

There is no convincing evidence that working memory training is effective: A reply to Au et al. (2014) and Karbach and Verhaeghen (2014)

Monica Melby-Lervåg¹ · Charles Hulme^{1,2}

Published online: 17 June 2015
© Psychonomic Society, Inc. 2015

Abstract The possible cognitive benefits of working memory training programs have been the subject of intense interest and controversy. Recently two meta-analyses have claimed that working memory training can be effective in enhancing cognitive skills in adulthood (Au et al. *Behavioural Brain Research* 228:(1) 107–115, 2014) and stemming cognitive decline in old age (Karbach & Verhaeghen *Psychological Science* 25:2027–2037, 2014). The current article critically evaluates these claims. We argue that these meta-analyses produce misleading results because of (1) biases in the studies included, (2) a failure to take account of baseline differences when calculating effect sizes, and (3) a failure to emphasize the difference between studies with treated versus untreated control groups. We present new meta-analyses and conclude that there is no convincing evidence that working memory training produces general cognitive benefits.

Keywords Cognitive training · Working memory

Introduction

The possible cognitive benefits of working memory training have been the subject of extensive recent research and a

considerable amount of controversy (see Shipstead, Redick & Engle, 2012). Working memory training programmes typically involve short periods of fairly intensive training on computerized tasks designed to tax working memory capacity. Theoretically, such training has been related to the idea that the capacity of working memory may place constraints on a wide range of cognitive functions including reasoning ability. This carries the implication that if working memory capacity can be increased by training it will produce a wide range of cognitive benefits. Most provocatively, it has been claimed that such training can produce widespread cognitive benefits including increases in scores on standardized tests of intelligence such as Raven's matrices (e.g., Klingberg, 2010). However, effects from studies of working memory training have been highly variable, and it has been suggested that meta-analysis can play an important role in clarifying our understanding of its effects (Melby-Lervåg & Hulme, 2013). In our earlier meta-analysis (Melby-Lervåg & Hulme, 2013) we concluded that working memory training produced effects on tasks that were trained directly but did not produce the sorts of "far transfer" effects to measures such non-verbal reasoning that some had claimed.

Two recent meta-analyses, however, claim that Working Memory training can be effective in enhancing cognitive skills in adulthood (Au et al., 2014) and stemming cognitive decline in old age (Karbach & Verhaeghen, 2014). Au et al. (2014), who focused on the effects of n-back training, state: "We conclude that short-term cognitive training on the order of weeks can result in beneficial effects in important cognitive functions as measured by laboratory tests" (p 1) and "Since Gf is a fundamental cognitive skill that underlies a wide range of life functions, even small improvements can have profound societal ramifications" (p 10). More strongly, in the latest such meta-analysis, Karbach and Verhaeghen (2014) conclude that "executive functions training and working memory training in

Electronic supplementary material The online version of this article (doi:10.3758/s13423-015-0862-z) contains supplementary material, which is available to authorized users.

✉ Charles Hulme
c.hulme@ucl.ac.uk

¹ Faculty of Educational Sciences, University of Oslo, Oslo, Norway

² Division of Psychology and Language Sciences, University College London, Chandler House, 2 Wakefield St., London WC1N 2PF, UK

old age is highly effective” and “can be useful tools for intervention in ... old age” (p 2035). Unfortunately, we believe such conclusions are unwarranted. In this commentary we will outline the claims put forward in these two recent meta-analyses and discuss why we believe their claims are unjustified. We highlight three main problems in these meta-analyses: (1) the basis for inclusion of studies, (2) the methods for calculation of a mean effect size, (3) the importance of making a clear distinction between treated and untreated control groups. Based on these concerns we present a reanalysis of the studies dealt with in these two meta-analyses. In our re-analyses we focus purely on the effects of Working Memory training on nonverbal reasoning as assessed by a variety of widely used measures of IQ with good psychometric properties. Our re-analyses indicate that working memory training does not produce reliable improvements on such measures.

The inclusion of studies

The first issue concerns the selection and coding of studies. Neither of the meta-analyses cited above provide information about individual study characteristics and what measures and effect sizes have been coded. In the Au et al. meta-analysis, there is no information about when the search started or ended, and 7 out of the 19 publications included in the meta-analysis are not listed in the reference list. The Karbach and Verhaeghen meta-analysis does not include a flow chart of studies through the review (so that the number of articles identified and abstracts reviewed and studies excluded is not clear). These issues make the meta-analyses less than transparent. Because a meta-analysis can potentially have large impact on a field, transparency and the ability to reproduce findings are important. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement (Moher et al. 2009, <http://www.prisma-statement.org>) is a consensus statement developed by an international group of researchers in health care for the conduct and reporting of systematic reviews and meta-analyses. This consensus statement strongly recommends the inclusion of detailed information about search and the characteristics of studies and outcomes coded so that the results from a meta-analysis can be reproduced.

For the meta-analysis by Au et al. (2014), their inclusion criteria were (1) studies that trained the participants on some form of adaptive n-back task, (2) included a control group, (3) used some form of fluid intelligence (Gf) outcome measure, (4) had participants between 18 and 50 years, (5) used a training program where n-back training could be isolated, and (6) training duration was more than a week. Their search for studies that matched these criteria was restricted to Google Scholar and PubMed (see p 3). However, Psych INFO and PubMed do not overlap completely, so that Psych Info can produce additional hits. Databases in education such as ERIC might also have added additional studies. By undertaking a more

comprehensive search in Psych Info and ERIC we detected several studies that are not included in the meta-analysis by Au et al., or listed as excluded studies in the flow chart, but that seemingly match the criteria for the review. For example: (1) Anguera et al. (2012) trained college students for 4–5 days a week with n-back training and measured effects on a card rotation Gf test. (2) Nussbaumer et al. (2013) trained healthy undergraduate students for 7.5 h and measured effects on Raven. (3) Colom, Quiroga et al. (2010) trained healthy undergraduates for 18 sessions. With a total sample of only 19 publications, the missing data these studies represent could affect the overall results.

The inclusion criteria relating to the design and the type of control conditions that were accepted for studies in the Karbach and Verhaeghen meta-analysis are not clear. However, from the reference list it is apparent that they included studies with no control group (e.g., Dotson, Sozda, et al. 2012; Dulaney & Rogers, 1994). The main part of their first analysis and the figures related to this were based on gain scores from pretest to posttest, either from studies without a control group or by not taking into account control group data. This is potentially misleading, since without control group data, pretest–posttest improvements do not provide any convincing evidence for the effectiveness of an intervention. Showing that an intervention is effective depends upon showing that participants in an intervention group make more progress between pretest and posttest than participants in a control group (see Shadish, Cook & Campbell, 2002). Thus, the first analysis of Karbach and Verhaeghen, which focuses purely on pretest–posttest gains without reference to changes in control participants, is highly misleading and can provide no evidence for the effectiveness of working memory training. Also, the electronic search by Karbach and Verhaeghen is restricted to the databases PsycInfo and PsycArticles. This is a limited search, and not searching databases such as Medline and ERIC may have led to studies being missed. We undertook a more comprehensive search also in ERIC, Medline and Google Scholar. This revealed two new studies (Bürki, Ludwig, et al., 2014; Xin, Lai et al., 2014) which were published after the Karbach and Verhaeghen meta-analysis was completed. Both of these studies examine working memory training in adults with a mean age over 60 using a control group. These studies appear to fulfill the inclusion criteria outlined in their paper and in our re-analysis below we report analyses both with, and without, these two new studies added.

Calculation of a mean effect size

The second, and arguably most serious, problem with both the Au et al. (2014) and Karbach and Verhaeghen (2014) meta-analyses is the manner in which effect sizes have been calculated. Neither paper adjusts for pretest differences on a by-study level when calculating a mean effect size. Instead, they

note that for the sample of studies as a whole there are no differences between the training and control groups at baseline. Failing to account for baseline differences can potentially give very misleading results. In the field of cognitive training, many studies fail to use random assignment of participants to conditions and have small sample sizes. This can result in large pretest differences between groups that have to be taken into account when calculating the effect size from an intervention. Even if there are no differences between the training and control groups at base line for the full sample of studies, many of the analyses done in these two papers are on subsets of studies. For example, in the Karbach and Verhaeghen meta-analysis, the analyses of studies concerning working memory training are based on a sample of between 5 and 12 studies. In such a small sample of studies, not correcting for baseline differences can have serious consequences for the results, even if there are no baseline differences for the full sample of studies.

To be more specific about examples of such bias, in the Karbach and Verhaeghen meta-analysis (their second analysis), they calculated effect sizes based merely on the standardized mean difference at posttest between groups (see p 2029). In one of the papers included in the meta-analysis (von Bastian, Langer et al. 2013), the effect size based on posttest differences only is +0.10 standard deviation units on Raven's matrices in favor of the trained group compared to an active control group. However, when baseline differences are taken into account, the effect size is actually -0.4 standard deviation units (the active control group actually make *more* gains on the Raven's matrices than the treatment group). Similarly, in the study of Salminen et al. (2012), which is included in the Au et al. meta-analysis, the treated group start off with higher scores than the control group and remain ahead of the control group at posttest. This results in a positive effect size for this study in the Au et al. meta-analysis. However, only the control group show improvements between pretest and posttest, while the treated group did not, which means that this study actually should yield a negative effect size ($g = 0.69$, see Table s1). These examples demonstrate that ignoring baseline differences can lead to seriously erroneous conclusions.

Distinction between treated and untreated controls

A third critical issue is that when evaluating cognitive training it is essential to make a clear distinction between studies using treated versus untreated controls. The better studies in this area use a treated control group who receive a task of similar content and identical duration to the treatment group (for example, adaptive working memory training versus non-adaptive practice on the same task, see Harrison, Shipstead et al. 2013). Such a design controls for numerous non-specific effects that may lead to improvements at posttest (for example, increased motivation, belief that you have benefitted from training,

familiarity with a computer that may be used to present the posttest measures).

We would argue that only studies with an appropriate treated control group can provide convincing support for a specific effect of cognitive training (just as in drug trials which typically use an inactive placebo pill to compare with the effects of an "active" pill containing a new drug). Studies have shown that participants randomized to a no-treatment control condition tend to improve less than those who receive a control condition of study defined treatment, or control treatment outside the study (Mohr, Spring et al. 2009). It follows that studies using untreated controls are likely to over-estimate the true effects of working memory training (or indeed any other form of training).

In line with this argument, in the meta-analysis by Au et al. (2014) there is a difference between studies with treated and untreated controls (see Table 1, p 5). The difference is not significant, but there are only 12 studies in each category. In fact, the effect size for studies using a treated control group is close to zero ($g = 0.08$) but moderate in size for the untreated controls ($g = 0.28$). In the Au et al. paper this finding is interpreted as follows: "In the analysis of control groups, however, the present direction of effects actually suggests that passive control groups could end up outperforming active control groups (passive vs. active: $g = 0.28$ vs. $g = 0.08$; Table 1), which runs opposite to the direction suggested by the idea that Hawthorne or expectancy effects drive improvements in both active control and treatment groups" (p 9). These statements by Au et al., however, appear to be based on a misunderstanding. The pattern they report (a larger effect of working memory training in studies with untreated controls compared to studies with treated controls) is exactly what is expected if expectancy effects are operating to facilitate performance.

In the Karbach and Verhaeghen study it is also clear that the effects of working memory and executive function training on far transfer measures are not significant when trained groups are compared to treated controls ($p = .056$). Thus, these findings underline the importance of separating studies with treated controls from those with only untreated controls.

Reanalyses of data

Au et al. (2014). We decided to repeat the meta-analysis by Au et al. after correcting for the short-comings discussed above. In the first analysis we did a replication of the analysis with the same studies that were included in the original dataset; however, we were unable to retrieve three studies included by Au et al in the meta-analysis. After contact with the authors, information for one of these studies was provided. The two remaining missing studies were coded with the overall effect size and standard errors reported by Au et al. (for Katz et al.

(submitted for publication) $g = 0.054$; Jaeggi et al. (2010) $g = -0.019$), and thus are included in our calculation of a mean effect size. In the second analysis, we added studies that were not detected by their search, but that met their inclusion criteria, and are published in the same period as studies that are included in their meta-analysis. This led to the addition of studies by Anguera et al. 2012, Colom et al. 2010 and Nussbaumer et al. 2013. In addition, we added a study by Bürki et al. (2014) that was published after Au et al. published their meta-analysis. Outcomes were coded from each study in line with the supplemental online information provided by Au et al. (2014), with a list of fluid intelligence outcomes used (Table S2). Notably, Au et al. coded reading comprehension as a measure of fluid intelligence. This measure is questionable as a measure of fluid intelligence, and several latent variable studies show that reading comprehension loads on a different factor than other tests commonly used to measure fluid intelligence, such as Raven or Cattell (see Francis, Snow et al. 2006). We therefore excluded measures of reading comprehension. For a full list of included studies and their characteristics and measures coded from each study in this reanalysis, see Table S1 Appendix 1. Furthermore, in our analysis we separated studies with treated controls from studies with only untreated controls, and computed effect sizes for training effects after correcting for possible differences between groups at pretest.

We did a meta-analysis on these studies to examine far transfer effects to measures of non-verbal reasoning for all studies (studies with either untreated or treated controls). Our first analysis was restricted to the studies that were originally included by Au et al. This analysis shows the following pattern: overall we find a small but significant pooled effect size of $g = 0.13$ 95 % CI [0.03, 0.20], $P = 0.01$, $k = 39$. For treated controls, we found a non-significant effect size $g = 0.09$ 95 % CI [-0.07, 0.25], $p = 0.26$, $k = 16$; and for untreated controls there was a larger and significant effect size, $g = 0.21$ 95 % CI [0.05, 0.36], $p = 0.01$, $k = 21$. Notably, even though this analysis is based on the same studies as reported by Au et al, we coded 39 different independent comparisons of n-back training versus control training from these studies (see Table S1 for what we have coded from each study), while Au et al. report only 24. Because of the lack of transparency in the Au et al. paper the reasons for this discrepancy are unclear.

Furthermore, if we extend our analysis to include the studies that were not included by Au et al. we get the same pattern: for all studies, a small but significant pooled effect size of $g = 0.10$ 95 % CI [0.1, 0.19], $p = 0.02$, $k = 45$, for treated controls, $g = 0.06$ 95 % CI [-0.06, 0.18], $p = 0.30$, $k = 20$, and for untreated controls $g = 0.19$ 95 % CI [0.04, 0.35], $p = 0.02$, $k = 22$. Thus our reanalysis (whether based either on exactly the same papers as Au et al., or after adding new studies) corresponds fairly closely to the results they reported for treated controls. For all studies (treated and untreated

controls) we found an overall effect size of $g = .13$ (their studies) or $g = .10$ (after adding some new studies), while the overall effect found by Au et al. was $g = 0.24$. Most critically, when we focus on studies with treated controls neither of the effect sizes $g = .09$ and $.06$ for studies with treated control groups is significant.

Notably, in spite of a seemingly large variation in effect sizes (ranging from -0.69 to 0.88) statistically the overall heterogeneity was not significant, $Q(44) = 37.52$, $P = 0.74$, $I = 0$. This corresponds to what Au et al. found, and is perhaps not surprising since all these studies examine the same type of training on a similar population (healthy adults). However, in spite of this, we believe that the difference found between treated and untreated controls here, although tentative, is important, since the effect size for studies with a treated control group is only one-third of the size found with untreated controls.

Karbach and Verhaeghen (2014)

We decided to assess the effects of working memory training on arguably the most important measure of far transfer (non-verbal ability) in elderly participants. Our re-analyses correct for the short-comings identified above (only using studies with a control group, separating studies with treated controls from studies with only untreated controls, and computing effect sizes for training effects after correcting for possible differences between groups at pretest). Karbach and Verhaeghen (their Fig. 1c) merged together different tasks such as task switching and nonverbal reasoning in their analysis of far transfer effects (see p 3). Here we focus purely on effects of Working Memory training on nonverbal reasoning because measures of nonverbal reasoning and task switching appear to tap different constructs. Latent variable studies show that task switching and nonverbal reasoning typically load on different factors, and are not highly correlated (see Friedman, Miyake et al. 2006). Also, although they are related, Salthouse (1998) found that most of the relationship between a task switching construct and higher order cognition was shared with other variables. We conclude that the empirical support for merging task switching and nonverbal reasoning into one far transfer construct is weak. We therefore focus here on nonverbal reasoning as assessed by a variety of widely used measures of IQ with good psychometric properties (typically Raven's Matrices, the Cattell Culture Fair Intelligence Test, or performance subtests from the Wechsler Abbreviated Scale of Intelligence; see Table S2, Appendix 1 for details).

Although it is unclear precisely which studies were included in Karbach and Verhaeghen's meta-analysis of working memory training, from their supplemental online reference list we have identified 17 studies of working memory training that appear to have been included. Of these studies, two had no control group (Dotson, Sozda, et al. 2012; Dulaney & Rogers,

1994), one study (Brehmer, Westerberg, et al. 2012) is based on the same sample as Brehmer et al. 2011, and two studies include only memory-related measures with no measures of far transfer (Buschkuehl, Jaeggi et al. 2008 and Shing, Schmiedek, et al. 2012). We will not consider those studies further here. The remaining studies do have control groups and do include measures of far transfer (11 studies with 12 independent experiments concerning older adults; 4 studies with a treated, and 8 with an untreated, control group). In addition to these studies, we include studies that met the inclusion criteria used by Karbach and Verhaeghen that we detected in a more comprehensive and updated search (Bürki, Ludwig, et al. 2014, Xin, Lai et al. 2014). See Table S2, Appendix 1 for a full list of the studies and the effect sizes coded from each study.

We conducted a meta-analysis to examine far transfer effects to measures of non-verbal reasoning. In our analyses we corrected for pretest differences between the groups. Overall, for studies that were included in the Karbach and Verhaeghen analysis, we find a significant mean effect size $g = 0.21$ [0.01, 0.42], $p = 0.04$, $k = 12$. There is also significant heterogeneity between these studies, $Q(18) = 33.30$, $p = 0.02$, $I^2 = 45.95$, $\text{Tau}^2 = 0.061$. A closer look at these studies shows that one study is an outlier. Borella, Caretti et al. 2010 found an improvement of 1.14 standard deviation units in the training group relative to their treated control group after only three sessions of training. Since there are few studies, this outlier has a large influence on the overall effect size. After this is excluded, the overall effect size is small and close to zero, $g = 0.05$, 95 % CI [-0.13, 0.23], $p = 0.58$, $k = 11$. For the subset of studies with untreated controls we find a significant mean difference of $g = 0.24$ 95 % CI [0.02, 0.45], $p = 0.03$, $k = 8$ (when excluding the outlying study $g = 0.19$ 95 % CI [-0.04, 0.41], $p = 0.10$, $k = 7$). However, for studies with treated controls the difference between groups is close to zero and nonsignificant, $g = -0.02$ 95 % CI [-0.71, 0.68], $p = 0.96$, $k = 4$. When adding the two more recent studies revealed by our search, the pattern remains unchanged: overall there was a small effect size $g = 0.13$ [-0.03, 0.28], $p = 0.04$, $k = 17$ (with the outlier included), for untreated controls $g = 0.15$ [-0.02, 0.31], $p = 0.09$, $k = 11$ and treated controls $g = 0.02$ [-0.39, 0.43], $p = 0.92$, $k = 6$. Thus, based on studies with treated controls, there is absolutely no sign of a beneficial effect of working memory training on measures of nonverbal reasoning in elderly participants.

Conclusion

Our focus in the current paper has been on two recent meta-analyses (Au et al., 2014; Karbach & Verhaeghen 2014) that have examined the effects of working memory training. Perhaps the most provocative suggestion from these two papers (and others in the Working Memory training literature) is

that such training can produce widespread cognitive benefits, including increases in scores on standardized tests of intelligence such as Raven's matrices. In our re-analyses of the data presented in these papers we focused purely on the effects of Working Memory training on nonverbal reasoning as assessed by a variety of widely used measures of IQ with good psychometric properties. The conclusion from our re-analyses is stark: there is no evidence that working memory training increases performance on these measures of IQ. For the studies included in the Au et al. meta-analysis, the overall effect size for working memory training (compared to either untreated or treated control groups) is $g = .13$; the corresponding figure for studies in the Karbach and Verhaeghen meta-analysis is $g = .21$ (though it must be emphasized that this estimate reflects a large effect from one study which is an outlier—when this is removed the overall effect of working memory training for studies in their analysis is $g = .05$). The magnitude of effect size that is meaningful for psychological or educational practice is debatable, but at least two organizations set this bar at 0.25 standard deviation units for studies using a rigorous design [Promising Practices Network (PPN) 2007; What Works Clearinghouse, 2007, See Cooper, 2008). A rigorous design in this case refers to randomized controlled trials or quasi-experimental designs using a convincing comparison group, and a sample size that exceeds 30. Thus, even if we overlook issues related to type of control group, the overall mean effect size here is clearly small and according to some guidelines would be seen as unlikely to be practical significance.

However, one of the main conclusions from our analyses is that studies using an untreated control group appear to substantially over-estimate the “true” effects of working memory training (or indeed any other type of training). As we have shown, for both of the meta-analyses evaluated here (Au et al., 2014; Karbach & Verhaeghen, 2014), for measures of nonverbal reasoning (IQ), there is a large difference in results from studies that use treated (active) rather than untreated (passive) controls. Most strikingly, the effect sizes for measures of far transfer are close to zero in studies that have employed treated control groups. We believe that it is essential for future studies of working memory training to use suitable treated control groups, and that studies with merely an untreated control group (that confound specific effects of working memory training with effects that may be due to expectancy and other non-specific effects) should no longer be published.

Finally, we should note that, many of the studies included in the two meta-analyses we have considered (and which have been reanalysed by us here) contain small sample sizes that result in very low power. Bogg and Lasecki (2015) have recently drawn attention to the problem that low power causes for the meta-analysis reported by Au et al. (2014). The problem is that studies with low power are likely to be biased because only those with large or very large effect sizes will generate statistically significant results and so get published

(the so-called “file-drawer problem” that studies with non-significant effects remain unpublished). This is termed by Bogg and Lasecki a “winners curse” because such very large effects are unlikely to be true. These issues were also highlighted earlier by Kraemer, Gardner, Brooks and Yesavage (1998) who argued forcefully that, when conducting meta-analyses, authors should exclude studies that are underpowered as this will go a long way to removing the problem of misleading conclusions arising from the file drawer problem. Another strong recommendation that follows from these observations is that journals should stop publishing studies on working memory training that do not have adequate statistical power.

In conclusion, we have argued that the meta-analyses of Au et al. and Karbach and Verhaeghen do not provide any convincing support for their conclusions that “our work demonstrates the efficacy of several weeks of n-back training in improving performance on measures of Gf” or that “executive functions training and working memory training in old age is highly effective”. In contrast, our reanalysis of the studies reviewed in these two papers shows that there is no evidence that working memory training produces improvements on measures of non-verbal reasoning taken from well standardized measures of cognitive ability (IQ). We believe that this conclusion is of theoretical and practical importance and is in line with a recent consensus statement on the effects of brain training (<http://longevity3.stanford.edu/blog/2014/10/15/the-consensus-on-the-brain-training-industry-from-the-scientific-community/>).

References

References marked with * are included in the meta-analyses.

- * Anguera, J. A., Bernard, J. A., Jaeggi, S. M., Buschkuhl, M., Benson, B. L., Jennett, S., ..., Seidler, R. D. (2012). The effects of working memory resource depletion and training on sensorimotor adaptation. *Behavioural Brain Research*, *228*(1), 107–115. doi: 10.1016/j.bbr.2011.11.040
- Au, J., Sheehan, E., Tsai, N., Duncan, G., Buschkuhl, M., & Jaeggi, S. (2014). Improving fluid intelligence with training on working memory: A meta-analysis. *Psychonomic Bulletin and Review*, *22*(2), 366–377. doi:10.3758/s13423-014-0699-x
- Bogg, T., & Lasecki, L. (2015). Reliable gains? Evidence for substantially underpowered designs in studies of working memory training transfer to fluid intelligence. *Frontiers of Psychology*, *5*, 1589. doi:10.3389/fpsyg.2014.01589
- Borella, E., Carretti, B., Riboldi, F., & De Beni, R. (2010). Working memory training in older adults: Evidence of transfer and maintenance effects. *Psychology and Aging*, *25*(4), 767.
- * Borella, E., Carretti, B., Zanoni, G., Zavagnin, M., & De Beni, R. (2013). Working memory training in old age: An examination of transfer and maintenance effects. *Archives of Clinical Neuropsychology*, *28*(4), 331–347. doi:10.1093/arclin/act020
- * Borella, E., Carretti, B., Cantarella, A., Riboldi, F., Zavagnin, M., & De Beni, R. (2014). Benefits of training visuospatial working memory in young-old and old-old. *Developmental Psychology*, *50*(3), 815–828. doi:10.1037/a0034293
- 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*(4), 1110–1120.
- * 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*, 63. doi:10.1177/0956797614548725DS1
- Buschkuhl, M., Jaeggi, S.M., Hutchison, S., Perrig-Chiello, P., Däpp, C., Müller, M., ..., Perrig, W.J. (2008). Impact of working memory training on memory performance in old-old adults. *Psychology and Aging*, *23*(4), 743.
- * Bürki, C. N., Ludwig, C., Chicherio, C., & de Ribaupierre, A. (2014). Individual differences in cognitive plasticity: An investigation of training curves in younger and older adults. *Psychological Research*, *78*(6), 821–835. doi:10.1007/s00426-014-0559-3
- * Carretti, B., Borella, E., Zavagnin, M., & Beni, R. (2013). Gains in language comprehension relating to working memory training in healthy older adults. *International Journal of Geriatric Psychiatry*, *28*(5), 539–546. doi:10.1002/gps.3859
- * Chooi, W., & Thompson, L. A. (2012). Working memory training does not improve intelligence in healthy young adults. *Intelligence*, *40*(6), 531–542. doi:10.1016/j.intell.2012.07.004
- * Clouter, A. (2013). *The effects of dual n-back training on the components of working memory and fluid intelligence: An individual difference approach. Submitted in partial fulfillment for the requirements for the degree of Master of Science.* Nova Scotia: Dalhousie University.
- * Colom, R., Quiroga, M., Shih, P., Martínez, K., Burgaleta, M., Martínez-Molina, A., ..., Ramírez, I. (2010). Improvement in working memory is not related to increased intelligence scores. *Intelligence*, *38*, 497–505. doi:10.1016/j.intell.2010.06.008
- * Colom, R., Román, F. J., Abad, F. J., Shih, P. C., Privado, J., Froufe, M., ..., Jaeggi, S. (2013). Adaptive n-back training does not improve general ability at the construct level: Gains on individual tests suggest training may enhance visuospatial processing. *Intelligence*, *41*(5), 712–727. doi.org/10.1016/j.intell.2013.09.002
- Cooper, H. (2008). The search for meaningful ways to express the effects of interventions. *Child Development Perspectives*, *2*, 181–186. doi: 10.1111/j.1750-8606.2008.00063.x
- * Dahlin, E., Nyberg, L., Bäckman, L., & Neely, A. (2008). Plasticity of executive functioning in young and older adults: Immediate training gains, transfer, and long-term maintenance. *Psychology and Aging*, *23*(4), 720–730. doi:10.1037/a0014296
- Dotson, V.M., Sozda, C.N., Marsiske, M., & Perlstein, W.M. (2012). Within-session practice eliminates age differences in cognitive control. *Aging, Neuropsychology, and Cognition*, 1–10.
- Dulaney, C. L., & Rogers, W. A. (1994). Mechanisms underlying reduction in Stroop interference with practice for young and old adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(2), 470.
- Francis, D., Snow, C., August, D., Carlson, C. D., Miller, J., & Iglesias, A. (2006). Measures of reading comprehension: A latent variable analysis of the diagnostic assessment of reading comprehension. *Scientific Studies of Reading*, *10*(3), 301–322.
- Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science*, *17*, 172–179.
- 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(12), 2409–2419. doi:10.1177/0956797613492984
- * Heinzl, S., Schulte, S., Onken, J., Quynh-Lam, D., Riemer, T. G., Heinz, A., ..., Rapp, M.A. (2014). Working memory training improvements and gains in non-trained cognitive tasks in young and older adults. *Aging, Neuropsychology, and Cognition*, 21(2), 146–173. doi:10.1080/13825585.2013.790338
- * Jaeggi, S. M., Studer-Luethi, B., Buschkuhl, 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. doi:10.1016/j.intell.2010.09.001
- * Jaeggi, S. M., Buschkuhl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 6829–6833. doi:10.1073/pnas.0801268105
- * Jaeggi, S. M., Buschkuhl, M., Shah, P., & Jonides, J. (2014). The role of individual differences in cognitive training and transfer. *Memory & Cognition*, 42(3), 464–480. doi:10.3758/s13421-013-0364-z
- Karbach, J., & Verhaeghen, P. (2014). Making working memory work: A meta-analysis of executive-control and working memory training in older adults. *Psychological Science*, 25, 2027–2037.
- Klingberg, T. (2010). Training and plasticity of working memory. *Trends in Cognitive Sciences*, 14, 317–324. doi:10.1016/j.tics.2010.05.002
- Kraemer, H., Gardner, C., Brooks, J. O., & Yesavage, J. A. (1998). Advantages of excluding underpowered studies from meta-analysis: Inclusionist vs exclusionist viewpoints. *Psychological Methods*, 3, 23–31. doi:10.1037/1082-989X.3.1.23
- * Kundu, B., Sutterer, D. W., Emrich, S. M., & Postle, B. R. (2013). Strengthened effective connectivity underlies transfer of working memory training to tests of short-term memory and attention. *The Journal of Neuroscience*, 33(20), 8705–8715. doi:10.1523/JNEUROSCI.5565-12.2013
- Melby-Lervåg, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology*, 49, 270–291.
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med*, 6(6), e1000097. doi:10.1371/journal.pmed1000097
- Mohr, D.C., Spring, B., Freedland, K.E., Beckner, V., Arean, P., Hollon, S.D., ..., Kaplan, R. (2009). The selection and design of control conditions for randomized controlled trials of psychological interventions. *Psychotherapy and Psychosomatics*, 78 (5), 275–284
- * Nussbaumer, D., Grabner, R. H., Schneider, M., & Stern, E. (2013). Limitations and chances of working memory training. *Proceedings of the 35th Annual Conference of the Cognitive Science Society*, 3175–3180.
- * Oelhafen, S., Nikolaidis, A., Tullia, P., Blaser, D., Koenig, T., & Perrig, W. J. (2013). Increased parietal activity after training of interference control. *Neuropsychologia*, 51(13), 2781–2790. doi:10.1016/j
- * 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(2), 359–379. doi:10.1037/a0029082
- * Richmond, L. L., Morrison, A., Chein, J., & Olson, I. R. (2011). Working memory training and transfer in older adults. *Psychology and Aging*, 26, 813–822. doi:10.1037/a0023631
- * Rudebeck, S.R., Bor, D., Ormond, A., O'Reilly, J.X., & Lee, A.C.H. (2012). A potential spatial working memory training task to improve both episodic memory and fluid intelligence. *PLoS ONE* 7(11) doi: 10.1371/journal.pone.0050431
- * Salminen, T., Strobach, T., & Schubert, T. (2012). On the impacts of working memory training on executive functioning. *Frontiers In Human Neuroscience*, 6(6), 166. doi:10.3389/fnhum.2012.00166
- Salthouse, T. A. (1998). Independence of age-related influences on cognitive abilities across the life span. *Developmental Psychology*, 34, 851–864.
- * Schwarb, H. (2012). *Optimized cognitive training: Investigating the limits of brain training on generalized cognitive function. A dissertation presented to the academic faculty*. Georgia: Institute of Technology.
- * Schweizer, S., Hampshire, A., & Dalgleish, T. (2011). Extending brain-training to the affective domain: Increasing cognitive and affective executive control through emotional working memory training. *PLoS ONE*, 6(9), e24372. doi:10.1371/journal.pone.0024372
- * Smith, S. P., Stibric, M., & Smithson, D. (2013). Exploring the effectiveness of commercial and custom-built games for cognitive training. *Computers in Human Behavior*, 29(6), 2388–2393. doi:10.1016/j.chb.2013.05.014
- * Stepankova, H., Lukavsky, J., Buschkuhl, M., Kopecek, M., Ripova, D., & Jaeggi, S. M. (2014). The malleability of working memory and visuospatial skills: A randomized controlled study in old adults. *Developmental Psychology*, 50(4), 1049–1059. doi:10.1037/a0034913
- * Stephenson, C., & Halpern, D. F. (2013). Improved matrix reasoning is limited to training on tasks with a visuospatial component. *Intelligence*, 41, 341–357. doi:10.1016/j.intell.2013.05.006
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton-Mifflin.
- Shing, Y. L., Schmiedek, F., Lövdén, M., & Lindenberger, U. (2012). Memory updating practice across 100 days in the COGITO study. *Psychology and Aging*, 27(2), 451. doi:10.1037/a0025568
- Shipstead, Z., Redick, T. S., & Engle, R. W. (2012). Is working memory training effective? *Psychological Bulletin*, 138, 628–654. doi:10.1037/a0027473
- * Thompson, T. W., Waskom, M., 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(5), e63614. doi:10.1371/journal.pone.0063614
- * von Bastian, C. C., Langer, N., Jäncke, L., & Oberauer, K. (2013). Effects of working memory training in young and old adults. *Memory & Cognition*, 41, 611–624. doi:10.3758/s13421-012-0280-7
- * Xin, Z., Lai, Z., Li, F., & Maes, J. H. R. (2014). Near- and far-transfer effects of working memory updating training in elderly adults. *Applied Cognitive Psychology*, 28, 403–408. doi:10.1002/acp.3011
- * Zinke, K., Zeintl, M., Rose, N. S., Putzmann, J., Pydde, A., & Kliegel, M. (2013). Working memory training and transfer in older adults: Effects of age, baseline performance, and training gains. *Developmental Psychology*, 50(1), 304–315. doi:10.1037/a0032982
- * Zinke, K., Zeintl, M., Eschen, A., Herzog, C., & Kliegel, M. (2012). Potentials and limits of plasticity induced by working memory training in old-old age. *Gerontology*, 58(1), 79–87. doi:10.1159/000324240 007