

The Influence of Individual Differences in Cognitive Ability on Working Memory Training Gains

**Elizabeth A. Wiemers, Thomas S. Redick
& Alexandra B. Morrison**

Journal of Cognitive Enhancement

ISSN 2509-3290

Volume 3

Number 2

J Cogn Enhanc (2019) 3:174-185

DOI 10.1007/s41465-018-0111-2



Your article is protected by copyright and all rights are held exclusively by Springer Nature Switzerland AG. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".



The Influence of Individual Differences in Cognitive Ability on Working Memory Training Gains

Elizabeth A. Wiemers¹ · Thomas S. Redick¹ · Alexandra B. Morrison²

Received: 18 July 2018 / Accepted: 6 November 2018 / Published online: 16 November 2018
© Springer Nature Switzerland AG 2018

Abstract

Working memory training research has produced mixed results in terms of finding benefits beyond the trained tasks (i.e., transfer). One potential limitation is that the research thus far has failed to isolate the specific combination of factors that makes working memory training work best. Individual differences in cognitive ability at pretest may be an important factor, suggesting possible aptitude-by-treatment interactions. Baseline cognitive ability could be (a) positively related, (b) negatively related, or (c) unrelated to training task improvements. The relationship between ability and training gains is important given the idea that larger training improvements should lead to greater transfer. However, the majority of training studies tend to be under-powered to examine individual differences. We pooled studies conducted in related labs to increase power while minimizing differences between studies. In the studies that were identified for this project, young adults completed complex span training and working memory and/or fluid intelligence as pretest measures. The combined samples from seven studies resulted in a sample of 192 participants. Analyses focused on the relationship between pretest cognitive ability and training performance across training days. There was no evidence that individuals lower in cognitive ability improved more than high-ability subjects on the training tasks. Instead, we found a positive relationship for both working memory and fluid intelligence measured at pretest with the amount of training improvement. In addition, the association between pretest working memory and working memory training performance appears to be domain-general—verbal and visuospatial content do not produce differential relationships.

Keywords Cognitive training · Working memory · Fluid intelligence

Working memory (WM) capacity is a construct involving storage, active maintenance, and manipulation of information (see Baddeley 2012; Cowan 2017, for reviews). Individuals with high WM functioning often outperform individuals low in WM on a plethora of laboratory tasks requiring cognitive control. WM is also positively associated with educational assessments, and negatively correlated with mind-wandering frequency (Unsworth and Redick 2017). Due to positive relationships with many other higher-order cognitive constructs including fluid intelligence (Gf) and language comprehension, WM tasks have become a popular paradigm for cognitive training. That is, because higher WM is often associated with

better outcomes in a variety of contexts, researchers have pursued increasing WM via training with the belief that such WM improvements will lead to concomitant benefits in related, ecologically relevant behaviors (Shipstead et al. 2012). These goals vary from decreasing symptoms of ADHD (Klingberg et al. 2005) to helping reduce addictive tendencies (Houben et al. 2011) to improving intelligence (Jaeggi et al. 2008).

Working Memory Training and Transfer

Generally, WM training paradigms involve a pretest session, several days or weeks of training, and a posttest session. In the case of WM training, training often involves either complex span tasks or *n*-back tasks, both of which have been heavily used in the broader WM literature. Complex span tasks interleave processing and storage tasks. One commonly used complex span task, operation span, presents participants with math problems to

✉ Elizabeth A. Wiemers
ewiemers@purdue.edu

¹ Department of Psychological Sciences, Purdue University, 703 Third Street, West Lafayette, IN 47907, USA

² California State University, Sacramento, CA, USA

solve and to-be-remembered letters. Later, they are asked to report the letters in the order in which they were presented. The n -back task requires participants to update a running list of items, and respond whether the current item matches the item presented a given number (n) items back.

In WM training studies, the pretest and posttest sessions include various cognitive assessments. Tasks that are similar to the training programs, or thought to measure the same construct (WM), are considered “near transfer” tasks. For example, operation span training may include reading span at pretest and posttest. These measures are typically of interest to ensure the training “worked,” in that the construct—WM—was trained, rather than performance on the specific task—operation span—increasing simply due to practice. Typically, the critical transfer results for those hoping to improve the lives of individuals with alcohol use disorders or children with ADHD are the assessments that are less similar to the WM training materials that are sometimes referred to as “far transfer” tasks. Higher-order cognition tasks related to WM, such as intelligence and language comprehension, fall into this category.

Transfer results in WM training studies have been mixed—most studies observe significant improvement in tasks that are similar to the training materials, but not in other outcomes (see Melby-Lervåg et al. 2016; Soveri et al. 2017 for meta-analyses). There are several identified methodological concerns about WM training studies that must be accounted for, and could be the source of some of the differing results (Morrison and Chein 2011; Redick et al. 2015). One such limitation is that the research thus far has failed to isolate the specific combination of factors that would optimize WM training efficacy (Katz et al. 2016). One of the earliest WM training studies argued that the dose, or amount of training, is an important factor in determining whether far transfer is observed. Jaeggi et al. (2008) assessed Gf with Raven’s Advanced Progressive Matrices or Bochumer Matrizen test before and after participants completed 8–19 sessions of dual n -back training. Specifically, they concluded, “Furthermore, we demonstrate that the extent of gain in intelligence critically depends on the amount of training: the more training, the more improvement in Gf” (Jaeggi et al. 2008, p. 6829). Stepankova et al. (2014) provided additional support for this notion in a study of older adults, reporting that a training group that performed 20 n -back sessions showed more transfer to Gf compared to a group that completed 10 n -back sessions. However, meta-analyses have shown no effect of length of training on far transfer (Au et al. 2015; Melby-Lervåg et al. 2016; Sala and Gobet 2017), and Redick et al. (2013) discussed specific interpretative challenges that apply to Jaeggi et al. (2008).

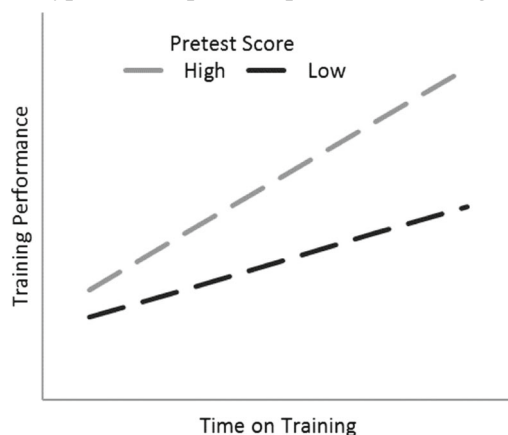
Individual Differences in Working Memory Training Research

While methodological factors contributing to transfer in WM training have been the focus of many reviews of the WM training literature (e.g., Redick et al. 2015), individual differences may also play an important role in the amount of transfer observed (Katz et al. 2016). Other research areas, particularly education psychology, have shown that individual differences in cognitive ability influence the effectiveness of interventions (aptitude-by-treatment interactions). For example, individual differences in WM moderated which of two types of lesson plan was more effective for at-risk students learning mathematical fractions (Fuchs et al. 2014). This type of interaction has the potential to be important for WM training efficacy, too. These possibilities have been discussed in the cognitive training literature as examples of the *compensatory* versus *magnification* accounts (Lövdén et al. 2012; Karbach et al. 2017). The compensatory perspective proposes that interventions such as cognitive training might serve to minimize cognitive ability differences by primarily benefitting individuals who were lower in cognitive functioning before the treatment. In contrast, the magnification account states that interventions will exacerbate differences in cognitive abilities, such that those high in ability will benefit even more from the intervention than those lower in ability.

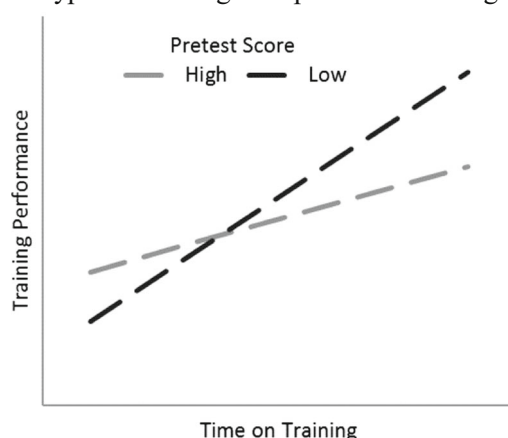
Indeed, although there is ongoing debate about the robustness of transfer after WM training, one common finding across WM training studies is that there is substantial between-subject variability in the amount of WM training improvement across sessions. Several recent studies have examined the interaction between pre-existing traits and subsequent training performance. Figure 1 depicts the three possibilities when looking at training gains with respect to pretest measures—they could be (a) positively related, (b) negatively related, or (c) unrelated to training task improvements. Guye et al. (2017) evaluated many non-cognitive factors including personality, hobbies, experience with technology, and cognition-related beliefs at pretest, but found that none of them influenced WM training gains. Studer-Luethi et al. (2012) found that young adults low in neuroticism and high in conscientiousness, measured at pre-test, had higher means on the n -back training task. Studer-Luethi et al. (2016) similarly showed that children lower in neuroticism achieved higher levels of performance on WM training tasks. Finally, Bürki et al. (2014) investigated age differences in training and found a positive relationship such that young adults showed greater improvements in n -back training than older adults.

As noted elsewhere (von Bastian and Oberauer 2014), relatively few studies have directly assessed the relationship between cognitive ability and training gains, which is particularly important given the idea that larger training improvements should lead to greater transfer (Jaeggi et al. 2011). Karbach

a Hypothetical positive pretest to training relationship



b Hypothetical negative pretest to training relationship



c Hypothetical null pretest to training relationship

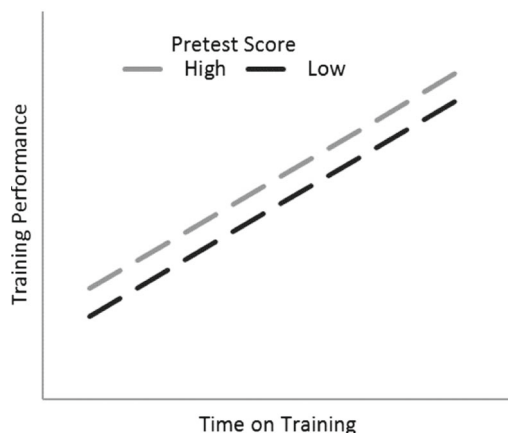


Fig. 1 a Hypothetical positive pretest to training relationship. b Hypothetical negative pretest to training relationship. c Hypothetical null pretest to training relationship

et al. (2017) observed that individuals lower in WM at pretest showed greater response time improvements across four sessions of task-switching training, but the training did not involve WM. With WM training, two studies are particularly relevant for the analyses conducted in the current research.

Jaeggi et al. (2011) investigated changes in Gf after n -back training with children. They found “children with an initially high level of Gf performance started with higher WM training levels...but showed less gain in training” (Jaeggi et al. 2011, p. 10084). Thus, cognitive ability appeared to have a negative relationship with WM training improvements (Fig. 1b). This pattern is consistent with the notion that those lower in cognitive ability have more room for improvement than those already high in cognitive ability (Titz and Karbach 2014). Conversely, Foster et al. (2017) investigated near and far transfer after complex span training with young adults. Foster et al. found that those at pretest higher in WM, measured by complex span, improved more on WM training tasks (Fig. 1a).

There are several potential reasons why the Jaeggi et al. and Foster et al. studies show conflicting results. 1) Jaeggi et al. used Gf as the cognitive ability predictor, whereas Foster et al. used WM. Although WM and Gf are highly correlated (cf. Redick et al. 2016), they are not identical constructs (Heitz et al. 2006; Shipstead et al. 2016), and thus may lead to different influences on WM training gains. 2) Jaeggi et al. studied children and Foster et al. studied young adults. Relationships between pretest cognitive ability scores and WM training gains could vary depending on the age of the sample. 3) Jaeggi et al. used n -back training and Foster et al. used complex span training. Although n -back and complex span tasks are both thought to measure WM functioning, the two tasks are less correlated than one might expect of two tasks thought to measure the same construct (Redick and Lindsey 2013). 4) Another factor to consider when interpreting the results of these two studies is the relatively small sample sizes of the relevant training groups; Jaeggi et al. tested 32 children and Foster et al.’s final sample of complex span training participants was $n = 40$. 5) Finally, the analysis approach differs with Jaeggi et al. using a median-split approach to create high- and low-Gf groups. In contrast, Foster et al., used an extreme-groups design to form high- and low-WM training groups based on participants’ scores falling in the upper or lower quartile of scores from a previous battery of complex span tests.

We note that prior investigations of the relationship between individual differences in training improvement and transfer improvement have frequently used correlations of difference scores, and such an approach has numerous statistical issues that render the results uninformative about causal relations (see Smoleń et al. 2018; Tidwell et al. 2014). However, despite the statistical limitations, the idea that the amount of training improvements is critical for transfer has pervaded the literature. Accordingly, it is the general idea that this pattern could be important for the efficacy of training studies that we consider.

Current Study

As stated above, there has been little research on the role of individual differences in cognitive ability and their impact on WM training gains. In addition, the studies that have examined this topic have been rather limited in their designs and analytical approaches, particularly as it relates to insufficient sample sizes for individual differences analyses and the use of correlated gain scores (Smoleń et al. 2018; Tidwell et al. 2014). In the present work, we do not investigate gain scores, and we do not relate training-task improvements to posttest performance. Rather, we use the pretest data only to predict training improvements. For the training data, instead of looking at the difference between starting and ending levels in training, we use modeling to include all of the training days. By doing so, we do not lose all the information from the intervening sessions, which may be important for characterizing the trajectory of training improvements (viz., linear or non-linear).

The current research uses a full, continuous, range of pretest scores, which can better answer the question of whether and how pretest performance affects training performance, instead of median-split or extreme-groups dichotomization of the pretest data. We also increase the number of participants compared to previous research by evaluating a multi-level model that aggregates data from several studies. In the present work, we specifically examine the degree to which pre-existing individual differences in WM and Gf influence WM training gains. In addition, we examined verbal and spatial WM training separately; at least one training study has explicitly argued that Gf far transfer occurs specifically after WM training with visuospatial tasks (Stephenson and Halpern 2013).

In the present work, we evaluate three possible patterns for the relationship between baseline ability and training gains. Following the logic of Jaeggi et al. (2011), we would expect that with increased scores at pretest, training trajectories would be flatter, showing less improvement, than for those with lower scores at pretest (Fig. 1b). However, based on Foster et al. (2017), we would expect the opposite pattern, with participants who exhibit higher WM and Gf scores at pretest producing greater training improvements compared to those lower in WM and Gf (Fig. 1a). There is also a third possibility, specifically that there is no relationship between cognitive ability at pretest and training gains (Fig. 1c).

Method

The current analyses were conducted on a combined dataset of seven separate WM training studies. Because most of the studies used training programs based on either Chein and Morrison (2010) or Harrison et al. (2013), the training task

procedures were very similar across studies. Samples of healthy, young adults who had completed complex span training and had pretest WM and/or Gf measures were identified for this project. Specifically, those studies that included training tasks from Chein and Morrison (2010) or Harrison et al. (2013) were targets due to the confidence in the similarity of the training tasks. After identifying candidate studies, we contacted the authors to obtain the necessary pretest and training data; in some cases, authors were unable to provide their data for use in this project, and subsequently could not be included. The combined sample from seven studies resulted in a total of $N = 192$ participants with between $n = 20$ to 39 coming from the complex span training group within each study. The participants were generally university students, though some included young adult community members recruited through flyers or announcements made via campus resources. Two studies included recruitment materials that explicitly mentioned training. All participants were between ages 18 to 30 years old and all samples included male and female participants. Each study had 10 to 20 training sessions between the pretest and posttest sessions, with four of the studies having 20 training sessions. Each training session was about 20 to 45 min long with the exception of Harrison et al. (2013) who allowed up to 45 min for each of the two training tasks. The studies are described in Table 1.

Measures

Although the pretest and training materials were similar across the studies in the dataset, differences in the scoring and/or number of tests meant that performance needed to be standardized a couple of ways. For the WM pretest, complex span and running memory span tasks were included (Broadway and Engle 2010). For the Gf pretest, only the matrix reasoning tests Raven's Advanced Progressive Matrices or Bochumer Matrizen-test were included. If the pretest battery included only a single WM or Gf task, a z -score was calculated. When multiple WM tasks were administered at pretest, composites were calculated by averaging the z -scores of the individual tasks.¹

For the training tasks, there was variation in the maximum possible levels of performance, even across some of the studies using the same task and acquired from the same lab. Therefore, the mean score for each participant for each day was divided by the maximum score attained by any participant within that study. This correction scales all scores into a proportion score from 0 to 1 while maintaining the relationships between the participants' scores within a study and the relationships from one training day to the next.

¹ No studies presented here used multiple Gf tasks; therefore, calculation of composites was not necessary.

Table 1 Information about included datasets

Authors	N subjects	N sessions	Pretest WM	Pretest Gf	Spatial training	Verbal training
Blacker et al. (2017)	25	20	Operation span	BOMAT	Symmetry span	N/A
Chein and Morrison (2010)	20	20	Verbal span, Symmetry span	RAPM	Symmetry span	Verbal span
Gunn et al. (2018)	39	15	Reading span, Rotation span, Running letter span, Running matrix span	N/A	Symmetry span	Operation span
Harrison et al. (2013)	21	20	Reading span, Rotation span, Running letter span, Running matrix span	RAPM	Symmetry span	Operation span
Minear et al. (2016)	32	20	Operation span, Symmetry span, Reading span, Rotation span	RAPM	N/A	Verbal span
Redick et al. (under review)	30	10	Running letter span, Running spatial span	N/A	N/A	Operation span
Richey et al. (2014)	25	10	NA	RAPM	Symmetry-comparison span	Verbal-comparison span

WM, working memory; Gf, fluid intelligence; RAPM, Raven's Advanced Progressive Matrices; BOMAT, Bochumer Matrizen-test; N/A, not applicable

The training tasks analyzed here are adaptive spatial and verbal complex span tasks. Each study included criteria for how participants could increase or decrease the list length of processing-and-storage items performed within a session. Three verbal complex span tasks were included. Three of the studies used operation span (Unsworth et al. 2005), described above. Two studies used a task we call verbal span (Chein and Morrison 2010), which also requires remembering letters presented, but the processing task involves making lexical decisions about strings of letters for 4 s. The third task used by the remaining study is similar to the Chein & Morrison task but instead presents two strings of letters and asks whether they are both words or both non-words; accordingly we refer to this task as verbal-comparison span (Richey et al. 2014). The spatial complex span tasks are similar in structure but involve remembering the locations of colored squares that appear in a grid while making symmetry judgments of black and white patterns. All but one used the typical version of symmetry span with a single black and white pattern presented for the symmetry judgment. The exception, which we called spatial-comparison span, involved presenting two black and white patterns simultaneously and asked whether they were both symmetrical or both asymmetrical (Richey et al. 2014).

Analyses

Typically, in training studies, null-hypothesis testing analyses involving training are restricted to gain scores or other singular achievement measures. However, there is a wealth of untapped information in the training day data, and a multi-level modeling approach allows for the inclusion of all of the daily training information as well as characterization of potential longitudinal trajectories. Multi-level modeling also allows for the inclusion of participants who may have missing data due to a technical error in a particular session, for example, as

well as the combination of these datasets despite the difference in number of training sessions.

Analyses focus on the relationship between cognitive abilities, measured at pretest, and training performance. For the analyses with pretest WM as the predictor, four of the six studies had spatial training tasks ($N = 105$), and five of the six studies had verbal training tasks ($N = 142$). For the analyses with pretest Gf as the predictor, four of the five studies had spatial training tasks ($N = 91$), and four of the five studies had verbal training tasks ($N = 98$). These two types of training are analyzed separately for each pretest measure with an otherwise identical model. The model for all four analyses was specified as

$$\begin{aligned} \text{TrainingAttainment}_{ij} = & (b_0 + b_{0j(i)}) + (b_1 + b_{1j})\text{TrainDay} \\ & + (b_2)\text{Pretest} \\ & + (b_3)\text{TrainDayQuadratic} \\ & + (b_4)\text{Pretest}^*\text{TrainDay} \\ & + (b_5)\text{Pretest}^*\text{TrainDayQuadratic} \\ & + \Gamma_{ijk}, \end{aligned}$$

where i equals study, j equals participant, and k equals time. The model is specified with participant nested in study. The slope is allowed to vary by training day, grouped by participant, and there are fixed effects for pretest score and training day and an interaction between pretest score and training day. The main effect for training day and the interaction term are included as both linear and quadratic. The inclusion of study allows the model to adjust for differences between training levels caused by differing difficulty as well as accounting for the different numbers of sessions, without imputing data. This also applies to subjects who had missing data for a given session. For the spatial and verbal training, respectively,

individual missing sessions accounted for 1% and < 1% of total possible sessions. The models were run using the lme4 package in CRAN R software (Bates et al. 2015).² To test the hypothesis that individual differences in pretest performance may affect attainment throughout training, the critical piece of this model is the interaction term between pretest score and training day.

Results

Working Memory Pretest

Figure 2a shows that the spatial training improvement slopes are positive, and steeper for those at +1 SD at pretest than for those at the mean WM score, both of which are steeper than those at -1 SD. The positive slopes are supported by a significant fixed effect of training day, $t(240.30) = 17.92$, $p < .001$. The positive estimate ($E = 0.024$, $SE = 0.001$) for this effect, reported in Table 2, indicates that participants at mean pretest level improved on the training about 2.4% per training day. The significant main effect for the quadratic term, $t(1690.2) = -14.41$, $p < .001$, with a negative estimate ($E = -0.001$, $SE = 0.000$) indicates that the slopes accelerate less over time, forming the quadratic shape. There was a significant fixed effect of pretest score, $t(110.90) = 2.35$, $p = .020$. The positive estimate ($E = 0.033$, $SE = 0.014$) indicates that participants who had higher WM scores at pretest achieved higher overall levels on the training tasks, a roughly 3.3% increase in training per 1 SD increase in pretest score. Critically, the interaction for pretest scores and training day was significant for both linear and quadratic estimates, $t(229.9) = 3.39$, $p < .001$, and $t(1677.40) = -2.32$, $p = .021$, respectively. This linear interaction term ($E = 0.006$, $SE = 0.002$) indicates that those who had higher WM at pretest also improved more (i.e., had a steeper slope) over the course of the training. The quadratic interaction term ($E = -0.0001$, $SE = 0.000$) indicates that the acceleration of gain across training days changes such that those with higher pretest scores show slightly less increase as time goes on. However, the overall pattern still suggests that high pretest scorers consistently gain the most throughout the training period. Predicted values calculated from these estimates for training days at mean level of pretest score and at ± 1 SD of the pretest score are reported in Table 3. Compared to the 21% improvement in training exhibited by those with a mean pretest WM score, those at -1 SD and +1 SD pretest WM score improved 15% and 27%, respectively.

² One of the four models, the WM pretest with verbal training model, was initially inconclusive with the default lmer settings. A control option was added with the Optimx package (Nash and Varadhan 2011) which allowed the model to function properly, with minimal adjustments. Fixed effect estimates were identical to the 7th decimal place. The mathematical model is identical.

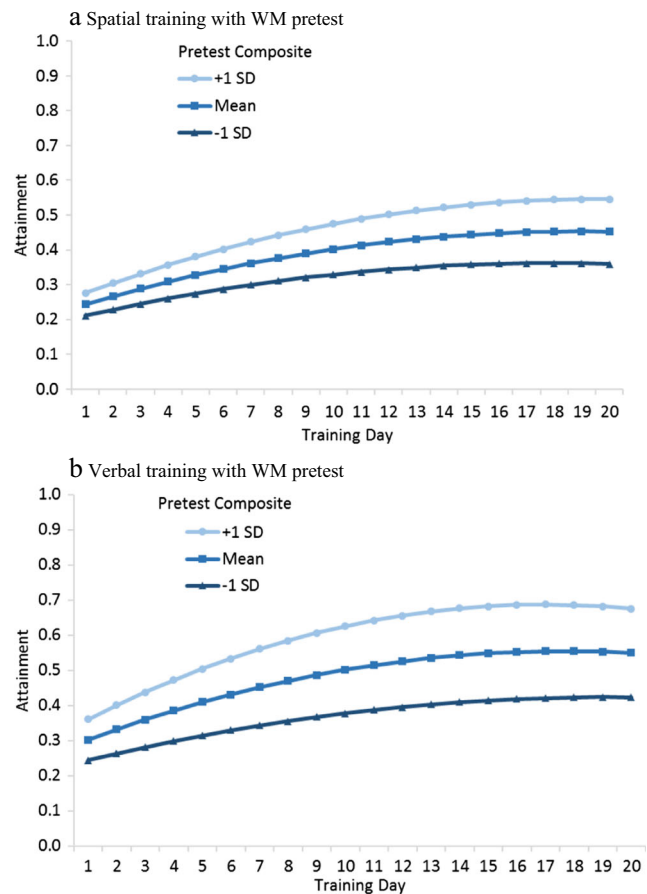


Fig. 2 a Spatial training with WM pretest. b Verbal training with WM pretest

The data for the verbal training appear to have a very similar pattern, shown in Fig. 2b, with steeper slopes for those with higher pretest WM scores. Again, the positive slopes were supported by a significant fixed effect of training day, for both linear, $t(268.80) = 19.60$, $p < .001$, and quadratic components, $t(2126.90) = -15.86$, $p < .001$. The positive estimate ($E = 0.030$, $SE = 0.002$) indicates that participants at mean WM pretest scores improved on training approximately 3.0% per training day, whereas the negative quadratic estimate ($E = -0.001$, $SE = 0.000$) indicates that there is a small decrease in that linear estimate with each successive day, forming the quadratic curve. There was a significant fixed effect of pretest composite, $t(162.20) = 4.96$, $p < .001$. The positive estimate ($E = 0.059$, $SE = 0.012$) indicates that participants who had higher WM scores at pretest had higher overall performance on the training tasks, about 5.9% for each 1 SD. Critically, and again supporting the pattern in Fig. 2b, there was a significant interaction for pretest composite scores and training day with the linear component, $t(267.40) = 5.20$, $p < .001$. This interaction ($E = 0.011$, $SE = 0.002$) indicates that those who scored higher on WM measures at pretest also improved more (i.e., had a steeper slope) over the course of the training sessions. Additionally, there was again a significant

Table 2 Model results

	Spatial Training			Verbal Training		
	Effect		<i>p</i>	Effect		<i>p</i>
Working memory						
Fixed effects		SE			SE	
Intercept	0.2434	0.011	< .001	0.3030	0.009	< .001
Training day	0.0236	0.001	< .001	0.0303	0.002	< .001
Training day quadratic	−0.0007	0.000	< .001	−0.0009	0.000	< .001
Pretest composite	0.0327	0.014	.020	0.0587	0.012	< .001
Interaction	0.0056	0.002	< .001	0.0105	0.002	< .001
Interaction quadratic	−0.0001	0.000	.021	−0.0004	0.000	< .001
Random effects		SD			SD	
Intercept - subject (study)	0.0042	0.065		0.0045	0.069	
Intercept - subject	0.0073	0.086		0.0045	0.069	
Training day	0.0001	0.010		0.0002	0.014	
Residual	0.0027	0.052		0.0048	0.070	
Fluid intelligence						
Fixed effects		SE			SE	
Intercept	0.2653	0.013	< .001	0.3086	0.012	< .001
Training day	0.0281	0.002	< .001	0.0234	0.001	< .001
Training day quadratic	−0.0006	0.000	< .001	−0.0007	0.000	< .001
Pretest score	0.0134	0.013	.298	0.0290	0.012	.017
Interaction	0.0065	0.002	< .001	0.0019	0.001	.193
Interaction quadratic	−0.0002	0.000	< .001	−0.0001	0.000	.020
Random effects		SD			SD	
Intercept - subject (study)	0.0084	0.091		0.0076	0.087	
Intercept - subject	0.0047	0.069		0.0041	0.064	
Training day	0.0002	0.014		0.0001	0.010	
Residual	0.0031	0.056		0.0042	0.065	

interaction using the quadratic component, $t(2139.30) = -4.92$, $p < .001$. This interaction ($E = -0.0003$, $SE = 0.000$) indicates that for increasing pretest values there is a slightly lower increase in slope with each day of training. Predicted values for training days at mean level of pretest composite and at ± 1 SD of the pretest composite are reported in Table 3. Similar to the spatial training, those at -1 SD at pretest improved less (18%) than those at mean at pretest (25%) and those at $+1$ SD at pretest (31%). The results with WM as a predictor of WM training changes are most consistent with the predictions depicted in Fig. 1a.

General Fluid Intelligence Pretest

Figure 3a shows a pattern for spatial training that looks like the patterns seen with the WM pretest analyses, with positive slopes that are steeper in those at higher pretest Gf levels. The positive slopes were supported by a significant linear fixed effect of training day, $t(141.20) = 16.06$, $p < .001$. The positive estimate ($E = 0.028$, $SE = 0.002$) indicates that participants with mean pretest Gf scores improved on training about

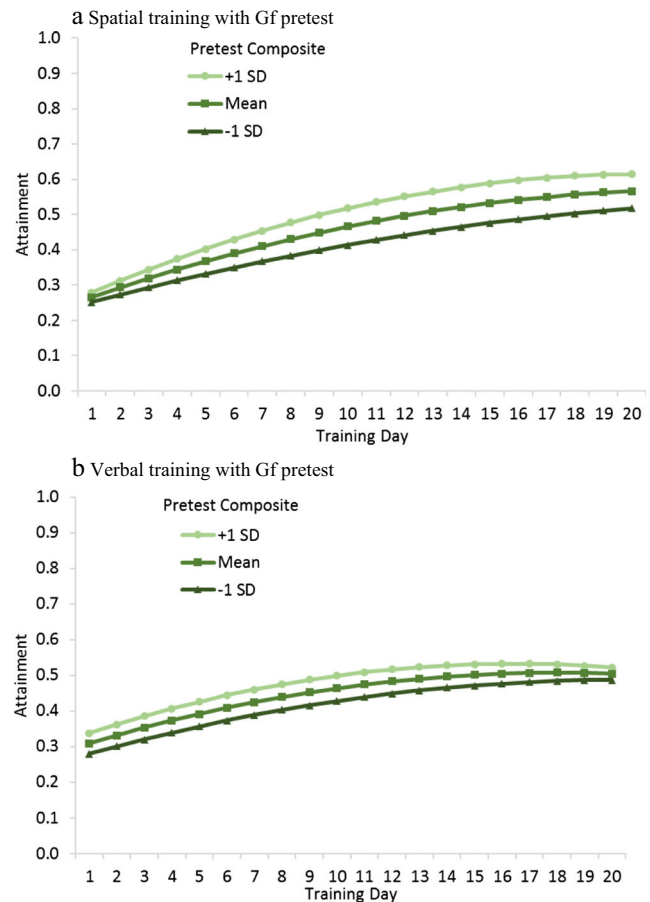
2.8% per training day. The main effect of training day with the quadratic component was also significant, $t(1417.90) = -12.32$, $p < .001$, with the negative estimate ($E = -0.001$, $SE = 0.000$) indicating the slight decrease in the estimate over training days resulting in the quadratic curve. Unlike the WM pretest composite, there was not a significant fixed effect of Gf pretest score, $t(96.90) = 1.05$, $p = .298$. Despite this lack of pretest main effect, there was a significant interaction for pretest scores and training day with both linear and quadratic components, $t(141.20) = 3.63$, $p < .001$ and $t(1417.90) = -4.59$, $p < .001$, respectively. These interactions (linear: $E = 0.006$, $SE = 0.002$; quadratic: $E = -0.0002$, $SE = 0.000$) indicate, as with the WM analyses, that slopes are steeper and accelerate slightly less with time for those with higher pretest scores. Predicted values calculated from these estimates for training days at mean level of Gf pretest score and at ± 1 SD of the pretest score are reported in Table 4. Those at -1 SD improved around 27%, which is less than those at mean (30%) and those at $+1$ SD (33%).

The verbal training slopes (Fig. 3b) for all levels of Gf pretest score look slightly less distinct from one another than

Table 3 Predicted values for representative ranges of WM pretest

Training day	Spatial training			Verbal training		
	Pretest composite range			Pretest composite range		
	−1 SD	Mean	+1 SD	−1 SD	Mean	+1 SD
1	0.21	0.24	0.28	0.24	0.30	0.36
2	0.23	0.27	0.30	0.26	0.33	0.40
3	0.24	0.29	0.33	0.28	0.36	0.44
4	0.26	0.31	0.36	0.30	0.39	0.47
5	0.27	0.33	0.38	0.31	0.41	0.50
6	0.29	0.34	0.40	0.33	0.43	0.53
7	0.30	0.36	0.42	0.34	0.45	0.56
8	0.31	0.38	0.44	0.36	0.47	0.58
9	0.32	0.39	0.46	0.37	0.49	0.61
10	0.33	0.40	0.47	0.38	0.50	0.63
11	0.34	0.41	0.49	0.39	0.51	0.64
12	0.34	0.42	0.50	0.40	0.53	0.66
13	0.35	0.43	0.51	0.40	0.53	0.67
14	0.35	0.44	0.52	0.41	0.54	0.68
15	0.36	0.44	0.53	0.41	0.55	0.68
16	0.36	0.45	0.54	0.42	0.55	0.69
17	0.36	0.45	0.54	0.42	0.55	0.69
18	0.36	0.45	0.54	0.42	0.55	0.69
19	0.36	0.45	0.55	0.42	0.55	0.68
20	0.36	0.45	0.55	0.42	0.55	0.67
Total change	0.15	0.21	0.27	0.18	0.25	0.31

they have in the previous three models, all improving to nearly the same place, though not necessarily at equivalent rates. Confirming the positive direction of the slopes, there was a significant fixed effect of training day, $t(229.00) = 16.05$, $p < .001$. The positive estimate ($E = 0.023$, $SE = 0.001$) indicates that participants' training levels improved across training. Similar to all other models, there was also a significant main effect of training day as a quadratic component, $t(1566.40) = -12.24$, $p < .001$. The negative estimate ($E = -0.0007$, $SE = 0.000$) indicates again that there is a slight decrease in the effect with time, creating the quadratic curve. The highest Gf scorers at pretest achieved the highest training performance, which was confirmed with a significant fixed effect of Gf pretest, $t(107.20) = 2.42$, $p = .017$. The positive estimate ($E = 0.029$, $SE = 0.012$) indicates that participants who had higher scores at pretest had higher overall achievement on training tasks, with about a 2.9% increase per 1 SD. Unlike the previous three models, the interaction with the linear component was not significant, $t(229.30) = 1.30$, $p = .193$. However, the interaction between the Gf pretest score and the quadratic training day component was significant, $t(1566.30) = -2.34$, $p = .020$. The negative estimate ($E = -0.0001$, $SE = 0.000$) indicates, as it does in all 3 other

**Fig. 3** a Spatial training with Gf pretest. b Verbal training with Gf pretest

models, that there is a slightly lower slope increase for the higher pretest scores as training days progress. While the higher and lower pretest performers start and end at nearly the same levels, the trajectories differ. Predicted values for training days at mean level of Gf pretest score and at ± 1 SD of the pretest score are reported in Table 4. All participants improved very similarly, with those at -1 SD improving 21%, those at mean improving 19%, and those at $+1$ SD improving 18%.

Discussion

In the present work, several datasets were combined to examine how individual differences in cognitive ability influence WM training gains. We found that individuals high in WM and Gf before undergoing WM training showed the greatest training gains. None of the four analyses provide any evidence that individuals lower in cognitive ability would improve more on training tasks compared to those higher in cognitive ability, similar to Fig. 1b. Three of the four models show evidence for the opposite pattern, with individuals higher in cognitive ability gaining more, similar to Fig. 1a. The fourth

Table 4 Predicted values for representative ranges of Gf pretest

Training day	Spatial training			Verbal training		
	Pretest score range			Pretest score range		
	−1 SD	Mean	+1 SD	−1 SD	Mean	+1 SD
1	0.25	0.27	0.28	0.28	0.31	0.34
2	0.27	0.29	0.31	0.30	0.33	0.36
3	0.29	0.32	0.34	0.32	0.35	0.39
4	0.31	0.34	0.37	0.34	0.37	0.41
5	0.33	0.37	0.40	0.36	0.39	0.43
6	0.35	0.39	0.43	0.37	0.41	0.44
7	0.37	0.41	0.45	0.39	0.42	0.46
8	0.38	0.43	0.48	0.40	0.44	0.47
9	0.40	0.45	0.50	0.42	0.45	0.49
10	0.41	0.47	0.52	0.43	0.46	0.50
11	0.43	0.48	0.54	0.44	0.47	0.51
12	0.44	0.50	0.55	0.45	0.48	0.52
13	0.45	0.51	0.57	0.46	0.49	0.52
14	0.47	0.52	0.58	0.46	0.50	0.53
15	0.48	0.53	0.59	0.47	0.50	0.53
16	0.49	0.54	0.60	0.48	0.50	0.53
17	0.50	0.55	0.60	0.48	0.51	0.53
18	0.50	0.56	0.61	0.48	0.51	0.53
19	0.51	0.56	0.61	0.49	0.51	0.53
20	0.52	0.57	0.61	0.49	0.50	0.52
Total change	0.27	0.30	0.33	0.21	0.19	0.18

model showed no linear interaction between cognitive ability and training trajectory, though it did suggest the higher performers maintain higher scores throughout training and gain more quickly earlier in training. Further, the relationship between WM and Gf pretest scores and WM training appears to be domain-general, in that we find essentially the same results for both verbal and visuospatial training paradigms.

The current results show a “rich-get-richer” effect as it relates to individual differences in WM and Gf before training (Redick et al. 2015; von Bastian and Oberauer 2014). Notably, these results run counter to the aim of using WM training as a compensatory strategy where those with lower initial WM could expand their capacity. While not directly examined in the present study, these results may also inform which individuals are more likely to show far transfer from WM training. It has been suggested that those who improve most on the training tasks are also the ones most likely to exhibit far transfer (e.g., Jaeggi et al. 2011). If this training-transfer gain hypothesis put forth by Jaeggi et al. (2011) is correct, according to our results, individuals who are already higher in WM or Gf would be expected to show more far transfer than those with lower baseline WM or Gf. This is important because many of the commercial applications of

WM training are marketed toward helping those who are low in WM (Redick et al. 2015).

Expectations about *who* will benefit from training and *how* warrant further attention, though, when considering the rich history of the skill-acquisition literature. Schmidt and Bjork (1992) reviewed the literature on “desirable difficulty” in training and transfer, showing that variables such as schedules of practice (e.g., variable) and frequency of feedback (e.g., intermittent) can depress training gains but lead to larger transfer improvements at posttest. Therefore, one could make a prediction that would contradict the hypothesis that larger WM training improvements would be associated with greater far transfer. That is, low-cognitive ability at pretest, predicting lower training gains throughout training, may yet lead to larger training gains. Most currently used WM training paradigms are adaptive, so everyone should be at a level that is difficult for them, independent of their cognitive ability level. So, it is possible that the role difficulty would play in the WM training literature could be quite different from skill acquisition research.

As Table 1 notes, not all studies have both pretest measures and/or both training measures. Therefore, all four models are based on different subsets of participants from the seven studies. However, all four samples are still each much larger than the single-study analyses typical of the literature (for exception, see Guye et al. 2017). We had anticipated incorporating study into the model, separated from participant to ensure that the influence of the slight variations in procedures across studies was minimal. However, the models would not converge with this variable. Our interpretation is that the variance in study is too low to include in the model, and all important variance is sufficiently captured by participant or participant nested in study, both of which are included.

One further consideration involves the pretest measures investigated. Foster et al. (2017) used WM pretest performance, whereas Jaeggi et al. (2011) used Gf pretest performance. The present work evaluated both pretest measures separately, and the WM pretest results agree with the extreme-groups results from Foster et al. However, the Gf pretest results from the present work were not in line with the results from Jaeggi et al. The Gf pretest in the current study was predictive of performance on the WM training tasks but in the opposite direction than shown by Jaeggi et al., and there was no evidence of a negative relationship with training.

There are also a few reasons Jaeggi et al. might have found a different pattern, as noted in the introduction. Particularly, the statistical correlations between training gain scores and transfer pretest-posttest difference scores has been discussed elsewhere as inappropriate and uninformative (Smoleń et al. 2018; Tidwell et al. 2014), which could indicate that the improvements shown in studies using those methods are not reflective of true improvements. Ignoring the statistical issues, if the same results had been found with appropriate methods, a

direct comparison is still imperfect given Jaeggi et al. used *n*-back training, rather than complex span training. As such, there is a possibility that training on *n*-back and complex span, though both thought to measure WM, could have different relationships with Gf. We also note that our samples included only healthy, young adults, while Jaeggi et al. (2011) tested children. Future research is required to determine if the relationship between pre-existing cognitive ability and training improvements is moderated by the age composition of the sample.

The slightly weaker results for Gf pretest compared to WM pretest may seem concerning given the positive relationship between WM and Gf. However, WM and Gf are not isomorphic (Heitz et al. 2006) and it is, therefore, not entirely surprising that Gf would not predict WM training improvements as well as complex span tasks do. Because the specific aspects of WM that are being trained in these studies are not clear, it is then possible that the specific aspects of WM that may be trainable, or that are targeted by complex span training, are not those that overlap with the processes driving matrix reasoning performance—the measure of Gf in the present work.

Our results are incompatible with a compensatory account stating that the low performers should improve more throughout training. Rather, the opposite pattern is found, with the higher performers gaining more. This finding is consistent for both WM and Gf pretest measures and for both verbal and visuospatial training. This cumulative (as opposed to compensatory) finding should not be entirely surprising. Matthew effects, where lower performers show slower gains or, in some extreme cases, worse performance over time, have been documented in educational settings for a long time. Shaywitz et al. (1995) found a significant Matthew effect in IQ for children tested at kindergarten through 6th grade. There are many situations in which achievement begets higher achievement. However, our results, and other examples of this pattern do not necessarily have direct implications for transfer if the assumptions of Jaeggi et al. are not supported.

In the present work, we show that WM and Gf have relationships to training gains regardless of domain (spatial vs. verbal), but this does not imply that transfer would occur. Specifically, we note that although Foster et al. (2017) found that high-WM participants showed larger training gains than low-WM participants, the participants' pre-existing WM ability had no bearing on the amount and presence of transfer. Regardless of how much pre-existing ability aided training, transfer was not similarly affected. In a study of third-grade children, Rode et al. (2014) came to a similar conclusion—students initially higher in WM and academic achievement produced larger WM training gains, yet this did not translate into significant transfer on most outcomes compared to a passive-control group. These

findings are not particularly surprising, given meta-analyses that show transfer is minimal at best, and generally restricted to tasks very similar the training tasks (e.g., Melby-Lervåg et al. 2016; Sala and Gobet 2017). That is, WM training generally leads to improved task performance on the training task(s) and sometimes very similar tasks, but has not modified the domain-general ability underlying task performance. Participants' effectiveness in improving training task performance is less relevant if the underlying ability is not really being improved, or if the strategies developed to facilitate improvement during training are not applicable to the transfer tests used (Dunning and Holmes 2014; Morrison et al. 2016).

Intelligence researchers may be less surprised by our results, which can be restated as showing that individuals higher in cognitive ability (WM and Gf) are more able to learn, consistent with numerous other examples showing such a pattern outside of the cognitive training literature (e.g., Daneman and Green 1986; Kyllonen and Stephens 1990; Shute 1991). Ideally, we would have comparable data from multiple sessions of an active-control task in order to investigate whether the same cognitive predictors show the same relationship to training gains in a non-WM training task. However, several of the datasets we included here do not have comparable control groups that permit such an analysis. Three of the studies use adaptive visual search tasks (Gunn et al. 2018; Harrison et al. 2013; Redick et al. *under review*), but the other four studies used different active-control tasks or included only passive-control groups. This remains a question for future research, although we note that Foster et al. (2017) did not find a differential amount of improvement over 20 days on adaptive visual-search control tasks for the high- and low-WM participants.

To conclude, the current results show that individuals already higher in WM or Gf are the ones that show greater WM training gains. This result is particularly concerning for those who wish to train lower performing students up to the level of their higher performing peers, under the assumption that individuals who show greater WM training gains should subsequently exhibit more transfer.

Acknowledgements The research described here was presented at the 2018 annual meeting of the Midwestern Psychological Association and the International Society for Intelligence Research. The data and script used for the models are available from the authors at <https://osf.io/peswn/>. The authors thank Dr. Kara Blacker, Dr. Jason Chein, Dr. Tyler Harrison, Dr. Meredith Minear, Dr. Elizabeth Richey, and their coauthors, for providing additional data and clarifications from their published studies included in these analyses. The authors also thank Dr. Sean Lane for his assistance with the multi-level modeling analyses.

Funding Information While writing this manuscript, EAW and TSR were supported by the National Science Foundation (Award # 1632403) and the National Institutes of Health (Award # 2R01AA013650-11A1).

Compliance with Ethical Standards

Conflict of Interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

*Indicates Data Source

- Au, J., Sheehan, E., Tsai, N., Duncan, G. J., Buschkuhl, M., & Jaeggi, S. M. (2015). Improving fluid intelligence with training on working memory: a meta-analysis. *Psychonomic Bulletin & Review*, 22, 366–377. <https://doi.org/10.3758/s13423-014-0699-x>.
- *Blacker, K. J., Negoita, S., Ewen, J. B., & Courtney, S. M. (2017). N-back versus complex span working memory training. *Journal of Cognitive Enhancement*, 1(4), 434–454. <https://doi.org/10.1007/s41465-017-004-1>.
- Baddeley, A. (2012). Working memory: theories, models, and controversies. *Annual Review of Psychology*, 63, 1–29. <https://doi.org/10.1146/annurev-psych-120710-100422>.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>.
- Broadway, J. M., & Engle, R. W. (2010). Validating running memory span: Measurement of working memory capacity and links with fluid intelligence. *Behavior Research Methods*, 42(2), 563–570. <https://doi.org/10.3758/BRM.42.2.563>.
- Bürki, C. N., Ludwig, C., Chicherio, C., & de Ribaupierre, A. (2014). Individual differences in cognitive plasticity: an investigation of training curves in younger and older adults. *Psychological Research*, 78, 821–835. <https://doi.org/10.1007/s00426-014-0559-3>.
- *Chein, J., & Morrison, A. M. (2010). Expanding the mind's workspace: training and transfer effects with a complex working memory span task. *Psychonomic Bulletin & Review*, 17(2), 193–199. <https://doi.org/10.3758/PBR.17.2.193>.
- Cowan, N. (2017). The many faces of working memory and short-term storage. *Psychonomic Bulletin & Review*, 24, 1158–1170. <https://doi.org/10.3758/s13423-016-1191-6>.
- Daneman, M., & Green, I. (1986). Individual differences in comprehending and producing words in context. *Journal of Memory and Language*, 25(1), 1–18. [https://doi.org/10.1016/0749-596X\(86\)90018-5](https://doi.org/10.1016/0749-596X(86)90018-5).
- Dunning, D. L., & Holmes, J. (2014). Does working memory training promote the use of strategies on untrained working memory tasks? *Memory & Cognition*, 42(6), 854–862. <https://doi.org/10.3758/s13421-014-0410-5>.
- Foster, J. L., Harrison, T. L., Hicks, K. L., Draheim, C., Redick, T. S., & Engle, R. W. (2017). Do the effects of working memory training depend on baseline ability level? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(11), 1677–1689. <https://doi.org/10.1037/xlm0000426>.
- Fuchs, L. S., Schumacher, R. F., Sterba, S. K., Long, J., Namkung, J., Malone, A., et al. (2014). Does working memory moderate the effects of fraction intervention? An aptitude-treatment interaction. *Journal of Educational Psychology*, 106(2), 499–514. <https://doi.org/10.1037/a0034341>.
- *Gunn, R. L., Gerst, K. R., Wiemers, E. A., Redick, T. S., & Finn, P. R. (2018). Predictors of effective working memory training in those with alcohol use disorders. *Alcoholism: Clinical and Experimental Research*. <https://doi.org/10.1111/acer.13892>.
- Guye, S., De Simoni, C., & von Bastian, C. C. (2017). Do individual differences predict change in cognitive training performance? A latent growth curve modeling approach. *Journal of Cognitive Enhancement*, 1, 374–393. <https://doi.org/10.1007/s41465-017-0049-9>.
- *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. <https://doi.org/10.1177/0956797613492984>.
- Heitz, R. P., Redick, T. S., Hambrick, D. Z., Kane, M. J., Conway, A. R. A., & Engle, R. W. (2006). Working memory, executive function, and general fluid intelligence are not the same. *Behavioral and Brain Sciences*, 29, 134–135. <https://doi.org/10.1017/S0140525X06319036>.
- Houben, K., Wiers, R. W., & Jansen, A. (2011). Getting a grip on drinking behavior: training working memory to reduce alcohol abuse. *Psychological Science*, 22(7), 968–975. <https://doi.org/10.1177/0956797611412392>.
- 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. <https://doi.org/10.1073/pnas.0801268105>.
- Jaeggi, S. M., Buschkuhl, M., Jonides, J., & Shah, P. (2011). Short- and long-term benefits of cognitive training. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 10081–10086. <https://doi.org/10.1073/pnas.1103228108>.
- Karbach, J., Könen, T., & Spengler, M. (2017). Who benefits the most? Individual differences in the transfer of executive control training across the lifespan. *Journal of Cognitive Enhancement*, 1, 394–405. <https://doi.org/10.1007/s1465-017-0054-z>.
- Katz, B., Jones, M. R., Shah, P., Buschkuhl, M., & Jaeggi, S. M. (2016). Individual difference and motivational effects. In T. Strobach & J. Karbach (Eds.), *Cognitive training: an overview of features and applications* (pp. 157–166). Switzerland: Springer.
- Klingberg, T., Fernell, R., Olesen, P. J., Johnson, M., Gustafsson, P., Dahlström, K., et al. (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(2), 177–186. <https://doi.org/10.1097/00004583-200502000-00010>.
- Kyllonen, P. C., & Stephens, D. L. (1990). Cognitive abilities as determinants of success in acquiring logic skill. *Learning and Individual Differences*, 2(2), 129–160. [https://doi.org/10.1016/1041-6080\(90\)90020-H](https://doi.org/10.1016/1041-6080(90)90020-H).
- Lövdén, M., Brehmer, Y., Li, S.-C., & Lindenberger, U. (2012). Training-induced compensation versus magnification of individual differences in memory performance. *Frontiers in Human Neuroscience*, 6. <https://doi.org/10.3389/fnhum.2012.00141>.
- *Minear, M., Brasher, F., Guerrero, C. B., Brasher, M., Moore, A., & Sukeena, J. (2016). A simultaneous examination of two forms of working memory training: evidence for near transfer only. *Memory & Cognition*, 44, 1014–1037. <https://doi.org/10.3758/s13421-016-06016-9>.
- Melby-Lervåg, M., Redick, T. S., & Hulme, C. (2016). Working memory training does not improve performance on measures of intelligence or other measures of “far transfer”: evidence from a meta-analytic review. *Perspectives on Psychological Science*, 11(4), 512–534. <https://doi.org/10.1177/1745691616635612>.
- 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(1), 46–60. <https://doi.org/10.3758/s13423-010-0034-0>.

- Morrison, A. B., Rosenbaum, G. M., Fair, D., & Chein, J. M. (2016). Variation in strategy use across measures of verbal working memory. *Memory & Cognition*, 44(6), 922–936. <https://doi.org/10.3758/s13421-016-0608-9>.
- Nash, J. C., & Varadhan, R. (2011). Unifying optimization algorithms to aid software system users: Optimx for R. *Journal of Statistical Software*, 43(9), 1–14. <https://doi.org/10.18637/jss.v043.i09>.
- Redick, T. S., & Lindsey, D. R. B. (2013). Complex span and n-back measures of working memory: a meta-analysis. *Psychonomic Bulletin & Review*, 20(6), 1102–1113. <https://doi.org/10.3758/s13423-013-0453-9>.
- Redick, T. S., Shipstead, Z., Harrison, T. L., Hicks, K. L., Fried, D. E., Hambrick, D. Z., et al. (2013). No evidence of intelligence improvement after working memory training: a randomized, placebo-controlled study. *Journal of Experimental Psychology: General*, 142(2), 359–379. <https://doi.org/10.1037/a0029082>.
- Redick, T. S., Shipstead, Z., Wiemers, E. A., Melby-Lervåg, M., & Hulme, C. (2015). What's working in working memory training? An educational perspective. *Educational Psychology Review*, 27, 617–633. <https://doi.org/10.1007/s10648-015-9314-6>.
- Redick, T. S., Shipstead, Z., Meier, M. E., Montroy, J. J., Hicks, K. L., Unsworth, N., et al. (2016). Cognitive predictors of a common multitasking ability: contributions from working memory, attention control, and fluid intelligence. *Journal of Experimental Psychology: General*, 145(11), 1473–1492. <https://doi.org/10.1037/xge0000219>.
- *Redick, T. S., Wiemers, E. A., & Engle, R. W. (under review). The role of proactive interference, variability, and strategy use in working memory training and transfer.
- *Richey, J. E., Phillips, J. S., Schunn, C. D., & Schneider, W. S. (2014). Is the link from working memory to analogy causal? No analogy improvements following working memory training gains. *PLoS One*, 9(9). <https://doi.org/10.1371/journal.pone.0106616>.
- Rode, C., Robson, R., Purviance, A., Geary, D. C., & Mayr, U. (2014). Is working memory training effective? A study in a school setting. *PLoS One*, 9(8), e104796. <https://doi.org/10.1371/journal.pone.0104796>.
- Sala, G., & Gobet, F. (2017). Working memory training in typically developing children: a meta-analysis of the available evidence. *Developmental Psychology*, 53(4), 671–685. <https://doi.org/10.1037/dev0000265>.
- Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: common principles in three paradigms suggest new concept for training. *Psychological Science*, 3(4), 207–217. <https://doi.org/10.1111/j.1467-9280.1992.tb00029.x>.
- Shaywitz, B. A., Holford, T. R., Holahan, J. M., Fletcher, J. M., Stuebing, K. K., Francis, D. J., & Shaywitz, S. E. (1995). A Matthew effect for IQ but not for reading: results from a longitudinal study. *Reading Research Quarterly*, 30(4), 894–906. <https://doi.org/10.2307/748203>.
- Shipstead, Z., Redick, T. S., & Engle, R. W. (2012). Is working memory training effective? *Psychological Bulletin*, 138(4), 628–654. <https://doi.org/10.1037/a0027473>.
- Shipstead, Z., Harrison, T., & Engle, R. (2016). Working memory capacity and fluid intelligence. *Perspectives on Psychological Science*, 11(6), 771–799. <https://doi.org/10.1037/a002747>.
- Shute, V. J. (1991). Who is likely to acquire programming skills? *Journal of Educational Computing Research*, 7(1), 1–24. <https://doi.org/10.2190/VQJD-TIYD-5WVB-RYPJ>.
- Smoleń, T., Jastrzębski, J., Estrada, E., & Chuderski, A. (2018). Most evidence for the compensation account of cognitive training is unreliable. *Memory & Cognition*. Advance online publication. <https://doi.org/10.3758/s13421-018-0839-z>.
- Soveri, A., Antfolk, J., Karlsson, L., Salo, B., & Laine, M. (2017). Working memory training revisited: a multi-level meta-analysis of n-back training studies. *Psychonomic Bulletin & Review*, 24, 1077–1096. <https://doi.org/10.3758/s13423-016-1217-0>.
- 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 older adults. *Developmental Psychology*, 50(4), 1049–1059. <https://doi.org/10.1037/a0034913>.
- Stephenson, C. L., & Halpem, D. F. (2013). Improved matrix reasoning is limited to training on tasks with a visuospatial component. *Intelligence*, 41(5), 341–357. <https://doi.org/10.1016/j.intell.2013.05.006>.
- Studer-Luethi, B., Jaeggi, S. M., Buschkuhl, M., & Perrig, W. J. (2012). Influence of neuroticism and conscientiousness on working memory training outcome. *Personality and Individual Differences*, 53(1), 44–49. <https://doi.org/10.1016/j.paid.2012.02.012>.
- Studer-Luethi, B., Bauer, C., & Perrig, W. J. (2016). Working memory training in children: effectiveness depends on temperament. *Memory & Cognition*, 44(2), 171–186. <https://doi.org/10.3758/s13421-015-0548-9>.
- Tidwell, J. W., Dougherty, M. R., Chrabaszez, J. R., Thomas, R. P., & Mendoza, J. L. (2014). What counts as evidence for working memory training? Problems with correlated gains and dichotomization. *Psychonomic Bulletin & Review*, 21, 620–628. <https://doi.org/10.3758/s13423-013-0560-7>.
- Titz, C., & Karbach, J. (2014). Working memory and executive functions: Effects of training on academic achievement. *Psychological Research*, 78, 852–868. <https://doi.org/10.1007/s00426-013-0537-1>.
- Unsworth, N., & Redick, T.S. (2017). Working memory and intelligence. In J. Wixted (Ed.), *Cognitive psychology of memory*. Vol. 2 of learning and memory: a comprehensive reference, 4 vols., 2nd edition (J. Byrne, editor). Oxford: Elsevier.
- 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. <https://doi.org/10.3758/BF03192720>.
- von Bastian, C. C., & Oberauer, K. (2014). Effects and mechanisms of working memory training: a review. *Psychological Research*, 78, 803–820. <https://doi.org/10.1007/s00426-013-0524-6>.