	-0.54	0.40	1/4.08

Note. Covariates presented in bold showed effects significantly different from zero. BFs reported only for the interaction terms. *CI* = credible interval; BF_{10} = Bayes factor in favor of H_1; Active = mean change between pre- and post-test in the active control group; Active*WM Group = difference in change between the WM training group and the active control group; Active*Passive = difference in change between the passive and the active control group; SP = storage and processing; RI = relational integration; V = verbal; N = numerical; FP = figural-pattern; FI = figural-icon.

Table 3 (Continued)

Parameters Estimates from Bayesian Generalized Linear Mixed-Effects Models for Storage and Processing Tasks.

Tasks/Covariates	Estimate	*SE*	*z*	*p*	*CI* low	*CI* high	BF_{10}
SP Figural-Pattern Task							
Intercept	**-1.07**	**.06**	**-17.03**	**<.001**	**-1.19**	**-0.94**	
Active	**0.35**	**.16**	**0.23**	**025**	**0.04**	**0.66**	
Active*SPV	0.27	.23	0.16	244	-0.18	0.71	1/2.83
Active*SPN	-0.20	23	0.89	373	-0.65	0.25	1/2.64
Active*SPFP	**2.08**	**.24**	**8.71**	**<.001**	**1.61**	**2.55**	**> 1,000**
Active*SPFI	0.32	.22	1.46	.144	-0.11	0.75	1/2.00
Active*RIV	0.01	.23	0.06	.953	-0.43	0.46	1/4.64
Active*RIN	0.11	.23	0.48	.634	-0.33	0.55	1/4.55
Active*RIFP	-0.17	.25	-0.70	.482	-0.66	0.31	1/3.04
Active*RIFI	0.24	.23	1.05	.293	-0.21	0.69	1/3.27
Active*Passive	-0.25	.24	-1.05	.294	-0.71	0.22	1/2.13
SP Figural-Icon Task							
Intercept	**0.36**	**.06**	**6.20**	**<.001**	**0.25**	**0.47**	
Active	0.10	.15	0.65	.513	-0.20	0.39	
Active*SPV	0.41	.23	1.82	.069	-0.03	0.87	1/1.15
Active*SPN	0.13	.21	0.58	.563	-0.30	0.55	1/4.52
Active*SPFP	0.38	.21	1.74	.081	-0.05	0.81	1/1.31
Active*SPFI	**0.72**	**.22**	**3.30**	**<.001**	**0.29**	**1.14**	**50.00**
Active*RIV	-0.02	.22	-0.08	.937	-0.44	0.41	1/4.79
Active*RIN	0.36	.22	1.62	.104	-0.07	0.80	1/1.66
Active*RIFP	0.14	.23	0.60	.550	-0.31	0.59	1/4.14
Active*RIFI	0.33	.23	1.45	.146	-0.11	0.77	1/1.95
Active*Passive	-0.10	.22	-0.46	.649	-0.53	0.33	1/3.96

Note. Covariates presented in bold showed effects significantly different from zero. BFs reported only for the interaction terms. *CI* = credible interval; BF_{10} = Bayes factor in favor of H_1; Active = mean change between pre- and post-test in the active control group; Active*WM Group = difference in change between the WM training group and the active control group; Active*Passive = difference in change between the passive and the active control group; SP = storage and processing; RI = relational integration; V = verbal; N = numerical; FP = figural-pattern; FI = figural-icon.

Table 4

Parameters Estimates from Bayesian Generalized Linear Mixed-Effects Models for Relational Integration Tasks.

Tasks/Covariates	Estimate	*SE*	*z*	*p*	*CI* low	*CI* high	BF_{10}
RI Verbal Task							
Intercept	2.35	.04	62.87	.000	**2.28**	**2.42**	
Active	**0.26**	**.10**	**2.56**	**.010**	**0.06**	**0.45**	
Active*SPV	-0.16	.15	-1.10	.269	-0.45	0.13	1/3.53
Active*SPN	-0.02	.15	-0.13	.896	-0.30	0.27	1/6.74
Active*SPFP	0.02	.14	0.17	.862	-0.25	0.30	1/7.04
Active*SPFI	0.10	.14	0.69	.488	-0.18	0.38	1/5.67
Active*RIV	**0.40**	**.15**	**2.58**	**.009**	**0.10**	**0.70**	**4.55**
Active*RIN	0.13	.15	0.86	.392	-0.17	0.42	1/4.75
Active*RIFP	0.10	.16	0.66	.506	-0.20	0.41	1/5.21
Active*RIFI	**0.32**	**.16**	**2.02**	**.043**	**0.01**	**0.62**	**1.18**
Active* Passive	-0.10	.14	0.71	.481	-0.39	0.18	1/5.45
RI Numerical Task							
Intercept	**2.67**	**.04**	**5.90**	**.000**	**2.59**	**2.75**	
Active	**0.27**	**.12**	**0.34**	**.019**	**0.04**	**0.50**	
Active*SPV	0.07	.17	0.40	.686	-0.26	0.40	1/5.73
Active*SPN	-0.14	.16	0.89	.371	-0.46	0.17	1/3.96
Active*SPFP	-0.09	.16	0.55	.580	-0.41	0.23	1/5.17
Active*SPFI	-0.03	.16	0.18	.858	-0.34	0.28	1/6.22
Active*RIV	0.20	.17	0.18	.238	-0.13	0.52	1/3.09
Active*RIN	**0.54**	**.19**	**0.84**	**.005**	**0.17**	**0.91**	**11.11**
Active*RIFP	-0.03	.17	0.18	.856	-0.37	0.31	1/5.69
Active*RIFI	0.26	.18	0.45	.148	-0.09	0.61	1/2.05
Active* Passive	0.04	.17	0.24	.807	-0.28	0.36	1/5.94

Note. Covariates presented in bold showed effects significantly different from zero. BFs reported only for the interaction terms. *CI* = credible interval; BF_{10} = Bayes factor in favor of H_1; Active = mean change between pre- and post-test in the active control group; Active*WM Group = difference in change between the WM training group and the active control group; Active*Passive = difference in change between the passive and the active control group; SP = storage and processing; RI = relational integration; V = verbal; N = numerical; FP = figural-pattern; FI = figural-icon.

Table 4 (Continued)

Parameters Estimates from Bayesian Generalized Linear Mixed-Effects Models for Relational Integration Tasks.

Tasks/Covariates	Estimate	*SE*	*z*	*p*	*CI low*	*CI high*	BF_{10}
RI Figural-Pattern Task							
Intercept	**2.29**	**.03**	**86.57**	**.000**	**2.23**	**2.34**	
Active	**0.17**	**.08**	**2.00**	**.045**	**0.00**	**0.33**	
Active*SPV	0.09	.12	0.72	.474	-0.15	0.33	1/6.51
Active*SPN	-0.02	.12	-0.19	.854	-0.25	0.21	1/8.30
Active*SPFP	0.06	.12	0.51	.611	-0.17	0.29	1/7.64
Active*SPFI	-0.01	.11	-0.09	.922	-0.23	0.21	1/8.69
Active*RIV	-.03	.12	-0.23	.817	-0.26	0.21	1/8.30
Active*RIN	0.07	.12	0.59	.554	-0.17	0.31	1/7.21
Active*RIFP	0.24	.13	1.83	.068	-0.02	0.51	1/1.40
Active*RIFI	0.10	.13	0.83	.406	-0.14	0.35	1/5.67
Active*Passive	-0.02	.12	-0.16	.870	-0.26	0.22	1/8.22
RI Figural-Icon Task							
Intercept	**3.96**	**.09**	**42.43**	**.000**	**3.78**	**4.14**	
Active	0.22	.19	1.15	.249	-0.16	0.61	
Active*SPV	0.62	.32	1.95	.052	-0.01	1.25	1/1.07
Active*SPN	0.11	.28	0.39	.693	-0.43	0.65	1/3.84
Active*SPFP	**0.73**	**.29**	**2.51**	**.012**	**0.16**	**1.30**	**2.56**
Active*SPFI	0.49	.29	1.70	.088	-0.07	1.06	1/1.62
Active*RIV	0.58	.31	1.90	.058	-0.02	1.18	1/1.18
Active*RIN	0.40	.31	1.26	.207	-0.22	1.01	1/2.36
Active*RIFP	0.50	.31	1.60	.109	-0.11	1.11	1/1.66
Active*RIFI	**1.00**	**.34**	**2.98**	**.003**	**0.34**	**1.65**	**9.09**
Active*Passive	0.56	.30	1.86	.063	-0.03	1.14	1/1.28

Note. Covariates presented in bold showed effects significantly different from zero. BFs reported only for the interaction terms. *CI* = credible interval; BF10 = Bayes factor (in favor of H_1); Active = mean change between pre- and post-test in the active control group; Active*WM Group = difference in change between the WM training group and the active control group; Active*Passive = difference in change between the passive and the active control group; SP = storage and processing; RI = relational integration; V = verbal; N = numerical; FP = figural-pattern; FI = figural-icon.

Cognitive Strategy Survey Report

Table 5 displays the frequency of strategies reported. For storage and processing tasks, the dominant strategy was verbalization, reported by 82.5% of participants before training and 86.5% after training; for relational integration tasks, it was visualization (80.1% at pre-training and 71.3% at post-training). Notably, storage and processing showed an increase in verbalization and relational integration showed a decrease in visualization strategies between pre- and post-test. As expected, whereas most participants used visualization in the storage and processing figural-pattern task (62.60% at pre and 49.10% at post-test), verbalization was the primary reported strategy in the storage and processing figural-icon task (53.20% at pre and 59.60% at post-test).[2] In contrast, most participants used verbalization only in the verbal relational integration task (54.40% at pre and 63.70% at post-test) but not in any of the other relational integration tasks. In addition, some participants reported to have used either both strategies, an alternative/different approach, and no strategy in both storage and processing, and relational integration tasks (see Table 5). For example, alternative strategies reported included naming the shape in storage and processing figural-pattern task, focusing on the potential combination of the stimuli or monitoring the movement of the stimuli in the relational integration task.

When shifting attention from descriptive to exploratory inferential statistics to test to what extent strategy reports can explain the training effects, the results tell a similar story. We conducted a multinomial logistic regression to examine whether task-specific training would contribute to use task-specific strategies (Tables A4 - A5). Group was the predictor,

---

[2] The Wilcoxon Sign-rank test revealed that there was significant difference in the use of cognitive strategies between figural-icon and figural-pattern storage and processing tasks ($z = -4.349$, $p < .001$). Median strategy use was verbalization for figural-icon, whereas it was visualization for figural-pattern.

with the figural-pattern group as the reference category, and type of strategy was the criterion, with visual strategy serving as the reference category. This analysis compared the figural-pattern group to the other groups (i.e., verbal, numerical, and figural-icon), and visual strategy use to the use of other strategies (i.e., verbal, both, alternate, or no strategy), with all other variables held constant in the model. Note that, due to the uneven distribution of strategy use in some groups, and the small numbers of participants in some of the cells, some of the odds ratios are inconclusive. Largely though, results revealed no significant differences in verbal strategy relative to visual strategy use between the figural-pattern group and the verbal and numerical groups, for both WM operations. The only difference occurred when comparing the storage and processing figural-pattern to the figural-icon group in the storage and processing figural-icon task, where the multinomial logit was 11.98 times higher for using a verbal strategy relative to a visual strategy when being a member of the figural-icon group. Taken together and this exception aside, task-specific training did not contribute to use task-specific strategies.

Table 5

Percentages of Using Different Strategies across WM Tasks.

Storage and processing tasks										
Types of strategies	SP verbal		SP numerical		SP figural-icon		SP figural-pattern		SP total	
	Pre-test	Posttest	Pre-test	Posttest	Pre-test	Posttest	Pre-test	Posttest	Pre-test	Posttest
Visualize	9.9%	11.1%	8.2%	8.8%	19.3%	17.0%	62.6%	49.1%	17.5%	13.5%
Verbalize	62.6%	75.4%	73.1%	77.2%	53.2%	59.6%	14%	25.1%	82.5%	86.5%
Both strategies	27.5%	11.7%	18.7%	12.9%	27.5%	20.5%	19.3%	8.7%		
No strategy	0.0%	0.6%	0.0%	0.6%	0.0%	0.6%	4.1%	2.3%		
Alternate strategy	0.0%	1.2%	0.0%	0.6%	0.0%	2.3%	2.6%	4.8%		
Relational integration tasks										
	RI verbal		RI numerical		RI figural-icon		RI figural-pattern		RI total	
	Pre-test	Posttest	Pre-test	Posttest	Pre-test	Posttest	Pre-test	Posttest	Pre-test	Posttest
Visualize	19.3%	17.0%	48.5%	53.2%	59.1%	56.1%	77.8%	81.7%	81.1%	71.3%
Verbalize	54.4%	63.7%	22.2%	25.1%	10.5%	15.2%	2.3%	1.8%	18.9%	28.7%
Both strategies	5.3%	7.6%	5.8%	7.0%	7.6%	13.5%	0.6%	4%		
No strategy	9%	3%	9.4%	5.3%	14%	4.7%	10.5%	7.7%		
Alternate strategy	12%	8.8%	14.1%	9.4%	8.8%	10.5%	8.8%	4.7%		

Note. SP = storage and processing; RI = relational integration.

Discussion

In this study, we aimed to investigate why transfer of WM training is more likely to occur on tasks with similar materials than on other types of tasks. Specifically, we tested the hypothesis that transfer occurs if training and transfer tasks afford the same strategies. We addressed this question by training and assessing the same WM operations (storage and processing and relational integration) with a range of materials that we expected to afford different strategies (verbalization vs. visualization). We designed the tasks based on the facet model of WM (Oberauer et al., 2003; see also Hilbert et al., 2017). To facilitate adoption of verbalization strategies in the figural-icon task, we used icons as stimuli that were easy to verbalize in contrast to the less easily verbalizable patterns in the figural-pattern task. In line with our expectations, more participants stated to have used a verbal strategy with the former compared to the latter in the storage and processing tasks. However, this was not the case for the figural-icon and figural-pattern relational integration tasks. For both operations, the newly developed theory-based figural stimuli for measuring WM capacity shared more variance with the respective latent factors than the original verbal, numerical, and figural-pattern stimuli in Oberauer et al.'s (2003) study.

The higher loading of the storage and processing figural-icon task on storage and processing suggests that this task differs from other storage and processing tasks regarding retrieval of information − recall and recognition. This task requires participants to perform a recognition task in which they have to choose the correct set of icons from a number of candidates, whereas other storage and processing tasks involve freely recalling the words/numbers/patterns. The ability of recognition is an essential determinant for storage and processing tasks (measured with complex span tasks; Lilienthal et al., 2015).

In line with our hypotheses, we found consistent evidence of gains in the tasks trained for most training groups. Contrary to our hypotheses, however, there was no evidence for transfer of training to untrained but structurally similar tasks, even when they afforded verbalization or visualization to a similar extent. Finally, no transfer was observed between the two WM operations, storage and processing, and relational integration, except one significant effect from figural-pattern storage and processing training to the figural-icon relational integration task. However, given that the evidence for this effect was ambiguous only, and that there was no effect from training the figural-icon relational integration task to performance in the figural-pattern storage and processing task, it seems safe to assume that these gains reflected only random data fluctuation.

Furthermore, to address the still ongoing debate over what constitutes an adequate control group (e.g., Au et al., 2020), we incorporated both an active and a passive control group. Whereas comparing the WM training groups with the active control group indicated little differences in performance gains from pre- to post-test, we observed better performance at post-test in the WM training groups relative to the passive control group – a result that could have led to falsely inferring WM training-induced transfer effects. These findings highlight the importance of including an active control group.

Replication of the Facet Model

We focused on the facet model of WM (Oberauer et al., 2003), as this study is a continuation of Hilbert et al. (2017). The Bayesian confirmatory factor analysis of the two-factor model (Figure 3) showed that storage and processing and relational integration were correlated but distinct factors, thus, replicating the original model with the new task materials included. However, as the strategy reports showed, this distinction between storage and processing and relational integration might be confounded with the two types of strategies people use in these tasks; specifically, participants reported to have predominantly used verbalization in the storage and processing tasks, and visualization in the relational integration tasks. A possible explanation for this difference in strategies used is that information is presented sequentially in the storage and processing tasks, thereby encouraging people to spontaneously encode information semantically and construct episodes (Wyer, 2004). In contrast, in relational integration tasks, information was arranged in a grid and participants were instructed to consider the spatial relations between the items (e.g., three identical last digits appear either in a row, column, or diagonal line in a 3 × 3 matrix). Verbalization strategies – which also draw on semantics – may have been more useful for the storage and processing tasks (exception for the figural-pattern storage and processing task), and visualization for the relational integration tasks (except for the verbal relational integration task). The two factors could, therefore, reflect differences in domain-specific processing rather than in operational facets of WM.

Training Gains in Accuracy but not Detection Performance

Relative to previous WM training studies, we found some unusual patterns of WM performance changes over 12 days of training (see Hilbert et al., 2017 for similar findings). Whereas the storage and processing figural-pattern group and the active control group showed improvements in performance, the other storage and processing groups' performance remained relatively stable (Figure 4). Although accuracy scores were used for the storage and processing tasks, training performance did not improve as much as expected. Specifically, performance of the storage and processing numerical group questions the training effectiveness. This training group showed performance decreases until the first few sessions, but after that performance increased. Possibly, because we deal more often with verbal, numerical, and icon (easy to verbalize) materials compared to visuo-spatial material in everyday life, participants of these groups already function close to the optimal level (accuracy close to .80) and have less room for improvement in the beginning of training. The improved storage and processing figural-pattern task and the OLMT task reflect participant's capacity to adapt to the training environment of these tasks.

In contrast, the relational integration groups mean performance significantly decreased over the course of training (Figure 4). This decrease in task performance is likely due to how difficulty was adjusted by the adaptive algorithm. Specifically, whereas accuracy – the proportion of correct responses – was used to trigger adjusting task difficulty, we used detection performance for the analysis of training performance. Detection performance, assessed by the discriminability index, incorporates both hits (correct responses when targets are present) and false alarms (wrong responses when targets are absent), and it can vary even if accuracy remains stable. For example, if both hit and false-alarm rates are reduced, accuracy will remain the same, but detection performance will drop. Indeed, exploring change in accuracy over the course of relational integration training showed the expected steady increase in performance. This was due to an increase in responding correctly when the target was present, with wrong responses remaining on a stable level. With increasing task difficulty of the present task, item complexity and speed demands were raised, manipulations which have been shown to result in drops in detection performance in comparable paradigms such as multiple-objects tracking tasks (Howard & Holocombe, 2008). However, as long as overall accuracy did not fall below 75%, the adaptive algorithm would not reduce task difficulty even if detection performance would decrease, resulting in the unusual pattern of training performance we observed[3].Therefore, future studies of relational integration training should base their adaptive algorithm on detection performance rather than accuracy.

Material- and Operation-Specific Performance Gains

Consistent with previous literature on WM training (e.g., Hilbert et al., 2017; Himi, 2018; Himi et al., 2018; Redick et al., 2015; von Bastian & Eschen, 2016), the present results largely indicate improvement on identical tasks to those used at training from pre- to post-test (Figures 5 and 6) relative to the active control group, suggesting that training leads to task-specific improvement ('stimuli-specific expertise', De Simoni & von Bastian, 2018). However, unlike previous studies demonstrating transfer from complex span training to other span-based measures (e.g., Harrison et al., 2013; Hilbert et al., 2017; but see Minear et al., 2016), no evidence was found for transfer of training to untrained tasks, even when those tasks involved the same narrow ability but used different stimuli materials, which has been referred to as the 'curse of specificity' (Green & Bavelier, 2012, p. 198; see Tables 3 and 4). However, these previous studies differed from the present study in several critical aspects. Specifically, Harrison et al. (2013) used the core training procedure (multiple training regimens - adaptive operation and symmetry span tasks), thereby increasing variability and

[3] While we took the accuracy scores for the relational integration training tasks, performance increased over training.

minimizing automatization, which may have led to transfer to the reading span task. The present findings also seem to contradict the results reported by Hilbert et al. (2017), although they used highly similar training tasks. However, Hilbert et al. did not test baseline cognitive performance, which may have led to biased effect sizes (Melby-Lervåg et al., 2016).

The most obvious potential explanation for the absence of transfer effects in storage and processing and relational integration tasks is that intensive practice on specific content domains (narrow training) may simply not expand domain-general WM capacity. In this regard, it is important to keep in mind that the variance observed in gains can substantially differ from the variance in the (baseline) scores (Hayes et al., 2015). More specifically, the latent-variable analysis (Figure 3) revealed that the latent storage and processing and relational integration factors accounted for 28% to 66% and 18% to 35%, respectively, of the variance in each of their four indicators. This suggests that each of the single tasks possesses unique variances which are not explained by domain-general storage and processing and relational integration. If training-related changes are specific to the variance not captured by the domain-general factor, no domain-general transfer will occur. A recent meta-analysis concluded that benefits of training with specific materials only might indeed be limited to training gains (Schwaighofer et al., 2015).

Moving beyond capacity-driven transfer, we hypothesized that these highly task-specific gains are due to the acquisition of strategies. We deliberately designed the tasks to facilitate transfer from verbal and numerical to figural-icons and vice versa, and we predicted that transfer would occur for tasks affording similar strategies (see also Gathercole et al., 2019). However, our results suggest that strategy use and affordability cannot fully explain the lack of transfer either. The present data did neither support that training can expand general cognitive ability (transfer effects to untrained different tasks) nor that the acquisition of strategies improved cognitive efficiency – (transfer effect on an untrained similar task). An alternative possibility is that trainees take advantage of statistical regularities in their particular training tasks (Brady et al., 2009; von Bastian et al., 2022). These regularities enable participants to compress the presented information and use their available WM capacity more efficiently.

When a person is confronted with a novel cognitive task, that person's response primarily relies on their preferred meta-cognitive strategy, using prior knowledge from long-term memory. Over the course of the training, one typically develops a strategy that involves the task-surface feature of this particular task and relies less on semantic and episodic aspects of long-term memory. In this study, participants may have reported their cognitive style, instead of an actual training task-specific strategy, while the specific strategy in the task can be adapted irrespective of the habitual cognitive processing style (Hilbert et al., 2015). This could explain the absence of a significant relationship between task-specific training and the participants' reported strategy use. The language of the present self-report strategy questionnaire (fixed-choice format) might restrict the participants response and, therefore, lead to biases. However, the strategy variable of the current study is categorized relatively broadly as verbal and visual strategy to increase the precision of the self-report. Moreover, previous studies report relatively acceptable reliability estimate (i.e., Kappa) of strategy reports in WM task (Waris et al., 2021). However, a possibly useful asset for future studies might be the use of nondirective measurement procedures by asking participants an open-ended question ('How do you do the task?'). Alternatively, future research could use experience-sampling methods – repeated assessment of strategy use over the course of time – to determine the applied strategy.

Considering the cognitive-routine framework (Gathercole et al., 2019), we cannot rule out that participants used strategies they had already acquired prior to training and, therefore, did not develop novel routines. Therefore, future research would benefit from assessing not only which strategies people use but also whether they would have used those or similar routines before in other contexts. Alternatively, whereas all three tasks (verbal, numerical, and figural-icon) afforded verbal strategies, they might have differed in the exact types of verbal strategies that can be applied. More specifically, to understand this issue a real-life example might be helpful: In the numerical task, the digit ‘9’ indicates not only a number, but may also hold some meaning (e.g., by using verbal strategies such as associating the digit “9” with the month September and, thereby associating with some personal semantic meaning, such as having been born in September). This makes it more distinct from other digits such as ‘7’ or ‘8’. Inherently, such a strategy would be helpful to remember more digits but would be highly stimulus-specific and, therefore, not transferable to other materials, even if all digits generally afford a similar verbal strategy. More generally, training with one of these three tasks may have led to using or developing a new yet highly specific verbal strategy that could not be applied to the other tasks affording other verbal strategies. Furthermore, in the present study, spontaneously generated-strategies may have mainly depended on task-surface features (e.g., grouping information with common feature), rather less relying on deep information processing such as self-reference elaborative type strategy (e.g., semantic connection with own life or personally experienced episodic aspects of long-term memory; Wyer et al., 2008) or the mnemonics strategy (e.g., chunking; for details, see McCabe et al., 2016). Future research would benefit from using a more fine-grained categorization of types of strategies.

Similarly, the theory of transfer (Barnett & Ceci, 2002) explains transfer as a function of the content of practiced elements (i.e., specific stimuli) and the context in which practice and transfer occurs (i.e., a situation). It seems that transfer only takes place if the stimuli and the context of the task are synchronized. This issue could be explained more clearly by differentiating between familiarity-based processing (recalling item regardless of context) and recollection-based processing (recalling item together with its context), as discussed by De Simoni and von Bastian (2018). Transfer is only observed when participants use recollection-based processing (dealing with interference). After critically inspecting the transposition errors (i.e., recalling a correct item at the wrong position; for a similar approach, see De Simoni & von Bastian, 2018) in a few datasets of verbal storage and processing task, we found that participants could accurately recall more items regardless of correct serial position, thus reflecting familiarity-based processing. People may be able to distinguish between stimuli more clearly, but may not be able to keep many of them in WM simultaneously, resulting in lack of transfer.

Further, it is also possible that the absence of transfer effects was observed because of low training intensity (i.e., the number of trials practiced in each session). The number of trials in the storage and processing and relational integration training sessions varied as a function of participants' performance and training group. Participants in the relational integration groups practiced more trials (on average 500 trials) in each session than those in the storage and processing groups (on average 50 trials), but the storage and processing tasks required more time to complete. We cannot rule out that a longer training regime would have led to transfer effects; however, the lack of even a trend of transfer effects leaves us pessimistic that a higher training intensity would have been more effective.

Finally, it is possible that individual differences may affect the amount of transfer observed. Previous studies have explored demographics (e.g., age and gender), baseline cognitive abilities, and personality (Bürki et al., 2014; Foster et al., 2017; Guye et al., 2017; Wiemers et al. 2019). However, only age and baseline cognitive ability have been shown to explain individual differences in training effects. In the present study, we explored whether baseline cognitive ability could explain the presence or absence of transfer effects by extending our comparisons of participants performing in the top third to those performing in the bottom third at pretest to all training groups (Figure A2)[4]. Whereas we found no group differences for the storage and processing verbal, numerical, and figural-icon groups, top and bottom performers in the figural-pattern storage and processing and the relational integration groups showed a different picture: In accordance with Jaeggi et al. (2011), individuals with lower initial WM capacity improved more strongly than high performers over the course of training. The findings suggest that high performing individuals show less benefit, possibly because they are already functioning at close to optimal level and thus have less room for improvement. Our results conform to compensatory accounts of cognitive change but are contrary to magnification effects reported in other studies (Foster et al., 2017; Guye et al., 2017; Wiemers et al., 2019). However, it is worth to mentioning that this kind of responder analysis (degree of transfer depends on how much improvement occur on the trained task) has been criticized in the training literature regarding the effectiveness of cognitive training (Tidwell et al., 2014).

Transfer Effect between WM Operations

Furthermore, the present results also provided no compelling evidence for a near transfer across the WM operations (storage and processing, and relational integration), except the transfer between relational integration figural-icon and storage and processing figural-pattern; however the evidence was highly ambiguous and observed unidirectionally only. The present findings are in line with previous studies (Hilbert et al., 2017; von Bastian &

[4] We rank-ordered participants by their pre-test performance for each experimental group separately. We identified participants as high performers when they were the top third and as low performers when they were in the bottom third of this rank order.

Oberauer, 2013). In the Bayesian confirmatory factor analysis (Figure 3), the storage and processing, and relational integration factors shared about 26% of the variance, indicating that training may tap the remaining 74% of the variance but not affect the shared variance (for a similar description, see Lange & Süß, 2015). Accordingly, even though storage and processing and relational integration are positively correlated, it is not necessarily the case that repeatedly practicing on specific tasks goes hand in hand with improvement in the common processes shared with other measures as well. Moreover, correlation does not entail causation, as exemplified by Harrison et al. (2013; see also Meiran et al., 2019): Although body weight and height are correlated, making somebody heavier would not necessarily make this person taller. Instead, the positive correlation may result from another underlying ability that is related to both constructs but is not trained. For example, people with higher intelligence scores would perform better in both tasks (hence the shared variance) but, because intelligence is not trained, no transfer is observed. In addition, the neural networks respond differently to various WM tasks (storage and processing: Chein et al., 2011; relational integration: Parkin et al., 2015), which might be another reason for the lack of near transfer.

Limitations and Future Directions

One limitation of our study is the small sample size, which resulted in lower statistical power than we originally aimed for to detect possible transfer effects. Low power not only reduces the likelihood of detecting a true effect but also leads to a low positive predictive value and potential overestimation of the magnitude of the effect (Button et al., 2013; Halsey et al., 2015). Cumming (2011) recommended using precision analysis (the size of the confidence intervals) instead of power analysis, as the confidence interval of a parameter indicates how close the estimated value is to the population value. The relatively small group sizes may have contributed to why we gathered only ambiguous evidence for some comparisons. That said, notably, the majority of effects were associated with Bayes factors indicating sufficient evidence for either the alternative or the null hypotheses (see Tables 3 - 4). Nevertheless, to achieve sufficient power and increased precision, future interventions trying to induce transfer effects should incorporate large-scale samples (cf. von Bastian et al., 2020).

Another potential limitation is the degree of participants attrition (22.97%), although it did not differ among the ten groups, $\chi^2(9) = 11.94$, $p = .217$, and was lower than in other training studies (e.g., 36.78% in Harrison et al., 2013; 43.84% in Redick et al., 2013). However, this dropout rate may reflect individual differences in motivational and metacognitive aspects between people who completed the intervention and those who dropped out. In addition, the metamemory framework of Nelson and Narens (1990) suggests that memory control processes contribute to the application of effective strategies. However, we did not include any measures to explicitly assess participants' motivational or meta-memory aspects, such as self-efficacy or self-monitoring. Yet, these factors could play a role in better understanding how WM training promotes improvements and may provide insight in future research.

Last, we employed a control group that actively engaged in a task that was non-adaptive (different from the WM training groups). Therefore, participants in this group completed a specific task every session (without increasing the difficulty level), which may have been monotonous. In this regard, Shipstead et al. (2012) recommended using an adaptive task for the control group to minimize the treatment difference in terms of the rigor of practice. Thus, expectations might differ between the WM training groups and the active control group. In addition, Boot et al. (2013) emphasized that failure to match expectations between training and active control groups weakens causal inference. However, this concern is somewhat mitigated given that each of three subtests of OLMT is built around a particular challenge (e.g., task-related effort, comparing performance with a superior opponent, or others) that contributes to motivating participants' performance, and that participants also received feedback about their performance. Indeed, the steadily increasing training curve of the OLMT group (see Figure 4c) suggests that the current task was effective in motivating the participants. Furthermore, the active control group also outperformed the other groups on some WM measures in the post-test. These findings align with practice-related improvement (Ackerman et al., 2010). The OLMT assesses processing speed (i.e., how many fields can be covered in 10s), which tends to be related to WM (Oberauer et al., 2003; Schmiedek et al., 2007). Therefore, the comparison of WM training groups with the OLMT group might underestimate the transfer effect (cf. von Bastian & Oberauer, 2014). Appropriate task for active control group might be used in future study.

Conclusion

This study made an essential step towards examining the role of cognitive strategies in training and transfer effects within the framework of Oberauer et al.'s (2003) facet model of WM - an important aspect of cognitive processing in general and WM in particular. Despite substantial gains in the trained tasks, our data showed little evidence for transfer of these gains between tasks that afford the same strategies. The present results are conclusive − the hypothesis that the lack of transfer can be explained (at least primarily) by strategy use was rejected. As this idea has been floating around in the literature for a while now (e.g., von Bastian & Oberauer, 2013; Laine et al., 2018; Forsberg et al., 2020), the findings of a direct test of this hypothesis are an important theoretical contribution. Instead, the results suggest that highly task-specific strategies and actions were trained that did not transfer to tasks that do not require them. Spontaneously developed task-specific strategies might involve the task-surface features of a certain task and, to a lesser extent, rely on semantic and episodic aspects of long-term memory, which might be associated with these features. Moreover, the comparison of the training groups with an active and a passive control group showed considerable differences in the transfer effects, thus highlighting the importance of distinguishing between active and passive controls. The most obvious practical implication of our findings is that if tasks rarely generalize across different domains, the most effective way to acquire a skill is to train exactly that particular skill. Finally, the present study advocates conducting further research by including a large-scale sample. Such replication of our work may permit to fully evaluate the effects of training on the performance of other WM tasks.

Context Paragraph

Working memory training can be used as an intervention to improve WM and related cognitive skills, and it may be particularly beneficial for people who have problems in everyday life. Prior studies found some evidence in support of, but also against, this claim. The extent to which transfer of training is task-specific (i.e., material-dependent) or process-specific (i.e., material-independent) is still subject to debate. One possible explanation for the inconsistent findings of past research is that transfer may only occur when cognitive strategies acquired during training can also be applied in the transfer tasks. To shed further light on this issue, we systematically varied the content and operation domains of WM assessed by training and transfer tasks and, thereby, examined the role of strategy by contrasting performance in tasks eliciting a verbal strategy to tasks that do not. Specifically, we expected that using a verbal strategy would be associated with transfer to tasks comprising easily verbalizable stimuli such as words, numbers, and icons but not to tasks with stimuli that are difficult to verbalize, such as figural patterns. The consistent absence of the expected effects throughout our analysis, leads us to the conclusion that WM training with a small number of tasks has little chance of having a significant impact on the broader WM domain. During training, a spontaneous cognitive strategy is likely to be developed for the specific task at hand, and establishes a salient relationship between to-be-remembered information and information already exists in memory. The current work provides evidence in favor of task-specific benefits, which suggests important practical implications for education and skill acquisition program to enhance particular cognitive or physical ability.

References

Ackerman, P. L., Kanfer, R., & Calderwood, C. (2010). Use it or lose it? Wii brain exercise practice and reading for domain knowledge. *Psychology and Aging, 25*(4), 753–766. https://doi.org/10.1037/a0019277

Ang, S. Y., Lee, K., Cheam, F., Poon, K., & Koh, J. (2015). Updating and working memory training: Immediate improvement, long-term maintenance, and generalisability to non-trained tasks. *Journal of Applied Research in Memory and Cognition, 4,* 121-128. http://dx.doi.org/10.1016/j.jarmac.2015.03.001

Au, J., Gibson, B. C., Bunarjo, K., Buschkuehl, M., & Jaeggi, S. M. (2020). Quantifying the difference between active and passive control groups in cognitive interventions using two meta-analytical approaches. *Journal of Cognitive Enhancement, 4*, 192–210. https://doi.org/10.1007/s41465-020-00164-6

Au, J., Sheehan, E., Tsai, N., Duncan, G. J., Buschkuehl, M., & Jaeggi,S. M. (2015). Improving fluid intelligence with training on working memory: A meta-analysis. *Psychonomic Bulletin & Review, 22,* 366–377. http://dx.doi.org/10.3758/s13423-014-0699-x

Bailey, H. R., Dunlosky, J., & Hertzog, C. (2014). Does strategy training reduce age-related deficits in working memory? *Gerontology*, *60*, 346–356. http://dx.doi.org/10.1159/000356699

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn?: A taxonomy for far transfer. *Psychological Bulletin, 128,* 612–637. http://dx.doi.org/10.1037//0033-2909.128.4.612

Barrett, L. F., Tugade, M. M., & Engle, R. W. (2004). Individual differences in working memory capacity and dual-process theories of the mind. *Psychological Bulletin, 130,* 553–573. http://dx.doi.org/10.1037/0033-2909.130.4.553

Boot, W. R., Simons, D. J., Stothart, C., & Stutts, C. (2013). The pervasive problem with placebos in psychology: Why active control groups are not sufficient to rule out placebo effects. *Perspectives on Psychological Science, 8,* 445–454. http://dx.doi.org/10.1177/1745691613491271

Brown, J. (1958). Some tests of the decay theory of immediate memory. *The Quarterly Journal of Experimental Psychology, 10,* 12–21. https://doi.org/10.1080/17470215808416249

Bühner, M., Krumm, S., & Pick, M. (2005). Reasoning = working memory ≠ attention. *Intelligence, 33,* 251–272. http://dx.doi.org/10.1016/j.intell.2005.01.002

Bürki C. N., Ludwig C., Chicherio C., & de Ribaupierre A. (2014). Individual differences in cognitive plasticity: An investigation of training curves in younger and older adults. *Psychological Research, 78*, 821–835. http://dx.doi.org/10.1007/s00426-014-0559-3

Bürkner, P.-C., & Gabry, J. (2021). brms: Bayesian regression models using 'Stan'. R package version 2.15.0 [Computer software]. https://cran.r-project.org/web/packages/brms/brms.pdf

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience, 14,* 365–376. http://dx.doi.org/10.1038/nrn3475

Chein, J. M., Moore, A. B., Conway, A. R. (2011). Domain-general mechanisms of complex working memory span. Neuroimage, 54(1), 550-559. http://dx.doi.org/10.1016/j.neuroimage.2010.07.067.

Clark, C. M., Lawlor-Savage, L., & Goghari, V. M. (2017). Working memory training in healthy young adults: Support for the null from a randomized comparison to active and passive control groups. *PLOS ONE*, *12*, e0177707. http://dx.doi.org/ 10.1371/journal.pone.0177707

Comblain, A. (1994). Working memory in Down's syndrome: Training the rehearsal strategy. *Down Syndrome Research and Practice, 2,* 123-126. http://dx.doi.org/ 10.3104/reports.42

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. http://dx.doi.org/10.3758/BF031 96772

Cumming, G. (2011). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis.* Abingdon, UK: Routledge.

De Simoni, C., & von Bastian, C. C. (2018). Working memory updating and binding training: Bayesian evidence supporting the absence of transfer. *Journal of Experimental Psychology: General, 147,* 829 – 858. http://dx.doi.org/10.1037/xge0000453

Dorie, V. (2015). blme: Bayesian linear mixed-effects models. R package version 1.0-4 [Computer software]. https://cran.r-project.org/web/packages/blme/blme.pdf

Dougherty, M.R., Hamovitz, T. & Tidwell, J.W. (2016). Reevaluating the effectiveness of n-back training on transfer through the Bayesian lens: Support for the null. *Psychonomic Bulletin & Review, 23*, 306–316. https:// dx.doi.org/10.3758/s13423-015-0865-9

Dunlosky, J., & Kane, M. J. (2007). The contributions of strategy use to working memory span: A comparison of strategy assessment methods. *Quarterly Journal of Experimental Psychology*, *60*, 1227–1245. http://dx.doi.org/10.1080/17470210600926075

Dunning, D. L., & Holmes, J. (2014). Does working memory training promote the use of strategies on untrained working memory tasks? *Memory & Cognition*, 854–862. http://dx.doi.org/10.3758/s13421-014-0410-5

Ericsson, K. A., Chase, W. G., & Faloon, S. (1980). Acquisition of a memory skill. *Science*, *208*, 1181–1182. http://dx.doi.org/10.1126/science.7375930

Fellman, D., Jylkkä, J., Waris, O., Soveri, A., Ritakallio, L., Haga, S., . . . Laine, M. (2020). The role of strategy use in working memory training outcomes. Journal of Memory and Language, 110, Article 104064.

Forsberg, A., Fellman, D., Laine, M., Johnson, W., & Logie, R. H. (2020). Strategy mediation in working memory training in younger and older adults. *Quarterly Journal of Experimental Psychology*, *73*(8), 1206–1226. https://doi.org/10.1177/1747021820915107

Foster, J. L., Harrison, T. L., Hicks, K. L., Draheim, C., Redick, T. S., & Engle, R. W. (2017). Do the effects of working memory training depend on baseline ability level? *Journal of Experimental Psychology: Learning Memory and Cognition*, *43*, 1677–1689. http://dx.doi.org/10.1037/xlm0000426

Gathercole, S. E., Dunning, D. L., Holmes, J., & Norris, D. (2019). Working memory training involves learning new skills. *Journal of Memory and Language, 105,* 19–42. https://doi.org/10.1016/j.jml.2018.10.003

Ginsburg, V., Archambeau, K., van Dijck, J.-P., Chetail, F., & Gevers, W. (2017). Coding of serial order in verbal, visual and spatial working memory. *Journal of Experimental Psychology: General*, *146*, 632 - 650. http://dx.doi.org/10.1037/xge0000278

Green, C. S., & Bavelier, D. (2012). Learning, attentional control, and action video games. *Current Biology, 22,* R197 – R206. https://doi.org/10.1016/j.cub.2012.02.012

Guye, S., De Simoni, C., & von Bastian, C.C. (2017). Do individual differences predict change in cognitive training performance? A latent growth curve modeling approach. *Journal of Cognitive Enhancement, 1*, 374-393. http://dx.doi.org/10.1007/s41465-017-0049-9

Guye, S., & von Bastian, C. C. (2017). Working memory training in older adults: Bayesian evidence supporting the absence of transfer. *Psychology and Aging*, *32*, 732–746. http://dx.doi.org/10.1037/pag0000206

Halsey, L., Curran-Everett, D., Vowler, S., Drummond, G. B. (2015). The fickle *P* value generates irreproducible results. *Nature Methods, 12*, 179–185. https://doi.org/10.1038/nmeth.3288

Harrison, T. L., Shipstead, Z., Hicks, K. L., Hambrick, D. Z., Redick, T. S., & Engle, R. W. (2013). Working memory training may increase working memory capacity but not fluid intelligence. *Psychological Science, 24,* 2409–2419. http://dx.doi.org/10.1177/0956797613492984

Hayes, T. R., Petrov, A. A., & Sederberg, P. B. (2015). Do we really become smarter when our fluid-intelligence test scores improve? *Intelligence, 48,* 1–14. http://dx.doi.org/10.1016/j.intell.2014.10.005

Hilbert, S., Bühner, M., Sarubin, N., Koschutnig, K., Weiss, E., Papousek, I., Reishofer, G., Magg, M. & Fink, A. (2015). The influence of cognitive styles and strategies in the digit span backwards task: Effects on performance and neuronal activity. *Personality and Individual Differences*, *87*, 242–247. http://dx.doi.org/10.1016/j.paid.2015.08.012

Hilbert, S., Nakagawa, T. T., Schuett, S., & Zihl, J. (2014). Mirror reading of words and numbers: Practice and transfer effects. *Visual Cognition, 22*(2), 173–192. http://doi.org/10. 1080/13506285.2014.889261

Hilbert, S., Schwaighofer, M., Zech, A., Sarubin, N., Arendasy, M., & Bühner, M. (2017). Working memory tasks train working memory but not reasoning: A material- and operation-specific investigation of transfer from working memory practice. *Intelligence*, *61*, 102–114. http://dx.doi.org/10.1016/j.intell.2017.01.010

Hilbert, S., Stadler, M., Lindl, A., Naumann, F. & Bühner, M. (2019). Analyzing longitudinal intervention studies with linear mixed models. *Testing, Psychometry, Methodology in Applied Psychology, 26*, 101–119.

Himi, S. A. (2018). Understanding cognitive structure of multitasking behavior and working memory training effects. [Doctoral Dissertation, Ludwig-Maximilians-Universität München. Electronic thesis of LMU Munich. https://edoc.ub.uni-muenchen.de/23189/1/Himi_Samsad_Afrin.pdf

Himi, S. A., Bühner, M., & Hilbert, S. (2018, September 18-22). Multitasking behavior and its related constructs: Executive functions, working memory capacity, relational integration, and divided attention [Paper presentation]. 51th Conference of German Psychological Association (DGP), Frankfurt, Germany.

Himi, S. A., Bühner, M., Schwaighofer, M., Klapetek, A., & Hilbert, S. (2019). Multitasking behavior and its related constructs: Executive functions, working memory capacity, relational integration, and divided attention. *Cognition*, *189*, 275–298. https://doi.org/10.1016/j.cognition.2019.04.010

Himi, S. A., Stadler, M., Bühner, M., & Hilbert, S. (2022). Use of cognitive strategies in working memory training [Manuscript in preparation]. Department of Psychology, University of Munich.

Himi, S. A., Stadler, M., von Bastian, C. C., Bühner, M., & Hilbert, S. (2022, July). Limits of near trasfer: Content- and operation-specific effects of working memory training. Retrieved from https://osf.io/jwr9h/?view_only=21e95876f9a5423da3ab8917d1e3451d

Howard, C. J., & Holcombe, A. O. (2008). Tracking the changing features of multiple objects: Progressively poorer perceptual precision and progressively greater perceptual lag. *Vision Research, 48*(9), 1164-1180. https://doi.org/10.1016/j.visres.2008.01.023

Hu, L.-t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences*, *105*(19), 6829 – 6833. https://doi.org/10.1073/pnas.0801268105

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Shah, P. (2011). Short- and long-termbenefits of cognitive training. *Proceedings of the National Academy of Sciences of the United States of America, 108*, 10081–10086. https://doi.org/10.1073/pnas.1103228108

Klingberg T, Forssberg H, & Westerberg H. (2002). Training of working memory in children with ADHD. *Journal of Clinical Experimental Neuropsychology, 24*(6), 781-91. http://dx.doi.org/10.1076/jcen.24.6.781.8395

Kruschke, J. K. (2013). Bayesian estimation supersedes the t test. *Journal of Experimental Psychology: General, 142*(2), 573–603. https://doi.org/10.1037/a0029146

Laine, M., Fellman, D., Waris, O., & Nyman, T. J. (2018). The early effects of external and internal strategies on working memory updating training. *Scientific Reports, 8*, 4045. http://dx.doi.org/10.1038/s41598-018-22396-5

Lange, S., & Süß., H. M. (2015). Experimental evaluation of near- and far-transfer effects of an adaptive multicomponent working memory training. *Applied Cognitive Psychology, 29,* 502–514. http://dx.doi.org/10.1002/acp.3126

Linares, R., Borella, E., Lechuga, M. T., Carretti, B., & Pelegrina, S. (2019). Nearest transfer effects of working memory training: A comparison of two programs focused on working memory updating. *PLOS ONE*, *14*(2), e0211321. http://dx.doi.org/10.1371/journal.pone.0211321

Lilienthal, L., Rose, N. S., Tamez, E., Myerson, J., & Hale, S. (2015). Individuals with low working memory spans show greater interference from irrelevant information because of poor source monitoring, not greater activation. *Memory & Cognition, 43,* 357 – 366. http://dx.doi.org/10.3758/s13421-014-0465-3

Meiran, N., Dreisbach, G., & von Bastian, C. C. (2019). Mechanisms of working memory training: Insights from individual differences. *Intelligence, 73*, 78-87. https://doi.org/10.1016/j.intell.2019.01.010

Merkle, E. C., Rosseel, Y., Goodrich, B., Garnier-Villarreal, M., Jorgensen, T. D. (2020). blavaan: Bayesian structural equation models via parameter expansion. R package version .3-10 [Computer software]. https://cran.r-project.org/web/packages/blavaan/index.html

McCabe, J. A., Redick, T. S., & Engle, R. W. (2016). Brain-training pessimism, but applied-memory optimism. *Psychological Science in the Public Interest*, *17*, 187–191. http://dx.doi.org/10.1177/1529100616664716

Melby-Lervåg, M., Redick, T., & Hulme, C. (2016). Working memory training does not improve performance on measures of intelligence or other measures of "far transfer": evidence from a meta-analytic review. *Perspectives on Psychological Science*, *11,* 512–534. http://dx.doi.org/10.1177/1745691616635612

Minear, M., Brasher, F., Guerrero, C. B., Brasher, M., Moore, A., & Sukeena, J. (2016). A simultaneous examination of two forms of working memory training: Evidence for near transfer only. *Memory & Cognition*, *44,* 1014–1037. http://dx.doi.org/10.3758/s13421-016-0616-9

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. http://dx.doi.org/10.1006/cogp.1999.0734

Morrison, A. B., Rosenbaum, G. M., Fair, D., & Chein, J. M. (2016). Variation in strategy use across measures of verbal working memory. *Memory & Cognition*, *44,* 922–936. http://dx.doi.org/10.3758/s13421-016-0608-9

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 125–173). Academic Press.

Norris, D. G., Hall, J., & Gathercole, S. E. (2019). Can short-term memory be trained? *Memory & cognition*, *47*(5), 1012–1023. https://doi.org/10.3758/s13421-019-00901-z

Oberauer, K., Süß, H. -M., Wilhelm, O., & Wittman, W. W. (2003). The multiple faces of working memory. *Intelligence, 31,* 167–193. http://dx.doi.org/10.1016/S0160-2896(02)00115-0

Olesen, P. J., Westerberg, H., & Klingberg, T. (2004). Increased prefrontal and parietal activity after training of working memory. *Nature Neuroscience*, *7,* 75–79. http://dx.doi.org/10.1038/nn1165

Parkin, B. L., Hellyer, P. J., Leech, R., & Hampshire, A. (2015). Dynamic network mechanisms of relational integration. *The Journal of Neuroscience, 35*(20), 7660–7673. https://doi.org/10.1523/JNEUROSCI.4956-14.2015

R Development Core Team. (2015). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org/

Redick, T. S., Shipstead, Z., Harrison, T. L., Hicks, K. L., Fried, D. E., Hambrick, D. Z., … Engle, R. W. (2013). No evidence of intelligence improvement after working memory training: A randomized, placebo-controlled study. *Journal of Experimental Psychology: General*, *142,* 359–379. http://dx.doi.org/10.1037/a0029082

Redick, T. S., Shipstead, Z., Wiemers, E. A., Melby-lervåg, M., & Hulme, C. (2015). What's working in working memory training? An educational perspective. *Educational Psychology Review, 27,* 617–633. http://doi.org/10.1007/s10648-015-9314-6

Schmidt-Atzert, L. (2004). *Manual objective achievement motivation test* (OLMT). Mödling, Austria: Schuhfried Gmbh.

Schmiedek, F., Oberauer, K., Wilhelm, O., Süß, H.-M., & Wittmann, W. W. (2007). Individual differences in components of reaction time distributions and their relations to working memory and intelligence. *Journal of Experimental Psychology: General, 136*(3), 414–429. https://doi.org/10.1037/0096-3445.136.3.414

Schwaighofer, M., Fischer, F., & Bühner, M. (2015). Does working memory training transfer? A meta-analysis including training conditions as moderators. *Educational Psychologist*, *50,* 138–166. http://dx.doi.org/10.1080/00461520.2015.1036274

Shipstead, Z., Redick, T. S., & Engle, R. W. (2012). Is working memory training effective? *Psychological Bulletin*, *138,* 628–654. http://dx.doi.org/10.1037/a0027473

Simons, D. J., Boot, W. R., Charness, N., Gathercole, S. E., Chabris, C. F., Hambrick, D. Z., & Stine-Morrow, E. A. L. (2016). Do “brain-training” programs work? *Psychological Science in the Public Interest, 17,* 103–186. http://dx.doi.org/ 10.1177/1529100616661983

Sprenger, A. M., Atkins, S. M., Bolger, D. J., Harbison, J. I., Novick, J. M., Chrabaszcz, J.S., Weems, S. A., Smith, V., Bobb, S., Bunting, M., F., & Dougherty, M. R. (2013). Training working memory: Limits of transfer. *Intelligence, 41,* 638–663. https://doi.org/10.1016/j.intell.2013.07.013

Stieff, M., Werner, S., DeSutter, D., Franconeri, S., & Hegarty, M. (2020). Visual chunking as a strategy for spatial thinking in STEM. *Cognitive Research: Principles and Implications, 5,* 18. https://doi.org/10.1186/s41235-020-00217-6

Taatgen, N. A. (2013). The nature and transfer of cognitive skills. *Psychological Review, 120*(3), 439–471. https://doi.org/10.1037/a0033138

Tidwell, J. W., Dougherty, M. R., Chrabaszcz, J. R., Thomas, R. P., & Mendoza, J. L. (2014). What counts as evidence for working memory training? Problems with correlated gains and dichotomization. *Psychonomic Bulletin & Review, 21*(3), 620-628. https://doi.org/10.3758/s13423-013-0560-7

von Bastian, C.C., Belleville, S., Udale, R.C., Reinhartz, A., Essounni, M., Strobach, T. (2022). Mechanisms underlying training-induced cognitive change. *Nature Reviews Psychology, 1*, 30–41. https://doi.org/10.1038/s44159-021-00001-3

von Bastian, C. C., & Eschen, A. (2016). Does working memory training have to be adaptive? *Psychological Research, 80,* 181–194. http://dx.doi.org/10.1007/s00426-015-0655-z von Bastian, C. C., Guye, S., & De Simoni, C. (2020). How strong is the evidence for the effectiveness of working memory training? In J. M. Novick, M. F. Bunting, M. R. Dougherty, & R. W. Engle (eds.), *Cognitive and working memory training: Perspectives from psychology, neuroscience, and human development.* Oxford Scholarship Online. http://dx.doi.org/10.1093/oso/9780199974467.001.0001

von Bastian, C. C., Langer, N., Jäncke, L., & Oberauer, K. (2013). Effects of working memory training in young and old adults. *Memory & Cognition*, *41,* 611–624. http://dx.doi.org/10.3758/s13421-012-0280-7

von Bastian, C. C., & Oberauer, K. (2013). Distinct transfer effects of training different facets of working memory capacity. *Journal of Memory and Language, 69,* 36–58. http://dx.doi.org/10.1016/j.jml.2013.02.002

von Bastian, C. C., & Oberauer, K. (2014). Effects and mechanisms of working memory training: A review. *Psychological Research*, *78,* 803–820. http://dx.doi.org/10.1007/s00426-013-0524-6

Waris, O., Jylkka, J., Fellman, D., & Laine, M. (2021). Spontaneous strategy use during a working memory updating task. *Acta Psychologica, 212*, 103211. https://doi.org/10.1016/j.actpsy.2020.103211

Wetzels, R., & Wagenmakers, E. -J. (2012). A default Bayesian hypothesis test for correlations and partial correlations. *Psychonomic Bulletin & Review*, 1057–1064. http://dx.doi.org/10.3758/s13423-012-0295-x

Wickham, H. (2009). *ggplot2: Elegant graphics for data analysis.* Springer-Verlag, New York.

Wickham, H. (2021). tidyverse: Easily Install and Load the 'Tidyverse'. R package version 1.3.1[Computer software]. https://cran.r-project.org/web/packages/tidyverse/tidyverse.pdf

Wiemers, E. A., Redick, T. S., & Morrison, A. B. (2019). The influence of individual differences in cognitive ability on working memory training gains. *Journal of Cognitive Enhancement, 3*(2), 174-185. http://dx.doi.org/10.1007/s41465-018-0111-2

Wyer, R. S. (2004). Social comprehension and judgment: The role of situation models, narratives and implicit theories. Erlbaum.

Wyer, R. S., Jr., Hung, I. W., & Jiang, Y. (2008). Visual and verbal processing strategies in comprehension and judgment. *Journal of Consumer Psychology, 18*(4), 244–257. https://doi.org/10.1016/j.jcps.2008.09.002

Zimmermann, K., von Bastian, C. C., Röcke, C., Martin, M., & Eschen, A. (2016). Transfer after process-based object-location memory training in healthy older adults. *Psychology and Aging, 31,* 798–814. http://dx.doi.org/10.1037/pag0000123

(Appendices as follow)

Appendix A

Table A1

Significance Testing Results for Baseline Differences among the Groups.

Tasks	*F*	*df*	*p*
Storage and Processing Verbal	0.743	9	.669
Storage and Processing Numerical	0.521	9	.858
Storage and Processing Figural-Pattern	1.191	9	.304
Storage and Processing Figural-Icon	0.691	9	.716
Relational Integration Verbal	1.054	9	.400
Relational Integration Numerical	1.208	9	.293
Relational Integration Figural-Pattern	0.887	9	.539
Relational Integration Figural-Icon	0.794	9	.623

Note. *F* = *F*-value of the independent measures ANOVA; *df* = degrees of freedom.

Table A2

Parameters Estimates from Linear Mixed-Effects Models for Storage and Processing Tasks.

Tasks/Covariates	Estimate	*SE*	*z*	*p*	BF_{10}
SP Verbal Task					
Intercept	**-0.54**	**0.05**	**-9.95**	**<.001**	
Passive	0.13	0.16	0.83	.41	
Passive*SPV	**1.06**	**0.23**	**4.66**	**<.001**	**> 1,000**
Passive*SPN	0.31	0.22	1.42	.156	1/2.99
Passive*SPFP	**0.65**	**0.22**	**2.95**	**.003**	**7.14**
Passive*SPFI	**0.49**	**0.21**	**2.30**	**.021**	**1.27**
Passive*RIV	0.41	0.22	1.86	.063	1/1.69
Passive*RIN	**0.70**	**0.22**	**3.13**	**.002**	**11.11**
Passive*RIFP	0.45	0.23	1.92	.056	1/1.36
Passive*RIFI	0.38	0.23	1.67	.10	1/2.09
Passive*Active	0.50	0.21	2.31	.021	1.30
SP Numerical Task					
Intercept	**0.68**	**0.07**	**9.77**	**<.001**	
Passive	0.09	0.18	0.50	.614	
Passive*SPV	0.33	0.26	1.31	.758	1/2.24
Passive*SPN	**0.92**	**0.27**	**3.42**	**<.001**	**100.00**
Passive*SPFP	0.23	0.25	0.92	.571	1/3.41
Passive*SPFI	0.14	0.24	0.57	.571	1/4.34
Passive*RIV	0.46	0.26	1.79	.358	1/1.11
Passive*RIN	0.30	0.26	1.15	.074	1/2.65
Passive*RIFP	-0.09	0.26	-0.35	.252	1/3.49
Passive*RIFI	0.03	0.26	0.12	.907	1/4.47
Passive*Active	0.07	0.24	0.31	.728	1/4.68
SP Figural-Pattern Task					
Intercept	**-1.07**	**0.06**	**-1.702**	**< .001**	
Passive	0.10	0.18	0.56	0.57	
Passive*SPV	**0.51**	**0.25**	**2.80**	**0.04**	**1/1.04**
Passive*SPN	0.04	0.25	0.18	0.86	1/4.18
Passive*SPFP	**2.33**	**0.26**	**9.07**	**< .001**	**> 1,000**
Passive*SPFI	**0.57**	**0.24**	**2.39**	**0.02**	**1.79**
Passive*RIV	0.26	0.25	1.06	0.29	1/3.52
Passive*RIN	0.36	0.24	1.45	0.15	1/2.6
Passive*RIFP	0.07	0.26	0.28	0.78	1/3.99
Passive*RIFI	0.49	0.25	1.98	0.05	1/1.17
Passive*Active	0.25	0.24	1.05	0.29	1/3.65
SP Figural-Icon Task					
Intercept	**0.36**	**0.06**	**6.19**	**.000**	
Passive	0.00	0.17	-0.01	.989	
Passive*SPV	**0.52**	**0.24**	**2.14**	**.032**	**1.39**
Passive*SPN	0.23	0.23	0.98	.327	1/3.58
Passive*SPFP	**0.48**	**0.23**	**2.08**	**.037**	**1.28**
Passive*SPFI	**0.82**	**0.23**	**3.55**	**< .001**	**50.00**
Passive*RIV	0.08	0.23	0.36	.718	1/4.64
Passive*RIN	0.46	0.2	1.97	.049	1/1.01
Passive*RIFP	0.24	0.24	0.98	.327	1/3.36
Passive*RIFI	0.43	0.24	1.80	.071	1/1.30
Passive*Active	0.10	0.22	0.45	.649	1/4.70

Note. Covariates presented in bold showed effects significantly different from zero. BFs reported only for the interaction terms. BF_{10} = evidence in support of H_1; Passive = mean change between pre- and post-test in the passive group; Passive*WM Group = mean difference in change between the working memory training group and the passive group; Passive*Active = difference in change between the active and the passive group; SP = storage and processing; RI = relational integration; V = verbal; N = numerical; FP = figural-pattern; FI = figural-icon.

Table A3

Parameters Estimates from Linear Mixed-Effects Models for Relational Integration Tasks.

Tasks/Covariates	Estimate	*SE*	*z*	*p*	BF_{10}
RI Verbal Task					
Intercept	**2.34**	**0.04**	**62.88**	**<.001**	
Passive	0.15	0.11	1.45	.147	
Passive*SPV	-0.06	0.15	-0.39	.691	1/5.90
Passive*SPN	0.08	0.15	0.56	.579	1/5.90
Passive*SPFP	0.13	0.15	0.87	.387	1/5.15
Passive*SPFI	0.20	0.15	1.37	.172	1/2.99
Passive*RIV	**0.50**	**0.16**	**3.17**	**.002**	**25.00**
Passive*RIN	0.23	0.15	1.49	.136	1/2.44
Passive*RIFP	0.21	0.16	1.29	.198	1/3.05
Passive*RIFI	**0.42**	**0.16**	**2.60**	**.009**	**3.85**
Passive*Active	0.10	0.14	0.71	.480	1/5.63
RI Numerical Task					
Intercept	**2.67**	**0.04**	**65.89**	**<.001**	
Passive	**0.31**	**0.12**	**2.56**	**.010**	
Passive*SPV	0.03	0.17	0.16	.872	1/5.68
Passive*SPN	-0.18	0.17	-1.11	.266	1/3.17
Passive*SPFP	-0.13	0.17	-0.78	.436	1/4.31
Passive*SPFI	-0.07	0.16	-0.42	.675	1/5.45
Passive*RIV	0.16	0.17	0.92	.357	1/3.69
Passive*RIN	**0.50**	**0.19**	**2.56**	**.011**	**5.88**
Passive*RIFP	-0.07	0.18	-0.40	.688	1/5.26
Passive*RIFI	0.22	0.18	1.19	.234	1/2.59
Passive*Active	-0.04	0.17	-0.25	.807	1/5.93
RI Figural-Pattern Task					
Intercept	**2.29**	**0.03**	**86.58**	**<.001**	
Passive	0.15	0.09	1.59	.112	
Passive*SPV	0.11	0.13	0.83	.405	1/5.80
Passive*SPN	0.00	0.13	-0.02	.987	1/8.28
Passive*SPFP	0.08	0.13	0.64	.522	1/6.84
Passive*SPFI	0.01	0.12	0.07	.943	1/8.54
Passive*RIV	-0.01	0.13	-0.06	.952	1/8.02
Passive*RIN	0.09	0.13	0.72	.473	1/6.21
Passive*RIFP	0.26	0.14	1.89	.059	1/1.28
Passive*RIFI	0.12	0.13	0.95	.345	1/5.05
Passive*Active	0.02	0.12	0.16	.870	1/8.16
RI Figural-Icon					
Intercept	**3.96**	**0.09**	**42.44**	**<.001**	
Passive	**0.78**	**0.23**	**3.42**	**<.001**	
Passive*SPV	0.07	0.34	.20	.063	1/3.12
Passive*SPN	-0.45	0.30	1.49	.866	1/1.43
Passive*SPFP	0.17	0.31	0.5 6	.634	1/2.69
Passive*SPFI	-0.06	0.31	0.21	.213	1/3.63
Passive*RIV	0.03	0.33	0.08	.933	1/3.26
Passive*RIN	-0.16	0.34	-0.48	.579	1/3.06
Passive*RIFP	-0.06	0.33	-0.17	.137	1/3.35
Passive*RIFI	0.44	0.35	1.24	.836	1/1.12
Passive*Active	-0.56	0.30	-1.86	.842	1.43

Note. Covariates presented in bold showed effects significantly different from zero. BFs reported only for the interaction terms. BF_{10} = evidence in support of H_1; Passive = mean change between pre- and post-test in the passive group; Passive*WM Group = mean difference in change between the working memory training group and the passive group; Passive*Active = difference in change between the active and the passive group; SP = storage and processing; RI = relational integration; V = verbal; N = numerical; FP = figural-pattern; FI = figural-icon.

Table A4

Multinomial Logistic Regression for Strategy Use of the Storage and Processing Groups.

	Effect	Estimate	*SE*	*z*	Odds ratio	*p value*
Strategy use in SPV task						
$P = \frac{(\text{Yi} = \text{verbal})}{(\text{Yi} = \text{visual})}$	Intercept	2.013	0.75	2.67	7.49	.007
	verbal vs figural-pattern	9.20	75.66	0.12	9942.14	.903
	numerical vs figural-pattern	0.54	1.28	0.43	1.73	.669
	Figural-icon vs figural-pattern	0.06	1.06	0.05	1.06	.953
$P = \frac{(\text{Yi} = \text{both})}{(\text{Yi} = \text{visual})}$	Intercept	-0.69	1.22	-0.56	.49	.569
	verbal vs figural-pattern	10.44	75.67	0.13	34416.11	.890
	numerical vs figural-pattern	1.78	1.68	1.06	5.99	.288
	figural-icon vs figural-pattern	0.69	1.58	0.4	2.00	.660
Strategy use in SPN task						
$P = \frac{(\text{Yi} = \text{verbal})}{(\text{Yi} = \text{visual})}$	Intercept	2.01	0.75	2.68	7.50	.007
	verbal vs figural-pattern	9.39	83.36	0.11	12045.51	.910
	numerical vs figural-pattern	0.69	1.27	0.54	1.99	.587
	figural-icon vs figural-pattern	-0.40	0.98	-0.41	0.66	.680
$P = \frac{(\text{Yi} = \text{both})}{(\text{Yi} = \text{visual})}$	Intercept	-0.69	1.22	-0.56	0.50	.571
	verbal vs figural-pattern	10.63	83.37	0.12	41696.64	.898
	numerical vs figural-pattern	0.69	1.87	0.37	1.99	.711
	figural-icon vs figural-pattern	0.28	1.52	0.18	1.33	.850
Strategy use in SPFP task						
$P = \frac{(\text{Yi} = \text{verbal})}{(\text{Yi} = \text{visual})}$	Intercept	0.18	0.60	0.30	120.02	.763
	verbal vs figural-pattern	-0.36	0.85	-0.42	0.69	.670
	numerical vs figural-pattern	-1.88	0.97	-1.92	0.15	.053
	figural-icon vs figural-pattern	-0.47	8.11	-5.79	0.63	.562
$P = \frac{(\text{Yi} = \text{both})}{(\text{Yi} = \text{visual})}$	Intercept	.00	0.63	0.00	1.00	.999
	verbal vs figural-pattern	-0.18	0.87	-0.20	0.83	.834
	numerical vs figural-pattern	-1.29	0.90	-1.43	0.27	.152
	figural-icon vs figural-pattern	-0.69	8.80	-7.87	0.49	.430
$P = \frac{(\text{Yi} = \text{alternate})}{(\text{Yi} = \text{visual})}$	Intercept	-.	0.83	-1.09	0.40	.273
	verbal vs figural-pattern	-25.08	NaN	NaN	0.00	NaN
	numerical vs figural-pattern	-29.06	NaN	NaN	0.00	NaN
	figural-icon vs figural-pattern	-22.88	.00	-1627190.0	0.00	.000
$P = \frac{(\text{Yi} = \text{no strategy})}{(\text{Yi} = \text{visual})}$	Intercept	-1.51	668.35	-0.02	0.00	.982
	verbal vs figural-pattern	-1.80	523.78	-0.00	0.17	.997
	numerical vs figural-pattern	12.71	668.35	0.01	331083.6	.985
	figural-icon vs figural-pattern	13.72	6.68	0.02	910484.2	.984
Strategy use in SPFI task						
$P = \frac{(\text{Yi} = \text{verbal})}{(\text{Yi} = \text{visual})}$	Intercept	0.28	0.54	0.53	1.33	.593
	verbal vs figural-pattern	0.11	0.84	0.139	1.12	.888
	numerical vs figural-pattern	2.27	1.16	1.94	9.74	.052
	figural-icon vs figural-pattern	**2.48**	**1.163**	**2.13**	**11.98**	**.032**
$P = \frac{(\text{Yi} = \text{both})}{(\text{Yi} = \text{visual})}$	Intercept	-0.69	0.70	-0.97	0.50	.327
	verbal vs figural-pattern	1.09	0.95	1.147	3.00	.251
	numerical vs figural-pattern	1.79	1.35	1.32	5.99	.186
	figural-icon vs figural-pattern	1.79	1.35	1.32	5.99	.186
$P = \frac{(\text{Yi} = \text{alternate})}{(\text{Yi} = \text{visual})}$	Intercept	-1.78	1.0	-1.65	0.17	.097
	verbal vs figural-pattern	-12.18	541.37	-0.02	0.00	.982
	numerical vs figural-pattern	-8.93	213.17	-0.04	0.00	.960
	figural-icon vs figural-pattern	-9.70	313.83	-0.03	0.00	.975

Note. The differences in verbal strategy relative to visual strategy use between the figural-pattern group and verbal/numerical/figural-icon groups were of interest for the current study. Covariate presented in bold is significant. Some coefficients are ambiguous due to the uneven distribution of the strategies in the respective group. SP = storage and processing; V = verbal; N = numerical; FP = figural-pattern; FI = figural-icon.

Table A5

Multinomial Logistic Regression for Cognitive Strategy Use of the Relational Integration Groups.

	Effect	Estimate	*SE*	*z*	Odds ratio	*p value*
Strategy use in RIV task						
$P = \frac{(\mathrm{Yi} = \mathrm{verbal})}{(\mathrm{Yi} = \mathrm{visual})}$	Intercept	0.69	0.71	0.98	1.99	.327
	verbal vs figural-pattern	1.94	1.25	1.55	7.00	.120
	numerical vs figural-pattern	16.82	0.56	30.21	2.018925e+07	.000
	figural-icon vs figural-pattern	0.81	0.00	0.769	2.25	.441
$P = \frac{(\mathrm{Yi} = \mathrm{both})}{(\mathrm{Yi} = \mathrm{visual})}$	Intercept	-0.41	0.91	-0.44	0.67	.656
	verbal vs figural-pattern	1.09	1.52	0.72	3.00	.471
	numerical vs figural-pattern	-10.08	0.00	-3.806e+13	0.00	.000
	figural-icon vs figural-pattern	1.09	1.25	0.873	3.04	.382
$P = \frac{(\mathrm{Yi} = \mathrm{alternate})}{(\mathrm{Yi} = \mathrm{visual})}$	Intercept	0.00	0.82	0.00	-0.56	.999
	verbal vs figural-pattern	-18.99	0.00	-7.694718e+08	0.15	.000
	numerical vs figural-pattern	-6.71	0.00	-1.253807e+10	-0.13	.000
	figural-icon vs figural-pattern	-21.91	0.00	-1.147978e+10	0.00	.000
$P = \frac{(\mathrm{Yi} = \mathrm{no\ strategy})}{(\mathrm{Yi} = \mathrm{visual})}$	Intercept	-21.96	0.54	-3.992810e+01	0.00	.000
	verbal vs figural-pattern	-4.73	0.00	-6.649871e+11	0.00	.000
	numerical vs figural-pattern	36.70	0.56	65.92	8.694164e+15	.000
	figural-icon vs figural-pattern	21.26	0.77	27.39	1.22511e+09	.000
Strategy use in RIN task						
$P = \frac{(\mathrm{Yi} = \mathrm{verbal})}{(\mathrm{Yi} = \mathrm{visual})}$	Intercept	-0.56	0.63	-0.89	0.57	.371
	verbal vs figural-pattern	0.15	0.82	0.19	1.16	.850
	numerical vs figural-pattern	-0.13	0.88	-0.15	0.87	.878
	figural-icon vs figural-pattern	0.27	0.83	3.29	1.31	.742
$P = \frac{(\mathrm{Yi} = \mathrm{both})}{(\mathrm{Yi} = \mathrm{visual})}$	Intercept	-16.46	0.42	-38.92	0.00	.000
	verbal vs figural-pattern	14.96	0.69	21.49	3127309	.000
	numerical vs figural-pattern	14.38	0.86	16.70	1759109	.000
	figural-icon vs figural-pattern	14.38	8.61	16.70	1759130	.000
$P = \frac{(\mathrm{Yi} = \mathrm{alternate})}{(\mathrm{Yi} = \mathrm{visual})}$	Intercept	-1.25	1.07	-1.56	0.29	.118
	verbal vs figural-pattern	-21.48	0.00	-1.162082e+10	0.00	.000
	numerical vs figural-pattern	-0.13	1.33	-0.12	1.75	.905
	figural-icon vs figural-pattern	-20.93	1.51	-6.91	0.00	.000
$P = \frac{(\mathrm{Yi} = \mathrm{no\ strategy})}{(\mathrm{Yi} = \mathrm{visual})}$	Intercept	-1.95	1.07	-1.82	0.14	.068
	verbal vs figural-pattern	-17.07	0.00	-8.35	0.00	.000
	numerical vs figural-pattern	0.56	1.33	0.42	1.75	.673
	figural-icon vs figural-pattern	-0.13	1.50	-8.86	0.88	.920

Note. The differences in verbal strategy relative to visual strategy use between the figural-pattern group and verbal/numerical/figural-icon groups were of interest for the current study. Some coefficients are ambiguous due to the uneven distribution of the strategies in the respective group. RI = relational integration; V = verbal; N = numerical; FP = figural-pattern; FI = figural-icon.

Table A5 (continued)

Multinomial Logistic Regression for Cognitive Strategy Use of the Relational Integration Groups.

	Effect	Estimate	*SE*	*z*	Odds ratio	*p value*
Strategy use in RIFP task						
$P = \frac{(\text{Yi} = \text{verbal})}{(\text{Yi} = \text{visual})}$	Intercept	-2.30	1.04	-2.19	0. 99	.028
	verbal vs figural-pattern	-16.68	3540.87	-0.00	0.00	.990
	numerical vs figural-pattern	-17.42	0.00	-49169.27	0.00	.000
	figural-icon vs figural-pattern	0.35	1.29	0.28	1.42	.782
$P = \frac{(\text{Yi} = \text{both})}{(\text{Yi} = \text{visual})}$	Intercept	-1.60	0.77	-2.07	0.19	.037
	verbal vs figural-pattern	-1.02	1.29	-0.79	0.36	.425
	numerical vs figural-pattern	23.81	NaN	NaN	0.00	NaN
	figural-icon vs figural-pattern	-23.02	0.00	-6.261980e+07	0.00	.000
$P = \frac{(\text{Yi} = \text{alternate})}{(\text{Yi} = \text{visual})}$	Intercept	-2.30	1.05	-2.19	0.01	.028
	verbal vs figural-pattern	-0.33	1.47	-0.22	0.71	.819
	numerical vs figural-pattern	-18.43	0.00	-2653130.45	0.00	.000
	figural-icon vs figural-pattern	-17.75	0.00	0.00	0.00	.000
$P = \frac{(\text{Yi} = \text{no strategy})}{(\text{Yi} = \text{visual})}$	Intercept	-2.30	-19.34	-39.72	0.00	.000
	verbal vs figural-pattern	-16.68	16.70	21.66	1797088	.000
	numerical vs figural-pattern	-17.42	16.57	21.55	1572453	.000
	figural-icon vs figural-pattern	0.35	-10.54	-4.507638e+	0.00	.000
Strategy use in RIFS task						
$P = \frac{(\text{Yi} = \text{verbal})}{(\text{Yi} = \text{visual})}$	Intercept	-1.25	0.80	-1.56	0.29	.118
	verbal vs figural-pattern	-0.45	1.11	-0.40	0.63	.684
	numerical vs figural-pattern	0.15	1.04	0.14	1.16	.882
	figural-icon vs figural-pattern	-0.45	1.11	-0.41	0.63	.684
$P = \frac{(\text{Yi} = \text{both})}{(\text{Yi} = \text{visual})}$	Intercept	-1.94	1.06	-1.82	0.14	.068
	verbal vs figural-pattern	0.65	1.25	0.51	1.90	.605
	numerical vs figural-pattern	-0.25	1.50	-0.16	0.78	.867
	figural-icon vs figural-pattern	0.24	1.31	0.18	1.27	.854
$P = \frac{(\text{Yi} = \text{alternative})}{(\text{Yi} = \text{visual})}$	Intercept	-1.25	0.80	-1.56	0.29	.118
	verbal vs figural-pattern	-1.14	1.31	-0.86	0.31	.384
	numerical vs figural-pattern	-0.25	1.11	-0.22	0.78	.822
	figural-icon vs figural-pattern	-16.56	2231.50	-0.01	0.00	.994
$P = \frac{(\text{Yi} = \text{no stateny})}{(\text{Yi} = \text{visual})}$	Intercept	-1.25	0.80	-1.56	0.29	.118
	verbal vs figural-pattern	-16.47	2128.67	-0.01	0.00	.993
	numerical vs figural-pattern	-0.25	1.12	-0.22	0.78	.822
	figural-icon vs figural-pattern	-1.14	1.31	-0.86	0.32	.384

Note. The differences in verbal strategy relative to visual strategy use between the figural-pattern group and verbal/numerical/figural-icon groups were of interest for the current study. Some coefficients are ambiguous due to the uneven distribution of the strategies in the respective group. RI = relational integration; V = verbal; N = numerical; FP = figural-pattern; FI = figural-icon.

Figure A1

Training Performance of Relational Integration task During 12 Training Sessions.

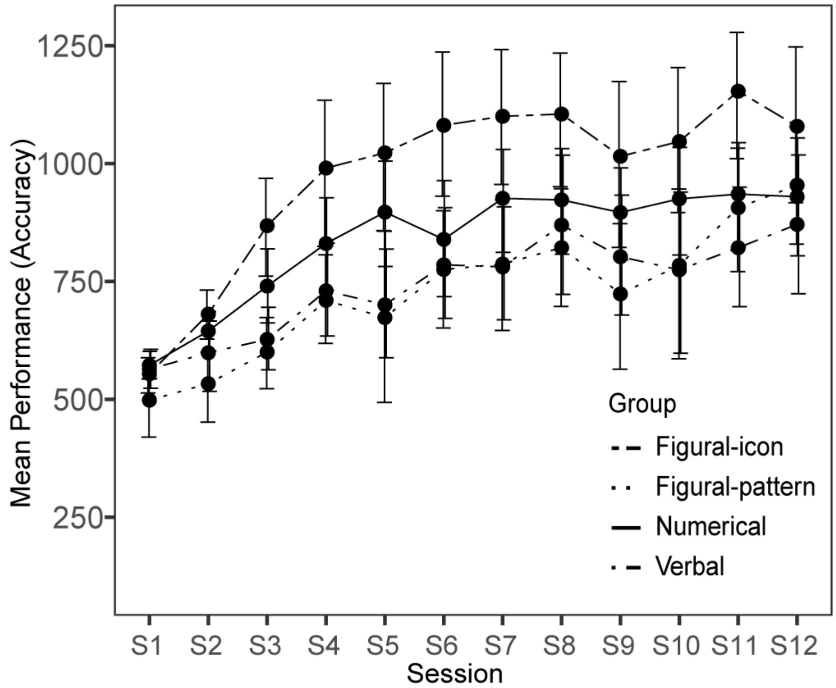


Figure A2

The 12 Training Sessions of High and Low Performers based on Pre-Test Storage and Processing and Relational Integration Task Performance.