HHS Public Access
Author manuscript
J Cogn Enhanc. Author manuscript; available in PMC 2021 August 04.

Published in final edited form as:
J Cogn Enhanc. 2020 March ; 4(1): 100–120. doi:10.1007/s41465-019-00134-7.

Divergent Research Methods Limit Understanding of Working Memory Training

Valentina Pergher[1], **Mahsa Alizadeh Shalchy**[2], **Anja Pahor**[2], **Marc M. Van Hulle**[1], **Susanne M. Jaeggi**[3], **Aaron R. Seitz**[2,3]

[1]Department of Neurosciences, Laboratory for Neuro- and Psychophysiology, KU Leuven-University of Leuven, Leuven, Belgium

[2]Department of Psychology, University of California, Riverside, CA, USA

[3]School of Education, School of Social Sciences, Department of Cognitive Sciences, University of California, Irvine, CA, USA

Abstract

Working memory training has been a hot topic over the last decade. Although studies show benefits in trained and untrained tasks as a function of training, there is an ongoing debate on the efficacy of working memory training. There have been numerous meta-analyses put forth to the field, some finding overall broad transfer effects while others do not. However, discussion of this research typically overlooks specific qualities of the training and transfer tasks. As such, there has been next to no discussion in the literature on what training and transfer tasks features are likely to mediate training outcomes. To address this gap, here, we characterized the broad diversity of features employed in N-back training tasks and outcome measures in published working memory training studies. Extant meta-analyses have not taken into account the diversity of methodology at this level, primarily because there are too few studies using common methods to allow for a robust meta-analysis. We suggest that these limitations preclude strong conclusions from published data. In order to advance research on working memory training, and in particular, N-back training, more studies are needed that systematically compare training features and use common outcome measures to assess transfer effects.

Keywords

Transfer; Working memory training; N-back; Cognitive functions; Meta-analysis

Introduction

A longstanding debate has regarded the extent to which training can improve our basic cognitive functions (Katz et al. 2018). Here, we address this issue in reference to working memory (WM), defined as a limited-capacity system responsible for temporary storage and

Valentina Pergher valentina.pergher@kuleuven.be.
Valentina Pergher, Mahsa Alizadeh Shalchy and Anja Pahor contributed equally to the work. MarcM. Van Hulle SusanneM. Jaeggi and Aaron R. Seitz are share senior authorship
[1]Note that only studies that assessed transfer are reported here.

manipulation of relevant information (Baddeley 2003, 2012).WM isimportant for a wide range of complex cognitive activities, such as reading or problem solving (Shah and Miyake 1999). In the last decade, there has been a considerable amount of literature focused on WM training (Jaeggi et al. 2008; Von Bastian and Oberauer 2014; Morrison and Chein 2011; Klingberg 2012). For example, WM training on a given task can transfer to improvements in untrained working memory tasks (Blacker et al. 2017; Lilienthal et al. 2013; Chein and Morrison 2010; Borella et al. 2010), as well as tasks pertaining to other cognitive domains such as fluid intelligence (Jaeggi et al. 2008; Heinzel et al. 2017; Chein and Morrison 2010; Borella et al. 2010). While there are numerous reports of transfer in the literature, there is also substantial evidence for failure of transfer (Thompson et al. 2013; Jackson et al. 2012). The field has reached a point in which there is a battle of meta-analyses lingering with roughly half of them finding evidence of transfer while the others do not (see Table 1 for variety of individual studies upon which these meta-analyses are based). The lack of explanation regarding this variability not only casts a shadow on WM training research but also poses a significant hurdle when evaluating the effectiveness of WM training.

One of the most common measures of WM is the N-back task, an updating task that requires multiple processes (storage, maintenance, and manipulation of information) and is predictive of inter-individual differences in higher cognitive functions (Jaeggi et al. 2010a, b). Since the N-back task is also one of the most prominent tasks used in WM training studies, here, we limit our discussion on WM training to interventions using N-back tasks.

However, with as many studies using the N-back task, there are as many variants in methodology. These range from the adopted training approaches (e.g., varying in terms of task timing, types of stimuli, number of stimulus streams, adaptive algorithms, feedback provided, number of training sessions, blind/not blind; see Fig. 1 for illustration; Table 1) to the transfer tasks that are rarely consistent from one study to the next with over 120 different transfer tasks used across the 57 experiments reviewed in 51 studies (see Fig. 1 for illustration and Table 1 for details). For example, across these experiments, 31 different tasks assess aspects of WM and short-term memory (STM), including N-back and other updating tasks, simple span tasks, and various complex WM tasks. Another 29 tasks assess aspects of fluid intelligence, the content of which is predominantly visuospatial (matrix reasoning, block design, figure weights, paper folding, form board, surface development, space relations, abstract reasoning, mental rotation, card rotation, TONI, etc.) followed by verbal (letter sets, inference test, nonsense syllogisms, inductive reasoning PMA-R, verbal analogies, reading comprehension), and quantitative (number series) (cf. Table 2). With many unique combinations of training methodologies and transfer tasks, and no model to interpret these differences (Katz et al. 2018), we are left with the difficulty of understanding what approaches might give rise to which cognitive outcomes and what features might determine the boundary conditions of N-back training.

To date, discrepant findings regarding transfer effects reported by meta-analytic studies, focusing primarily on healthy adults, have been discussed in regard to important moderators such as population demographics, training dose, training type (e.g., single task, multiple tasks), training task (e.g., single N-back, dual N-back), training modality (visual, auditory, both), stimulus content (verbal, nonverbal), type of transfer tasks, design type (randomized/not randomized), type of control group (active/passive), attrition rate, training location, supervision, instructional support, feedback, and publication bias (Au et al. 2015; Soveri et al. 2017a, b; Melby-Lervåg and Hulme 2013; Melby-Lervåg et al. 2016; Schwaighofer et al. 2015). While these moderators are certainly relevant, the details of procedures employed in each training study, such as trained and transfer tasks features, which may mediate learning, have been largely ignored.

In this qualitative review, we examine a variety of design factors previously overlooked in N-back training that bear potential to affect learning and transfer, such as task timing and adaptive procedures, types of stimuli, and sensory modality. A summary of all training features can be found in Table 1. Interestingly, only 8 experiments relied on the same training method, whereas 49 experiments had unique training conditions (Fig. 1). In addition, we discuss issues pertaining to the size of the transfer battery and the inconsistency in transfer tasks across studies, and how these factors can affect the findings and their interpretation. The novelty of this review is to highlight the fact that different training protocols and transfer tasks might differentially affect training efficacy and transfer results.

Training Task Features

We highlight six training task attributes (types of N-back task and stimulus modality, task timing, adaptive threshold, feedback, and intervention length) that commonly vary across implementations of N-back training studies. In addition to these, numerous other factors varied across studies within training tasks, such as the number of blocks for each training session, response types (e.g., requiring participants to respond to targets only or also to non-targets), and how feedback was provided (visual/auditory). Within participants, there are additional factors that might determine training outcome, such as N-back levels achieved, used strategies, or motivation. Note that in many cases, details of the procedures that might be important are simply not reported (see Table S1, Supplemental Material). Another source of variation is the inclusion of training procedures that go beyond the N-back task, thereby targeting additional cognitive processes. For example, Li et al. (2008) incorporated mental spatial shifting in the N-back training procedure and Mohammed et al. (2017) used a 2D game version of the N-back task that required navigational skills. In four studies, participants trained on other types of updating WM tasks in addition to the N-back, which precludes understanding of the individual contributions of these training tasks to transfer (Maraver et al. 2016; Waris et al. 2015; Kühn et al. 2013; Loosli et al. 2016).

N-back Task Type—Single vs. Dual

A main area of variation is the use of single or dual N-back training. Conducting multiple N-back tasks simultaneously places different demands on attentional and WM resources as compared with a single N-back. For example, Jaeggi et al. (2003) showed that single and dual N-back tasks differ at the behavioral level with longer reaction times and more errors on dual N-back tasks compared with single N-back. On the other hand, no differentiation between single and dual N-back tasks was observed at the neural level: prefrontal activation increased with higher load irrespective of task type. This may explain why single N-back training seems to be as effective as dual N-back training (Jaeggi et al. 2008; Jaeggi et al.

2010a). In the current sample, 30 out of the 57 experiments adopted single N-back training (13 reporting transfer within WM, 11 reporting transfer beyond WM, and 6 reporting no transfer[1]) and 27 experiments employed dual N-back training (8 reporting transfer within WM, 9 reporting transfer beyond WM, and 10 reporting no transfer). While this may suggest that dual N-back training is more likely to yield transfer within and beyond WM, as compared with single N-back, which seems more likely to show transfer within WM, it should be noted that not all studies assessed both types of transfer. Within the single N-back studies, 2 experiments tested untrained WM tasks, 10 experiments tested for far transfer (6 experiments focusing on fluid intelligence), and 18 experiments tested both. Within the dual N-back studies, 1 experiment tested untrained WM tasks, 9 tested for far transfer (4 experiments using fluid intelligence), and 17 experiments tested both. Even though the single vs dual N-back dichotomy is the most powered of available comparisons, the differences between study methodologies, as described below, largely preclude strong meta-analytic conclusions.

Stimulus Modalities

While WM is often discussed as a domain-general process (Kane et al. 2004), there is substantial evidence that stimuli presented in different modalities (i.e. visual, spatial or auditory stimuli) are processed differently in WM. Owen et al. (2005) showed changes in brain activation between different N-back modalities, specifically for location and for non-verbal stimuli. Similarly, Crottaz-Herbette et al. (2004) found differences in neural activation for auditory and non-spatial WM tasks. The authors used, in a randomized order, a visual and an auditory N-back task. The stimuli were either single-digit numbers (0–9) presented visually at the center of the screen, or binaurally in case of the auditory version. The results showed bilateral suppression of the superior and middle temporal (auditory) cortex during visual (non-spatial) WM, and changes in the occipital (visual) cortex during auditory WM, suggesting that although similar prefrontal and parietal regions are involved in both auditory and visual WM, there are important modality differences in the way neural signals are generated and processed.

For the current review, we define modalities used to categorize the N-back stimuli as follows: (1) "spatial N-back" is a single N-back task that requires the processing of spatial locations of visual stimuli; (2) "visual N-back" describes a single N-back task that requires the processing of visual stimuli (objects, colors, or letters) irrespective of their spatial location; and (3) "audio N-back" describes a single N-back in which stimuli are presented in the auditory domain (e.g., letters, numbers, or other sounds). Dual N-back stimulus modalities are categorized as combinations of the three types of modalities described above: (1) "audio-spatial N-back" involves concurrent processing of auditory stimuli and spatial locations of visual stimuli; (2) "audio-visual N-back" requires simultaneous processing of auditory stimuli and visual stimuli irrespective of their spatial location; and (3) "visual-spatial N-back" requires the processing of both visual stimuli and the spatial locations of these stimuli. In addition, "visual/spatial gaming N-back" refers to a gamified (dual) N-back

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s41465-019-00134-7) contains supplementary material, which is available to authorized users.

task that involves processing of different types of visual stimuli presented at different locations.

In our sample, we find that training task modalities vary widely, with 26 using auditory stimuli (7 reporting transfer within WM, 11 reporting transfer beyond WM, and 8 reporting no transfer), 13 using visual stimuli (non-spatial) (5 reporting transfer within WM, 6 reporting transfer beyond WM, and 2 reporting no transfer), and 18 using spatial stimuli (9 reporting transfer within WM, 3 reporting transfer beyond WM, and 6 reporting no transfer). Within those using auditory stimuli, 2 experiments employed a single audio N-back, 22 used dual audio/spatial N-back, and 2 used audio/visual N-back for training. The variety of the auditory stimuli is further highlighted by some studies using letters or syllables for the audio/spatial sub-group, others using words or other type of sounds for the audio/visual sub-group. Overall, N-back training tasks implement a variety of stimuli (shapes, objects, letters, numbers, etc.) in different modalities (visual, auditory, with or without a spatial component) (see Fig. 1), which can be problematic for cross-study comparisons of transfer effects.

Task Timing

Another training feature rarely considered as a relevant factor impacting WM training is the timing between stimuli in the N-back tasks. Inter-stimulus intervals (ISI) can have an important impact on the time available to process each stimulus and to engage in strategies such as rehearsal or grouping and comparison. The use of these strategies can modify performance levels, give rise to very different experiences during training, and thus likely impact learning outcomes (Laine et al. 2018). Strüber and Polich (2002) showed that during an oddball task, in which participants needed to press a button every time the visual target stimulus appeared, shorter ISIs were associated with smaller P300 amplitudes. They suggested that long ISIs enable a “recovery cycle” that can reduce task difficulty. To date, ISI has not been considered a factor relevant to WM training.

In the papers that we reviewed, we screened 57 experiments across single and dual N-back training and found 46 experiments that reported long ISIs (between 1800 and 2500 ms; 18 reporting transfer within WM, 14 reporting transfer beyond WM and 14reporting no transfer), 9 that used short ISIs (between 500 and 1800 ms; 8 reporting transfer within WM, and 1 reporting transfer beyond WM), while 2 experiments did not report ISI information (and did not report any transfer either).

Adaptive Threshold

The extent to which training adapts to participants’ abilities is another factor that can have a substantial impact on learning and transfer. For example, in the case of perceptual learning, transfer is greatly impacted by task difficulty with more difficult/precise tasks giving rise to more specificity of learning than found through training involving easier/less-precise stimulus judgements (Hung and Seitz 2014; Ahissar and Hochstein 1997). Most N-back training studies utilize adaptive training by adjusting the level of task difficulty based on individual performance, and it has been shown that adapting the difficulty level of the task is engaging for the participant (Jaeggi et al. 2014). Moreover, Holmes et al. (2009) showed that WM training gains were significantly greater for an adaptive training group compared with a non-adaptive training group, although others have failed to observe any effects of adaptivity on learning outcome (von Bastian and Eschen 2016).

In the papers that we reviewed, we distinguished experiments based on the adaptive threshold used to pass to the next difficulty level: most experiments used a threshold of 90% correct responses (non-forgiving), whereas others used a threshold of 65% or 80% (forgiving). Of 46 experiments, 12 adopted a threshold lower than 90% to achieve the next level (7 reporting transfer within WM, 1 reporting transfer beyond WM, and 4 reporting no transfer), while 34 adopted a threshold of 90% correct (16 reporting transfer within WM, 10 reporting transfer beyond WM, and 8 reporting no transfer). Finally, 3 experiments adapted task difficulty by changing the ISI length (not considered here).

Feedback

Feedback plays an important role in the process of learning, particularly in complex cognitive tasks and in monitoring goal progress (West et al. 2001). Feedback is usually delivered based on participants' accuracy and/or response speed and is typically designed to encourage participants to optimize their performance to achieve better learning and/or greater reward (Simen et al. 2009). Feedback can indeed facilitate learning, as demonstrated by cognitive training and perceptual learning research (Abe et al. 2011; Seitz et al. 2006).

Out of the 57 experiments reviewed, 25 experiments employed some type of feedback (11 reporting transfer to untrained WM tasks, 6 reporting transfer beyond WM, and 8 reporting no transfer) while 32 experiments either did not provide feedback or did not explicitly report the use of feedback (16 reporting transfer within WM, 9 reporting transfer beyond WM, and 7 reporting no transfer). Of those experiments employing feedback, 22 gave information about when the feedback was provided: at the end of each block ($N = 9$), at the end of each session ($N = 9$), after each trial ($N = 4$). Thus, despite the critical role of feedback in motivation and learning (Burgers et al. 2015), the majority of studies ($N = 32$) do not describe whether or what type of feedback was employed.

Intervention Length

There is evidence that longer training leads to more learning in terms of more pronounced changes in brain regions involved in WM function (Dahlin et al. 2008; Lövdén et al. 2010). Hempel et al. (2004) highlighted the role of visual spatial N-back training length, showing specific brain activation increases with improved performance after 2 weeks of training, and conversely, activation decreases at the time of consolidation of performance gains after 4 weeks. These results are consistent with the hypothesis that WM training duration affects training results (Jaeggi et al. 2008; Stepankova et al. 2013), although the appropriate amount of training for a given procedure for a given participant is not well established.

In our sample, of the 57 experiments that measured both transfer to WM and beyond WM, 47 used training equal or longer than 10 sessions (29 reporting transfer within WM, 12 reporting transfer beyond WM, and 6 reporting no transfer), and 10 experiments used fewer than 10 sessions (5 reporting transfer within WM, 2 reporting transfer beyond WM, and 3 reporting no transfer).

Transfer Task Features

In addition to the parameters of the training tasks, it is important to consider the details of the outcome measures. Across 57 experiments, 122 different transfer tasks were employed (see Table 2), which speaks to the issue of variability in transfer tasks. The number of outcome measures per study ranged from 1 to as many as 20. Using large test batteries can give rise to participant fatigue and decreased participant engagement (Ackerman and Kanfer 2009), and it can also lead to issues with multiple comparison. In addition, unexpected cognitive benefits may occur as a function of assessing multiple tasks at once, wherein the transfer battery could act as a form of training (Salthouse and Tucker-Drob 2008; see also Green et al. 2019; Morrison and Chein 2011). However, using only one or a few outcome measures can limit opportunity to estimate latent factors. Most of the studies investigated transfer effects using a large variety of tests designed to measure more than one cognitive ability, within and beyond WM. In particular, across all the experiments, 9 focused on just one cognitive function (or task type), 11 experiments focused on two, 9 on three, and 28 on four or more cognitive functions. As follows, we give an overview of how these outcome measures varied across experiments:

Transfer within the domain of WM was assessed with 31 different tasks, including various *simple span* measures (Corsi block, digit span, grid span) and *complex span* tasks (operation span, symmetry span, etc.), *updating* tasks (N-back, running span, numerical updating, etc.), and *other* types of WM tasks such as delayed match to sample tasks and sequencing tasks. Fourteen experiments did not assess WM according to our classification (denoted as *N/A* in Fig. 1), 21 experiments reported using WM measures that fall under one of the four categories mentioned above, and 22 experiments reported using WM tasks that include at least two of these categories (denoted as *multiple* in Fig. 1). Out of the experiments that used only one WM task type, 3 experiments used simple span tasks, another 4 used complex span tasks, 13 used updating tasks, and 1 experiment used a WM task classified as “other” (for details, see “WM task type” in Fig. 1). Out of the 43 experiments that measured WM, 13 experiments reported using only verbal/numerical WM tasks and 3 reported using only visual/spatial WM tasks; however, most used WM tasks that covered both verbal/numerical and visual/spatial domains ($N = 27$; see “WM task domain” in Fig. 1).

In sum, even though they all measure some aspects of WM, these 31 different tasks are likely to measure a number of cognitive skills, a fact often overlooked by extant meta-analyses. While some distinctions have been made in terms of task type (untrained N-back vs. WM tasks in Soveri et al. 2017a, b) and task domain (verbal vs. visuospatial WM in Melby-Lervåg and Hulme 2013; Melby-Lervåg et al. 2016; Schwaighofer et al. 2015), such categorization does not capture the full range of cognitive demands imposed by different WM tasks and may even mask improvements in a subgroup of tasks. Performance on N-back tasks only correlates weakly with performance on complex span tasks (Redick and Lindsey 2013) therefore it makes sense to consider updating and span tasks separately. Furthermore, even if two research groups use the same task with similar types of stimuli, the tasks may still differ in the choice of timing parameters, instructions, feedback, etc., as is often the case with custom-built tasks.

Transfer beyond WM, in particular to fluid intelligence, was assessed with 27 different tasks. Forty-eight out of fifty-seven experiments reported assessing fluid intelligence. These tasks were categorized as: *matrix reasoning tests* (including any type of Raven's matrices or Bochum Matrices Test Advanced (BOMAT)), *spatial visualization tests* (paper folding, mental rotation, card rotation, surface development test, form board, block design, spatial relations), *deduction tests* (nonsense syllogisms, inferences), *induction tests* (number series, inductive reasoning PMA-R, letter sets, abstract reasoning DAT, verbal analogies), and *other tests* (reading comprehension, figure weights). Approximately half of the experiments reported the use of batteries that contain multiple tests (e.g., WASI) or the use of multiple tests that include at least two of the categories described above (e.g., matrix reasoning and deduction), which were classified as *Multiple* ($N = 26$). The remaining experiments included matrix reasoning tests (21 experiments) and spatial visualization tests (1 experiment) (see "Fluid intelligence task type" in Fig. 1, and Table 2). Moreover, in terms of "task domain," fluid intelligence tests were categorized as: figural, verbal, or numerical (Beauducel et al. 2001). Most experiments ($N = 39$) reported using tests with figural content, and even though no experiments used only verbal or only numerical tests, 9 experiments reported using a combination of figural/verbal or figural/numerical tests. While matrix reasoning was the most common type of test used to assess fluid intelligence, which allows for a certain level of comparison across experiments, using just one type of test is not sufficient to estimate fluid intelligence at the latent level. When combined with other fluid intelligence tasks, which vary substantially in terms of the cognitive processes that are required to solve the task (i.e., visuospatial transformation, induction, deduction, attention, working memory), and the degree to which these overlap with the cognitive processes targeted during training, estimating training-related changes in the construct of fluid intelligence across studies becomes challenging.

In addition to the two cognitive domains described above, studies also used other transfer measures representing a wide range of cognitive functions (not reported in Fig. 1; for further details see Table 2). Specifically, 4 different tasks were used to assess long-term memory (LTM), 1 task to assess false memory, 4 different tasks to assess visual search, 11 to assess crystallized/general intelligence, 3 different tasks for reading, 4 for math, 10 different tasks for processing speed, 4 for decision making/problem solving, 17 different tasks for attention/ cognitive control, 1 for motor learning, 2 for multitasking, and 1 for divergent thinking (for further details see Table 2).

Overall, this diversity of transfer tasks measured across studies raises serious issues of the extent to which the same underlying cognitive outcomes are assessed across studies and thus, limits the interpretation of the extant literature.

Test Reliability

An important factor that might impact transfer is task reliability, especially test-retest reliability (Jaeggi et al. 2014). However, for most of the 122 of tasks used, no reliability measures are reported, and it is unclear whether standard forms or custom forms of the tasks are employed, making it difficult to find information on the reliability in the extant literature. It is not uncommon for WM measures to show weak or inconsistent test-retest reliability (e.g., Jaeggi et al. 2010a, b), which could mask transfer effects: the lower the reliability, the lower the chances for transfer (Jaeggi et al. 2014). Comparing transfer effects on tasks that differ substantially in their reliability may be misleading if this factor is not taken into account. Unfortunately, only a few fluid intelligence tasks have reliable parallel test versions and the commonly used method of splitting tests in half can reduce the reliability and validity of the tests (Jaeggi et al. 2014). Recent efforts to develop multiple parallel reasoning tests may mitigate these types of problems in future intervention studies (Pahor et al. 2018; Kyllonen et al. 2018).

Overall, the diversity of transfer tests and batteries used across studies poses a challenge as these outcome measures vary in their degree of similarity with the trained task, and furthermore, their reliability and their validity in measuring the factor of interest are often unclear.

Control Group

It has also been argued that the type of control group plays a significant role in whether transfer is observed. The impact of control groups is related to the degree of similarity between the N-back training and the control interventions, and/or to the differential participant engagement and motivation, and/or participant expectations (Green et al. 2019). For example, Tsai et al. (2018) suggested that placebo effects might represent an additional factor that contributes to improvements achieved during cognitive training due to alterations in participant expectations. However, literature on WM training is mixed both in regard to what control conditions are employed, some using active controls and others passive controls, and also the extent to which the control type seems to alter the magnitude of observed transfer (Au et al. 2015). A simple reason for this is that the features and the effects of the control condition are likely to be more nuanced than what can be captured by simple distinction into active or passive controls. Participant recruitment and population, as well as other factors like engagement and self-perceived improvements might considerably contribute to the extent to which expectations may impact training outcomes.

In our sample, 52 experiments included at least one control group: 22 experiments included only an active control group, 17 experiments included only a passive control group, and 13 experiments included both. Among the 35 experiments that included an active control group, 7 experiments used vocabulary or knowledge-based training, 8 used commercial games such as Tetris, Angry Birds, and Bejeweled, 9 used a variant of N-back training (typically non-adaptive and/or low-difficulty), 8 non-WM training (e.g., processing speed training), and 3 experiments, all belonging to one study, employed alternative WM training (spatial STM). These active control conditions differ in their cognitive and perceptual demands and similarity to the experimental condition, as well as most likely in the induced expectations about performance improvement due to training, again making it difficult to compare results across studies.

Discussion and Future Directions

Although reports on N-back training are steadily increasing, the mechanisms of transfer and the factors that might impact them are still unclear. We suggest that this lack of clarity is due to the variety of training procedures implemented and the selection of transfer measures gauging training outcomes. Despite numerous meta-analyses aimed to understand the effectiveness of N-back training (Au et al. 2015; Soveri et al. 2017a, b; Melby-Lervåg and Hulme 2013; Melby-Lervåg et al. 2016; Schwaighofer et al. 2015), there is still disagreement about the extent of transfer after N-back training. Here we show that N-back training studies, while seemingly similar, employ a wide variety of training features, and in addition, they assess transfer effects with a large and diverse selection of outcome measures. To highlight this variety, we characterized some of the factors that might be important for learning, such as type of N-back, stimulus modalities, task timing, adaptive threshold, feedback and intervention length (see Table 1). Given the small sample size of certain training task features and the extensive variability of methods in the literature, we can only speculate whether these factors are meaningful mediators and moderators. The sheer number of transfer tasks used to assess working memory and other cognitive functions further complicates the matter. At this point, in order to achieve a better understanding of the factors that might interfere with transfer outcomes, we suggest that further training studies and meta-analyses should evaluate more carefully the choice of training features (type of stimuli, ISI, intervention length, etc.), transfer measures (for WM, fluid intelligence, LTM, etc.), the type of control groups, and characteristics within the individuals (educational background, strategies, expectation, etc.) before making inferences. Furthermore, training features, transfer tasks, and individual differences need to be systematically addressed, as the large variability represents a severe issue that limits quantitative conclusions.

We suggest that there are several factors that are leading to this diversity of methods, which we argue limit progress in the field. First, there is the conceptual understanding of WM or fluid intelligence as domain-general processes. This view presumably leads researchers to overlook the importance of domain and task specificity, assuming that it does not matter how a specific exercise or test on WM is given (type, modality, etc.), as all approaches would impact the same cognitive process. Although there is still an ongoing debate about the relationship between specificity of cognitive functions and domain-general processes, emphasis should be given to the fact that all the tests used to investigate these constructs are only partially correlated with the underlying construct. Thus, different training approaches, even if related to the underlying construct, may lead to distinct transfer outcomes due to task specific learning. The second factor is related to the relative nascence of the field. With any new discovery, it makes sense to conduct studies to address the validity of the results and thus using a variety of methods can be vital to explore the space of possibilities. However, this variance of methods produces the inferential problems in making comparisons across studies.

As a first step to address these issues, researchers should both align training and outcome measures across studies and also conduct large-scale comparative studies. As a field, we need to reach some consensus about the training features that may be most conducive to learning, and thus, worth further study. Moreover, a core set of pre-post-measures should be defined both within the WM domain and beyond the WM domain. While studies should necessarily differ in attributes, some uniformity across studies with common tests that have known reliability and stability will allow for comparison with other studies, and researches will still have the option to expand the test battery based on their particular study goals. This would give more power to meta-analyses to address the question whether WM training is worthwhile (and more importantly, for whom it might work and under which circumstances). We recognize that unifying training and transfer task features may be difficult to achieve in practice and so another approach is to conduct larger scale comparative studies with sample sizes sufficient to directly examine unique combinations of training and transfer. Addressing these issues will elevate our understanding about what approaches do or do not lead to improvements in untrained tasks, as well as the specific domains that are most susceptible to the effects of WM training.

Another important step is bridging the gap between lab tests of cognitive functions and tests that reflect the use of cognitive functions in daily life. To enter the next stage of maturity in the field, new approaches that facilitate comparisons of different training approaches and outcomes are needed, to address issues of robustness, reproducibility and broader generality of findings outside of a limited set of laboratory conditions. To accomplish this, we need to become aware of which WM processes are differently required in daily life activities, and which training condition would be hypothesized to transfer to these conditions. To whatever extent existing tests of cognitive functions predict cognitive functions in daily life, this relationship may not hold after training on task structures that specifically resemble the cognitive tests. For example, if performance on two tasks is correlated, but they do not rely upon the exact same mechanisms, then a change in one may not predict a change in the other.

In conclusion, we suggest that it is time for WM training research to retool. Methods employed to date have been valuable to identify a broad set of issues that need to be considered in order to understand the true benefits and limitations of WM training. However, to move the field forward, it will be necessary to conduct large-scale studies that are targeted to uncover how particular training features and transfer measures may lead to differential learning and generalization of that learning. Furthermore, individual differences that may moderate these training effects need to be considered, together with a standard set of reliable outcome measures to better understand the profiles of transfer, and how these are reflected in daily-life activities, going beyond the simple question of whether or not near or far transfer occurs.

Method

To identify candidate papers, we searched Google Scholar, Google, and PubMed for relevant research reports in the last decade, between 2008 and 2018. The search terms used were “N-back training” and “updating training”/”N-back training game” and “updating training game.” In Google, citation marks were used to reduce noise in the research. The first run resulted in 12,100 hits in Google Scholar for N-back training, 675,000 for updating training, 2730 for N-back training game and 127,000 for updating training game. We found 219 hits in PubMed for N-back training, 1501 for updating training, 6 hits for N-back training game and 9 for updating training game. In Google, the hits were 46,300 for N-back training, 71,400 for updating training, 2170 for N-back training game, and no results found for updating training game. We screened all hits in the databases (Google Scholar, PubMed and Google) thereby limiting ourselves to the first 150 ranked ones. For a study to be included at this stage, it needed to meet the following criteria:

1. Cognitive training that included game or no-game version of single or dual N-back task
2. Studies with at least one training group
3. Sample of healthy adults (mean age range 19–69 years old)
4. N-back training equal to or longer than 3 sessions
5. Focused on transfer to WM and/or other cognitive domains

Search hits were screened in the mentioned ranking, and papers already evaluated in previous databases were not considered in the following screening. Our inclusion criteria decreased the number of the studies to 45 on Google Scholar, 6 on PubMed and 0 on Google for N-back training, updating training, N-back training game, and updating training game. In total, our research resulted in 51 studies (excluding the number of overlapping studies) (Fig. 2). Of these 51 studies, 5 studies included more than one N-back training group, which we considered separately, giving rise to a total of 57 experiments.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding This research was supported by NIMH R01 MH111742 to ARS and SMJ, NIH/NIA 1K02AG054665 to SMJ, and a research grant to VP from the Belgian Fund for Scientific Research-Flanders (G088314N), by research grants to MMVH from the Financing program (PFV/10/008), an interdisciplinary research project (IDO/12/007), an industrial research fund project (IOF/HB/12/021) and a special research fund project (C24/ 18/098) of the KU Leuven, the Belgian Fund for Scientific Research-Flanders (G088314N, G0A0914N, G0A4118N), the Interuniversity Attraction Poles Programme—Belgian Science Policy (IUAP P7/11), the Flemish Regional Ministry of Education (Belgium) (GOA 10/019), and the Hercules Foundation (AKUL 043).

SMJ has an indirect conflict of interest with the MIND Research Institute whose interests are related to this work.

References

Abe M, Schambra H, Wassermann EM, Luckenbaugh D, Schweighofer N, & Cohen LG (2011). Reward improves long-term retention of a motor memory through induction of offline memory gains. Current Biology, 21(7), 557–562. [PubMed: 21419628]

Ackerman PL, & Kanfer R (2009). Test length and cognitive fatigue: an empirical examination of effects on performance and test-taker reactions. Journal of Experimental Psychology: Applied, 15(2), 163. [PubMed: 19586255]

Ahissar M, & Hochstein S (1997). Task difficulty and the specificity of perceptual learning. Nature, 387(6631), 401. [PubMed: 9163425]

Anguera JA, Bernard JA, Jaeggi SM, Buschkuehl M, Benson BL, Jennett S, ... & Seidler RD (2012). The effects of working memory resource depletion and training on sensorimotor adaptation. Behavioural Brain Research, 228(1), 107–115. [PubMed: 22155489]

Au J, Sheehan E, Tsai N, Duncan GJ, Buschkuehl M, & Jaeggi SM (2015). Improving fluid intelligence with training on working memory: a meta-analysis. Psychonomic bulletin & review, 22(2), 366–377. [PubMed: 25102926]

Baddeley A (2003). Working memory: looking back and looking forward. Nature reviews neuroscience, 4(10), 829. [PubMed: 14523382]

Baddeley A (2012). Working memory: theories, models, and controversies. Annual review of psychology, 63, 1–29.

Beauducel A, Brocke B, & Liepmann D (2001). Perspectives on fluid and crystallized intelligence: facets for verbal, numerical, and figural intelligence. Personality and individual Differences, 30(6), 977–994.

Beavon P (2012). Improving memory using N-back training. Retrieved from https://ro.ecu.edu.au/theses_hons/65. Accessed 29 Apr 2019.

Blacker KJ, Negoita S, Ewen JB, & Courtney SM (2017). N-back versus complex span working memory training. Journal of Cognitive Enhancement, 1(4), 434–454. [PubMed: 29430567]

Borella E, Carretti B, Riboldi F, & De Beni R (2010). Working memory training in older adults: evidence of transfer and maintenance effects. Psychology and aging, 25(4), 767. [PubMed: 20973604]

Burgers C, Eden A, van Engelenburg MD, & Buningh S (2015). How feedback boosts motivation and play in a brain-training game. Computers in Human Behavior, 48, 94–103.

Bürki CN, 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(6), 821–835. [PubMed: 24652343]

Buschkuehl M, Hernandez-Garcia L, Jaeggi SM, Bernard JA, & Jonides J (2014). Neural effects of short-term training on working memory. Cognitive, Affective, & Behavioral Neuroscience, 14(1), 147–160.

Chein JM, & Morrison AB (2010). Expanding the mind's workspace: training and transfer effects with a complex working memory span task. Psychonomic Bulletin & Review, 17(2), 193–199. [PubMed: 20382919]

Chooi W-T, & Thompson LA (2012). Working memory training does not improve intelligence in healthy young adults. Intelligence, 40(6), 531–542.

Clark CM, Lawlor-Savage L, & Goghari VM (2017a). Functional brain activation associated with working memory training and transfer. Behavioural Brain Research, 334, 34–49. [PubMed: 28750832]

Clark CM, Lawlor-Savage L, & Goghari VM (2017b). Working memory training in healthy young adults: support for the null from a randomized comparison to active and passive control groups. PLoS ONE, 12(5), e0177707.

Clouter A (2013). The effects of dual n-back training on the components of working memory and fluid intelligence: an individual differences approach. Retrieved from https://dalspace.library.dal.ca/handle/10222/36238. Accessed 29 Apr 2019.

Colom R, Román FJ, Abad FJ, Shih PC, Privado J, Froufe M, ... & Karama S (2013). Adaptive n-back training does not improve fluid intelligence at the construct level: gains on individual tests suggest that training may enhance visuospatial processing. Intelligence, 41(5), 712–727.

Crottaz-Herbette S, Anagnoson RT, & Menon V (2004). Modality effects in verbal working memory: differential prefrontal and parietal responses to auditory and visual stimuli. Neuroimage, 21(1), 340–351. [PubMed: 14741672]

Dahlin E, Neely AS, Larsson A, Bäckman L, & Nyberg L (2008). Transfer of learning after updating training mediated by the striatum. Science, 320(5882), 1510–1512. [PubMed: 18556560]

Green CS, Bavelier D, Kramer AF, Vinogradov S, Ansorge U, Ball KK, ... Witt CM (2019). Improving methodological standards in behavioral interventions for cognitive enhancement. Journal of Cognitive Enhancement, 1–28.

Heinzel S, Schulte S, Onken J, Duong QL, Riemer TG, Heinz A, & Rapp MA (2014). Working memory training improvements and gains in non-trained cognitive tasks in young and older adults. Neuropsychology, Development, and Cognition. Section B, Aging, Neuropsychology and Cognition, 21(2), 146–173.

Heinzel S, Lorenz RC, Pelz P, Heinz A,Walter H, Kathmann N, … & Stelzel C (2016). Neural correlates of training and transfer effects in working memory in older adults. Neuroimage, 134, 236–249. [PubMed: 27046110]

Heinzel S, Rimpel J, Stelzel C, & Rapp MA (2017). Transfer effects to a multimodal dual-task after working memory training and associated neural correlates in older adults—a pilot study. Frontiers in Human Neuroscience, 11, 85. [PubMed: 28286477]

Hempel A, Giesel FL, Garcia Caraballo NM, Amann M, Meyer H, Wüstenberg T, et al. (2004). Plasticity of cortical activation related to working memory during training. American Journal of Psychiatry, 161(4), 745–747.

Hogrefe AB, Studer-Luethi B, Kodzhabashev S, & Perrig WJ (2017). Mechanisms underlying n-back training: response consistency during training influences training outcome. Journal of Cognitive Enhancement, 1(4), 406–418.

Holmes J, Gathercole SE, & Dunning DL (2009). Adaptive training leads to sustained enhancement of poor working memory in children. Developmental Science, 12(4), F9–F15. [PubMed: 19635074]

Hung SC, & Seitz AR (2014). Prolonged training at threshold promotes robust retinotopic specificity in perceptual learning. Journal of Neuroscience, 34(25), 8423–8431. [PubMed: 24948798]

Hussey EK, Harbison J, Teubner-Rhodes SE, Mishler A, Velnoskey K, & Novick JM (2017). Memory and language improvements following cognitive control training. Journal of Experimental Psychology: Learning, Memory, and Cognition, 43(1), 23.

Jackson JJ, Hill PL, Payne BR, Roberts BW, & Stine-Morrow EA (2012). Can an old dog learn (and want to experience) new tricks? Cognitive training increases openness to experience in older adults. Psychology and Aging, 27(2), 286–292. [PubMed: 22251379]

Jaeggi SM, Seewer R, Nirkko AC, Eckstein D, Schroth G, Groner R, & Gutbrod K (2003). Does excessive memory load attenuate activation in the prefrontal cortex? Load-dependent processing in single and dual tasks: functional magnetic resonance imaging study. NeuroImage, 19(2), 210–225. [PubMed: 12814572]

Jaeggi SM, Buschkuehl M, Jonides J, & Perrig WJ (2008). Improving fluid intelligence with training on working memory. Proceedings of the National Academy of Sciences, 105(19), 6829–6833.

Jaeggi SM, Studer-Luethi B, Buschkuehl M, Su YF, Jonides J, & Perrig WJ (2010a). The relationship between n-back performance and matrix reasoning—implications for training and transfer. Intelligence, 38(6), 625–635.

Jaeggi SM, Buschkuehl M, Perrig WJ, & Meier B (2010b). The concurrent validity of the N-back task as a working memory measure. Memory, 18(4), 394–412. [PubMed: 20408039]

Jaeggi SM, Buschkuehl M, Shah P, & Jonides J (2014). The role of individual differences in cognitive training and transfer. Memory and Cognition, 42(3), 464–480. [PubMed: 24081919]

Jonasson C (2014). Defining boundaries between school and work: teachers and students' attribution of quality to school-based vocational training. Journal of Education and Work, 27(5), 544–563.

Kane MJ, Hambrick DZ, Tuholski SW, Wilhelm O, Payne TW, & Engle RW (2004). The generality of working memory capacity: a latent-variable approach to verbal and visuospatial memory span and reasoning. Journal of Experimental Psychology: General, 133(2), 189. [PubMed: 15149250]

Katz B, Jaeggi SM, Buschkuehl M, Shah P, & Jonides J (2018). The effect of monetary compensation on cognitive training outcomes. Learning and Motivation, 63, 77–90.

Klingberg T (2012). Is working memory capacity fixed? Journal of Applied Research in Memory and Cognition, 1(3), 194–196.

Kühn S, Schmiedek F, Noack H, Wenger E, Bodammer NC, Lindenberger U, & Lövden M (2013). The dynamics of change in striatal activity following updating training. Human Brain Mapping, 34(7), 1530–1541. [PubMed: 22331673]

Kundu B, Sutterer DW, Emrich SM, & Postle BR (2013). Strengthened effective connectivity underlies transfer of working memory training to tests of short-term memory and attention. Journal of Neuroscience, 33(20), 8705–8715. [PubMed: 23678114]

Küper K, & Karbach J (2016). Increased training complexity reduces the effectiveness of brief working memory training: evidence from short-term single and dual n-back training interventions. Journal of Cognitive Psychology, 28(2), 199–208.

Kyllonen P, Hartman R, Sprenger A, Weeks J, Bertling M, McGrew K, et al. (2018). General fluid/inductive reasoning battery for a high-ability population. Behavior Research Methods, 1–16. [PubMed: 29340969]

Laine M, Fellman D, Waris O, & Nyman TJ (2018). The early effects of external and internal strategies on working memory updating training. Scientific Reports, 8(1), 4045. [PubMed: 29511316]

Lawlor-Savage L, & Goghari VM (2016). Dual n-back working memory training in healthy adults: a randomized comparison to processing speed training. PloS ONE, 11(4), e0151817.

Li SC, Schmiedek F, Huxhold O, Röcke C, Smith J, & Lindenberger U (2008). Working memory plasticity in old age: practice gain, transfer, and maintenance. Psychology and Aging, 23(4), 731. [PubMed: 19140644]

Lilienthal L, Tamez E, Shelton JT, Myerson J, & Hale S (2013). Dual N-back training increases the capacity of the focus of attention. Psychonomic Bulletin & Review, 20(1), 135–141. [PubMed: 23184506]

Loosli SV, Falquez R, Unterrainer JM, Weiller C, Rahm B, & Kaller CP (2016). Training of resistance to proactive interference and working memory in older adults: a randomized double-blind study. International Psychogeriatrics, 28(3), 453–467. [PubMed: 26478277]

Lövdén M, Schaefer S, Noack H, Kanowski M, Kaufmann J, Tempelmann C, et al. (2010). Performance-related increases in hippocampal N-acetylaspartate (NAA) induced by spatial navigation training are restricted to BDNF Val homozygotes. Cerebral Cortex, 21(6), 1435–1442. [PubMed: 21071619]

Maraver MJ, Bajo MT, & Gomez-Ariza CJ (2016). Training on working memory and inhibitory control in young adults. Frontiers in Human Neuroscience, 10, 588. [PubMed: 27917117]

Marček V (2015). Effectiveness of n-back cognitive training: quantitative and qualitative aspects (Doctoral dissertation, Masarykova univerzita, Fakulta sociálních studií).

Melby-Lervåg M, & Hulme C (2013). Is working memory training effective? A meta-analytic review. Developmental Psychology, 49(2), 270. [PubMed: 22612437]

Melby-Lervåg M, Redick TS, & 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(4), 512–534. [PubMed: 27474138]

Minear M, Brasher F, Guerrero CB, 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(7), 1014–1037. [PubMed: 27129921]

Mohammed S, Flores L, Deveau J, Hoffing RC, Phung C, Parlett CM, et al. (2017). The benefits and challenges of implementing motivational features to boost cognitive training outcome. Journal of Cognitive Enhancement, 1(4), 491–507. [PubMed: 30221244]

Morrison AB, & Chein JM (2011). Does working memory training work? The promise and challenges of enhancing cognition by training working memory. Psychonomic Bulletin & Review, 18(1), 46–60.

Nagle A, Riener R, & Wolf P (2015). High user control in game design elements increases compliance and in-game performance in a memory training game. Frontiers in Psychology, 6, 1774. [PubMed: 26635681]

Owen AM, McMillan KM, Laird AR, & Bullmore E (2005). N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. Human Brain Mapping, 25(1), 46–59. [PubMed: 15846822]

Pahor A, Stavropoulos T, Jaeggi SM, & Seitz AR (2018). Validation of a matrix reasoning task for mobile devices. Behavior Research Methods, 1–12. [PubMed: 29340969]

Preece D (2012). The effect of working memory (n-back) training on fluid intelligence. Retrieved from https://ro.ecu.edu.au/theses_hons/54. Accessed 29 Apr 2019.

Qiu F, Wei Q, Zhao L, Lin L (2009). Study on Improving Fluid Intelligence through Cognitive Training System Based on Gabor Stimulus. The 1st International Conference on Information Science and Engineering (ICISE2009).

Redick TS, & Lindsey DR (2013). Complex span and n-back measures of working memory: a meta-analysis. Psychonomic Bulletin & Review, 20(6), 1102–1113. [PubMed: 23733330]

Redick TS, Shipstead Z, Harrison TL, Hicks KL, Fried DE, Hambrick DZ, & Engle RW (2013). No evidence of intelligence improvement after working memory training: a randomized, placebo-controlled study. Journal of Experimental Psychology. General, 142(2), 359–379. [PubMed: 22708717]

Rudebeck SR, Bor D, Ormond A, O'Reilly JX, & Lee ACH (2012). A potential spatial working memory training task to improve both episodic memory and fluid intelligence. PLoS ONE, 7(11), e50431.

Salminen T, Strobach T, & Schubert T (2012). On the impacts of working memory training on executive functioning. Frontiers in Human Neuroscience, 6, 166. [PubMed: 22685428]

Salminen T, Frensch P, Strobach T, & Schubert T (2015). Age-specific differences of dual n-back training. Aging, Neuropsychology, and Cognition, 23(1), 18–39.

Salthouse TA, & Tucker-Drob EM (2008). Implications of short-term retest effects for the interpretation of longitudinal change. Neuropsychology, 22(6), 800. [PubMed: 18999354]

Schwaighofer M, Fischer F, & Bühner M (2015). Does working memory training transfer? A meta-analysis including training conditions as moderators. Educational Psychologist, 50(2), 138–166.

Schwarb H, Nail J, & Schumacher EH (2015). Working memory training improves visual short-term memory capacity. Psychological Research, 80(1), 128–148. [PubMed: 25656161]

Schweizer S, Hampshire A, & Dalgleish T (2011). Extending brain-training to the affective domain: increasing cognitive and affective executive control through emotional working memory training. PloS ONE, 6(9), e24372.

Seitz AR, Nanez JE, Holloway S, Tsushima Y, & Watanabe T (2006). Two cases requiring external reinforcement in perceptual learning. Journal of Vision, 6(9), 9–9.

Shah P, & Miyake A (1999). Models of working memory: An introduction. In Miyake A & Shah P (Eds.), Models of working memory: Mechanism of active maintenance and executive control (pp. 1–26). New York: Cambridge University Press.

Shahar N, & Meiran N (2015). Learning to control actions: transfer effects following a procedural cognitive control computerized training. PloS ONE, 10(3), e0119992.

Simen P, Contreras D, Buck C, Hu P, Holmes P, & Cohen JD (2009). Reward rate optimization in two-alternative decision making: empirical tests of theoretical predictions. Journal of Experimental Psychology: Human Perception and Performance, 35(6), 1865. [PubMed: 19968441]

Smith SP, Stibric M, & Smithson D (2013). Exploring the effectiveness of commercial and custom-built games for cognitive training. Computers in Human Behavior, 29(6), 2388–2393.

Soveri A, Karlsson E, Waris O, Grönholm-Nyman P, & Laine M (2017a). Pattern of near transfer effects following working memory training with a dual n-back task. Experimental Psychology, 64(4), 240. [PubMed: 28922999]

Soveri A, Antfolk J, Karlsson L, Salo B, & Laine M (2017b). Working memory training revisited: A multi-level meta-analysis of n-back training studies. Psychonomic Bulletin & Review, 24(4), 1077–1096.

Stepankova H, Lukavsky J, Buschkuehl M, Kopecek M, Ripova D, & Jaeggi SM (2013). The malleability of working memory and visuospatial skills: a randomized controlled study in older adults. Developmental Psychology, 50(4), 1049–1059. [PubMed: 24219314]

Stephenson CL, & Halpern DF (2013). Improved matrix reasoning is limited to training on tasks with a visuospatial component. Intelligence, 41(5), 341–357.

Strüber D, & Polich J (2002). P300 and slow wave from oddball and single-stimulus visual tasks: inter-stimulus interval effects. International Journal of Psychophysiology, 45(3), 187–196. [PubMed: 12208526]

Thompson TW, Waskom ML, Garel KL, Cardenas-Iniguez C, Reynolds GO, Winter R, & Gabrieli JD (2013). Failure of working memory training to enhance cognition or intelligence. PLoS ONE, 8(5), e63614.

Tsai N, Buschkuehl M, Kamarsu S, Shah P, Jonides J, & Jaeggi SM (2018). (Un) Great expectations: the role of placebo effects in cognitive training. Journal of Applied Research in Memory and Cognition, 7(4), 564–573. [PubMed: 31660288]

Urbánek T, & Marček V (2015). Investigating the effectiveness of working memory training in the context of Personality Systems Interaction theory. Psychological Research, 80(5), 877–888. [PubMed: 26208631]

Vartanian O, Jobidon M-E, Bouak F, Nakashima A, Smith I, Lam Q, & Cheung B (2013). Working memory training is associated with lower prefrontal cortex activation in a divergent thinking task. Neuroscience, 236, 186–194. [PubMed: 23357116]

von Bastian CC, & Eschen A (2016). Does working memory training have to be adaptive? Psychological Research, 80(2), 181–194. [PubMed: 25716189]

von Bastian CC, & Oberauer K (2014). Effects and mechanisms of working memory training: a review. Psychological Research, 78(6), 803–820. [PubMed: 24213250]

Waris O, Soveri A, & Laine M (2015). Transfer after working memory updating training. PloS ONE, 10(9), e0138734.

West RL, Welch DC, & Thorn RM (2001). Effects of goal-setting and feedback on memory performance and beliefs among older and younger adults. Psychology and Aging, 16(2), 240. [PubMed: 11405312]

Zajac-Lamparska L, & Trempala J (2016). Effects of working memory and attentional control training and their transfer onto fluid intelligence in early and late adulthood. Health Psychology Report, 4(1), 41–53.

Zhao X, Xu Y, Fu J, & Maes JH (2018). Are training and transfer effects of working memory updating training modulated by achievement motivation? Memory & Cognition, 46(3), 398–409. [PubMed: 29185201]

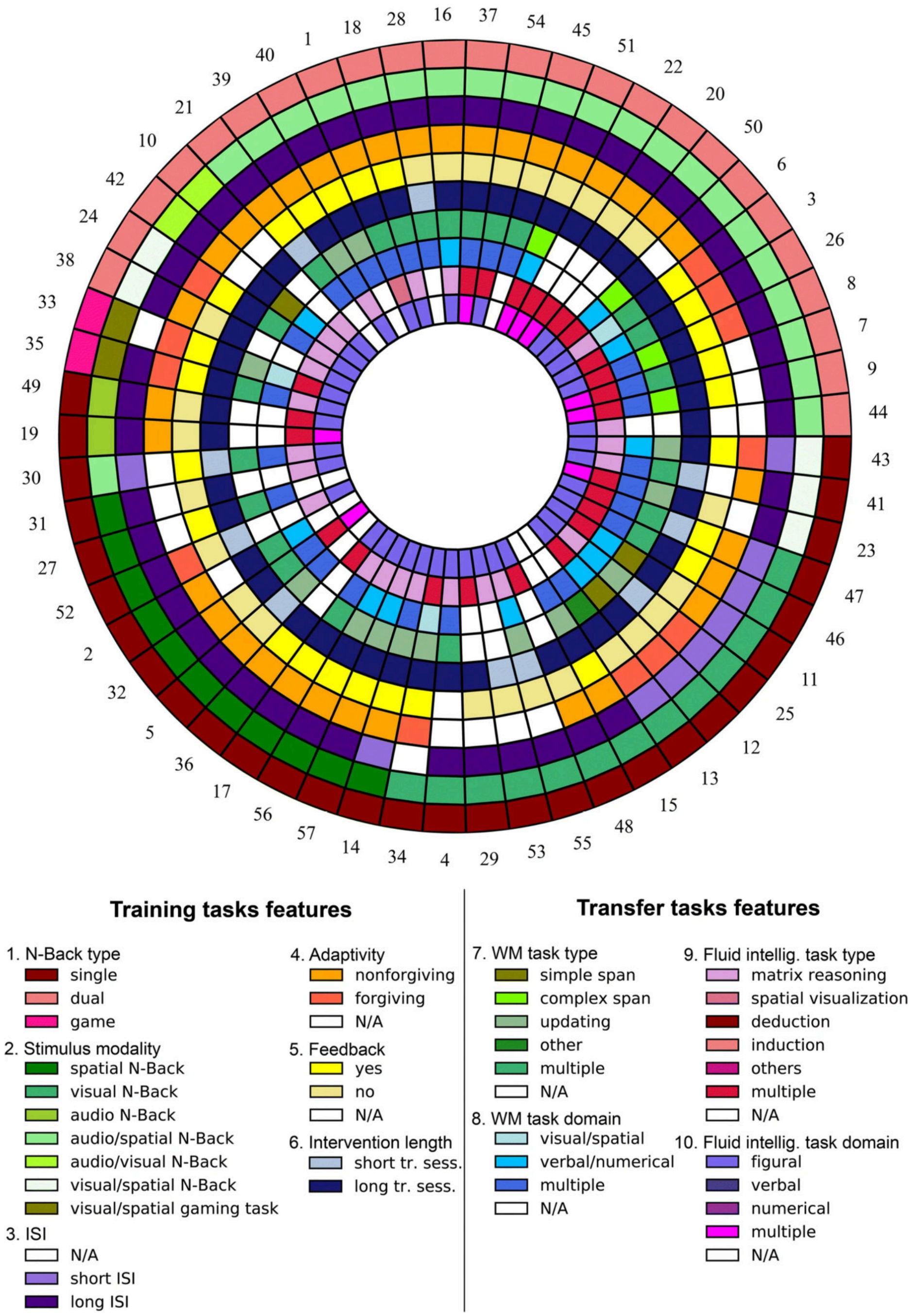

Fig. 1.

Diversity of training and transfer procedures. Each circle contains 57 sectors, each one corresponding to an N-back training group included in this review (see Table 1). The six outer circles reflect training task features whereas the four inner circles reflect transfer task features. Starting from the outer circle, each sector is colored in terms of N-back type (1) stimulus modality, (2) inter-stimulus interval (ISI), (3) adaptivity (forgiving vs. non-forgiving), (4) feedback, (5) intervention length (short < 10 sessions ≤ long), (6) WM (transfer) task type, (7) WM (transfer) task domain, (8) fluid intelligence (transfer) task type, (9) and fluid intelligence (transfer) task domain (10)

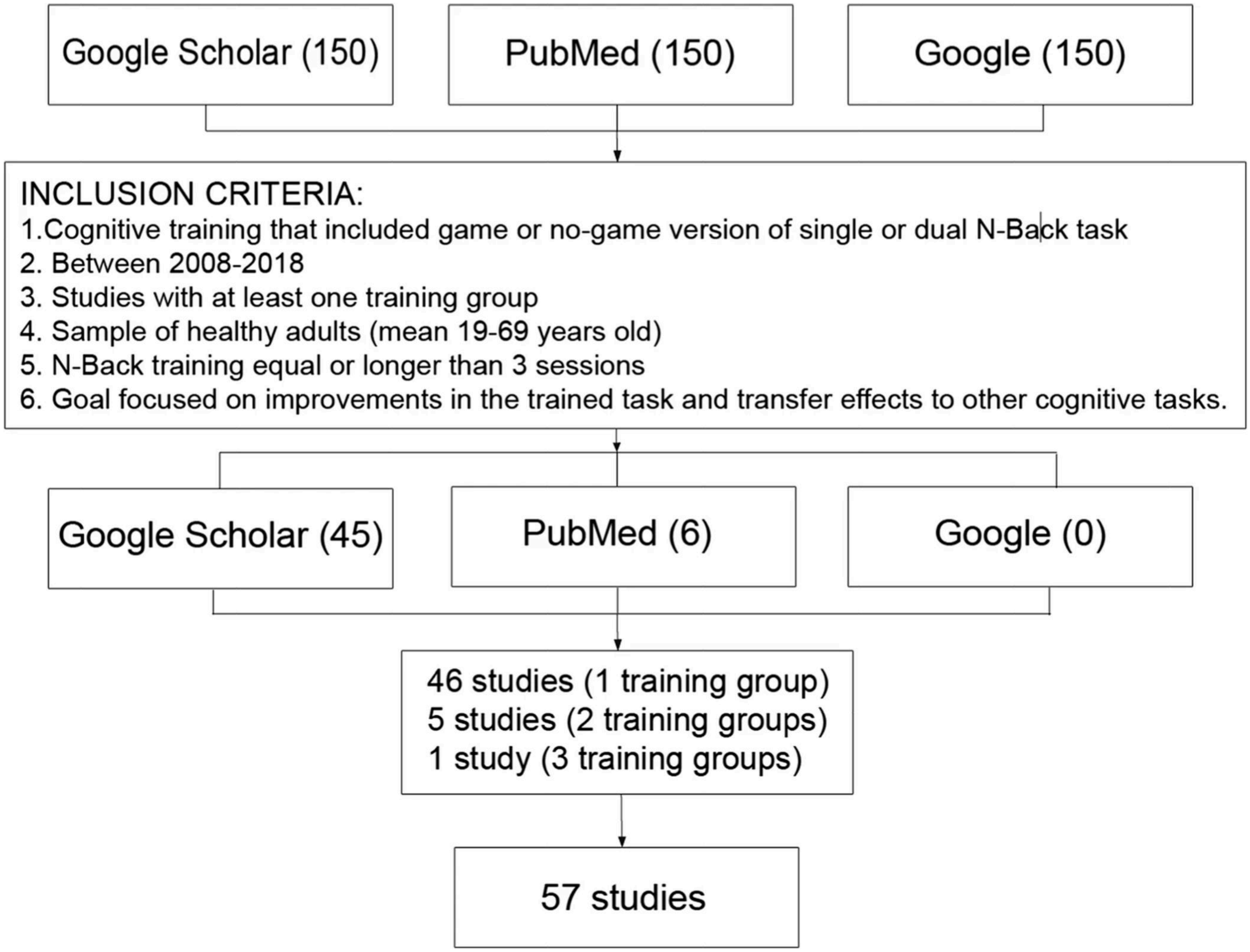

Fig. 2.
Search for literature and screening process

Table 1

Summary of training features from the 56 studies selected: transfer effect tests, N-back type, N-back modality, first/last-session N-back level, blocks, feedback, inter-stimulus interval (ISI), intervention length/training sessions (short < 10 sessions ≤ long), adaptivity, control group, years of education, single/double blind, strategies, motivation/expectation, payment, main results, and transfer tasks

Studies	Transfer effects tests	Single/dual	N-back modality	First/last-session N-back level (group average)	Blocks	Feedback	ISI	Training sessions	Adaptivity
1. Anguera et al. (2012)	Within and beyond WM	Dual	Audio/spatial	3-back to 5-back	9 blocks (20 trials/ block)	Yes	Long	Long	Non-forgiving
2. Beavon (2012)	Within and beyond WM	Single	Spatial	3-back to 5-back	15 blocks (20 trials/ block)	N/A	Long	Long	Non-forgiving
3. Blacker et al. (2017)	Within and beyond WM	Dual	Audio/spatial	1-back to 2-back	20 blocks (20 trials/ block)	Yes	Long	Long	Forgiving
4. Bürki et al. (2014)	Within and beyond WM	Single	Visual	3-back to 5-back	15 block (30 trials/ block)	N/A	Long	Long	N/A
5. Buschkuehl et al. (2014)	Within WM	Single	Spatial	4-back to 6-back	15 blocks (20 + *n* trials/block)	No	Long	Short	Non-forgiving
6. Chooi and Thompson (2012)	Beyond WM	Dual	Audio/spatial	2-back to 4-back	N/A	N/A	Long	Long	Non-forgiving
7. Clark et al. 2017a, b)	Beyond WM	Dual	Audio/spatial	2-back to 6-back	4 blocks (20 trials/ block)	Yes	Long	Long	N/A
8. Clouter (2013)	Within and beyond WM	Dual	Audio/spatial	2-back to 3-back	20 blocks (20 trials/ block)	Yes	Long	Long	Forgiving
9. Colom et al. (2013)	Beyond WM	8 single sessions followed by 16 dual sessions	Audio/spatial	Dual sessions: 2-back to 5-back	N/A	Yes	Long	Long	N/A
10. Qiu et al. (2009)	Beyond WM	Dual	Audio/visual	2-back to 3-back	20 blocks (20 + *n* trails/block)	N/A	Long	Long	Non-forgiving
11. Heinzel et al. (2014)	Within and beyond WM	Single	Visual	N/A	27 blocks (20–28 trials/block)	No	Short	Long	Non-forgiving
12. Heinzel et al. (2016)	Within and beyond WM	Single	Visual	N/A	36 blocks	No	Short	Long	Forgiving
13. Heinzel et al. (2017)	Within WM	Single	Visual	N/A	36 blocks	No	Short	Long	Forgiving
14. Hogrefe et al. (2017)	Within WM	Single	Spatial	N/A	10 blocks (20 trials/ each)	Yes	Short	Long	Non-forgiving
15. Hussey et al. (2017)	Within WM	Single	Visual	N/A	2 blocks (15 trials/ block)	Yes	Long	Long	Non-forgiving

Studies	Transfer effects tests	Single/dual	N-back modality	First/last-session N-back level (group average)	Blocks	Feedback	ISI	Training sessions	Adaptivity
16. Jaeggi et al. (2008)	Beyond WM	Dual	Audio/spatial	3-back to 5-back	20 blocks (20 trials/ block)	No	Long	Long	Non-forgiving
17. Jaeggi et al. 2010a, b); study 2	Within and beyond WM	Single	Spatial	3-back to 7-back	9 blocks (20 trials/ block)	Yes	Long	Long	Non-forgiving
18. Jaeggi et al. 2010a, b); study 2	Within and beyond WM	Dual	Audio/spatial	2-back to 4-back	9 blocks (20 trials/ block)	Yes	Long	Long	Non-forgiving
19. Jaeggi et al. (2014)	Within and beyond WM	Single	Audio	3-back to 6-back	15 blocks (20 trials/ block)	No	Long	Long	Non-forgiving
20. Jaeggi et al. (2014)	Within and beyond WM	Dual	Audio/spatial	2-back to 4-back	15 blocks (20 trials/ block)	No	Long	Long	Non-forgiving
21. Jonasson (2014)	Within WM	Dual	Audio/spatial	N/A	15 rounds	Yes	Long	Short	Non-forgiving
22. Katz et al. (2018)	Within and beyond WM	Dual	Audio/spatial	2-back to 4-back	15 blocks (20 trials/ block)	No	Long	Long	Non-forgiving
23. Kühn et al. (2013)	Within WM	Single	Visual/spatial (numerical updating and spatial N-back)	N/A	8 blocks (39 trials/ block)	No	Long	Long	N/A
24. Kundu et al. (2013)	Within WM	dual	Visual/spatial	N/A	25 blocks (25 trials/ block)	Yes	Long	Long	Forgiving
25. Küper and Karbach (2016)	Within and beyond WM	Single	Visual	3-back to 5-back	15 blocks (20 trials/ block)	No	Short	Short	Forgiving
26. Lawlor-Savage and Goghari (2016)	Within and beyond WM	Dual	Audio/spatial	1-back to 4-back	15 blocks (20 trials/ block)	Yes	Long	Long	Forgiving
27. Li et al. (2008)	Within WM	Single	Spatial (mental spatial shifting and updating)	N/A	4 blocks (22 trials/ block)	Yes	Long	Long	N/A
28. Lilienthal et al. (2013)	Within WM	Dual	Audio/spatial	2-back to 4-back	20 blocks (20 trials/ block)	No	Long	short	Non-forgiving
29. Loosli et al. (2016)	Within and beyond WM	Single	Visual	2-back	(100 trials/block)	No	Long	Long	N/A
30. Maraver et al. (2016)	Within and beyond WM	Single	Audio/spatial (N-back, WM search, WM updating)	1-back to 3-back	Not fixed (18 trials/ block)	Yes	Short	Short	Non-forgiving
31. Marček (2015)	Beyond WM	Single	Spatial	N/A	20 blocks (20 trials/ block)	No	Long	Long	N/A
32. Minear et al. (2016)	Within and beyond WM	Single	Spatial	4-back to 5-back[a]	15 blocks (20 trials/ block)	No	Long	Long	Non-forgiving
33. Mohammed et al. (2017); game N-back	Within and beyond WM	Game	Visual/spatial gaming task	3-back to 4-back[a]	8–15 blocks (20–40 trials/block)	Yes	N/A	Long	Forgiving

Studies	Transfer effects tests	Single/dual	N-back modality	First/last-session N-back level (group average)	Blocks	Feedback	ISI	Training sessions	Adaptivity
34. Mohammed et al. (2017); standard N-back;	Within and beyond WM	Single	Visual	3-back to 5-back[a]	8–15 blocks (20–40 trials/block)	Yes	N/A	Long	Forgiving
35. Nagle et al. (2015)	Within and beyond WM	Game	Visual/spatial gaming task	2-Back to 3-Back	4 blocks (15 trials/ block)	Yes	Long	Long	Forgiving
36. Preece (2012)	Beyond WM	Single	Spatial	3-back to 5-back	15 rounds	Yes	Long	Long	Non-forgiving
37. Redick et al. (2013)	Within and beyond WM	Dual	Audio/spatial	2-back to 4-back	20 blocks (20 trials for each)	No	Long	Long	Non-forgiving
38. Rudebeck et al. (2012)	Within and beyond WM	Dual	Visual/spatial	LG: 2-back to 2-back; HG: 2-back to 3-back	12 blocks (30 trials for each)	No	Long	Long	Non-forgiving
39. Salminen et al. (2012)	Within and beyond WM	Dual	Audio/spatial	2-back to 5-back	20 blocks (22 trials for each)	Yes	Long	Long	Non-forgiving
40. Salminen et al. (2015)	Within WM	Dual	Audio/spatial	2-back to 5-back	20 blocks (22 trials for each)	Yes	Long	Long	Non-forgiving
41. Schwarb et al. (2015)	Within and beyond WM	Single	Visual/spatial	4-back to 6-back	18 blocks (20 trials/ block)	N/A	Long	Short	Non-forgiving
42. Schweizer et al. (2011)	Beyond WM	Dual	Audio/visual	4-back to 7-back	20 blocks (20 trials/ block)	N/A	Long	Long	Non-forgiving
43. Shahar and Meiran (2015)	Beyond WM	Single	Visual/spatial	1-back to 3-back	10.7 blocks (64 trials/ block)	Yes	Short	Long	Forgiving
44. Smith et al. (2013)	Beyond WM	Dual	Audio/spatial	3-back- 5-back	(20 trials/block)	N/A	Long	Long	N/A
45. Soveri et al. 2017a, b)	Within WM	Dual	Audio/spatial	3-back to 4-back	20 blocks (20 trials/ block)	No	Long	Long	Non-forgiving
46. Stepankova et al. (2013); long intervention	Within and beyond WM	Single	Visual	2-back to 3-back	20 blocks (20 trials/ block)	Yes	Short	Long	Non-forgiving
47. Stepankova et al. (2013); short intervention	Within and beyond WM	Single	Visual	2-back to 4-back	20 blocks (20 trials/ block)	Yes	Short	Short	Non-forgiving
48. Stephenson and Halpem (2013); single visual N-back	Beyond WM	Single	Visual	N/A	(20 trials/block)	No	Long	Long	Non-forgiving
49. Stephenson and Halpem (2013); single audio N-back	Beyond WM	Single	Audio	N/A	(20 trials/block)	No	Long	Long	Non-forgiving
50. Stephenson and Halpem (2013); dual N-back	Beyond WM	dual	Audio/spatial	N/A	(20 trials/block)	No	Long	Long	Non-forgiving
51. Thompson et al. (2013)	Within and beyond WM	Dual	Audio/spatial	3-back to 5-back	30 blocks (20 trials/ block)	No	Long	Long	Non-forgiving
52. Urbänek and Marcek (2015)	Beyond WM	Single	Spatial	2-back to 4-back	15 blocks (20 trials/ block)	No	Long	short	Forgiving
53. Vartanian et al. (2013)	Beyond WM	Single	Visual	N/A	4	No	Long	Short	N/A

Studies	Transfer effects tests	Single/dual	N-back modality	First/last-session N-back level (group average)	Blocks	Feedback	ISI	Training sessions	Adaptivity
54. Waris et al. (2015)	Within and beyond WM	Dual	Audio/spatial (selective updating task, moving figures, dual N-back)	2-back to 4-back	(20 trials/block)	No	Long	Long	Non-forgiving
55. Zajac-Lamparska and Trempala (2016)	Beyond WM	Single	Visual	N/A	(20 trials/block)	No	Long	Short	N/A
56. Zhao et al. (2018); HAM group	Within and beyond WM	Single	Spatial	3-back to 6-back	15 blocks (20 trials/block)	Yes	Long	Long	Non-forgiving
57. Zhao et al. (2018); LAM group	Within and beyond WM	Single	Spatial	2-back to 4-back	15 blocks (20 trials/block)	Yes	Long	Long	Non-forgiving

Control group	Years of education info	Single/double-blind info	Strategies info	Motivation/expectation info	Payment	Main results (transfer effects)	Transfer tasks
Active (knowledge training)	No	No	No	Yes	Low	Transfer to the 3-back task and visuospatial tasks. No transfer to visuomotor task.	N-back, operation span (OSPAN), card rotation, digit symbol substitution (WAIS-R), visuomotor adaptation
Active (knowledge training)	No	No	No	No	N/A	No transfer effects to STM and WM.	Numbers reversed (WJ-III), auditory WM (WJ-III)
Active (adaptive non-WM task called permuted rule operations)	Yes	No	No	Yes	High	Transfer to near WM task (objects N-back). No far-transfer to fluid intelligence.	N-back, symmetry span, spatial locations, and relations, BOMAT
Active (implicit sequence learning training), passive	Yes	No	Yes	No	N/A	Near-transfer effects to a similar spatial N-back task.	N-back, numerical updating, reading span, RSPM, RAPM, Stroop, letter and number comparison (pattern comparison), SRT
Active (knowledge training)	No	No	No	No	Low	Near transfer to N-back task with different stimuli.	N-back
Active (dual 1-back training), passive	Yes	No	No	No	Low	No transfer effects.	OSPAN, vocabulary (Mill-Hill, PMA), word beginning and ending, Colorado perceptual speed, identical pictures, finding A's, card rotation (ETS), paper folding (ETS), mental rotation (Shepard-Metzler), RAPM
Active (processing speed training), passive	Yes	Yes (single blindi	No	Yes	High	No transfer effects.	RAPM, WAIS-IV, CCFT, lexical decision, OSPAN, spatial delayed response task
Active (dual 1-back training)	No	Yes (single blindi	No	Yes	N/A	Transfer effects to fluid intelligence, WM and conflict resolution.	CFIT, Stroop, Monty Hall Problem, OSPAN, symmetry span
Passive	No	No	No	Yes	N/A	No transfer to fluid intelligence, crystallized intelligence, WM and attention control.	RAPM, abstract reasoning (DAT-AR), inductive reasoning (PMA-R), WM: reading span, computation span, dot matrix

Control group	Years of education info	Single/ double-blind info	Strategies info	Motivation/ expectation info	Payment	Main results (transfer effects)	Transfer tasks
Passive	No	No	No	No	N/A	Transfer effects to fluid intelligence, larger for the training group.	RSPM
Passive	Yes	No	No	No	N/A	Transfer to verbal fluency and digit symbol, digit span Fwd, digit symbol, and CERAD Del Recall.	Digit span Fwd (WAIS), Digit span Bwd (WAIS), Recall (CERAD) (immediate and delayed), digit symbol (WAIS), verbal fluency (COWAT), RSPM, figural relations (LPS)
Passive	Yes	No	No	No	N/A	Transfer to Sternberg task, processing speed, executive functions and fluid intelligence.	Digit span Fwd (Weschler), digit span Bwd (Weschler), digit symbol substitution (Weschler), D2, Stroop, verbal fluency (COWAT), RSPM, figural relations (LPS)
Passive	Yes	No	No	No	N/A	Near-transfer to N-back.	Visual and auditory single tasks, dual task
Active (N-back training with no immediate feedback), passive	No	No	No	No	N/A	Transfer effects to the N-back task and to numerical memory updating.	N-back, task switching, Flanker, Stroop, numerical updating, RAPM
Active (adaptive N-back training without lures; non-adaptive 3-Back training without lures)	Yes	Yes (double blindi	No	No	High	Transfer to untrained memory and language conditions.	N-back, recognition memory, Verb generation, Stroop, garden-path recovery, relative clause processing
Passive	No	No	No	No	N/A	Transfer to fluid intelligence based on training amount.	RAPM, BOMAT
Passive	No	No	No	No	High	Transfer effects to fluid intelligence.	N-back, OSPAN, RAPM, BOMAT
Passive	No	No	No	No	High	Transfer effects to fluid intelligence.	N-back, OSPAN, RAPM, BOMAT
Active (knowledge training)	No	No	No	Yes	High	Transfer to fluid intelligence.	RAPM, CFIT, BOMAT, surface development test (ETS), space relations (DAT), form board test (ETS), interference test (ETS), reading comprehension (AFOQT), verbal analogies, digit symbol (WAIS)
Active (knowledge training)	No	No	No	Yes	N/A	Transfer to fluid intelligence.	RAPM, CFIT, BOMAT, surface development test (ETS), space relations (DAT), form board test (ETS), interference test (ETS), reading comprehension (AFOQT), verbal analogies, digit symbol (WAIS)
Active (face-name recall training)	No	No	No	No	N/A	No transfer effects.	N-back, OSPAN, addition, trail making (TMT), dual task
Active (knowledge training)	No	No	No	Yes	High	Transfer to visuospatial composite.	RAPM, CFIT, BOMAT, surface development test (ETS), space relations (DAT), form board test (ETS), interference test (ETS), reading comprehension (AFOQT), verbal analogies, digit symbol (WAIS)
Active (numerical updating and N-back training with fixed difficulty level)	No	No	No	No	High	Improvements on untrained working-memory tasks.	N-back, spatial updating, figurai and numerical reasoning (Berlin Intelligence Structure Test)

Control group	Years of education info	Single/ double-blind info	Strategies info	Motivation/ expectation info	Payment	Main results (transfer effects)	Transfer tasks
Active (adaptive Tetris)	No	No	No	No	N/A	Transfer to stimulus processing, STM and visual search performance	Visual-array comparison task, visual search, OSPAN, Stroop, RAPM
Passive	Yes	No	No	No	Low	Near-transfer to a WM updating task. No far-transfer to switch costs, Stroop and matrix reasoning tasks.	N-back, task switching, Stroop, RAPM, digit symbol substitution, Spot-a-word
Active (1-Back training)	No	Yes (double blind)	No	No	N/A	No transfer to fluid intelligence.	Digit span (WAIS), symbol search (WAIS), coding (WAIS), OSPAN, RAPM, CFIT
Passive	No	No	No	No	High	Transfer to a more complex spatial N-back task and numerical N-back task. No far transfer to complex span.	N-back, OSPAN, rotation span, decision speed
Active (non-adaptive dual N-back), passive	No	No	No	No	N/A	Transfer effects from adaptive training to a running span task (focus of attention).	Cued recall, focus-switching, grid span, operation span, running span
Active (recent-probes and N-back training with low proactive interference)	Yes	Yes (double blindi	No	No	Low	No transfer effects to untrained tasks.	Verb generation, paired associates, Stroop, digit symbol substitution (WAIS-R), TONI
Active (processing speed training), passive	No	No	No	Yes	Low	Transfer effects to reasoning for the inhibitory control training group.	N-back, Stroop, OSPAN, stop signal, AX-CPT, RAPM
Active (non-adaptive Sudoku)	No	No	Yes	Yes	Low	No transfer for the single N-back group, transfer to RAPM for the control group (triple N-back task).	RAPM, BOMAT
Active (non-adaptive N-back training, real time strategy video game)	No	No	Yes	Yes	High	Near-transfer effects to a different N-back task for both adaptive and non-adaptive N-back training group.	N-back, Speed of processing, dot judgment, array matching, (letter-digit-arrow-circle-reading-operation-letter-number-rotation-alignment) span, attention network, Simon, nonsense Syllogisms (ETS), inference tests (ETS), RPM, CFIT, mathematical aptitude (ETS)
None	No	No	No	Yes	Low	Transfer to untrained N-back, Far transfer to DRM free recall (falsely remembered), DRM (recognition), Space relations, Surface development, Form board, Delay discounting	N-back, AX-CPT, DRM, space relations (DAT), surface development (ETS), form board test (ETS), BOMAT, learning from lectures, Math, delay discounting
None	No	No	No	Yes	Low	Transfer to untrained N-back, far transfer to DRM free recall (falsely remembered), space relations, surface development, form board, Math, delay discounting	N-back, AX-CPT, DRM, space relations (DAT), surface development (ETS), form board test (ETS), BOMAT, learning from lectures, Math, delay discounting
None	No	No	No	Yes	N/A	No transfer effects.	RAPM, digit span Fwd, digit span Bwd, N-back
Active (vocabulary and knowledge training)	No	No	No	No	N/A	No transfer effects to fluid intelligence compared with the control group.	Figure weights (WAIS), RAPM
Active (visual search training), passive	No	No	Yes	Yes	High	No transfer effects in fluid intelligence, WM, crystallized intelligence and perceptual speed tasks.	RAPM, RSPM, CFIT, paper folding, letter sets, number series, inference, verbal analogies, SynWin, control tower, ATClab, symmetry span, running letter span, vocabulary, general knowledge, letter and number comparison

Author Manuscript
Author Manuscript
Author Manuscript

Control group	Years of education info	Single/ double-blind info	Strategies info	Motivation/ expectation info	Payment	Main results (transfer effects)	Transfer tasks
Passive	Yes	No	No	No	N/A	Transfer effects to episodic memory and fluid intelligence.	BOMAT, recognition memory
Passive	No	No	No	No	N/A	Transfer to WM updating task, switching situation task and attentional processing. No transfer to reasoning or dual N-back task.	Updating (AV numbers, VS black bars that was shown in 4 different locations), dual task, Task switching, attentional blink, RAPM
Passive	Yes	No	No	No	Low	Transfer effects to a WM updating task for both young and older adults.	Updating (AV numbers, VS black bars, one block with VS another block with AV), task switching, attentional blink
Passive	No	No	No	No	Low	Transfer effects to visual short-term memory capacity.	OSPAN, Symmetry span, RAPM, motion interference, rapid decision-making, change detection, short-term recall
Active (feature matching training)	Yes	No	/	No	N/A	Transfer to fluid intelligence. Transfer to emotional Stroop task only for affective training group.	RPM, Stroop, digit span
Passive	No	Yes (double blind)	No	No	Low	Transfer to processing speed compared with the control group.	Digits updating task, shape/digit classification task, Stroop, stop-signal, RAPM
Active (strategy video game training), passive	No	No	No	No	N/A	No transfer effect to fluid intelligence.	RAPM
Active (non-adaptive game Bejeweled 2)	Yes	No	No	Yes	Low	Transfer effects to different single N-back task and to a WM updating task. No transfer effects for active or passive groups.	N-back, verbal running span, visuospatial running span, number substitution, digit span Fwd, digit span Bwd, Corsi Block
Passive	Yes	No	No	No	N/A	Transfer effects to WM and visuospatial skills.	Digit span (WMS), letter number sequencing (WMS-III), block design (WAIS), matrix reasoning (WAIS)
Passive	Yes	No	No	No	N/A	Transfer effects to WM and visuospatial skills.	Digit span (WMS), letter number sequencing (WMS-III), block design (WAIS), matrix reasoning (WAIS)
Active (spatial matrix span training), passive	Yes	No	No	No	N/A	Transfer to 4 fluid intelligence tests (APM, Cattell, WASI, BETA-III).	RAPM, CFIT, WASI, BETA, mental rotation, paper folding, vocabulary, lexical decision
Active (spatial matrix span training), passive	Yes	No	No	No	N/A	Transfer to 3 fluid intelligence tests (APM, Cattell, WASI).	RAPM, CFIT, WASI, BETA, mental rotation, Paper folding, Vocabulary, Lexical decision
Active (spatial matrix span training), passive	Yes	No	No	No	N/A	Transfer to 4 fluid intelligence tests (APM, Cattell, WASI, BETA-III).	RAPM, CFIT, WASI, BETA, mental rotation, paper folding, vocabulary, lexical decision
Active (multiple object tracking training), passive	No	No	Yes	Yes	High	No transfer effects to fluid intelligence and other cognitive tasks.	Operation span, reading span, RAPM, block design (WASI), WAIS-III, reading comprehension (Nelson Denny), digit-symbol coding (WAIS-III), visual matching (Woodcock-Johnson III), pair cancelation (Woodcock-Johnson III)
Active (non-adaptive Sudoku)	No	No	No	No	Low	No transfer effects to fluid intelligence.	RAPM, BOMAT

Control group	Years of education info	Single/ double-blind info	Strategies info	Motivation/ expectation info	Payment	Main results (transfer effects)	Transfer tasks
Active (reaction time training)	No	No	No	No	N/A	Transfer effects to fluid intelligence.	RAPM, alternate uses task (AUT)—test of divergent thinking
Active (Angry Birds, Bejeweled 2, Peggie)	No	No	No	Yes	Low	Near-transfer effects in a different N-back task, WM updating and in a WM task.	Verbal running span, digit span, Corsi block, set shifting, visuospatial running span, CFIT, Simon, number substitution, numerical updating, N-back
Active (attentional control training), passive	Yes	No	No	No	N/A	Weak transfer effects to fluid intelligence.	Attentional control, RSPM, N-back
None	No	No	No	Yes	N/A	Near-transfer effects to WM. No far-transfer effects to executive functions and fluid intelligence.	N-back task, running digit span, Go/no-go, Stroop, task switching, RAPM.
None	No	No	No	Yes	N/A	Near-transfer effects to WM. No far-transfer effects to executive functions and fluid intelligence.	N-back task, running digit span, Go/no-go, Stroop, task switching, RAPM.

Legend: *HAM*, high achievement motivated group; *LAM*, low achievement motivated group; *WM*, working memory; *STM*, short-term memory; *COWAT*, Controlled Oral Word Association Test; *DAT*, Differential Aptitude Test; *WAIS*, Wechsler Adult Intelligence Scale; *WMS*, Wechsler Memory Scale; *WASI*, Wechsler Abbreviated Scale of Intelligence; *WJ*, Woodcock-Johnson; *ETS*, Educational Testing Service Kit; *RSPM*, Raven's Standard Progressive Matrices; *RAPM*, Raven's Advanced Progressive Matrices; *CFIT*, Culture Fair Intelligence Test; *LPS*, Leistungspriifsystem; *PMA-R*, Primary Mental Abilities Battery; *TONI*, Test of Non-verbal Intelligence; *WASI*, Wechsler Abbreviated Scales of Intelligence; *BOMAT*, Bochumer Matrizen Test; *AFOQT*, Air Force Officer Qualifying Test; *BIST*, Berlin Intelligence Structure Test; *CERAD*, Consortium to Establish a Registry for Alzheimer's Disease; *SRT*, Simple Reaction Time; *AX-CPT*, AX-continuous performance task

[a] Group average maximum N-back level

Table 2

Transfer tasks categorized by cognitive domain

COGNITIVE DOMAIN	NUMBER OF STUDIES
WORKING MEMORY	
Simple Span	
Corsi block	2
Digit span (forward, backward)	14
Grid span	1
Arrow/circle span	2
Complex span	
Operation span (OSPAN)	13
Symmetry span	4
Reading span	4
Dot matrix	1
Computation span	1
Rotation span	2
Alignment span	1
Updating	
N-Back	19
Numerical updating	6
Spatial updating	3
Running digit span	1
Running letter span	2
Verbal running span	2
Visuo-spatial running span	2
Visuospatial and auditory-verbal updating	1
Number substitution	2
Other	
Letter number sequencing (WAIS)	1
Letter number sequencing (WMS-III)	1
Auditory WM (WJ-III)	1
Spatial delayed response task	1
Visual array comparison task	2
Change detection task	1
Short term recall task	1
Cued recall span task	1
Focus-switching task	1
Delayed match to sample (single and dual)	2
Spatial locations and relations	1
LTM	
Recall (CERAD) (delayed, immediate)	2
Recognition memory	2

COGNITIVE DOMAIN	NUMBER OF STUDIES
Paired associates	1
Learning from lectures	2
FALSE MEMORY	
Deese-Roediger-McDermott	2
VISUAL SEARCH	
Visual search	1
Symbol search	1
Finding A's	1
Identical pictures	1
FLUID INTELLIGENCE *	
Letter sets	1
Inference	1
Space relations (DAT)	5
Abstract reasoning (DAT-AR)	1
Matrix resoning (BETA-III)	1
Matrix reasoning (WAIS)	1
Block design (WAIS/WASI)	2
Figure weights (WAIS)	1
Nonsense syllogisms (ETS)	1
Inference tests (ETS)	1
Paper folding (ETS)	3
Surface development test (ETS)	5
Form board test (ETS)	5
Interference test (ETS)	4
RSPM	8
RAPM	27
BOMAT	11
Space relations	1
Figural relations (LPS)	2
Inductive reasoning (PMA-R)	1
CFIT	10
TONI	1
Number series	1
Mental rotation (Shepard-Metzler)	2
Figural and numerical reasoning (BIST)	1
Verbal analogies	4
Reading comprehension (AFOQT)	4
Card rotation	2
CRYSTALLIZED / GENERAL INTELLIGENCE	
Verbal fluency (COWAT)	2
Lexical decision	2
Word beginning and ending	1

COGNITIVE DOMAIN	NUMBER OF STUDIES
Verb generation	2
Vocabulary (Mill-Hill, PMA)	4
General knowledge	1
WAIS-IV	1
Spot a word	1
Similarities (WASI)	1
Vocabulary (WASI)	1
Extended Range Vocabulary Test (ETS)	1
READING	
Nelson-Denny Comprehension	1
Lexical Decision Test	1
Nelson-Denny Reading Rate	1
MATH	
Mathematical aptitude (ETS)	1
Arithmetic aptitude test (ETS)	1
Addition	1
Math	2
PROCESSING SPEED	
Letter and number comparison (pattern comparison)	4
Simon	2
Coding (WAIS)	2
Visual matching (WJ- III)	1
Colorado Perceptual Speed Test	1
Shape/Digit Classification	1
SRT	2
Decision speed	1
Dot judgement	1
Digit symbol substitution (WAIS-R)	8
DECISION MAKING / PROBLEM-SOLVING	
Monty Hall problem	1
Rapid decision making	1
Delay discounting	2
Relative clause processing	1
ATTENTION / COGNITIVE CONTROL	
Garden path recovery	1
Set shifting	2
Trail making test	1
Stroop	12
Task switching	5
Focus switching	1
Attentional blink	2

COGNITIVE DOMAIN	NUMBER OF STUDIES
Pair cancellation (WJ-III)	1
Stop-signal	2
Go/no go	1
Flanker	1
Attention network	1
Motion interference	1
AX-CPT	3
D2	1
Attentional control	1
Visuomotor adaptation	1
MOTOR LEARNING	
Control tower	1
MULTITASKING	
Synwin	1
Atclab	1
DIVERGENT THINKING	
Alternate Uses Task	1

* Fluid intelligence classification was based on Au et al. (2015), Table S3.

Legend: WM = working memory; LTM = long-term memory; COWAT = Controlled Oral Word Association Test; DAT = Differential Aptitude Test; WAIS = Wechsler Adult Intelligence Scale; WMS = Wechsler Memory Scale; WASI = Wechsler Abbreviated Scale of Intelligence; WJ = Woodcock-Johnson; ETS = Educational Testing Service Kit; RSPM = Raven's Standard Progressive Matrices; RAPM = Raven's Advanced Progressive Matrices; CFIT = Culture Fair Intelligence Test; LPS = Leistungsprüfsystem; PMA-R = Primary Mental Abilities Battery; TONI = Test of Nonverbal Intelligence; WASI = Wechsler Abbreviated Scales of Intelligence; BOMAT = Bochumer Matrizen test; AFOQT = Air Force Officer Qualifying Test; BIST = Berlin Intelligence Structure Test; CERAD = Consortium to Establish a Registry for Alzheimer's Disease; SRT = Simple Reaction Time; Ax-CPT = Ax-continuous performance task