EDUCATIONAL PSYCHOLOGIST, *50*(2), 138–166, 2015

ISSN: 0046-1520 print / 1532-6985 online
DOI: 10.1080/00461520.2015.1036274

Routledge
Taylor & Francis Group

Does Working Memory Training Transfer? A Meta-Analysis Including Training Conditions as Moderators

Matthias Schwaighofer, Frank Fischer, and Markus Bühner
Department of Psychology
Ludwig Maximilians University of Munich, Germany

A meta-analysis was undertaken to reexamine near- and far-transfer effects following working-memory training and to consider potential moderators more systematically. Forty-seven studies with 65 group comparisons were included in the meta-analysis. Results showed near-transfer effects to short-term and working-memory skills that were sustained at follow-up with effect sizes ranging from $g = 0.37$ to $g = 0.72$ for immediate transfer and $g = 0.22$ to $g = 0.78$ for long-term transfer. Far-transfer effects to other cognitive skills were small, limited to nonverbal ($g = 0.14$) and verbal ($g = 0.16$) ability and not sustained at follow-up. Several moderators (e.g., duration of training sessions, supervision during training) had an influence on transfer effects, including far-transfer effects. We present principles for how best to improve working memory through training in the narrow-task paradigm and conjecture how best to improve basic cognitive functions in complex activity contexts.

In the past decade, interest among researchers and practitioners in working memory (WM) training has surged. The reason for this increased interest is that the idea of improving WM is typically associated with the idea that WM training effects might generalize to other cognitive functions. The results of a recent meta-analysis, however, cast doubt on the basic assumption that WM training produces transfer to other measures such as fluid intelligence (Melby-Lervåg & Hulme, 2013). Only 23 studies were included in Melby-Lervåg and Hulme's meta-analysis, but many more studies that examine transfer effects of WM training have been published since that meta-analysis. In addition, from an educational perspective it seemed strange that training conditions—or, more generally, the learning environment—were not considered systematically as moderators of transfer effects. Important from a theoretical perspective are to what extent and under which conditions training of a basic cognitive function such as WM would also lead to improvements in other cognitive functions (e.g., fluid intelligence). Of similar importance is the question of whether these improvements of other cognitive functions are practically relevant—for example, whether successful WM training would lead to better mathematical abilities. In this article, we present a meta-analysis that includes recently published studies and focuses on several potential moderators of transfer effects that have not been systematically investigated—most important are the training conditions. Before presenting our analysis, we explain how WM is conceptualized and how it is considered to influence performance in complex cognitive tasks. We also review the existing literature about the rationales behind WM training.

Correspondence should be addressed to Matthias Schwaighofer, Department of Psychology, Ludwig-Maximilians-Universität (LMU) München, Leopoldstr. 13, 80802 Munich, Germany. E-mail: M.Schwaighofer@psy.lmu.de

CONCEPTUALIZATION OF WM

WM has been conceptualized as "a limited capacity system, which temporarily maintains and stores information [and] supports human thought processes by providing an interface between perception, long-term memory and action" (Baddeley, 2003, p. 829). Short-term memory (STM) involves only the temporary storage of information (Shipstead, Redick, & Engle, 2012), whereas WM additionally requires the manipulation of information. Several models of WM exist (e.g., Baddeley, Allen, & Hitch, 2011; Cowan, 1995; Unsworth & Engle, 2007a). Although different terms are used within the various conceptualizations of WM, they share some commonalities. Most important, the majority of

WM researchers assume a limited capacity system and an attentional control component. These components consequently play a role in most tasks that measure working memory capacity (WMC).

Consistent with the distinction between STM and WM, Redick, Broadway, et al. (2012) differentiated between simple span tasks, such as digit span and Corsi Blocks, and complex span tasks, such as operation and symmetry span. Simple span tasks measure only the capacity of a short-term storage system, whereas complex span tasks also involve the processing of information, and therefore seem appropriate to assess WMC (Redick, Broadway, et al., 2012). For instance, in the automated operation span task, a mathematical operation is presented at first and the participant has to decide whether the result presented is right or wrong. Subsequently, a letter is presented, and after a variable number of operation-letter pairs, the participant has to remember the letters in their original order (Unsworth, Heitz, Schrock, & Engle, 2005). Unsworth and Engle (2007b) argued that "simple and complex span tasks largely measure the same basic subcomponent processes (e.g., rehearsal, maintenance, updating, controlled search) but differ in the extent to which these processes operate in a particular task" (p. 1038). The distinction between simple and complex span tasks made by Redick, Broadway, et al. (2012), however, is useful for separating measures of STM and WM. As Shipstead, Redick, and Engle (2010) noted, simple span tasks that exceed the limit of STM through lists that are longer than the capacity of STM require WM components. Therefore, Shipstead et al. (2010) suggested that complex span tasks are predictive of higher cognitive abilities, even with short lists, because they involve the processing component from the beginning of the task. Several training studies have differentiated between tasks measuring the verbal and visuospatial modalities of STM and WM using material with information stored in one of the two respective storages. Simple span tasks and complex span tasks differ in the material used in the tasks and capture different modalities of STM or WM, that is, verbal or visuospatial modality.

Besides the question of what distinguishes STM from WM, the nature of individual differences in WMC has been intensively investigated. Differences in WMC measured with complex span tasks reflect differences in executive attention (Kane, Conway, Hambrick, & Engle, 2007). Executive attention "reflects the ability to temporarily maintain goal-relevant information in primary memory and to retrieve information from secondary memory" (Redick, Broadway, et al., 2012). Primary memory can be seen as a capacity-limited system in which information could be held through continuous allocation of attention. Secondary memory, in contrast, refers to a larger, long-term system in which information lost from primary memory can be retrieved by means of a controlled, cue-dependent search. The assignment of cues selected for search and the way they are combined for search also require attention. Following this conceptualization, differences in WMC are related to differences in the active maintenance of information through continuous allocation of attention and retrieval of relevant information in the presence of interference through other stimuli (Unsworth & Engle, 2007a). According to this logic, domain-general differences in attention capabilities are important for the explanation of WMC differences. This conceptualization can be found in most WM training studies (Melby-Lervåg & Hulme, 2013). Thus, a higher WMC seems to be associated with better attention capabilities. Differences in these capabilities can, in turn, be related to differences in learning outcomes. Indeed, a large body of research examines the correlates of WMC and other cognitive abilities.

RELEVANCE OF WM FOR COMPLEX COGNITIVE TASKS

Attention capabilities seem to be important for the explanation of differences in WMC. Thus, unsurprisingly, differences in WMC are considered to be related to differences in performance in several basic attention tasks (Unsworth & Engle, 2007a). Differences in WMC have also been associated with a variety of higher cognitive achievements, such as mathematical abilities (Gathercole, Pickering, Knight, & Stegmann, 2004), reading comprehension (Daneman & Merikle, 1996), and chemistry performance (Tsaparlis, 2005). WMC has been found to be associated with problem solving (Bühner, Kröner, & Ziegler, 2008), multitasking (Bühner, König, Pick, & Krumm, 2006), and knowledge acquisition from hypermedia materials similar to Wikipedia (Banas & Sanchez, 2012). In addition, a strong association between WMC and fluid intelligence has been demonstrated repeatedly (e.g., Engle, Tuholski, Laughlin, & Conway, 1999; Oberauer, Süß, Wilhelm, & Wittmann, 2008; Redick, Unsworth, Kelly, & Engle, 2012).

TRAINING OF WM

Several studies on WM training have addressed the question of whether cognitive training can improve intellectual abilities and therefore may lead to transfer effects for other than the trained tasks. Considering the causes for differences in WMC, as mentioned in the preceding section, WM training should result in better executive attention. The improvement in the domain-general attention capability should lead to performance improvements in tasks meant to determine the capacity of WM (i.e., how many elements can be held in WM at a time) as well as tasks measuring other cognitive abilities that require executive attention (e.g., Raven's matrices; see Melby-Lervåg & Hulme, 2013). Improvements in WMC tasks and other types of tasks are referred to as *near-transfer effects* and *far-transfer* *effects*, respectively. In the present review, near-transfer effects include improvements in tasks that require short-term and working-memory components; far-transfer effects include any improvements attributable to WM training on tasks that require abilities beyond short-term and working-memory components (e.g., tasks that require mathematical abilities; an example is the arithmetic test used in the study by Van der Molen, Van Luit, Van der Molen, Klugkist, & Jongmans, 2010, which requires applying arithmetic operations such as addition, multiplication, or division).

Morrison and Chein (2011) stated that so-called core training is more suited than strategy training to reach far-transfer effects. Strategy training includes effective procedures to support encoding, retention, and/or recall from WM. In contrast, core training aims to limit the use of domain-specific strategies, to minimize automation, and to use tasks that are adaptive and typically demand frequent updating (Morrison & Chein, 2011). The difficulty levels of adaptive tasks continuously adjust to the performance of participants (Jaeggi, Studer-Luethi, et al., 2010).

The tasks used in WM training programs usually include maintaining and processing very simple stimuli from a narrow domain (e.g., blue squares, single digits) that can be considered to be relatively independent of prior knowledge in other domains. Thus, the trained tasks share only superficial commonalities (e.g., maintaining simple stimuli). A typical task used in core training is the dual n-back task. In this task, participants see a series of blue squares, each presented at one of eight different locations on a screen. The participants have to indicate when the position of a blue square matches the blue square presented n stimuli before. Spoken letters are presented simultaneously with the blue squares, and participants also have to indicate when a spoken letter matches a letter presented n stimuli before (Chooi & Thompson, 2012). Although stimuli used in the trained tasks are from a narrow domain, the idea of WM training interventions typically is that these tasks should improve a domain-general WM mechanism such as executive attention, which should be important for other cognitive abilities (e.g., mathematical abilities), and ultimately should affect solving problems and making decisions in completely different and much more complex domains.

Regarding the effectiveness of WM training, some optimistic perspectives have been offered. Klingberg (2010) concluded on the basis of selected studies that "WM training can induce improvements in performance in nontrained tasks that rely on WM and control of attention" (p. 322). Based on some recently published studies with children, Titz and Karbach (2014) argued that WM training can be beneficial for academic abilities, specifically those for language; reading; and, to a smaller extent, mathematics. Several published WM training studies demonstrate near- as well as far-transfer effects. For example, Brehmer, Westerberg, and Bäckman (2012) observed near-transfer effects to verbal and visual short-term memory. Alloway, Bibile, and Lau (2013) reported near-transfer effects to verbal and visuospatial WM, for example; Various authors showed far-transfer effects, inter alia, regarding mathematical reasoning (e.g., Holmes, Gathercole, & Dunning, 2009), reading comprehension (e.g., Chein & Morrison, 2010), and fluid intelligence (e.g., Jaeggi, Buschkuehl, Jonides, & Perrig, 2008; Jaušovec & Jaušovec, 2012).

However, other studies have raised several concerns about WM training (Melby-Lervåg & Hulme, 2013; Morrison & Chein, 2011; Shipstead et al., 2012). Morrison and Chein's (2011) review argued that there is a lack of control with respect to effort/expectancy effects (e.g., better results of experimental groups due to the learners' expectations of improvement) on outcome measures. Morrison and Chein noted that these and similar effects are not controlled by using adequate measures, such as self-reports and measures of motivation and commitment. To control for expectancy effects, Shipstead et al. (2012) suggested the inclusion of a control group that receives training that does not tax WM but is still adaptive, similar to the typical WM training tasks. Another limitation concerns the measurement of transfer effects. Morrison and Chein (2011) stressed that training paradigms, as well as tasks used for the assessment of transfer effects, are highly variable. These authors also raised concerns about the demonstration of transfer effects using tasks that capture only one aspect of the construct (e.g., matrix reasoning as an aspect of fluid intelligence in the studies by Jaeggi et al., 2008, and Jaeggi, Studer-Luethi, et al., 2010). Shipstead et al. (2012) stated that the abilities of interest should be measured with several instruments. Furthermore, near-transfer effects to WM components are often demonstrated with tasks that measure STM. More valid WM tasks, such as complex span tasks, have not been used consistently. Some studies failed to show a transfer to complex span tasks, and near-transfer effects might be attributed to task-specific overlaps between trained and transfer tasks (Shipstead et al., 2012). Even when transfer to complex span tasks has been demonstrated and can be traced back to improved executive attention, the specific mechanisms leading to it are not yet well understood (Morrison & Chein, 2011; Titz & Karbach, 2014; Von Bastian & Oberauer, 2014).

Besides mechanisms leading to transfer, the conditions of training that might have an influence on training outcomes have not been considered much (Von Bastian & Oberauer, 2014). Klingberg (2010) emphasized that the roles of variables such as duration and spacing of training to yield transfer and long-term improvements are not yet well understood. In addition, factors related to the time configuration of training (e.g., duration of single training sessions, frequency of training per week, and time interval between single sessions) have not yet been systematically considered. For example, with regard to the training interval, distributed learning was reported to be more effective than massed learning (e.g., Bloom & Shuell, 1981).

Concerning WM-training, Penner et al. (2012) found an advantage of distributed training (two times a week for 8 weeks) compared to massed training (four times a week for 4 weeks). Given a large number of WM training studies, it remains unclear whether the training interval has a moderating influence on transfer effects following WM training. An investigation of the mentioned variables related to the time configuration of training could help with a better understanding how attention capabilities related to WMC should be challenged over time to optimize transfer effects. Concerning the training itself, a need exists to investigate the influence of the modality (verbal, visuospatial, or both) of the trained tasks on these effects (Titz & Karbach, 2014). Studies differ in terms of the identification of the modalities that lead to higher transfer effects, which would help in designing optimal training tasks. From a theoretical perspective, of interest is the extent to which transfer effects are moderated through verbal, visuospatial, or both modalities.

Melby-Lervåg and Hulme (2013) investigated transfer effects of WM training and some moderators of these effects. These authors conducted a meta-analysis on the basis of 23 studies and 30 group comparisons to examine near- and far-transfer effects of WM training. They found short-term effects of WM training on verbal and visuospatial WM and limited evidence for sustained effects on visuospatial WM at follow-up. In particular, the effect size for the immediate transfer effect to verbal WM was 0.79, and the effect size for the short-term transfer effect to visuospatial WM was 0.52. The effect size for the long-term transfer effect to visuospatial WM was 0.41. Melby-Lervåg and Hulme found small immediate transfer effects to attention (measured with the Stroop task) and nonverbal ability but found no transfer effects to verbal ability, arithmetic, and word decoding. They reported significant heterogeneity among studies for immediate and sustained transfer effects to verbal WM. Studies also varied significantly with respect to immediate transfer effects to visuospatial WM and nonverbal ability.

Melby-Lervåg and Hulme (2013) investigated the following moderators: age (younger children, older children, young adults, older adults), training dose (total duration of training), design type (randomized or nonrandomized trials), type of control group (active or passive), learner status (participants with or without learning disabilities), and intervention type (kind of training program). Age was a significant moderator of the transfer effects to verbal WM, with larger gains in verbal WM in younger children compared to older children. Melby-Lervåg and Hulme noted that it is possible that WM training is more effective in early years when brain plasticity is particularly high. However, older adults might profit relatively more from WM training because WMC declines with age (Salthouse, Pink, & Tucker-Drob, 2008). Therefore, it seems possible that differences between age groups level out when considering a larger age span than in the analysis of Melby-Lervåg and Hulme. Concerning intervention type, the effect sizes of one of several commercial training programs on visuospatial WM were higher than those of noncommercial programs, whereas there was no difference between the commercial programs. Of interest, the type of control group turned out to be a moderator for transfer effects to nonverbal ability, with larger gains of training on nonverbal ability for passive versus active control groups. The authors stated that there is probably no systematic heterogeneity among studies on measures of far transfer that could be explained by moderator variables. They concluded that the idea of WM training and hence attaining improvements on other cognitive abilities in children and healthy adults is, at the least, doubtful.

The remaining question is whether the conclusions drawn by Melby-Lervåg and Hulme (2013) are still appropriate against the background of several recently published studies. Further potential WM training conditions are likely to have an influence on transfer effects of WM training, but they have not yet been systematically considered. Age as a potential moderator might have an influence on transfer effects due to differences in WMC across the life span (see previous paragraph). With respect to the trained task, in addition to the trained modality, feedback on the results varies greatly among studies. In some studies, the participants were provided with information only on the correctness of their answers—that is, mere knowledge of the results (e.g., Heinzel et al., 2014; Van der Molen et al., 2010). In other studies, participants also received more elaborate feedback, such as about individual improvement (Alloway et al., 2013; Egeland, Aarlien, & Saunes, 2013). The consideration of feedback on the trained tasks seems to be important because the type of feedback might be differentially effective for learning and transfer (Hattie & Timperley, 2007). If participants receive motivating feedback, they might focus their attention more on the trained task and thus yield larger training gains.

Other variables concern the training process from the beginning to the end of the training. An obviously important variable might be supervision—that is, whether the activities of the participants in the training were monitored and guided by someone else. In some studies, experimenters or other persons such as parents monitored the participants as to whether they were training properly (e.g., Borella, Carretti, Riboldi, & De Beni, 2010; Holmes & Gathercole, 2013). Thus, participants might focus their attention more strongly on the trained tasks. In other studies, experimenters were present without monitoring or intervening while the participants were training (e.g., Horowitz-Kraus & Breznitz, 2009; Salminen, Strobach, & Schubert, 2012). The mere presence of other persons has been shown to have a detrimental effect on task performance in complex tasks (for a review, see Aiello & Douthitt, 2001). One of the areas of interest in the present analysis is therefore to investigate the effects of supervision, meaning that someone present during the training *also* monitors and ensures that the participants follow the training instruction properly.

Another process variable that needs consideration is instructional support (i.e., whether participants receive additional explanations and help during the training). In some studies, participants received instructions on how to perform the training tasks only *at the beginning* of the training (e.g., Jaeggi et al., 2008; Thorell, Lindqvist, Bergman Nutley, Bohlin, & Klingberg, 2009), whereas in other studies, participants received additional instructions about the trained tasks *during* the training (e.g., Heinzel et al., 2014; Jaušovec & Jaušovec, 2012). Instructional support in the form of worked examples is supposed to be effective when prior knowledge is low (e.g., Kalyuga, 2007). With respect to WM training, participants do not know how to perform a task properly at the beginning of the training. Additional instructions could help participants to properly work on the training tasks without having to invest attentional resources for understanding the task. Thus, the effectiveness of the training could be improved. We included instructional support as a moderator variable to examine whether it has an impact on transfer of training effects.

A further variable that has not yet been considered systematically is the location of the training. Most of the WM training studies took place in a laboratory, but in some studies, participants trained elsewhere, such as at home (Brehmer et al., 2011; Brehmer et al., 2012). Plausibly, the laboratory offers fewer sources of interference and side-tracking than other places, and participants could focus on the trained task. The location of the training was included as a moderator in the present study to investigate whether training in or outside the laboratory has a different influence on training outcomes.

These training conditions might explain substantial amounts of variability in the effect of training studies. Thus, identifying moderators among training conditions might contribute to improved training designs for optimized outcomes.

In summary, although analyses on the effectiveness of WM training as a tool to enhance cognitive functions exist (Klingberg, 2010; Titz & Karbach, 2014), there are several shortcomings in the prior research. The studies included in the meta-analysis of Melby-Lervåg and Hulme (2013) were collected until November 5, 2011. Since then, several new published studies have presented new empirical evidence, which raises the question of whether the overall conclusion of transfer effects of WM training in prior studies are still valid. Moreover, Melby-Lervåg and Hulme did not consider the influence of several training conditions on training outcomes and transfer. It is plausible that the specific conditions under which WM training takes place substantially moderate training effectiveness and transfer. However, the effects of the modality of the trained tasks, supervision, location, and other training conditions have not yet been systematically investigated. The present meta-analysis therefore focuses on training conditions as further potential moderators. Uncovering the influence of certain training conditions on transfer effects could be helpful to optimize training in order to yield practically relevant transfer effects (e.g., to school-relevant skills). If certain forms of single moderators yield higher transfer effects than others (e.g., supervised training is better than training in the mere presence of other persons), the effectiveness of WM training could be improved.

THE CURRENT REVIEW

Research Questions and Hypotheses

We derived our research questions and hypotheses regarding transfer effects of WM training and some moderators based on the results of the meta-analysis of Melby-Lervåg and Hulme (2013). In addition, we differentiated between near-transfer effects to STM and WM. As stated previously, STM and WM tasks differ only in the extent to which the same basic subcomponent processes, such as rehearsal and updating, are involved (Unsworth & Engle, 2007b). The improvement of these subcomponent processes through WM training could result in improvements in tasks measuring WM *and* STM. We therefore suppose that if there are transfer effects to WM, then there will also be transfer effects to STM, and vice versa. We formulated the following research question and hypotheses regarding transfer effects of WM training.

RQ1: Which near- and far-transfer effects follow WM training?

Our expectations concerning the first research question are related to the results of the meta-analysis of Melby-Lervåg and Hulme (2013). In their analysis, WM training yielded short-term transfer effects on verbal and visuospatial WM and long-term transfer effects on visuospatial WM at follow-up. Accordingly, we expect near-transfer effects of WM training to STM and WM components (verbal and visuospatial domain) at posttest (Hypothesis 1.1), no transfer to STM and WM components in the verbal domain at follow-up (Hypothesis 1.2), and a transfer to STM and WM components in the visuospatial domain at follow-up (Hypothesis 1.3). Regarding far-transfer effects, small immediate transfer effects to inhibition (labelled as attention) and nonverbal ability, but no other far-transfer effects, were found (Melby-Lervåg & Hulme, 2013). We therefore expect a small immediate far-transfer effect to nonverbal ability that is not sustained at follow-up (Hypothesis 1.4), and no transfer effects to verbal ability, word decoding, and mathematical abilities (Hypothesis 1.5).

Notably, inhibition (the ability to inhibit prepotent responses) was not included in our analyses for several reasons, which we describe later, in the section on coding of transfer measures. The second research question is related to potential moderators of transfer effects following WM training.

RQ2: Which training conditions have a moderating influence on transfer effects following WM training?

Melby-Lervåg and Hulme (2013) found that age was a moderator of transfer effects to verbal WM, and training dose (total duration of training) was not a moderator of any transfer effect. Thus, we assume that age is a moderator of transfer effects to verbal WM with younger participants having larger training improvements compared to older participants (Hypothesis 2.1), and training dose (total duration of training) is not a moderator of transfer effects following WM training (Hypothesis 2.2). In addition, we considered a variety of other potential moderators of transfer effects. One group of these moderators is related to the time configuration of WM trainings. In studies showing transfer effects of WM training (e.g., Alloway et al., 2013; Chein & Morrison, 2010), participants trained for certain amounts of time in single training sessions and for several days a week. We hypothesize that WM needs to be challenged for a longer period in single training sessions and for several days a week to yield transfer effects. Thus, we assume that the duration of a single training session is a moderator of transfer effects following WM training. Specifically, we expect that the longer the duration of a single training session, the larger the effect size (Hypothesis 2.3). Because more frequent training should lead to higher gains from training, we assume that the frequency of training per week is a moderator of transfer effects following WM training. We hypothesize that the more training sessions per week, the larger the effect size (Hypothesis 2.4). Based on research about the benefits of distributed learning (e.g., Bloom & Shuell, 1981) and distributed WM training (Penner et al., 2012), we expected that the training interval is a moderator of transfer effects following WM training. The training interval was 1 or 2 days in most of the training studies. We hypothesize that a training interval of 2 days leads to a larger effect size than a training interval of 1 day (Hypothesis 2.5).

Furthermore, we tested hypotheses regarding the characteristics of training conditions and the implementation of training. As Titz and Karbach (2014) noted, the influence of the trained modality on transfer effects needs to be investigated. Of theoretical interest is the extent to which the trained modality moderates transfer effects in order to determine whether one specific or both modalities of WM need to be trained to yield optimal transfer effects. We assume that the trained modality is a moderator of transfer effects following WM training. Because transfer measures of verbal STM and WM tax different modalities than visuospatial STM and WM, we assume that transfer effects to verbal STM and WM are larger for interventions in which participants trained with tasks measuring verbal WM. Analogously, transfer effects to visuospatial STM and WM should be higher for interventions in which participants trained with tasks measuring visuospatial WM. Far-transfer measures differ in the extent to which specific components of WM are taxed. Therefore, we have no directed hypothesis regarding the moderating influence of the trained modality on the variability among effect sizes for far-transfer measures (Hypothesis 2.6).

With respect to the training process, supervised training (e.g., monitoring whether participants train properly) could be beneficial for transfer effects. In contrast, research has repeatedly shown the detrimental effects of the mere presence of other persons on task performance in complex tasks (Aiello & Douthitt, 2001). Hence, we expect the transfer effects for supervised training to be larger than for training without the presence of other persons and training in the mere presence of other persons. Transfer effects for training without the presence of other persons are higher than for training in the mere presence of other persons (Hypothesis 2.7).

Training studies also differed with respect to instructional support. Worked examples, as one form of instructional support, have been shown to be effective when prior knowledge is low (e.g., Kalyuga, 2007). In WM training studies, participants initially do not know how to properly work on a task. The participants could work more properly on the trained task if explanations were given in addition to the instructions at the beginning of the training because no attentional resources have to be invested for finding out how the task has to be approached. Consequently, we assume that instructional support is a moderator of transfer effects following WM training. Transfer effects for training with additional instructional support beyond the explanations at the beginning of the training are greater than for training without these additional explanations (Hypothesis 2.8).

Feedback is another process variable that might have an influence on transfer effects following WM training. Different types of feedback have different effects on learning and transfer (Hattie & Timperley, 2007). In the context of WM training, motivating feedback could stimulate participants to be more attentive to the trained task and thus lead to higher gains from training. We hypothesize that training with feedback beyond mere knowledge of the results yields larger effect sizes than training that includes mere knowledge of the results only (Hypothesis 2.9).

Working memory training has been conducted in different places. Participants training in the laboratory with few sources of distraction might focus their attention on the trained task better than participants training in other places. Hence we expect that transfer effects of training in the laboratory are larger than transfer effects of training in school or at home (Hypothesis 2.10).

In addition to the hypotheses addressing the characteristics of training conditions and the implementation of WM training, we formulated hypotheses to check methodological quality characteristics of the studies. These hypotheses were based on the results of the meta-analysis by Melby-Lervåg and Hulme (2013). We hypothesize that intervention type (i.e., the kind of training program) is a moderator of transfer effects to visuospatial WM (Hypothesis 2.11). We also assume that type of control group is a moderator of transfer effects of WM training to nonverbal ability. The mean effect on nonverbal ability for the comparison of training groups with passive control groups is larger than for the comparison of training groups with active control groups (Hypothesis 2.12).

METHOD

Literature Search and Inclusion Criteria

We searched for studies of WM training published from 1981 to December 2013 through the electronic databases PsycINFO, PsycARTICLES, ERIC, and Medline. Key words were "working-memory training," which was used as a string in a general search (i.e., without restrictions, e.g., to subject headings). In addition, we scanned reference lists of articles related to WM training and asked the authors of the included studies to provide other (unpublished) references, if possible. Results of the databases yielded 579 records after duplicates were removed. Together with articles from other sources (e.g., reviews, author requests), we found about 600 records.

We applied the same inclusion criteria used by Melby-Lervåg and Hulme (2013) to make our analysis comparable concerning this matter and to be able to include the studies analyzed by those authors. Subsequently, the requirements of the studies to be included are listed here:

1. Studies must be randomized controlled trials or quasi-experiments with a treatment and either an active or passive control group tested at pre- and posttest.
2. The treatment group had to receive an intervention for at least 2 weeks based on an adaptive computerized program that aimed to train working memory skills (verbal, visuospatial, or both).
3. Participants could be of any language background and learner status, but studies of adults more than 75 years of age were excluded.
4. The studies must provide data so that an effect size can be computed for the transfer measures (Melby-Lervåg & Hulme, 2013).

Regarding the fourth point, we asked the authors of the studies to provide information for the effect size calculation if the corresponding data were not reported in their articles.

The current analysis includes 47 studies with 65 group comparisons from journals. Twenty-three of these studies with 30 group comparisons were adopted from the analysis of Melby-Lervåg and Hulme (2013). Two limitations in their meta-analysis concern the studies by Schmiedek, Lövdén, and Lindenberger (2010) and Klingberg, Forssberg, and Westerberg (2002). In the study by Schmiedek et al., participants were trained on tasks taxing WM, but also perceptual speed and episodic memory; in the study by Klingberg et al. (2002), participants trained their WM and in addition practiced a mixture of a reaction time and inhibition task. Because our focus was primarily on WM training and we wanted our analysis to be comparable to the one conducted by Melby-Lervåg and Hulme, we ran our analysis with and without the studies by Schmiedek et al. and Klingberg et al. (2002). The exclusion of these two studies affected near-transfer effects to verbal and visuospatial STM and WM, as well as far-transfer effects to nonverbal and verbal ability.

Coding of Transfer Measures

Table 1 shows the constructs for which transfer effects were analyzed with explanations of the coded measures and examples of tests to measure them. Supplementary Table S2 (online only) shows the transfer effects for each study with the respective transfer measures.

We coded measures for STM and WM (verbal and visuospatial), nonverbal and verbal ability, word decoding, and mathematical abilities. As noted previously, Redick, Broadway, et al. (2012) distinguished STM tasks that require only the temporary storage of information from WM tasks that also involve a processing component. Although Melby-Lervåg and Hulme (2013) proposed to consider this distinction in the coding of measures, they actually did not consistently implement their plan. In fact, many authors of WM training studies measured transfer to WM by using STM tasks (see also Shipstead et al., 2012). For example, Melby-Lervåg and Hulme coded visuospatial STM measures, such as the span-board task (Klingberg et al., 2005) and the grid task (Nutley et al., 2011), as measures of visuospatial WM. In the grid task, which is an adaptation of the Corsi Blocks task, participants are asked to remember a series of dots presented on a grid in serial order (Alloway, Gathercole, & Pickering, 2006). Thus, the grid and Corsi Blocks tasks have a storage component only and are hence examples of STM tasks in our conceptualization. Besides the suggestions of Redick, Broadway, et al. (2012) for a separation between STM and WM, other studies acknowledged that such simple span measures with visuospatial material are measures of visuospatial STM (e.g., Alloway et al., 2006; Miyake, Friedman, Rettinger, Shah, & Hegarty, 2001). We consequently coded simple span measures with visuospatial material (e.g., the grid task) as measures of visuospatial STM and measures with an

TABLE 1
Constructs for Which Transfer Effects Were Analyzed With Explanations of Coded Measures and Examples of Tests

Construct	*Explanation of Coded Measures*	*Examples of Tests (Study Authors)*
Verbal STM	Tasks using material with verbal information that have to be stored in STM	Digit recall forward (St Clair-Thompson et al., 2010), word span test (Nutley et al., 2011)
Visuospatial STM	Tasks using material with visuospatial information that have to be stored in STM	Span-board task (Klingberg et al., 2005); grid task (Lilienthal et al., 2013); Corsi blocks task (Hubacher et al., 2013)
Verbal WM	Tasks using material with verbal information that have to be stored and processed in WM	Listening recall (van der Molen et al., 2010); reading span (Richmond et al., 2011)
Visuospatial WM	Tasks using material with visuospatial information that have to be stored and processed in WM	Shape recall test (Alloway et al., 2006); symmetry span (Redick et al., 2013)
Nonverbal ability	Measures that predominantly require solving problems without verbal information in the tasks	Raven (Chein & Morrison, 2010); Culture Fair Test, scale 3 (Borella et al., 2010)
Verbal ability	Tests that mainly require verbal information in the tasks for solving problems	Composite score of the subtests similarities and vocabulary from the Wechsler Abbreviated Scale of Intelligence (Dunning et al., 2013); Regensburger Word Fluency Test (Penner et al., 2012)
Word decoding	Coded measures involve speed and quality of word and nonword reading, but not reading comprehension (such tests were coded as measures of verbal ability)	Average of word decoding speed and quality of coding from the test battery LOGOS (Egeland et al., 2013); nelson denny reading rate (Thompson et al., 2013)
Mathematical abilities	Tests that require solving mathematical problems (e.g., applying arithmetic operations)	Arithmetic test (Van der Molen et al., 2010), national Standard Assessment Test (Holmes et al., 2013)

Note. STM = short-term memory; WM = working memory.

additional processing component (e.g., the shape recall test used by Alloway et al., 2013) as measures of visuospatial WM.

Differentiating between measures of verbal STM and verbal WM led to a coding procedure that was slightly more complex than the procedure used in the earlier meta-analysis. Melby-Lervåg and Hulme (2013) coded backward digit span in the studies of Borella et al. (2010) and Van der Molen et al. (2010; together with listening recall, a complex span task) as measures of verbal WM. In the digit span task, participants have to recall a series of spoken digits in the same (forward) or reverse (backward) condition order presented (Alloway et al., 2006). The findings of the study by St Clair-Thompson (2010) suggested that backward digit span is a measure of (verbal) WM for children and (verbal) STM for adults, which is in line with other research (e.g., Colom, Abad, Rebollo, & Chun Shih, 2005; St Clair-Thompson & Allen, 2013). We therefore coded backward digit recall as a measure of (verbal) WM in children and as a measure of verbal STM in adults. Simple span measures such as forward digit span and word recall (as used, e.g., in the study by Dunning, Holmes, & Gathercole, 2013) were coded as measures of verbal STM; complex span tasks or measures with an additional processing component, such as reading span (as used, e.g., in the study by Richmond, Morrison, Chein, & Olson, 2011), were coded as measures of verbal WM independent of age. For example, Redick, Broadway, et al. (2012) showed that the additional processing component in the reading span task involves deciding whether a sentence is right or wrong in between memorizing single letters.

Melby-Lervåg and Hulme (2013) used the values of an n-back task reported in the study by Jaeggi, Studer-Luethi, et al. (2010) to assess transfer to verbal working memory. In the n-back task used by Jaeggi, Studer-Luethi, et al. (2010), participants have to indicate when a presented visual stimulus matches a visual stimulus shown n trials prior. In contrast to Melby-Lervåg and Hulme, we used the operation span task to assess transfer to verbal working memory because the concurrent validity of the n-back task as a measure of working memory is not yet commonly accepted (for an overview, see Jaeggi, Buschkuehl, Perrig, & Meier, 2010). In contrast to the n-back task, the operation span task has been shown to be a reliable and valid measure of working memory capacity (Redick, Broadway, et al., 2012).

Coding of far-transfer measures was identical to the procedure in the meta-analysis by Melby-Lervåg and Hulme (2013) except for the ability to inhibit prepotent responses (inhibition; Miyake et al., 2000). Inhibition was labeled as attention and measured with the Stroop task in the meta-analysis. In incongruent trials of the Stroop task, participants have to name the ink color of a word describing a different color (Melby-Lervåg & Hulme, 2013). Deviating from Melby-Lervåg and Hulme, we did not include inhibition in our analyses for a number of reasons. The operation span task as a complex span task is similar to the reading span task with respect to its structure and has been shown to be most strongly related to updating (Miyake et al., 2000). Concordantly, Krumm et al. (2009) argued that inhibition could not be separated from complex span tasks. Friedman, Miyake, Robinson, and Hewitt (2011) showed, on the basis of structural equation models, that there was evidence for one common factor of updating, shifting, and inhibition (for an overview, see Miyake & Friedman, 2012). After controlling for this common factor, there was no unique variance left for inhibition. Inhibition could therefore already be assessed to a high degree with complex span tasks. Furthermore, inhibition itself is recognized as a multidimensional construct (Krumm et al., 2009; Nigg, 2000) because several inhibitory skills may exist (Friedman & Miyake, 2004; Hedden & Yoon, 2006). Following this argumentation, we did not include the Stroop task as a measure of attention, although Melby-Lervåg and Hulme did. The coded measures of word decoding involve speed and quality of word and nonword reading but not reading comprehension (such tests were coded as measures of verbal ability). One exception regarding reading comprehension as an aspect of verbal ability occurred in the study by Dunning et al. (2013). A composite score for verbal IQ was preferred over the measure of reading comprehension. The composite score was chosen because it is assumed to have more weight in the assessment of verbal ability than a single score. We coded measures of mathematical abilities and didn't choose the term "arithmetic" used by Melby-Lervåg and Hulme, because we also wanted to cover other aspects of mathematical abilities measured by tests in some studies (e.g., the comparison of numbers in the study by Karbach, Strobach, & Schubert, 2014). However, most of the mathematical tests used in the included studies do cover arithmetic skills (e.g., addition, division).

Coding of Moderators

Table 2 shows the moderators included in the analyses. At first, we included the significant moderators (age, intervention type, and type of control group) in the analysis of Melby-Lervåg and Hulme (2013), but we then made the following modifications. Sample age was not normally distributed, but the common logarithm of age was. We therefore included the common logarithm of age rather than age itself in the moderator analysis to minimize a loss of information associated with the categorization of continuous moderators applied by Melby-Lervåg and Hulme. Although training dose was not a significant moderator in the meta-analysis by Melby-Lervåg and Hulme, we included this variable in our analysis because it is an important factor with substantial variation among the studies and was included in a dichotomized form in Melby-Lervåg and Hulme. We included training dose as a continuous moderator to get more detailed information about its influence on transfer effects following WM training. However, because training dose was not normally distributed, we included the common logarithm of training dose (which was normally distributed) in our moderator analysis.

Type of control group (active vs. passive) was considered as a moderator that is under the control of the experimenter only. This moderator was included to check whether the size of the effect of an intervention depends on the comparison of the intervention group with the type of control group. The type of control group was a significant moderator in the analysis by Melby-Lervåg and Hulme (2013). If there were an active and a passive control group in a study, we compared the training group with both control groups. This is in contrast to Melby-Lervåg and Hulme, who included comparisons of the active control and the training group with the passive control group. One exception is the study by Alloway et al. (2013), for which we included the active control group as a second treatment group because it involved adaptive WM training, only with a lower frequency. This second treatment group and the first treatment group (high-frequency WM training) were each compared with the passive control group. Active control groups received training that didn't tax WM or was not

TABLE 2
Moderator Variables Included in the Analyses

Moderator	*Description*	*Coding*
Age	Sample age (in years)	Common logarithm of age
Training dose	Total amount of training (in hours)	Continuous moderator
Session duration	Duration of single training sessions (in minutes)	Continuous moderator
Frequency of training per week	No. of training sessions per week (in days)	1 day, 3 days, 4–6 days
Training interval	Time interval between single sessions (in days; excluding weekends)	1 or 2 days
Modality	Trained modality of WM	Verbal or visuospatial domain or both domains
Supervision	If training is monitored by a person (e.g., experimenter) or if a person is just present or if no person is present	Supervision vs. mere presence vs. no presence
Instructional support	Additional instructional support beyond the explanations at the beginning or not	Instructional support or not
Feedback	If feedback beyond mere knowledge of results was provided or not	Elaborated feedback or not
Location	Location of training	Training in laboratory vs. training in school vs. training at home

Note. WM = working memory.

adaptive (constant difficulty), and therefore most probably was not intended to train WM. We chose a comparison different from that of Melby-Lervåg and Hulme because we did not want to include active control groups without an adaptive WM training as training groups (see inclusion criteria). Although comparing several treatment groups with one common control group has its problems, we chose this procedure to get information that could be lost with other procedures (see Wecker & Fischer, 2014, for a similar approach).

Other moderators included in the current analysis were as follows:

1. Session duration. This is the duration of single training sessions.
2. Frequency of training per week. We didn't always get exact information about the number of days participants trained per week (e.g., one or two training sessions per week) and therefore formed three categories for the frequency of training per week (1 or 2 days, 3 days, 4–6 days).
3. Training interval. This is the time interval between single sessions. Most of the training had intervals of 1 or 2 days, so we dichotomized this variable into a training interval of 1 day versus a training interval of 2 days.

The influence of these variables was examined to get detailed information about the role of time configuration on transfer effects of WM training.

The following moderators were chosen to investigate the influence of variables that are characteristic of training paradigms and the implementation of training: (a) modality, (b) supervision, (c) instructional support, (d) feedback: The feedback provided in the different studies occurred after completing a trial within a task, at the end of a training day, weekly, or after completion of the entire training. Feedback provided at the end of the whole training was not included in our analysis. Participants received feedback about the correctness of their answers in most of the training paradigms included in our analysis, so we were interested in the effects of feedback beyond the mere knowledge of results (or feedback about the task; see Hattie & Timperley, 2007). (e) Location.

Effect Size Calculations

We calculated Hedges's *g* with bias correction (Hedges & Olkin, 1985) as effect size for each measure. We used *g* because for small sample sizes it provides a somewhat better effect size estimate than does Cohen's *d*. The reason is that in the calculation of *g*, variances are pooled by using n–1, whereas in the calculation of Cohen's *d*, variances are pooled by using n (Grissom & Kim, 2005). To control for differences between treatment and control groups at pretest, we computed the difference between the gain scores from pre- to posttest (immediate transfer effects) and from pretest to follow-up test (long-term transfer effects), and used this difference as the numerator in the calculation of Hedges's *g*. The pooled standard deviation of the standard deviations of the particular training and control groups at pretest was used as the denominator of Hedges's *g*. Morris (2008) favored this kind of calculation for Hedges's *g* in Pretest-Posttest-Control Group Designs. Melby-Lervåg and Hulme (2013) obtained their effect size estimates similarly, except that they used the average of the standard deviations of the particular training and control groups in the denominator of Hedges's g. This difference sometimes led to slightly different results (e.g., 1.10 for the effect size in the verbal WM in contrast to 1.09 by Melby-Lervåg & Hulme in the study by Thorell et al., 2009).

The control of differences between treatment and control groups at pretest may produce significant, positive effect sizes even when the treatment group shows no improvement but the control group deteriorates, as was the case for the transfer effect to nonverbal ability in the study by Harrison et al. (2013). In a similar vein, a significant negative effect could result when the control group improves and the treatment group deteriorates, which was the case for the transfer effects to verbal STM and nonverbal ability in the study by Richmond et al. (2011). We therefore ran the analyses for nonverbal ability and verbal STM with and without the effect sizes of Harrison et al. (2013) and Richmond et al. (2011), respectively. If studies contained more measures of a construct, we averaged the means and standard deviations for the calculation of an effect size. This procedure was also reported by Melby-Lervåg and Hulme (2013).

Large differences in effect sizes occurred when different outcome measures were used for the calculations. For example, Melby-Lervåg and Hulme (2013) used the letter memory task as a measure of verbal WM in the study by Dahlin, Nyberg, Bäckman, and Stigsdotter Neely (2008). However, this task was one of the trained tasks in the study by Dahlin, Nyberg, et al., and we wanted to assess transfer to verbal WM with complex span tasks such as the computation span also used in this study. This resulted in a large difference between effect sizes (see Supplementary Table S2 [online only] for effect sizes in our analysis).

The studies were coded by the first author of this article, and all effect sizes were double-coded by a trained research assistant. If the results differed, they were recalculated by the first author. Afterward, the intraclass correlation of the interrater reliability was 1. The first author also coded and double-coded the moderators. Moderators of 33 group comparisons were then again coded by a trained research assistant. Interrater reliability for the categorical moderator variables was calculated by using Cohen's kappa and was 1 with an agreement rate of 100%. Interrater reliability for the continuous moderator variables was calculated by using intraclass correlation and was 1 with an agreement rate of 100%.

Random-Effects Models

Similar to Melby-Lervåg and Hulme (2013), we chose a random-effects model to test the overall effect size for significance. The overall effect itself was computed by weighting the effect sizes of the individual studies with the inverse variance weight (Lipsey & Wilson, 2001). Random-effects models are currently favored for meta-analysis (Schmidt & Hunter, 2003); fixed-effects models do not seem advisable, especially when effect sizes vary among studies (Overton, 1998). The selection of a random-effects model for our analysis implies that the included effect sizes differ from the mean of the population due to sampling error and a random variance component. Generally, random-effects models are more conservative than fixed-effects models (Lipsey & Wilson, 2001). The analyses were conducted in SPSS and used the meta-analysis macros developed by Wilson (2005).

Procedure

We tested our hypotheses about transfer effects of WM by examining whether the mean effect size for each transfer measure differed significantly from zero. To check for significant heterogeneity in effect sizes among studies, we calculated the Q-statistic for each transfer measure (Hedges & Olkin, 1985). In addition, we calculated I^2, which informs about the degree of heterogeneity, or how much of the total variance in effect sizes can be assigned to true variance among studies (Huedo-Medina, Sanchez-Meca, Martin-Martinez, & Botella, 2006). If the Q-statistic was significant, moderator analyses were conducted to explain the heterogeneity among studies.

We used mixed-effects models based on the method of moments for all analyses of the moderators. Mixed-effects models imply that a certain proportion of variance among studies can be explained through a moderator variable, but a significant variance among studies remains (Lipsey & Wilson, 2001). We chose a mixed-effects model because we assumed that single moderators can explain a certain proportion, but not all, of the variance among studies. Categorical moderators were analyzed by using special analyses of variance (ANOVAs) for meta-analyses, also called analogs to the one-way ANOVA (Lipsey & Wilson, 2001). If significant group differences existed between the categories of a moderator with several categories, an ANOVA for each pairwise comparison was run with a Bonferroni correction of the significance level alpha to control for multiple testing.

We examined the influence of continuous moderators on effect sizes by using weighted meta-regression analyses. In these analyses, we used the effect size weighted with the inverse variance as the dependent variable (Lipsey & Wilson, 2001).We ran moderator analyses for each single moderator, but we were unable to apply multiple meta-regressions that included all moderators to get information about the influence of one moderator controlling for the other moderators due to missing data for single moderators. This led to large decreases in sample sizes when all moderators were included in the analyses. The missing data concerned categorical moderators. We did not apply imputation techniques because currently no convincing procedures to handle missing categorical predictors in meta-analysis are available (Pigott, 2012, p. 88). We were able, however, to run multiple regression analyses with moderators, which were found to be significant in single regression analyses.

We created forest plots to visualize the heterogeneity of effect sizes with their confidence intervals (Walker, Hernandez, & Kattan, 2008) and to identify outliers. The impact of the outliers was examined via sensitivity analyses, in which the mean effect sizes were estimated after outliers had been removed.

To check for publication bias, we created funnel plots for each transfer effect. In these funnel plots, we plotted the respective effect size on the x-axis and a measure of precision (standard error) on the y-axis. The interpretation of a funnel plot is subjective, so we applied a slightly modified version of the Egger's test (Egger, Davey Smith, Schneider, & Minder, 1997; Sterne & Egger, 2005). In the original Egger's test, the effect size divided by its standard error is predicted by the inverse of the standard error. When the y-intercept differs significantly from zero, bias is indicated (Egger et al., 1997; Rothstein, 2008). When the effect size is predicted by the standard error in a weighted regression, the slope is the analog to the y-intercept in the original Egger's test (Rothstein, 2008). We used the effect size divided by its variance as a criterion and the standard error as a predictor (Sterne & Egger, 2005) in order to conduct the Egger's test with a (weighted) meta-regression by using the macro by Wilson (2005). In addition, we conducted a trim-and-fill analysis (Duval & Tweedie, 2000a, 2000b) for random-effects models when the funnel plot and Egger's test indicated a potential publication bias. In the trim-and-fill method (Duval & Tweedie, 2000a, 2000b), the number of missing studies to make the funnel plot symmetric is estimated via different estimators. One of these estimators is L_0, which has been shown to be more robust than the other estimators (Talebi, 2013). We consequently used $L_0$for our trim-and-fill analysis.

Concerning the handling of missing data, we chose the same procedure as Melby-Lervåg and Hulme (2013). Studies were included in all analyses for which enough data were available.

RESULTS

Supplementary Tables S1 and S2 (online only) summarize information about the studies included in the meta-analysis (e.g., sample characteristics and effect sizes for pretest–posttest as well as pretest–follow-up test differences in gains). As can be seen from these tables, the data from the included studies differ largely regarding sample characteristics and transfer effects.

Near- and Far-Transfer Effects Following Training of WM (RQ1)

To explore which transfer effects follow WM training (RQ1), we ran meta-analyses for each near- and far-transfer effect. Table 3 provides the main information for these analyses.

Verbal STM. Figure 1 shows forest plots for 32 immediate and nine long-term (delayed) transfer effect sizes. After the removal of outliers, the mean effect size ranged from $g = 0.25$, 95% CI (confidence interval) [0.11, 0.39], to $g = 0.42$, 95% CI [0.25, 0.58]. The mean effect size of g was 0.42 when the effect size of the study by Richmond et al. (2011) was excluded. The funnel plot and Egger's test indicated no publication bias. Thus, in line with Hypothesis 1.1, there was an immediate transfer effect of WM training to verbal STM.

We conducted follow-up testing an average of 8.11 months after the posttest. After the removal of outliers, the mean effect size ranged from $g = 0.15$, 95% CI [−0.07, 0.36], to $g = 0.27$, 95% CI [0.05, 0.47]. The funnel plot and Egger's test indicated no publication bias. Thus, in contrast to Hypothesis 1.2, there was a sustained transfer effect of WM training to verbal STM at follow-up.

Visuospatial STM. Figure 2 shows forest plots for 25 immediate and seven long-term (delayed) transfer effect sizes. Removal of the effect size of the study by Klingberg et al. (2002), because participants in that study were trained with a mixture of a reaction time task and inhibition task in addition to a WM task, resulted in a moderate to large mean effect, $g = 0.70$, 95% CI [0.50, 0.91]. After the removal of outliers, the mean effect size ranged from $g = 0.61$, 95% CI [0.43, 0.80], to $g = 0.76$, 95% CI [0.55, 0.97]. The funnel plot and Egger's test indicated no publication bias. Thus, in line with Hypothesis 1.1, there was an immediate transfer effect of WM training to visuospatial STM

We conducted follow-up testing an average of 4.86 months after the posttest. After the removal of outliers, the mean effect size ranged from $g = 0.63$, 95% CI [0.22, 1.03], to $g = 0.91$, 95% CI [0.47, 1.35]. The funnel plot and Egger's test indicated no publication bias. Thus, in line with Hypothesis 1.3, there was a sustained transfer effect of WM training to visuospatial STM.

Verbal WM. Figure 3 shows forest plots for 42 immediate and 11 long-term (delayed) transfer effect sizes. Removal of the effect sizes of the studies by Klingberg et al. (2002) and Schmiedek et al. (2010) resulted in a small to moderate mean effect, $g = 0.58$, 95% CI [0.34, 0.82], $p < .001$. Effect sizes varied significantly among studies, $Q(39) = 242.20$, $p < .001$, $I^2 = 83.90\%$. After the removal of outliers, the mean effect size ranged from $g = 0.30$, 95% CI [0.14, 0.46], to $g = 0.62$, 95% CI [0.39, 0.85]. The funnel plot indicated a slight potential bias with missing studies on the left side of the mean. Egger's test also indicated a publication bias, $b = 2.50$, 95% CI [0.67, 4.32], $p < .01$. A trim-and-fill analysis found no indication of missing studies; hence, no study was imputed. Thus, in line with Hypothesis 1.1, there was an immediate transfer effect of WM training to verbal WM.

We conducted follow-up testing an average of 8.36 months after the posttest. After removal of outliers, the mean effect size ranged from $g = 0.16$, 95% CI [−0.12,

TABLE 3
Near- and Far-Transfer Effects Following WM Training

Transfer Effect	*No. of Effect Sizes (k)*	*Effect Size g*	*95% CI* *LL*	*UL*	I^{2a}	*Q-Statistic*
Verbal STM (short-term/long-term)	32/9	$0.37^{***}/0.22^{*}$	0.19/0.02	0.56/0.42	66.81%	$93.40^{***}/4.57$
Visuospatial STM (short-term/long-term)	25/7	$0.72^{***}/0.78^{***}$	0.56/0.33	0.92/1.23	64.82%/76.64%	$68.20^{***}/25.68^{***}$
Verbal WM (short-term/long-term)	42/11	$0.55^{***}/0.35^{*}$	0.33/0.17	0.78/0.68	83.73%/66.98%	$252.00^{***}/20.29^{***}$
Visuospatial WM (short-term/long-term)	19/6	$0.63^{***}/0.41^{*}$	0.35/0.00	0.90/0.81	79.47%/61.87%	$87.69^{***}/13.11^{*}$
Nonverbal ability (short-term/long-term)	45/11	$0.14^{*}/0.02$	0.01/−0.17	0.27/0.20	53.90%	$95.44^{***}/10.14$
Verbal ability (short-term/long-term)	29/5	$0.16^{**}/0.26$	0.05/−0.33	0.27/0.86	65.84%	$30.69/11.71^{*}$
Word decoding (short-term/long-term)	14/5	0.08/0.21	−0.06/−0.21	−0.22/0.45		8.85/3.52
Mathematical abilities (short-term/long-term)	15/8	0.09/0.08	−0.09/−0.12	0.27/0.28	49.65%	$27.81^{*}/3.22$

Note. STM = short-term memory; WM = working memory; CI = confidence interval; LL = lower limit; UL = upper limit.
[a] I^2 was calculated only if the Q-statistic was significant.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

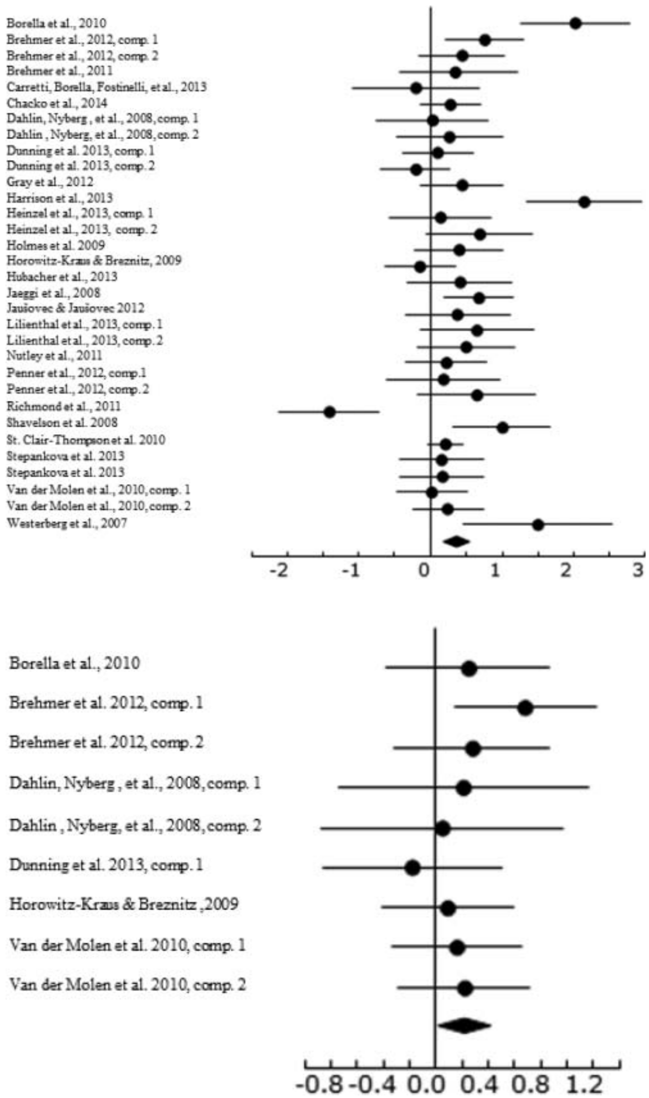

FIGURE 1 Forest plots for immediate (upper panel) and delayed (lower panel) transfer effects to verbal STM. *Note*. Points and horizontal lines show effect sizes (Hedges's g) and confidence intervals, respectively. The diamond displays the overall effect and its confidence interval (width of the diamond).

0.47], to $g = 0.44$, 95% CI [0.14, 0.75]. The funnel plot and Egger's test indicated no publication bias. Thus, in contrast to Hypothesis 1.2, there was a sustained transfer effect of WM training to verbal WM.

Visuospatial WM. Figure 4 shows forest plots for 19 immediate and six long-term (delayed) transfer effect sizes. After removal of the effect sizes of the study by Schmiedek et al. (2010), the effect was still moderate to large, $g = 0.69$, 95% CI [0.37, 1.00], $p < .001$. Effect sizes varied significantly among studies, $Q(16) = 80.48$, $p < .001$, $I^2 = 81.46\%$. After removal of the one outlier, the mean effect size was $g = 0.49$, 95% CI [0.30, 0.67]. The funnel plot indicated a potential bias with missing studies on the left side of the mean. Egger's test also indicated a publication bias, $b = 7.26$, 95% CI [4.70, 9.83], $p < .001$. After removal of the single outlier, however, the funnel plot and Egger's test indicated no publication bias. Therefore, no studies were imputed by trim-and-fill analysis. Thus, in line with Hypothesis 1.1, there was an immediate transfer effect of WM training to visuospatial WM.

We conducted follow-up testing an average of 6.83 months after the posttest. After the removal of the single outlier, the mean effect size was $g = 0.21$, 95% CI [−0.04, 0.47]. The funnel plot indicated a potential bias with missing studies on the left side of the mean. Egger's test also indicated a publication bias, $b = 7.06$, 95% CI [0.87, 13.26]. $p < .05$. After removal of the single outlier, however, the funnel plot and Egger's test indicated no publication bias. Therefore, no studies were imputed. Thus, in line with Hypothesis 1.3, the transfer effect of WM training on visuospatial WM was sustained.

Nonverbal ability. Figure 5 shows forest plots for 45 immediate and 11 long-term (delayed) transfer effect sizes. Effect sizes varied significantly among studies, $Q(44) = 95.44$, $p < .001$, $I^2 = 53.90\%$. After removal of the effect sizes of the studies by Klingberg et al. (2002) and Schmiedek et al. (2010), there was a nonsignificant small effect, $g = 0.01$, 95% CI [−0.03, 0.23], $p = .13$. Effect sizes varied significantly among studies, $Q(41) = 79.97$, $p < .001$, $I^2 = 48.73\%$. The additional exclusion of the effect sizes of the studies by Harrison et al. (2013) and Richmond et al. (2011) resulted in a small effect, $g = 0.12$, 95% CI [0.01, 0.22], $p < .05$. After removal of outliers, the mean effect size ranged from $g = 0.08$, 95% CI [−0.04, 0.20], to $g = 0.17$, 95% CI [0.06, 0.28]. After removal of outliers and the effect sizes of the studies by Klingberg et al. (2002) and Schmiedek et al. (2010), there was a nonsignificant small mean effect, $g = 0.08$, 95% CI [−0.01, 0.18], $p = .08$. The funnel plot and Egger's test indicated no publication bias. Thus, in line with Hypothesis 1.4, there was a small immediate transfer effect of WM training to nonverbal ability.

We conducted follow-up testing an average of 6.54 months after the posttest. After removal of outliers, the mean effect size was $g = -0.12$, 95% CI [−0.32, 0.09]. The funnel plot and Egger's test indicated no publication bias. Thus, in line with Hypothesis 1.4, there was no sustained transfer effect of WM training to nonverbal ability.

Verbal ability. Figure 6 shows forest plots for 29 immediate and five long-term (delayed) transfer effect sizes. After the removal of the effect sizes of the study by Schmiedek et al. (2010), there was a small effect, $g = 0.18$, 95% CI [0.06, 0.31], $p < .05$. Effect sizes did not vary significantly among studies, $Q(26) = 29.77$, $p = .28$. After removal of outliers, the mean effect size ranged from $g = 0.09$, 95% CI [−0.02, 0.19], to $g = 0.20$, 95% CI [0.08, 0.31]. The funnel plot indicated a potential bias with missing studies on the

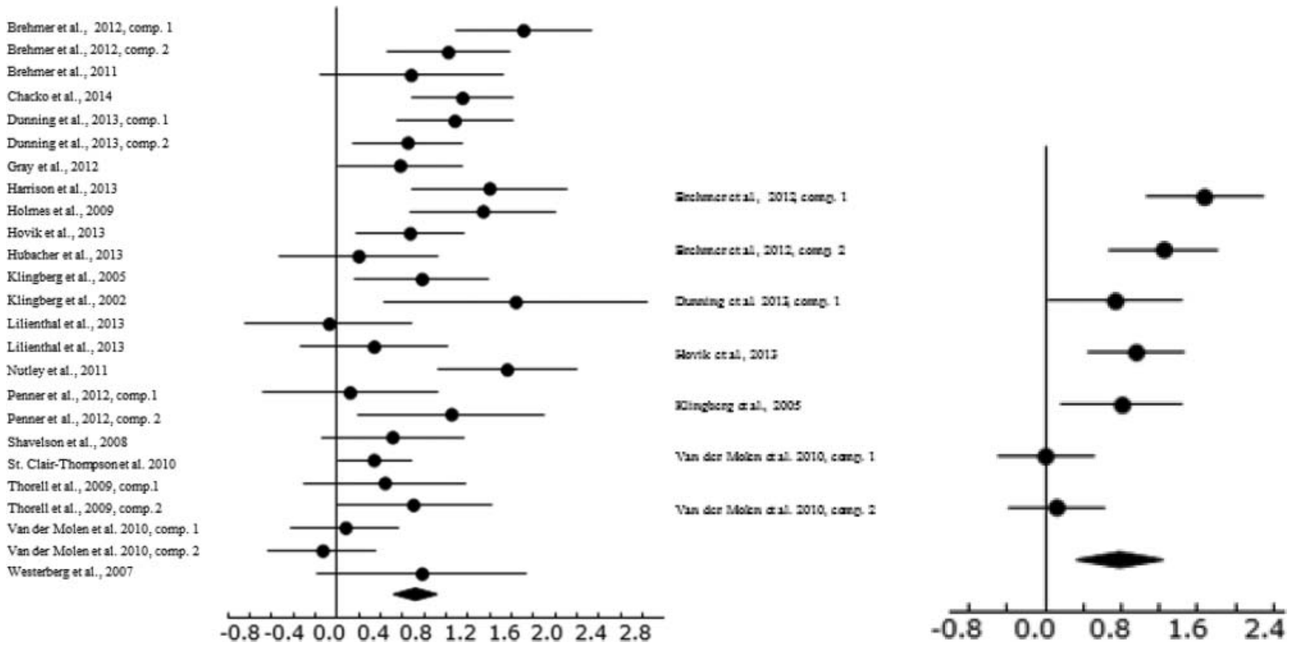

FIGURE 2 Forest plots for immediate (left panel) and delayed (right panel) transfer effects to visuospatial STM. *Note*. Points and horizontal lines show effect sizes (Hedges's *g*) and confidence intervals, respectively. The diamond displays the overall effect and its confidence interval (width of the diamond).

left side of the mean. Egger's test also indicated a publication bias, $b = -0.13$, 95% CI [-0.24, -0.01], $p < .05$. After a trim-and-fill analysis imputed one study, the adjusted mean effect size was $g = 0.14$, 95% CI [0.04, 0.25], $p < .05$. Thus, in contrast to Hypothesis 1.5, there was an immediate transfer effect of WM training to verbal ability.

We conducted follow-up testing an average of 12.8 months after the posttest. Effect sizes varied significantly

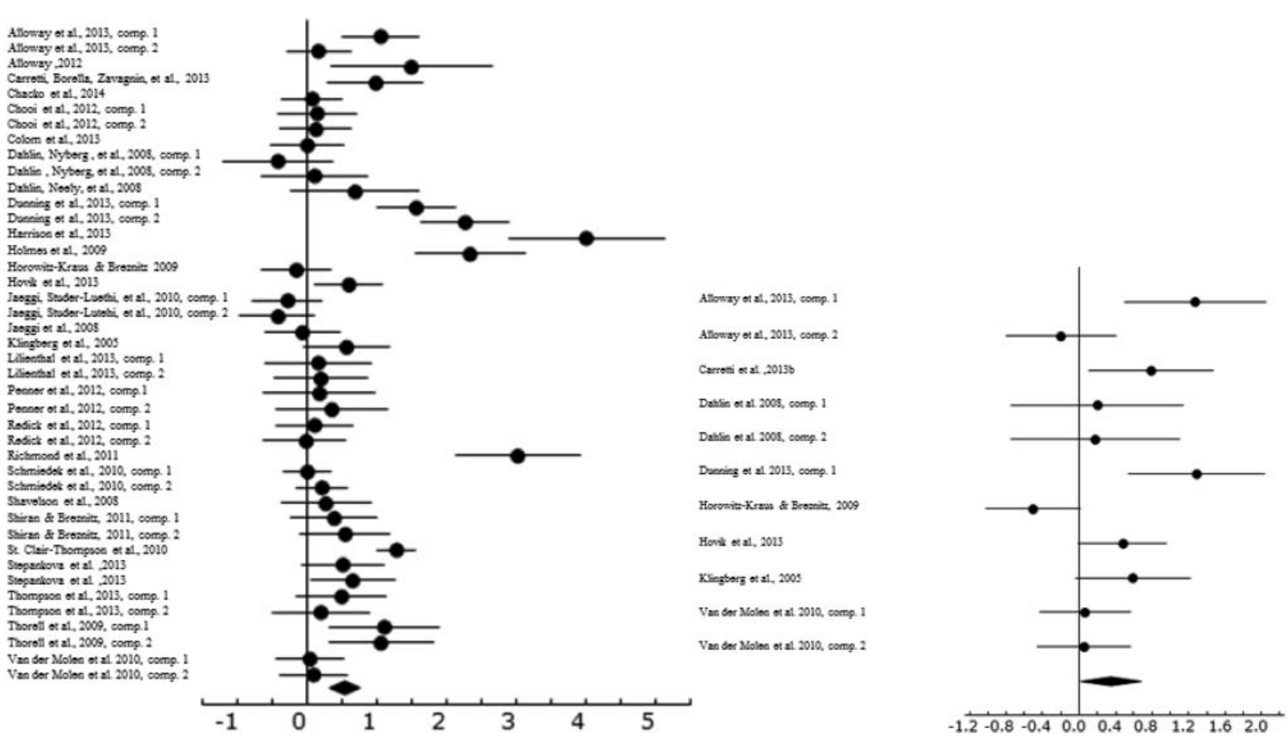

FIGURE 3 Forest plots for immediate (left panel) and delayed (right panel) transfer effects to verbal WM. *Note*. Points and horizontal lines show effect sizes (Hedges's *g*) and confidence intervals, respectively. The diamond displays the overall effect and its confidence interval (width of the diamond).

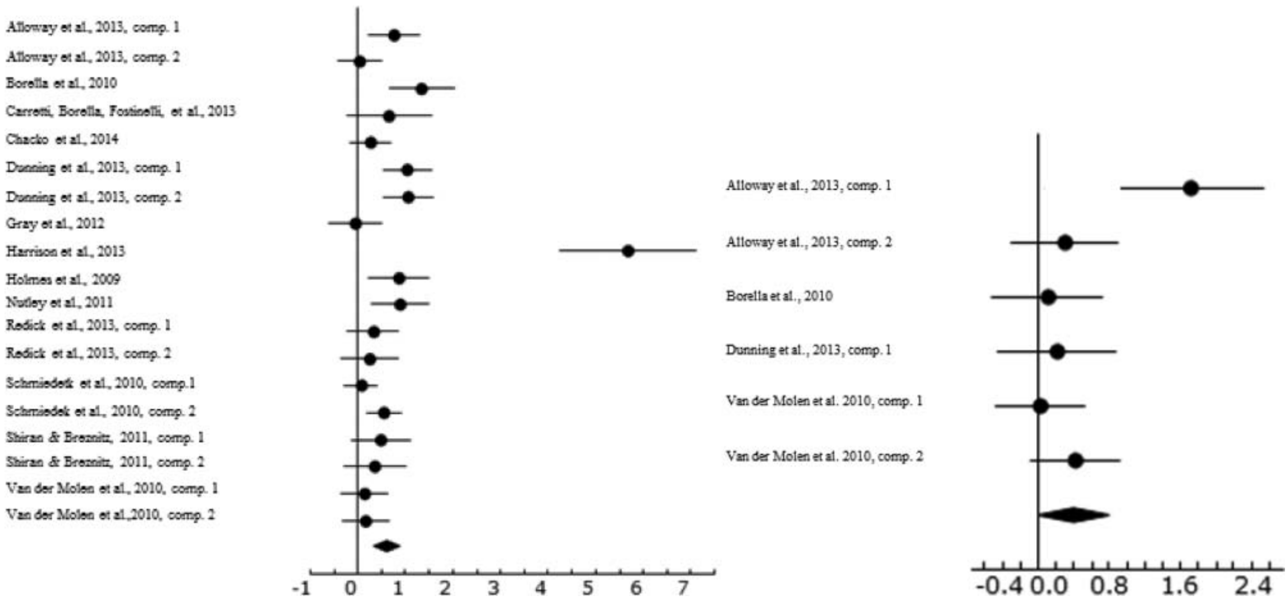

FIGURE 4 Forest plots for immediate (left panel) and delayed (right panel) transfer effects to visuospatial WM. *Note*. Points and horizontal lines show effect sizes (Hedges's *g*) and confidence intervals, respectively. The diamond displays the overall effect and its confidence interval (width of the diamond).

among studies, $Q(4) = 11.71$, $p < .05$, $I^2 = 65.84\%$. After removal of the one outlier, the mean effect size was $g = -0.06$, 95% CI [−0.44, 0.32]. The funnel plot and Egger's test indicated no publication bias. Thus, in line with Hypothesis 1.5, there was no sustained transfer effect of WM training to verbal ability.

Word decoding. Figure 7 shows forest plots for 14 immediate and five long-term (delayed) transfer effect sizes. After removal of outliers, the mean effect size ranged from $g = 0.04$, 95% CI [−0.11, 0.19], to $g = 0.09$, 95% CI [−0.05, 0.24]. The funnel plot and Egger's test indicated no publication bias. Thus, in line with Hypothesis 1.5, there

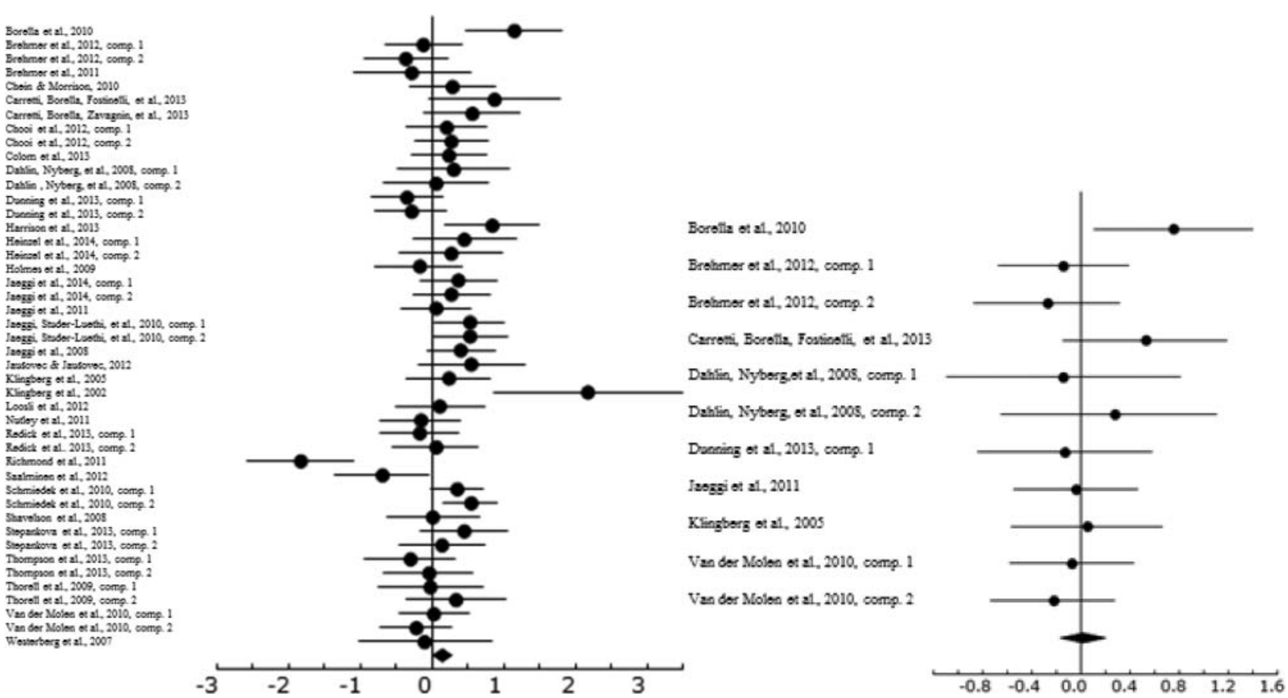

FIGURE 5 Forest plots for immediate (left panel) and delayed (right panel) transfer effects to nonverbal ability. *Note*. Points and horizontal lines show effect sizes (Hedges's *g*) and confidence intervals, respectively. The diamond displays the overall effect and its confidence interval (width of the diamond).

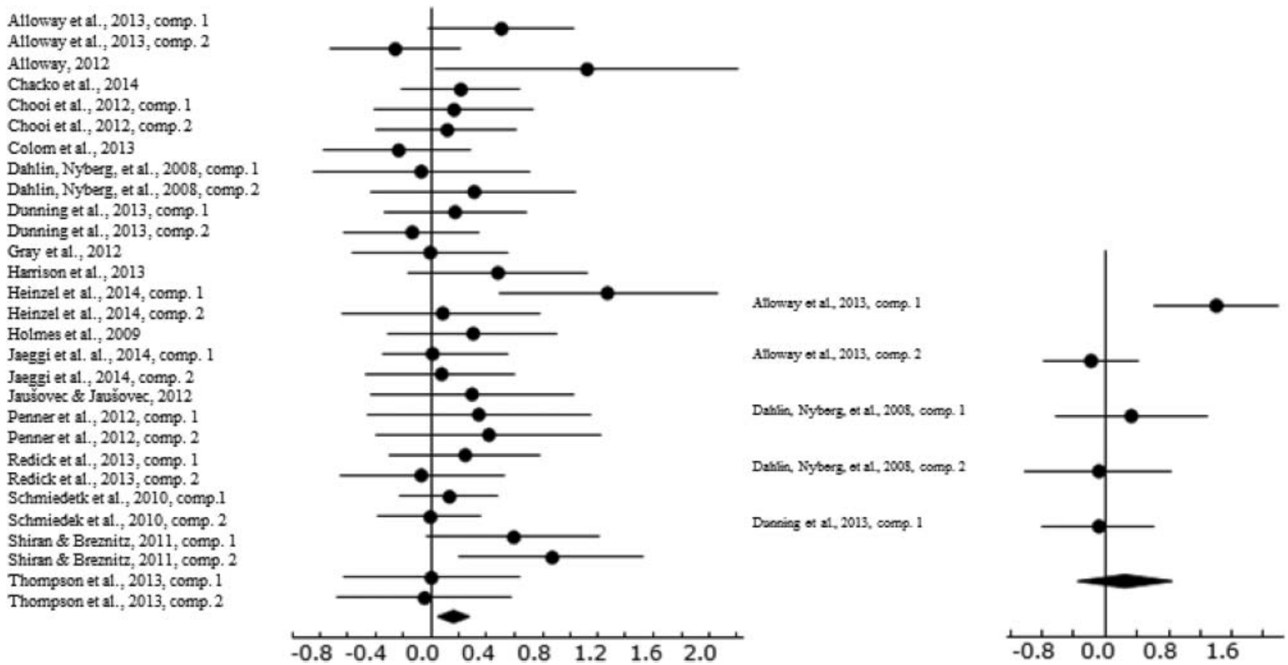

FIGURE 6 Forest plots for immediate (left panel) and delayed (right panel) transfer effects to verbal ability. *Note*. Points and horizontal lines show effect sizes (Hedges's g) and confidence intervals, respectively. The diamond displays the overall effect and its confidence interval (width of the diamond).

was no immediate transfer effect of WM training to word decoding.

We conducted follow-up testing an average of 6.2 months after the posttest. After the removal of the one outlier, the mean effect size was $g = 0.09$, 95% CI [−0.18, 0.36]. The funnel plot and Egger's test indicated no publication bias. Thus, in line with Hypothesis 1.5, there was no sustained transfer effect of WM training to word decoding.

Mathematical abilities. Figure 8 shows forest plots for 15 immediate and eight long-term (delayed) transfer effect sizes. After removal of outliers, the mean effect size

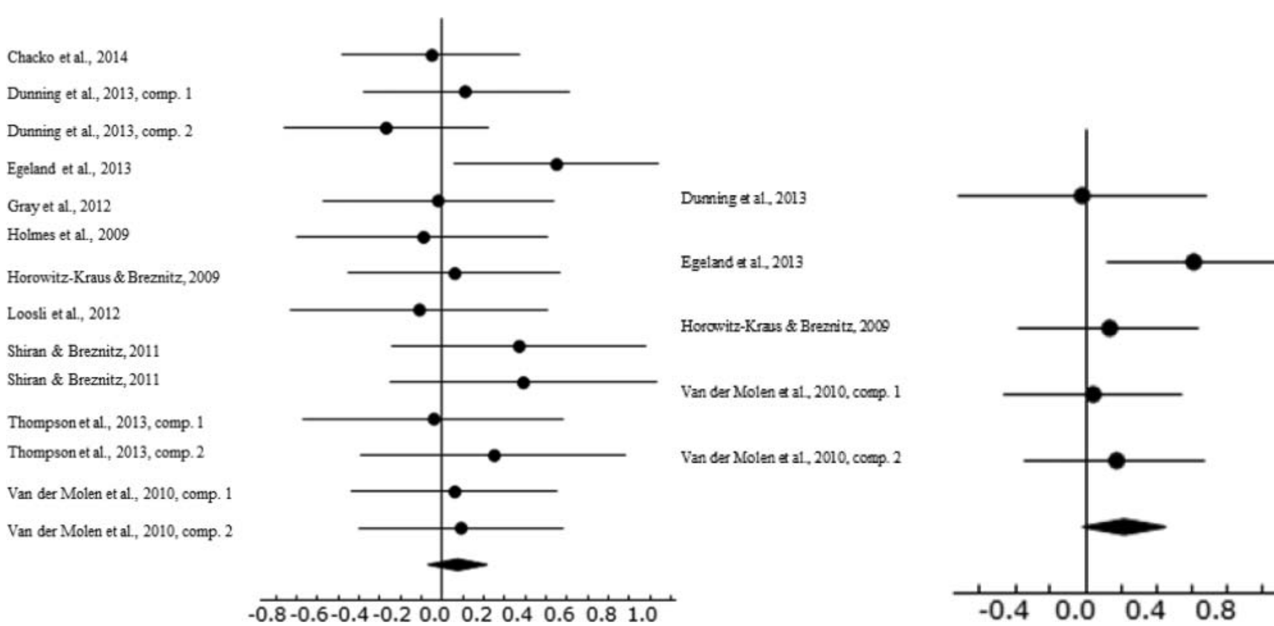

FIGURE 7 Forest plots for immediate (left panel) and delayed (right panel) transfer effects to word decoding. *Note*. Points and horizontal lines show effect sizes (Hedges's *g*) and confidence intervals, respectively. The diamond displays the overall effect and its confidence interval (width of the diamond).

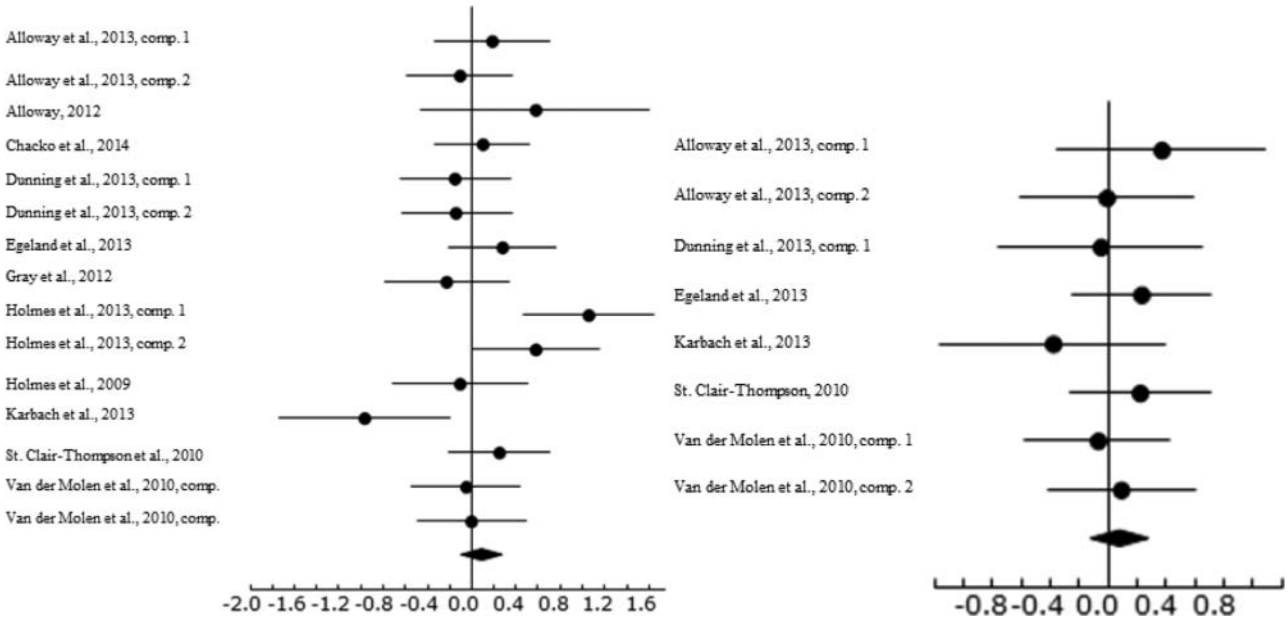

FIGURE 8 Forest plots for immediate (left panel) and delayed (right panel) transfer effects to mathematical abilities. *Note*. Points and horizontal lines show effect sizes (Hedges's g) and confidence intervals, respectively. The diamond displays the overall effect and its confidence interval (width of the diamond).

ranged from $g = 0.04$, 95% CI $[-0.11, 0.20]$, to $g = 0.13$, 95% CI $[-0.03, 0.30]$. Funnel plot and Egger's test indicated no publication bias. Thus, in line with Hypothesis 1.5, there was no immediate transfer effect of WM training to mathematical abilities.

We conducted follow-up testing an average of 6.13 months after the posttest. After removal of outliers, the mean effect size ranged from $g = 0.05$, 95% CI $[-0.15, 0.26]$, to $g = 0.11$, 95% CI $[-0.10, 0.32]$. The funnel plot and Egger's test indicated no publication bias. Thus, in line with Hypothesis 1.5, there was no sustained transfer effect of WM training to mathematical abilities.

Moderators of Transfer Effects Following Training of WM (RQ2)

To explore which training conditions have an influence on transfer effects following WM training (RQ2), we conducted moderator analyses for immediate transfer effects to verbal STM and verbal WM, immediate transfer effects to visuospatial STM and visuospatial WM, and immediate transfer effects to nonverbal ability and mathematical abilities. There were not enough studies to perform moderator analyses for the other transfer effects with a significant Q-statistic. The investigated moderators and the explained variance in heterogeneity between effect sizes for near- and far-transfer effects are shown in Table 4. This is followed by more detailed analyses of each moderator in the order of our initial hypotheses. We classify the effect sizes shown in Table 4 with respect to significance and effect size conventions for η^2 and R^2. For η^2, the conventions are .01 (small effect), .06 (medium effect), and .14 (large effect). For R^2, the conventions are .01 (small effect), .06 (medium effect), and .15 (large effect) (Murphy & Myors, 2004). These conventions were considered in addition to the significance of results because effect sizes could not be significant due to insufficient statistical power.

Age. In contrast to Hypothesis 2.1, age was not a significant moderator of any transfer effect. The values of effect sizes (R^2) were small and close to zero.

Training dose (total duration of training). To establish conservative hypothesis testing, we applied an alpha level of .20 to make it easier to reject the hypothesis. As can be seen in Table 4, contrary to Hypothesis 2.2, training dose explained variability in transfer effects to visuospatial STM ($\beta = .55$, $p < .01$, $k = 25$, $R^2 = .30$; large effect). The residual variance was not significant ($p = .41$). Training with a higher training dose (indicated by its common logarithm) yielded a larger mean effect size on visuospatial STM than training with a lower common logarithm of training dose. The common logarithm of training dose was not a significant moderator of transfer effects to visuospatial STM controlling for intervention type and two dummy-coded variables for location.

Session duration. As Table 4 shows, session duration explained variability in transfer effects to verbal STM. Session duration was the only significant moderator of transfer effects of WM training to verbal STM ($\beta = .32$, $p < .05$, k

TABLE 4
Effect Sizes η^2 and R^2 for Moderators of Near- and Far-Transfer Effects

	Verbal STM	*Visuospatial STM*	*Verbal WM*	*Visuospatial WM*	*Nonverbal Ability*	*Mathematical Abilities*
Age	.00	.00	.04	.00	.01	.00
Training dose	.01	**.30****	.01	.00	.00	.05
Session duration	**.10***	.05	.01	.04	.07	.03
Frequency of training per week	.04	.00	.04	.06	.10	.11
Training interval	.02	—	.11	.00	.09	—
Modality	.02	.01	.01	.03	.11	.15
Supervision	.06	.17	**.17****	**.16***	.01	.06
Instructional support	.01	.11	.01	.02	.00	.00
Feedback	.00	.05	.00	.00	.02	.11
Location	.08	**.24***	**.19****	.04	**.20****	—

Note. Effect size η was calculated for categorical moderators, effect size R^2 was calculated for continuous moderators. Significant values are printed in bold. Dashes indicate that a moderator analysis was not possible due to insufficient data. STM = short-term memory; WM = working memory.
*$p < .05$. **$p < .01$.

= 32, $R^2 = .10$; medium effect). Training with longer single training sessions yielded a larger mean effect size than training with shorter single training sessions. The residual variance was not significant ($p = .17$). Contrary to Hypothesis 2.3, session duration was not a significant moderator of other transfer effects of WM training. The values of effect sizes (R^2) were small to medium for transfer effects on which session duration had no significant influence.

Frequency of training per week. Contrary to Hypothesis 2.4, frequency of training per week was not a significant moderator of any transfer effect. The values of effect sizes (η^2) were small to medium.

Training interval. Contrary to Hypothesis 2.5, the training interval was not a significant moderator of any transfer effect. The values of effect sizes (η^2) were small to medium.

Modality. Contrary to Hypothesis 2.6, the trained modality was not a significant moderator of any transfer effect following WM training. However, the values of effect sizes (η^2) were small to large.

Supervision. As Table 4 shows, supervision explained variability in transfer effects to verbal and visuospatial WM. Supervision was a significant moderator of transfer effects to verbal WM, $Q(2) = 9.53$, $p < .01$. Eta-squared was large (.17). The residual variance was not significant ($p = .20$). Pairwise comparisons showed that supervised training yielded a larger mean effect size than training in the mere presence of other persons, $Q(1) = 7.75$, $p < .01$. In the multiple regression with intervention type, supervision, and location as predictors of transfer effects to verbal WM, training in the mere presence of other persons yielded smaller effects compared to training without the presence of other persons (reference category; $\beta = -.35$, $p < .05$).

Supervision was the only significant moderator for transfer effects to visuospatial WM, $Q(2) = 6.53$, $p < .05$. Eta-squared was large (.16). The residual variance was significant ($p < .01$). Pairwise comparisons showed that the mean effect size for supervised training was significantly larger than the mean effect size for training in the mere presence of other persons at an alpha level of .05, $Q(1) = 4.66$, $p = .03$. However, the difference between these two groups was not significant after Bonferroni correction of the alpha level (corrected alpha of .017). Thus, Hypothesis 2.7 was partly supported by the data. The values of effect sizes (η^2) were small to medium for transfer effects on which supervision had no significant influence.

Instructional support. Contrary to Hypothesis 2.8, instructional support was not a significant moderator of any transfer effect. The values of effect sizes (η^2) were small to medium.

Feedback. In contrast to Hypothesis 2.9, feedback was not a significant moderator of any transfer effect. The values of effect sizes (η^2) were small to large.

Location. We assumed that transfer effects of training in the laboratory should be larger than transfer effects of training at school or at home (Hypothesis 2.10). Location was a significant moderator of transfer effects to visuospatial STM, $Q(2) = 6.93$, $p < .05$. Eta-squared was large (.24). The residual variance was not significant ($p = .37$). Pairwise comparisons showed that training at home yielded a larger mean effect size than training at school, $Q(1) = 6.16$, $p = .013$. In the multiple regression with intervention type, supervision, and location as predictors of transfer effects to visuospatial STM, no dummy-coded variable for location was a significant moderator once other predictors were controlled for.

Location was also a significant moderator of transfer effects to verbal WM, $Q(2) = 12.14$, $p < .01$. Eta-squared was large (.19). The residual variance was significant ($p < .05$). Pairwise comparisons showed that training in school yielded a larger mean effect size than training in the laboratory, $Q(1) = 13.40$, $p < .001$. In the multiple regression with intervention type, supervision, and location as predictors of transfer effects to verbal WM, no dummy-coded variable for location was a significant moderator after controlling for the other predictors.

In addition, location was a significant moderator of transfer effects to nonverbal ability, $Q(2) = 10.04$, $p < .01$. Eta-squared was large (.20). The residual variance was not significant ($p = .31$), however. Pairwise comparisons showed that training in the laboratory yielded a larger mean effect size than training at school, $Q(1) = 8.13$, $p < .01$, as well as at home, $Q(1) = 6.43$, $p = .01$. In the multiple regression with intervention type and location as predictors of transfer effects to nonverbal ability, location was not a significant moderator after controlling for intervention type.

In summary, location was a significant moderator of several transfer effects, namely, visuospatial STM, verbal WM, and nonverbal ability. The pattern of influence of location on transfer effects was inconsistent for these constructs. Training in the laboratory was better than training at school and training at home only for transfer effects to nonverbal ability; therefore, our hypothesis received only partial evidence. The values of effect sizes (η^2) were small to medium for transfer measures on which location had no significant influence.

Hypotheses Addressing the Study Quality Characteristics

Intervention type. Intervention type was a significant moderator for transfer effects to visuospatial STM, $Q(1) = 9.68$, $p < .01$. Eta-squared was large (.28). The residual variance was not significant ($p = .33$), that is, after considering intervention type as a moderator of transfer effects to visuospatial STM, no significant variability in effect sizes among studies remained. Cogmed training as a commercial program stood out because it yielded a larger mean effect size than noncommercial training programs.

Intervention type was also a significant moderator of transfer effects to verbal WM, $Q(3) = 10.58$, $p < .05$. Eta-squared was large (.18). The residual variance was not significant ($p = .11$). Pairwise comparisons showed that Cogmed training yielded a larger mean effect size on verbal WM than n-back training, $Q(1) = 15.02$, $p < .001$. Jungle memory as a commercial training program yielded a larger mean effect size than n-back training only without Bonferroni correction, $Q(1) = 5.94$, $p = .015$.

Intervention type was also a significant moderator for transfer effects to nonverbal ability, $Q(2) = 6.44$, $p < .05$. Eta-squared was medium (.12). The residual variance was not significant ($p = .23$). Pairwise comparisons showed that Cogmed training yielded a smaller mean effect size than n-back training, $Q(1) = 8.02$, $p < .01$. Contrary to Hypothesis 2.11, intervention type was not a significant moderator of transfer effects of WM training to visuospatial WM.

Type of control group. We assumed that the mean effect on nonverbal ability for the comparison of training groups with passive control groups is larger than for the comparison of training groups with active control groups (Hypothesis 2.12). As can be seen in Table 4, type of control group explained variability in transfer effects to mathematical abilities. Type of control group was the only significant moderator for transfer effects to mathematical abilities, $Q(1) = 4.58$, $p < .05$. Eta-squared was large (.24), and the residual variance was not significant ($p = .32$). The mean effect size for comparisons of training groups with passive control groups was significantly larger than the mean effect size for comparisons of training groups with active control groups. Contrary to Hypothesis 2.12, type of control group was not a significant moderator of transfer effects of WM training to nonverbal ability. Because type of control group was a significant moderator for transfer effects only on mathematical abilities, and no other moderator explained variability in these transfer effects, type of control group could not be considered in multiple regression analyses for other moderators.

DISCUSSION

The results of our meta-analysis show that WM training yields both immediate and sustained *near-transfer effects* to both STM and WM components. We found small immediate far-transfer effects to nonverbal and verbal ability, but they are not sustainable. Training dose, session duration, supervision, and location are significant characteristics of the learning environments moderating transfer effects.

The results of our analyses clearly indicate that WMC is plastic and that WM training indeed taps into this potential. However, the core idea of WM training is to improve a domain-general attention capability that is also crucial for other cognitive abilities beyond WMC. This meta-analysis did not show sustained far-transfer effects of WM training to educationally relevant aspects such as verbal or mathematical abilities. Therefore the claim that WM training has practical benefits for learning or, more generally, education is not supported by the findings of this meta-analysis.

If this is a valid interpretation of the findings obtained in the field, there is a straightforward conclusion: We should bury all hopes that learning and education can be improved by boosting some general-purpose basic cognitive functions and redirect our resources for educational research and practice to more promising fields. We believe, however, that this would be premature. The findings could instead be interpreted as implicating that we have not even started to seriously design and vary the training conditions or, put more generally, the learning environment.

The remainder of this discussion therefore has two parts. In the first, we derive principles on conditions for improving specific WM aspects through training in what we call the *narrow task paradigm*. In the second, more speculative part, we suggest ways to considerably expand the theoretical framework of WM training by taking research on complex learning and transfer into account. These considerations lead us to some hypotheses for embedding WM training into *complex activity contexts*.

Principles for Effective WM Training in the Narrow Task Paradigm

The findings of this meta-analysis can be condensed to the following principles (also shown in Table 5).

We use the term "narrow task paradigm" for contexts that are intended to train only one specific basic cognitive function in an isolated task. We try to integrate the meta-analytic findings with earlier theorizing and study results in a series of eight principles.

1. Specificity Principle: WM training in the narrow task paradigm is effective with respect to the specifically trained WM component. This first principle is mainly based on the fact that we found only near-transfer effects to STM and WM components but hardly any far-transfer effects. With regard to the observed near-transfer effects, differences in near-transfer effects between the present analysis and the analysis by Melby-Lervåg and Hulme (2013) could be explained by the differentiation of transfer effects to STM and WM and the larger number of studies included in our analysis. Due to the larger sample sizes in our analysis, transfer effects are more precisely estimated than in the analysis by Melby-Lervåg and Hulme. Transfer effects for verbal STM and WM were different, which can be explained by task-specific overlap between the trained WM tasks and near-transfer WM tasks (see also Harrison et al., 2013). It seems that visuospatial STM and WM showed larger transfer effects than verbal STM and WM. We can only speculate about explanations of this effect. It might be that in the daily routine we deal more with verbal information than with visuospatial information. Thus, the verbal system might be more trained already, and thus be harder to improve, which results in lower transfer effects. Further studies are needed to provide a better understanding of this result.

2. Age Independence Principle: The effects of narrow task paradigm WM training are not dependent on age (at least for the age range 4–71 years). In contrast to the earlier meta-analysis by Melby-Lervåg and Hulme (2013), age was not found to be a significant moderator of any of the transfer effects in our meta-analysis in the range of 4 to 71 years. In contrast to the analysis by these authors, we did not categorize age but used the common logarithm of age in our analysis, therefore minimizing a loss of information associated with the categorization of continuous moderators. Melby-Lervåg and Hulme noted that it is possible that WM training is more effective in early years when brain plasticity is particularly high.

TABLE 5
Principles for Effective WM Training in the Narrow Task Paradigm

Principle No.	*Principle Name*	*Principle Formulation*
(1)	Specificity Principle	WM training in the narrow task paradigm is effective with respect to the specifically trained WM component.
(2)	Age Independence Principle	The effects of narrow task paradigm WM training are not dependent on age (at least for the age range 4–71 years).
(3)	The Longer-the-Better Principle	WM training effects in the narrow task paradigm increase with duration of single training sessions (at least in the range 6–60 min) and the total duration of the training.
(4)	Knowledge-of Results Principle	In narrow task paradigm WM training, simple and immediate feedback on knowledge of results works as well as elaborated feedback.
(5)	Strategy Prevention Principle	If narrow task paradigm WM training is adaptive to prevent the development or use of strategies to solve specific tasks, transfer effects will be larger compared to training that aims at or allows for the development or use of strategies to solve a task.
(6)	Ineffective Instruction Principle	In narrow task paradigm WM training, additional instructional support beyond explanations at the beginning of the training is neither effective for the trained task nor for transfer of the effects to other tasks
(7)	Supervision Principle	Effects of narrow task paradigm WM training are larger, if additionally present persons provide supervision and monitoring instead of just being in the same room.
(8)	Multimodality Principle	Bimodal WM training will yield larger effects than unimodal WM training.

Note. WM = working memory.

Because basic cognitive functions such as WMC decline with age (Salthouse et al., 2008), older adults might profit relatively more from WM training resulting in a nonsignificant overall effect of age. As other basic cognitive functions are also subject to an age-dependent decline (e.g., Deary et al., 2009), we propose that transfer effects of training in complex activity contexts are also independent of age within the range of 4 to 71 years. These assumptions do not imply that transfer effects outside this age span are dependent on age.

3. The Longer-the-Better Principle: WM training effects in the narrow task paradigm increase with duration of single training sessions (at least in the range 6–60 min) and the total time. Based on our moderator analyses, we assume that basic cognitive functions have to be trained for a certain amount of time in single training sessions to yield transfer effects. Although the time of single training sessions had an influence only on the transfer effect of WM training on verbal STM, the small to medium effects of session duration on the remaining transfer effects point to a positive influence of this variable on transfer effects. In a similar way, the effect sizes of the training dose (i.e., the total amount of training hours) on transfer effects suggest a positive influence of this moderator variable, assuming that the missing significance could be attributed to small sample sizes and low statistical power. Notably, the minimal duration of single training sessions and the minimal training dose to yield transfer effects are not yet clear.

4. Knowledge-of-Results Principle: In narrow task paradigm WM training, simple and immediate feedback on knowledge of results works as well as elaborated feedback. More elaborate forms of feedback thus cannot increase the WM training effect associated with simple and immediate feedback. This is consistent with prior research on feedback in general (Hattie & Timperley, 2007). Simple feedback, especially knowledge of results, has been reported to be effective for simple tasks and in the phase of skill automation (Hattie & Timperley, 2007). The results of the present meta-analysis seem to suggest that feedback beyond knowledge of results (e.g., process feedback) might not have an additional facilitating effect. We were not able to distinguish among the different types of feedback that go beyond knowledge of results, however. We therefore cannot exclude that these types of feedback have different directions of influence (positive or negative) on transfer effects of WM training, which meta-analytically levels out to a nonsignificant effect. Not much is known about what other kinds of feedback could be of any help in improving basic cognitive functions. There is a need for primary studies that test the hypothesis that feedback beyond knowledge of results might not have an additional facilitating effect directly by systematically varying different types of feedback.

5. Strategy Prevention Principle: If narrow task paradigm WM training is adaptive to prevent the development or use of strategies to solve specific tasks, transfer effects will be larger compared to training that aims at or allows for the development or use of strategies to solve a task. As Morrison and Chein (2011) noted, core training that uses tasks that continuously adapt to the performance of participants is more suited to reach far-transfer effects. Based on the results of our analysis (which included only studies with adaptive WM training), we assume that adaptive training of one specific basic cognitive function is best suited to improve the trained cognitive function.

6. Ineffective Instruction Principle: In narrow task paradigm WM training, additional instructional support beyond explanations at the beginning of the training is effective neither for the trained task nor for transfer of the effects to other tasks. Our meta-analysis showed that additional instructions during WM training had no influence on transfer effects. Training that is not intended to use strategies to solve a specific task should produce larger transfer effects (Morrison & Chein, 2011). Therefore, additional instructions during the training about strategies to solve the task should not be beneficial. Another explanation would be that instructions at the beginning of the task are enough to understand the trained tasks. The categories of instructional support were defined very broadly due to the large variation among studies regarding the characteristics of these moderators. No information enabling the construction of more elaborate categories was provided in the study descriptions; for example, qualitatively different instructional supports were coded with the same numbers in different studies. The presence of instructional support was coded if there was an instruction screen presented during training before each task as well as whether the experimenters answered questions (Lilienthal, Tamez, Shelton, Myerson, & Hale, 2013); it was also coded if instructions for the training task were given when participants wanted to check what they should do (Dahlin, Nyberg, et al., 2008). Possibly, for participants who have difficulties understanding a specific task, additional instructional support could be helpful. We were not able to explore this complex relationship with our meta-analysis. Therefore, it is unclear whether some categories of instructional support (e.g., explanations about how to properly work on the task when participants do not know what they should do) would uncover an influence of this variable.

7. Supervision Principle: Effects of narrow task paradigm WM training are larger, if additionally present persons provide supervision and monitoring instead of just being in the same room. The presence of a person who additionally supervises whether persons are training properly probably can motivate them to continue activities.

Supervised training yielded a larger mean effect size than training in the mere presence of other persons in the case of verbal and visuospatial WM. This is in line with research showing detrimental effects of the mere presence of other persons on task performance in complex tasks (for a review, see Aiello & Douthitt, 2001). In contrast to the mere presence of other persons, the presence of persons who additionally supervised the training might have motivated them to exert themselves. The values of effect sizes (η^2) were large for transfer effects to both verbal and visuospatial WM, suggesting an important role of supervision for the variability of the transfer effects. Supervision may also have an influence on the variability of other transfer effects. Insufficient statistical power could be one possible reason for nonsignificant medium to large effects of supervision on the variability of some transfer effects (e.g., verbal STM, mathematical abilities, visuospatial STM). Notably, it is conceivable that supervision during test phases could also be beneficial with respect to transfer effects. In addition, it is unclear whether tests phases were also supervised when training was supervised. We have no information regarding these aspects.

8. Multimodality Principle: Bimodal WM training will yield larger effects than unimodal WM training. This principle is somewhat more speculative than the others. The medium to large effect sizes obtained for transfer effects to nonverbal ability and mathematical abilities indicate that training both modalities of WM could be more beneficial than training one single modality. Training both modalities seems to be a more complex activity than training one modality. Notably, in most of the studies included in our analysis, both modalities of WM were trained. The inclusion of more studies in which only one modality of WM is trained could result in a different interpretation of the role of the trained modality for transfer effects.

The principles just mentioned are related to improvements of STM and WM components. The likelihood of transfer of effects from narrow task paradigm WM training to other activities increases with the similarity of training and transfer situation with respect to constraints and affordances. Inversely, transfer of effects from narrow task paradigm WM training to other cognitive functions decreases with decreasing overlap with respect to situational constraints and affordances in training and transfer situations. As situations requiring verbal and mathematical abilities, as well as fluid intelligence, have only minor overlap in task constraints and affordances with narrow task paradigm WM training tasks, transfer effects are unlikely to occur.

We found no far-transfer effects to word decoding and mathematical abilities. This is consistent with the results of Melby-Lervåg and Hulme (2013). No significant transfer effect to nonverbal ability remained after outliers were excluded. The immediate transfer effect to verbal ability was small and not sustained at follow-up. Therefore, we suggest concluding that isolated WM training programs of specific basic cognitive functions can improve performance beyond the specific task but in the confines of the trained cognitive function (i.e., WM).

Embedding WM Training in Complex Activity Contexts

Assuming that the meta-analytic results are partly due to the surprisingly small variability of the training conditions, we are now moving to somewhat more speculative grounds expressed by the hypotheses shown in Table 6.

We suggest using the term "complex activity contexts" for contexts in which a coordinated use of several basic cognitive functions is likely. We do not exclude the possible benefit of cognitive training in the clinical context (e.g., for persons with WM deficits). Impaired cognitive functions could be improved with isolated training of specific cognitive functions to deal with activities of everyday life, but for people with normal cognitive functioning, such training might yield no practically relevant benefits. One reason for the near-transfer effects in the absence of far-transfer effects could be the narrow paradigm of WM training (i.e., training of one specific cognitive function with a specific task). With respect to the theoretical model of WMC introduced earlier in this article, executive attention might have been improved for the specific characteristics of these narrow task contexts. However, complex cognitive performance (e.g., academic performance) is likely to require the interplay of several cognitive functions and involve working on diverse contents (e.g., mathematical formula, different types of text).

WM training research is built on the premise that improvement of basic cognitive functions is not comparable to conceptual learning or strategy learning because working on the training tasks in WM studies seems to be widely independent of knowledge. Obviously, differences between the two contexts exist. However, does this imply that mechanisms of learning transfer would be completely different with respect to the improvement of basic cognitive functions such as working memory or executive functions (Andersson, 2010)? Isolated training of specific cognitive functions in a narrow task context may have constellations of constraints and affordances for activities that are highly dissimilar from that in a complex activity context, such as solving a mathematical problem (Greeno, Smith, & Moore, 1993). This dissimilarity might be responsible for the lack in far-transfer effects.

People typically acquire complex skills (e.g., reading, writing) in complex activity contexts. Maybe basic WM functions could also be fostered in complex activity contexts. There is evidence that cognitive functions could be substantially enhanced in more complex activity contexts lasting for longer periods. For example, one study showed that children who were taught according to a curriculum that includes activities assumed to improve executive functions "showed better [executive functions] than peers

TABLE 6
Hypotheses for Effective Training of Basic Cognitive Functions in Complex Activity Contexts

Hypothesis No.	*Hypothesis Name*	*Hypothesis Formulation*
(1)	Activity Similarity Hypothesis	The transfer effects are more pronounced when working memory training is embedded in complex activity contexts as compared to the narrow task paradigm
(2)	Time and Transfer Hypothesis	The training effects in complex activity contexts increase with training duration of single sessions and with the total duration of training.
(3)	Adaptivity of Complex Learning Environments Hypothesis	In complex activity contexts, basic cognitive functions are enhanced only if they are challenged consistently (i.e., with increased cognitive function, the task has to be changed in a way to keep the demands for basic cognitive functions consistently high).
(4)	Cognitive Flexibility Hypothesis	Transfer of improved basic cognitive functioning to other tasks is more likely when two or more contexts are employed during training.
(5)	Detrimental Instruction Hypothesis	Instructional support in complex activity contexts reducing the demands on basic cognitive functions are detrimental for the improvement of these basic cognitive functions and hence for the transfer of effects.

attending other schools. They performed better in reading and math and showed more concern for fairness and justice" (Diamond & Lee, 2011, p. 5). Montessori classrooms have only one exemplar of all materials, and children therefore learn to wait; they also are engaged in instruction activities (Diamond & Lee, 2011). Such activities address cognitive basic functions such as working memory, shifting, and inhibition (Miyake et al., 2000). Working memory is addressed when a student listens to the instructions of two other students and has to keep the information in mind for later instruction. The student has to wait until the respective student has finished his or her instructions (inhibition as a basic cognitive function is addressed) and switch between the instructions of the two students (shifting as a basic cognitive function is addressed). In addition, these activities require coordination or interplay between the mentioned basic cognitive functions. Another example is schooling itself, which has been shown to enhance IQ (Ceci, 1991). Contexts that are more complex than WM training studies (such as curricula that include activities assumed to improve basic cognitive functions) address more cognitive functions and require the interplay of cognitive functions. The complex activities in these contexts may be more likely to be sufficient to yield transfer effects to complex cognitive performance.

Adele Diamond (2014) proposed a hypothetical model of how executive functions can be successfully improved. According to this model, executive functions such as working memory benefit most if they are continually challenged. Interventions are more successful if they also consider other variables that are related to executive functions. For example, in the light of negative consequences of stress and social isolation on executive functions, interventions should increase joy and feelings of social inclusion and support (Diamond, 2014).

Considering this seemingly high potential of embedding WM training in complex activity, we suggest five somewhat more speculative hypotheses to be tested in future research.

1. Activity Similarity Hypothesis: The transfer effects are more pronounced when WM training is embedded in complex activity contexts as compared to the narrow task paradigm. The rationale for this hypothesis is that basic cognitive functions can be used in a coordinated way during training, and affordances to use cognitive functions in coordinated ways are also typical of educationally more relevant transfer tasks. The reasoning behind this principle is that the lack of far transfer from WM training has one simple main reason: In the restricted action paradigm, a highly specific training task is targeting one specific cognitive function, namely, WM. However, academically relevant tasks do require several cognitive functions. For example, solving a moderately complex mathematical problem may require that relevant information to solve such a problem has to be kept in mind (taxing WM), irrelevant information has to be inhibited (taxing inhibition) and switching between aspects of the problem and information to solve it has to be done (taxing shifting). In addition, these activities have to be coordinated, requiring an "orchestration" of the basic cognitive functions involved. Complex activity contexts such as Montessori and add-ons to school curricula that are intended to tax several basic cognitive functions improved these functions and yielded transfer effects to academically relevant performance (for an overview, see Diamond & Lee, 2011).

2. Time and Transfer Hypothesis: The training effects in complex activity contexts increase with training duration of single sessions and with the total duration of training. With respect to complex activity contexts, we expect that the longer several basic cognitive functions and their interplay are taxed in these contexts in single training sessions and in total, the more these cognitive functions and other cognitive abilities and skills will improve. Complex activity contexts, such as school curricula that have shown to improve executive functions, challenge these functions throughout the whole day and for a longer period (for an overview, see Diamond, 2014). As is the case for WM training in the narrow task paradigm, the minimal duration of single training sessions and the minimal training dose to yield transfer effects are not yet clear.

3. Adaptivity of Complex Learning Environments Hypothesis: In complex activity contexts, basic cognitive functions are enhanced only if they are challenged consistently (i.e., with increased cognitive function, the task has to be changed in a way to keep the demands for basic cognitive functions consistently high). We thus suggest generalizing the according principle for the narrow task paradigm (see earlier, Principle 5, Strategy Prevention Principle) to the complex activity context.

4. Cognitive Flexibility Hypothesis: Transfer of improved basic cognitive functioning to other tasks is more likely when two or more contexts are employed during training. Established memory research (Tulving & Thomson, 1973), as well as more recent models on cognitive flexibility (Spiro, Coulson, Feltovich, & Anderson, 1988) and transfer of learning (e.g., Gick & Holyoak, 1983), suggests that learning is more likely to transfer when people have learned in multiple contexts. Although speculative, we suggest integrating this principle as an alternative to the transfer theory that was implicit in WM training studies (i.e., WM training improves executive attention, which then in turn is available in a multitude of different cognitive tasks beyond WM tasks). The latter model received no evidence in our meta-analytic findings.

5. Detrimental Instruction Hypothesis: Instructional support in complex activity contexts reducing the demands on basic cognitive functions are detrimental for the improvement of these basic cognitive functions and hence for the transfer of effects. Our meta-analysis has shown that additional instruction was not helpful in the narrow task paradigm. A clear introduction of the simple task at the beginning, simple feedback, together with some external monitoring during the training was enough. Of course, we do not claim that this generalizes to complex activity contexts. For these, we have ample empirical evidence that learning knowledge and skills benefit substantially from additional instructional guidance, more elaborate feedback, and increasing self-regulation. However, studies on these complex learning environments are typically targeted at learning concepts or strategies of a domain. Researchers have hardly ever controlled for effects of complex learning environments on basic cognitive functioning. There might be a tension between the improvement of basic cognitive functions and facilitating learning of domain knowledge and skills. For example, additional instruction by scaffolding aiming at reducing cognitive load might be effective for learning to engage in mathematical proofs (Kollar et al., 2014). Through reducing cognitive load, however, there is probably also a reduced taxing of the basic cognitive functions. With decreasing challenge, basic cognitive functions might be less likely to improve. Instructional support reducing the challenge could therefore be detrimental. Notably, if instructional support is reduced to a level so that learners are overstrained, they might not be able to properly carry out an activity or may stop carrying it out. Therefore, basic cognitive functions are probably less challenged.

Limitations of This Meta-Analysis

Some of the limitations of the meta-analysis of Melby-Lervåg and Hulme (2013) also apply to our analysis, so they are briefly mentioned here. One of these problems concerns a potential publication bias resulting from a predominant number of published studies in our meta-analysis. We were able to include only one study (Karbach et al., 2014) that had been published outside the period of our literature search, even though we asked many more of the authors of WM training studies to provide unpublished work. Unpublished studies and "gray" work might not have found significant transfer effects of WM training. A publication bias, in that positive results are more likely to be published, might lead to overestimations of the mean effect sizes. However, slight indications for a publication bias were found for only three transfer effects (verbal and visuospatial WM, and verbal ability).

Although we included more studies in our analysis than did Melby-Lervåg and Hulme (2013), we faced the same problem concerning the merging of a variety of different samples—namely, that group sizes were too small to perform analyses for separate subgroups.

Another limitation frequently seen in meta-analyses is the heterogeneity in how outcomes are measured (Walker et al., 2008). Due to differences in validities and reliabilities of outcome measures, transfer effects could have different degrees of evidence because these differences in methodological quality criteria are not considered in the calculation of the mean effect.

As mentioned earlier, a more specific problem of the current analysis is the calculation of single effect sizes using differences of gains between the training and control groups as the numerator in the calculation of Hedges's g. This may produce significant effect sizes even when treatment groups showed no improvement but control groups deteriorated. Also, a significant effect size could result when the control group improves and the treatment group deteriorates. However, these problems occurred for only two studies and were considered in the meta-analyses (effect size calculations with and without these studies).

Further limitations concern the single moderators used in the analysis. As discussed earlier, the categories of instructional support, feedback, and location were very broadly defined, allowing for a rough estimate of the influence of these variables on the variance in effect sizes among studies with a lack of differentiation (i.e., differentiated categories for these single moderators). Concerning training location, the specific conditions within one category of training location could differ widely. For training at home, we do not know important training conditions, such as whether training took place in a silent or noisy room or with distraction by a television. In addition, where transfer effects were assessed could also be relevant. For example, participants might be more able to focus on the transfer task in the laboratory compared to a classroom in a school, resulting in better transfer performance. However, we have hardly any information on where transfer effects were assessed in the analyzed studies. The frequency of training per week and training interval were also very broadly defined. Categorizations had to be made for these two variables because not enough information was available to analyze them as continuous moderators. In the case of the training interval, we did not have enough information at our disposal; for example, with respect to whether a break of 1 day between two training sessions also meant that the training was at the same time each day. Eventually, other potential moderators can be defined. For example, the influence of motivation on training outcomes has not yet been considered directly, but we expect that motivation can be relevant for transfer effects. Currently, information for the inclusion of this moderator (values on instruments measuring motivation) is missing in most of the studies considered in this review.

Directions for Future Research

We suggest that future research could address three different promising lines of research.

1. Using training as a paradigm for working memory research. This meta-analysis has shown that general executive attention that improves with training on one specific task is likely not the mechanism responsible for changes in WMC. Rather, these improvements of executive attention through training seem specific to a cognitive *function* (e.g., working memory). A refined training paradigm could help to investigate the very process of change in WMC more closely. After all, further establishing that working memory can be improved by dynamically changing environmental demands would open up highly interesting perspectives in several fields of cognitively oriented educational research. Most probably, this will be a theoretical perspective of orchestrated basic cognitive functions in which WM plays an important role.

2. Research on the moderating role of basic cognitive functioning in different learning environments. It seems important to further validate the role of basic cognitive functions, such as WM, shifting, or inhibition, in different types of learning environments (in the classroom and beyond). We clearly need more empirical research on the question under which instructional circumstances learners with weaker WMC, shifting, or inhibition capabilities are disadvantaged. For example, students with lower levels of basic cognitive functioning might be specifically disadvantaged in more taxing learning environments based on problem-solving (Hmelo-Silver, Duncan, & Chinn, 2007). For these students, learning environments based of direct instruction or highly scaffolded approaches such as worked-out examples (Renkl, 2014) might pose fewer problems. Of interest, although there has been a widely recognized discussion on the role of guidance in learning environments (Kirschner, Sweller, & Clark, 2006), there are hardly any systematic approaches to measure basic cognitive functioning (e.g., shifting and inhibition) as moderators of learning in differently guided learning environments.

3. Investigating the effects of different types of learning environments on basic cognitive functioning. It seems not unlikely that, over time, different types of learning environments might have different effects on basic cognitive functioning. It seems more promising to investigate ways to embed training of basic cognitive functions in different types of learning environments or even in a curriculum. Currently, we have only limited evidence coming from research on young children that an embedded training strategy might be effective (Diamond & Lee, 2011). For the development of domain-specific knowledge and skills in older children and adults, there might well be a negative relationship between optimally facilitating conceptual and strategic progress in a domain, and further improving basic cognitive functioning.

SUPPLEMENTAL MATERIAL

Supplemental data for this article can be accessed on the publisher's website.

REFERENCES

References marked with an asterisk indicate studies included in the meta-analysis.

Aiello, J. R., & Douthitt, E. A. (2001). Social facilitation from triplett to electronic performance monitoring. *Group Dynamics: Theory,* *Research, and Practice*, *5*, 163–180. http://dx.doi.org/10.1037/1089-2699.5.3.163

*Alloway, T. P. (2012). Can interactive working-memory training improve learning? *Journal of Interactive Learning Research*, *23*, 197–207.

*Alloway, T. P., Bibile, V., & Lau, G. (2013). Computerized working-memory training: Can it lead to gains in cognitive skills in students? *Computers in Human Behavior*, *29*, 632–638. http://dx.doi.org/10.1016/j.chb.2012.10.023

Alloway, T. P., Gathercole, S. E., & Pickering, S. J. (2006). Verbal and visuospatial short-term and working memory in children: Are they separable? *Child Development*, *77*, 1698–1716. http://dx.doi.org/10.1111/j.1467-8624.2006.00968.x

Andersson, U. (2010). Skill development in different components of arithmetic and basic cognitive functions: Findings from a 3-year longitudinal study of children with different types of learning difficulties. *Journal of Educational Psychology*, *102*, 115–134. http://dx.doi.org/10.1037/a0016838

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, *4*, 829–839. http://dx.doi.org/10.1038/nrn1201

Baddeley, A. D., Allen, R. J., & Hitch, G. J. (2011). Binding in visual working memory: The role of the episodic buffer. *Neuropsychologia*, *49*, 1393–1400. http://dx.doi.org/10.1016/j.neuropsychologia.2010.12.042

Banas, S., & Sanchez, C. A. (2012). Working memory capacity and learning underlying conceptual relationships across multiple documents. *Applied Cognitive Psychology*, *26*, 594–600. http://dx.doi.org/10.1002/acp.2834

Bloom, K. C., & Shuell, T. J. (1981). Effects of massed and distributed practice on the learning and retention of second-language vocabulary. *The Journal of Educational Research*, *74*, 245–248. http://dx.doi.org/10.1080/00220671.1981.10885317

*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*, 767–778. http://dx.doi.org/10.1037/a0020683

*Brehmer, Y., Rieckmann, A., Bellander, M., Westerberg, H., Fischer, H., & Bäckman, L. (2011). Neural correlates of training-related working-memory gains in old age. *Neuroimage*, *58*, 1110–1120. http://dx.doi.org/10.1016/j.neuroimage.2011.06.079

*Brehmer, Y., Westerberg, H., & Bäckman, L. (2012). Working-memory training in younger and older adults: Training gains, transfer, and maintenance. *Frontiers in Human Neuroscience*, *6*, 1–7. http://dx.doi.org/10.3389/fnhum.2012.00063

Bühner, M., König, C. J., Pick, M., & Krumm, S. (2006). Working memory dimensions as differential predictors of the speed and error aspect of multitasking performance. *Human Performance*, *19*, 253–275. http://dx.doi.org/10.1207/s15327043hup1903_4

Bühner, M., Kröner, S., & Ziegler, M. (2008). Working memory, visual-spatial–intelligence and their relationship to problem-solving. *Intelligence*, *36*, 672–680. http://dx.doi.org/10.1016/j.intell.2008.03.008

*Carretti, B., Borella, E., Fostinelli, S., & Zavagnin, M. (2013). Benefits of training working memory in amnestic mild cognitive impairment: specific and transfer effects. *International Psychogeriatrics*, *25*, 617–626. http://dx.doi.org/10.1017/S1041610212002177

*Carretti, B., Borella, E., Zavagnin, M., & de Beni, R. (2013). Gains in language comprehension relating to working-memory training in healthy older adults: Working-memory training in older adults. *International Journal of Geriatric Psychiatry*, *28*, 539–546. http://dx.doi.org/10.1002/gps.3859

Ceci, S. J. (1991). How much does schooling influence general intelligence and its cognitive components? A reassessment of the evidence. *Developmental Psychology*, *27*, 703–722. http://dx.doi.org/10.1037/0012-1649.27.5.703

*Chacko, A., Bedard, A. C., Marks, D. J., Feirsen, N., Uderman, J. Z., Chimiklis, A., & Ramon, M. (2014). A randomized clinical trial of Cogmed Working-memory training in school-age children with ADHD: A replication in a diverse sample using a control condition. *Journal of Child Psychology and Psychiatry*, *55*, 247–255. http://dx.doi.org/10.1111/jcpp.12146

*Chein, J. M., & Morrison, A. B. (2010). Expanding the mind's workspace: Training and transfer effects with a complex working memory span task. *Psychonomic Bulletin & Review*, *17*, 193–199. http://dx.doi.org/10.3758/PBR.17.2.193

*Chooi, W.-T., & Thompson, L. A. (2012). Working-memory training does not improve intelligence in healthy young adults. *Intelligence*, *40*, 531–542. http://dx.doi.org/10.1016/j.intell.2012.07.004

Colom, R., Abad, F. J., Rebollo, I., & Chun Shih, P. (2005). Memory span and general intelligence: A latent-variable approach. *Intelligence*, *33*, 623–642. http://dx.doi.org/10.1016/j.intell.2005.05.006

*Colom, R., Román, F. J., Abad, F. J., Shih, P. C., Privado, J., Froufe, M., & Jaeggi, S. M. (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*, 712–727. http://dx.doi.org/10.1016/j.intell.2013.09.002

Cowan, N. (1995). *Attention and memory: An integrated framework*. Oxford, England: Oxford University Press.

*Dahlin, E., Neely, A. S., Larsson, A., Bäckman, L., & Nyberg, L. (2008). Transfer of learning after updating training mediated by the striatum. *Science*, *320*, 1510–1512. http://dx.doi.org/10.1126/science.1155466

*Dahlin, E., Nyberg, L., Bäckman, L., & Stigsdotter Neely, A. (2008). Plasticity of executive functioning in young and older adults: Immediate training gains, transfer, and long-term maintenance. *Psychology and Aging*, *23*, 720–730. http://dx.doi.org/10.1037/a0014296

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review*, *3*, 422–433. http://dx.doi.org/10.3758/BF03214546

Deary, I. J., Corley, J., Gow, A. J., Harris, S. E., Houlihan, L. M., Marioni, R. E., . . . Starr, J. M. (2009). Age-associated cognitive decline. *British Medical Bulletin*, *92*, 135–152. http://dx.doi.org/10.1093/bmb/ldp033

Diamond, A. (2014). Want to optimize executive functions and academic outcomes? Simple, just nourish the human spirit. *Minnesota Symposia on Child Psychology*, *37*, 203–230. (NIHMS:605270)

Diamond, A., & Lee, K. (2011). Interventions shown to aid executive function development in children 4 to 12 years old. *Science*, *333*, 959–964. http://dx.doi.org/10.1126/science.1204529

*Dunning, D. L., Holmes, J., & Gathercole, S. E. (2013). Does working-memory training lead to generalized improvements in children with low working memory? A randomized controlled trial. *Developmental Science*, *16*, 915–925. http://dx.doi.org/10.1111/desc.12068

Duval, S., & Tweedie, R. (2000a). A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association*, *95*, 89–98. http://dx.doi.org/10.1080/01621459.2000.10473905

Duval, S., & Tweedie, R. (2000b). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*, 455–463. http://dx.doi.org/10.1111/j.0006-341X.2000.00455.x

*Egeland, J., Aarlien, A. K., & Saunes, B.-K. (2013). Few effects of far transfer of working-memory training in ADHD: A randomized controlled trial. *PLoS One*, *8*(10). Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2013-39053-001&site=ehost-live

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ (Clinical Research Ed.)*, *315*, 629–634. http://dx.doi.org/10.1136/bmj.315.7109.629

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. *Journal of Experimental Psychology: General*, *128*, 309. http://dx.doi.org/10.1037/0096-3445.128.3.309

Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: A latent-variable analysis. *Journal of Experimental Psychology. General*, *133*, 101–135. http://dx.doi.org/10.1037/0096-3445.133.1.101

Friedman, N. P., Miyake, A., Robinson, J. L., & Hewitt, J. K. (2011). Developmental trajectories in toddlers' self-restraint predict individual differences in executive functions 14 years later: A behavioral genetic analysis. *Developmental Psychology*, *47*, 1410–1430. http://dx.doi.org/10.1037/a0023750

Gathercole, S. E., Pickering, S. J., Knight, C., & Stegmann, Z. (2004). Working memory skills and educational attainment: Evidence from national curriculum assessments at 7 and 14 years of age. *Applied Cognitive Psychology*, *18*, 1–16. http://dx.doi.org/10.1002/acp.934

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, *15*, 1–38. http://dx.doi.org/10.1016/0010-0285(83)90002-6

*Gray, S. A., Chaban, P., Martinussen, R., Goldberg, R., Gotlieb, H., Kronitz, R., & Tannock, R. (2012). Effects of a computerized working-memory training program on working memory, attention, and academics in adolescents with severe LD and comorbid ADHD: a randomized controlled trial: Computerized working-memory training in adolescents with severe LD/ADHD. *Journal of Child Psychology and Psychiatry*, *53*, 1277–1284. http://dx.doi.org/10.1111/j.1469-7610.2012.02592.x

Greeno, J. G., Smith, D. R., & Moore, J. L. (1993). Transfer of situated learning. In D. K. Detterman & R. Sternberg (Eds.), *Transfer on trial: Intelligence, cognition, and instruction* (pp. 99–167). Norwood, NJ: Ablex.

Grissom, R. J., & Kim, J. J. (2005). *Effect size for research: A broad practical approach*. Mahwah, NJ: Erlbaum.

*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

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*, 81–112. http://dx.doi.org/10.3102/003465430298487

Hedden, T., & Yoon, C. (2006). Individual differences in executive processing predict susceptibility to interference in verbal working memory. *Neuropsychology*, *20*, 511–528. http://dx.doi.org/10.1037/0894-4105.20.5.511

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

*Heinzel, S., Schulte, S., Onken, J., Duong, Q.-L., Riemer, T. G., Heinz, A., & Rapp, M. A. (2014). Working-memory training improvements and gains in nontrained cognitive tasks in young and older adults. *Aging, Neuropsychology, and Cognition*, *21*, 146–173. http://dx.doi.org/10.1080/13825585.2013.790338

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist*, *42*, 99–107. http://dx.doi.org/10.1080/00461520701263368

*Holmes, J., & Gathercole, S. E. (2013). Taking working-memory training from the laboratory into schools. *Educational Psychology*, *34*, 440–450. http://dx.doi.org/10.1080/01443410.2013.797338

*Holmes, J., Gathercole, S. E., & Dunning, D. L. (2009). Adaptive training leads to sustained enhancement of poor working memory in children. *Developmental Science*, *12*, F9–F15. http://dx.doi.org/10.1111/j.1467-7687.2009.00848.x

*Horowitz-Kraus, T., & Breznitz, Z. (2009). Can the error detection mechanism benefit from training the working memory? A comparison between dyslexics and controls—An ERP study. *PLoS ONE*, *4*(9), e7141. http://dx.doi.org/10.1371/journal.pone.0007141

*Hovik, K. T., Saunes, B.-K., Aarlien, A. K., & Egeland, J. (2013). RCT of working-memory training in ADHD: Long-term near-transfer effects. *PLoS ONE*, *8*(12), e80561. http://dx.doi.org/10.1371/journal.pone.0080561

*Hubacher, M., Weiland, M., Calabrese, P., Stoppe, G., Stöcklin, M., Fischer-Barnicol, D., & Penner, I.-K. (2013). Working-memory training in patients with chronic schizophrenia: A pilot study. *Psychiatry Journal*, *2013*, 1–8. http://dx.doi.org/10.1155/2013/154867

Huedo-Medina, T. B., Sanchez-Meca, J., Martin-Martinez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I2 index? *Psychological Methods*, *11*, 193–206. http://dx.doi.org/10.1037/1082-989X.11.2.193

*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*, 6829–6833. http://dx.doi.org/10.1073/pnas.0801268105

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

Jaeggi, S. M., Buschkuehl, M., Perrig, W. J., & Meier, B. (2010). The concurrent validity of the N-back task as a working memory measure. *Memory*, *18*, 394–412. http://dx.doi.org/10.1080/09658211003702171

*Jaeggi, S. M., Buschkuehl, M., Shah, P., & Jonides, J. (2014). The role of individual differences in cognitive training and transfer. *Memory & Cognition*, *42*, 464–480. http://dx.doi.org/10.3758/s13421-013-0364-z

*Jaeggi, S. M., Studer-Luethi, B., Buschkuehl, M., Su, Y.-F., Jonides, J., & Perrig, W. J. (2010). The relationship between n-back performance and matrix reasoning — implications for training and transfer. *Intelligence*, *38*, 625–635. http://dx.doi.org/10.1016/j.intell.2010.09.001

*Jaušovec, N., & Jaušovec, K. (2012). Working-memory training: Improving intelligence—Changing brain activity. *Brain and Cognition*, *79*, 96–106. http://dx.doi.org/10.1016/j.bandc.2012.02.007

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review*, *19*, 509–539. http://dx.doi.org/10.1007/s10648-007-9054-3

Kane, M. J., Conway, A. R., Hambrick, D. Z., & Engle, R. W. (2007). Variation in working memory capacity as variation in executive attention and control. In A. Conway, C. Jarrold, M. Kane, A. Miyake, & J. Towse (Eds.), *Variation in working memory* (pp.21–48). New York, NY: Oxford University Press.

*Karbach, J., Strobach, T., & Schubert, T. (2014). Adaptive working-memory training benefits reading, but not mathematics in middle childhood. *Child Neuropsychology*. Advance online publication. http://dx.doi.org/10.1080/09297049.2014.899336

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, *41*, 75–86. http://dx.doi.org/10.1207/s15326985ep4102_1

Klingberg, T. (2010). Training and plasticity of working memory. *Trends in Cognitive Sciences*, *14*, 317–324. http://dx.doi.org/10.1016/j.tics.2010.05.002

*Klingberg, T., Fernell, E., Olesen, P. J., Johnson, M., Gustafsson, P., Dahlström, K., & Westerberg, H. (2005). Computerized training of working memory in children with ADHD—A randomized, controlled trial. *Journal of the American Academy of Child and Adolescent Psychiatry*, *44*, 177–186. http://dx.doi.org/10.1097/00004583-200502000-00010

*Klingberg, T., Forssberg, H., & Westerberg, H. (2002). Training of working memory in children with ADHD. *Journal of Clinical and Experimental Neuropsychology*, *24*, 781–791. http://dx.doi.org/10.1076/jcen.24.6.781.8395

Kollar, I., Ufer, S., Reichersdorfer, E., Vogel, F., Fischer, F., & Reiss, K. (2014). Effects of collaboration scripts and heuristic worked examples on the acquisition of mathematical argumentation skills of teacher students with different levels of prior achievement. *Learning and Instruction*, *32*, 22–36. http://dx.doi.org/10.1016/j.learninstruc.2014.01.003

Krumm, S., Schmidt-Atzert, L., Bühner, M., Ziegler, M., Michalczyk, K., & Arrow, K. (2009). Storage and non-storage components of working memory predicting reasoning: A simultaneous examination of a wide range of ability factors. *Intelligence*, *37*, 347–364. http://dx.doi.org/10.1016/j.intell.2009.02.003

*Lilienthal, L., Tamez, E., Shelton, J. T., Myerson, J., & Hale, S. (2013). Dual n-back training increases the capacity of the focus of attention. *Psychonomic Bulletin & Review*, *20*, 135–141. http://dx.doi.org/10.3758/s13423-012-0335-6

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.

*Loosli, S. V., Buschkuehl, M., Perrig, W. J., & Jaeggi, S. M. (2012). Working-memory training improves reading processes in typically developing children. *Child Neuropsychology*, *18*, 62–78. http://dx.doi.org/10.1080/09297049.2011.575772

Melby-Lervåg, M., & Hulme, C. (2013). Is working-memory training effective? A meta-analytic review. *Developmental Psychology*, *49*, 270–291. http://dx.doi.org/10.1037/a0028228

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science*, *21*, 8–14. http://dx.doi.org/10.1177/0963721411429458

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

Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General*, *130*, 621–640. http://dx.doi.org/10.1037/0096-3445.130.4.621

Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods*, *11*, 364–386. http://dx.doi.org/10.1177/1094428106291059

Morrison, A. B., & Chein, J. M. (2011). Does working-memory training work? The promise and challenges of enhancing cognition by training working memory. *Psychonomic Bulletin & Review*, *18*, 46–60. http://dx.doi.org/10.3758/s13423-010-0034-0

Murphy, K. R., & Myors, B. (2004). *Statistical power analysis: A simple and general model for traditional and modern hypothesis tests* (2nd ed.). Mahwah, NJ: Erlbaum.

Nigg, J. T. (2000). On inhibition/disinhibition in developmental psychopathology: Views from cognitive and personality psychology and a working inhibition taxonomy. *Psychological Bulletin*, *126*, 220–246. http://dx.doi.org/10.1037/0033-2909.126.2.220

*Nutley, S. B., Söderqvist, S., Bryde, S., Thorell, L. B., Humphreys, K., & Klingberg, T. (2011). Gains in fluid intelligence after training non-verbal reasoning in 4-year-old children: A controlled, randomized study. *Developmental Science*, *14*, 591–601. http://dx.doi.org/10.1111/j.1467-7687.2010.01022.x

Oberauer, K., Süβ, H.-M., Wilhelm, O., & Wittmann, W. W. (2008). Which working memory functions predict intelligence? *Intelligence*, *36*, 641–652 http://dx.doi.org/10.1016/j.intell.2008.01.007

Overton, R. C. (1998). A comparison of fixed-erediffects and mixed (random-effects) models for meta-analysis tests of moderator variable effects. *Psychological Methods*, *3*, 354–379. http://dx.doi.org/10.1037/1082-989X.3.3.354

*Penner, I.-K., Vogt, A., Stöcklin, M., Gschwind, L., Opwis, K., & Calabrese, P. (2012). Computerised working-memory training in healthy adults: A comparison of two different training schedules. *Neuropsychological Rehabilitation*, *22*, 716–733. http://dx.doi.org/10.1080/09602011.2012.686883

Pigott, T. D. (2012). *Advances in meta-analysis*. New York, NY: Springer Verlag.

Redick, T. S., Broadway, J. M., Meier, M. E., Kuriakose, P. S., Unsworth, N., Kane, M. J., & Engle, R. W. (2012). Measuring working memory capacity with automated complex span tasks. *European Journal of Psychological Assessment*, *28*, 164–171. http://dx.doi.org/10.1027/1015-5759/a000123

*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., Unsworth, N., Kelly, A. J., & Engle, R. W. (2012). Faster, smarter? Working memory capacity and perceptual speed in relation to fluid intelligence. *Journal of Cognitive Psychology*, *24*, 844–854. http://dx.doi.org/10.1080/20445911.2012.704359

Renkl, A. (2014). Toward an instructionally oriented theory of example-based learning. *Cognitive Science*, *38*, 1–37. http://dx.doi.org/10.1111/cogs.12086

*Richmond, L. L., Morrison, A. B., Chein, J. M., & Olson, I. R. (2011). Working-memory training and transfer in older adults. *Psychology and Aging*, *26*, 813–822. http://dx.doi.org/10.1037/a0023631

Rothstein, H. R. (2008). Publication bias as a threat to the validity of meta-analytic results. *Journal of Experimental Criminology*, *4*, 61–81. http://dx.doi.org/10.1007/s11292-007-9046-9

*Salminen, T., Strobach, T., & Schubert, T. (2012). On the impacts of working-memory training on executive functioning. *Frontiers in Human Neuroscience*, *6*. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2012-19447-001&site=ehost-live

Salthouse, T. A., Pink, J. E., & Tucker-Drob, E.M. (2008). Contextual analysis of fluid intelligence. *Intelligence*, *36*, 464–486. http://dx.doi.org/10.1016/j.intell.2007.10.003

Schmidt, F..L., & Hunter, J. E. (2003). History, development, evolution, and impact of validity generalization and meta-analysis methods 1975–2001. In K. R. Murphy (Ed.), *Validity generalization: A critical review* (pp. 31–65). Mahwah, NJ: Erlbaum.

*Schmiedek, F., Lövdén, M., & Lindenberger, U. (2010). Hundred days of cognitive training enhance broad cognitive abilities in adulthood: findings from the COGITO study. *Frontiers in Aging Neuroscience*, *2*, 1–10. http://dx.doi.org/10.3389/fnagi.2010.00027

*Shavelson, R. J., Yuan, K., Alonzo, A. C., Klingberg, T., & Andersson, M. (2008). On the impact of computerized cognitive training on working memory and fluid intelligence. In D. C. Berliner & H. Kuppermintz (Eds.), *Contributions of educational psychology to changing institutions, environments, and people* (pp. 1–11). New York, NY: Routledge.

Shipstead, Z., Redick, T. S., & Engle, R. W. (2010). Does working-memory training generalize? *Psychologica Belgica*, *50*, 245–276. http://dx.doi.org/10.5334/pb-50-3-4-245

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

*Shiran, A., & Breznitz, Z. (2011). The effect of cognitive training on recall range and speed of information processing in the working memory of dyslexic and skilled readers. *Journal of Neurolinguistics*, *24*, 524–537. http://dx.doi.org/10.1016/j.jneuroling.2010.12.001

Spiro, R. J., Coulson, R. L., Feltovich, P. J., & Anderson, D. K. (1988). Cognitive flexibility theory: Advanced knowledge acquisition in ill-structured domains. In *Proceedings of the Tenth Annual Conference of the Cognitive Science Society* (pp. 375–383). Hillsdale, NJ: Erlbaum.

St Clair-Thompson, H. L. (2010). Backwards digit recall: A measure of short-term memory or working memory? *European Journal of Cognitive Psychology*, *22*, 286–296. http://dx.doi.org/10.1080/09541440902771299

St Clair-Thompson, H. L., & Allen, R. J. (2013). Are forward and backward recall the same? A dual-task study of digit recall. *Memory & Cognition*, *41*, 519–532. http://dx.doi.org/10.3758/s13421-012-0277-2

*St Clair−Thompson, H., Stevens, R., Hunt, A., & Bolder, E. (2010). Improving children's working memory and classroom performance. *Educational Psychology*, *30*, 203–219. http://dx.doi.org/10.1080/01443410903509259

*Stepankova, H., Lukavsky, J., Buschkuehl, M., Kopecek, M., Ripova, D., & Jaeggi, S. M. (2014). The malleability of working memory and visuospatial skills: A randomized controlled study in older adults. *Developmental Psychology*, *50*, 1049–1059. http://dx.doi.org/10.1037/a0034913

Sterne, J. A. C., & Egger, M. (2005). Regression methods to detect publication and other bias in meta-analysis. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 99–110). Chichester, England: Wiley & Sons.

Talebi, M. (2013). Study of publication bias in meta-analysis using trim and fill method. *International Research Journal of Applied and Basic Sciences*, *4*, 31–36.

*Thompson, T. W., Waskom, M. L., Garel, K.-L. A., Cardenas-Iniguez, C., Reynolds, G. O., Winter, R., & Gabrieli, J. D. E. (2013). Failure of working-memory training to enhance cognition or intelligence. *PLoS One*, *8*, e63614. http://dx.doi.org/10.1371/journal.pone.0063614

*Thorell, L. B., Lindqvist, S., Bergman Nutley, S., Bohlin, G., & Klingberg, T. (2009). Training and transfer effects of executive functions in preschool children. *Developmental Science*, *12*, 106–113. http://dx.doi.org/10.1111/j.1467-7687.2008.00745.x

Titz, C., & Karbach, J. (2014). Working memory and executive functions: Effects of training on academic achievement. *Psychological Research*, *78*, 852–868. http://dx.doi.org/10.1007/s00426-013-0537-1

Tsaparlis, G. (2005). Non-algorithmic quantitative problem solving in university physical chemistry: A correlation study of the role of selective cognitive factors. *Research in Science & Technological Education*, *23*, 125–148. http://dx.doi.org/10.1080/02635140500266369

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*, 352–373. http://dx.doi.org/10.1037/h0020071

Unsworth, N., & Engle, R. W. (2007a). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, *114*, 104–132. http://dx.doi.org/10.1037/0033-295X.114.1.104

Unsworth, N., & Engle, R. W. (2007b). On the division of short-term and working memory: An examination of simple and complex span and their relation to higher order abilities. *Psychological Bulletin*, *133*, 1038–1066. http://dx.doi.org/10.1037/0033-2909.133.6.1038

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*, 498–505. http://dx.doi.org/10.3758/BF03192720

*Van der Molen, M. J., Van Luit, J. E. H., Van der Molen, M. W., Klugkist, I., & Jongmans, M. J. (2010). Effectiveness of a computerised working-memory training in adolescents with mild to borderline intellectual disabilities: Working-memory training in M-BID. *Journal of Intellectual Disability Research*, *54*, 433–447. http://dx.doi.org/10.1111/j.1365-2788.2010.01285.x

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

Walker, E., Hernandez, A. V., & Kattan, M. W. (2008). Meta-analysis: Its strengths and limitations. *Cleveland Clinic Journal of Medicine*, *75*, 431–439. http://dx.doi.org/10.3949/ccjm.75.6.431

Wecker, C., & Fischer, F. (2014). Where is the evidence? A meta-analysis on the role of argumentation for the acquisition of domain-specific knowledge in computer-supported collaborative learning. *Computers & Education*, *75*, 218–228. http://dx.doi.org/10.1016/j.compedu.2014.02.016

*Westerberg, H., Jacobaeus, H., Hirvikoski, T., Clevberger, P., Östensson, M.-L., Bartfai, A., & Klingberg, T. (2007). Computerized working-memory training after stroke–A pilot study. *Brain Injury*, *21*, 21–29. http://dx.doi.org/10.1080/02699050601148726

Wilson, D. B. (2005). *Meta-analysis macros for SAS, SPSS, and Stata*. Retrieved from http://mason.gmu.edu/̃dwilsonb/ma.html