Noack, Hannes; Lövdén, Martin; Schmiedek, Florian

On the validity and generality of transfer effects in cognitive training research

formal und inhaltlich überarbeitete Version der Originalveröffentlichung in:
formally and content revised edition of the original source in:
Psychological Research 78 (2014) 6, S. 773-789

On the Validity and Generality of Transfer Effects in Cognitive Training Research

Hannes Noack[1,2], Martin Lövdén[1,3], & Florian Schmiedek[1,4]

[1]Center for Lifespan Psychology, Max Planck Institute for Human Development, Germany

[2]Institut of Medical Psychology and Behavioral Neurobiology, Tübingen University, Germany

[3]Aging Research Center, Karolinska Institutet & Stockholm University, Sweden

[4]Center for Research on Education and Human Development (DIPF), German Institute for International Educational Research, Frankfurt am Main, Germany

Corresponding author: Hannes Noack, Institute for Medical Psychology and Behavioral Neurobiology, Tübingen University, Silcherstr. 5, 72076 Tübingen, Germany; e-mail: hannes.noack@uni-tuebingen.de

Abstract

Evaluation of training effectiveness is a long-standing problem of cognitive intervention research. The interpretation of transfer effects needs to meet two criteria, generality and specificity. We introduce each of the two, and suggest ways of implementing them. First, the scope of the construct of interest (e.g. working memory) defines the expected generality of transfer effects. Given that constructs of interest are typically defined at the latent level, data analysis should also be conducted at the latent level. Second, transfer should be restricted to measures that are theoretically related to the trained construct. Hence, the construct of interest also determines the specificity of expected training effects; to test for specificity, study designs should aim at convergent and discriminant validity. We evaluate the recent cognitive training literature in relation to both criteria. We conclude that most studies do not use latent factors for transfer assessment, and do not test for convergent and discriminant validity.

Introduction

Recently, positive transfer to performance on untrained tasks has been observed in response to working memory (e.g., Borella, Carretti, Riboldi, & De Beni, 2010; Jaeggi, Buschkuehl, Jonides, Perrig, 2008; Klingberg et al., 2005 Kuwajima & Sawaguchi, 2010) and executive control training (e.g., Forte et al., 2013; Karbach & Kray, 2009). Together with findings showing that cognitive engagement can result in changes at the neuronal level (Brehmer et al., 2011; Kühn et al., 2012; Lövdén et al., 2012; McNab et al., 2009; Olesen et al., 2004; Wenger et al., 2012), these observations have fostered optimism about the effectiveness of cognitive training procedures. This optimism is contrasted, however, by the cautious conclusions drawn by a recent meta-analysis (Melby-Lervåg & Hulme, 2013) and several systematic reviews (e.g. Lövdén, Bäckman, Lindenberger, Schäfer, & Schmiedek, 2010; Morrison & Chein, 2011; Noack, Lövdén, Schmiedek, & Lindenberger, 2009; Shipstead, Redick, & Engle, 2010; 2012). For example, Melby-Lervåg and Hulme (2013) argue that "…there was no convincing evidence of the generalization of working memory training to other skills." (p.270).

Here we propose that the validity of observed training gains is threatened from two sides, task-specific training gains and non-cognitive factors. We describe these two aspects in more detail, and suggest measures to minimize their effects. We then evaluate the recent training literature, asking if these measures have been commonly applied.

Generality of Transfer

Are training effects general? That is, are the training-related improvements meaningful in the sense that more than task-specific skills and strategies have been improved? This question is related to the old issue of separating task-specific from task-general effects of training (e.g. Baltes & Lindenberger, 1988; Lövdén et al., 2010). One way to answer this question is to try to determine and evaluate the distance of transfer tasks (Barnett & Ceci, 2002; Noack et al., 2009; Zelinski, 2009; cf. Baltes, Dittmann-Kohli, & Kliegl, 1986). In previous reviews (Lövdén et al., 2010; Noack et al., 2009), we suggested that the distance between trained and transfer tasks could be evaluated based on their relationship within a hierarchical structure of human cognitive ability (e.g., Carroll, 1993). One such model, introduced by Carroll, spans over three strata, starting at the less abstract level of narrow abilities at the bottom, going over the more abstract broad abilities in the middle, and arriving finally at a general cognitive ability at the top. Rather than using vaguely defined attributes like *near* and *far*, the taxonomy of transfer distance proposes to qualify transfer distance according to the highest stratum that must be passed to connect trained and transfer task (Noack et al., 2009). Transfer between simple reaction time tasks and choice reaction time tasks would, for example, represent transfer at the level of broad abilities because the two narrow abilities of simple reaction time and choice reaction time can only be connected by the broad ability of processing speed.

However, if evidence is restricted to one or two rather homogeneous tasks, more parsimonious explanations, such as the assumption of shared common elements among the tasks (Thorndike, 1906; Thorndike & Woodworth, 1901) that allow for the transfer of task-specific rather than task-general improvements, should be considered (e.g., Moody, 2009). This holds especially because tasks are always imperfect indicators of some cognitive ability and, inversely, cognitive abilities never explain the total variance of indicator tasks. Jensen (1998) noted, for example, that that the intelligence factor,*g,* predicts only around 65% of the variance in its best predictor, matrix reasoning, leaving another 35% that must be governed by other factors, which could alternatively account for observed transfer (see also Shipstead et al., 2012). Thus, if the aim is to make claims about some latent aspect (i.e., a cognitive process or ability) of the transfer task, and if the ultimate "…goal is not to train for the specific task but to impact the construct underlying the task" (McArdle & Prindle, 2008, p. 703; see also Baltes et al., 1986; Baltes & Lindenberger, 1988), then analyses must be conducted at the latent level, which first requires the definition of a *content domain* (Cattell, 1952) or a target construct that should be affected by training. Then, multiple heterogeneous transfer tasks (i.e., implementing different paradigms and tapping into different content dimensions) must be sampled from the theoretically determined task space (Little, Lindenberger, & Nesselroade, 1999; Lövdén et al., 2010). Finally, the appropriate analytic methods, such as structural equation modeling that forms a latent factor representing the common variance among its indicators, must be used (e.g., McArdle & Nesselroade, 1994; McArdle, 2009).

It is essential to note at this point that these kinds of analyses do not have to be restricted to psychometric hypotheses. The concept of latent constructs is independent of the theoretical concept that is being modeled. Latent constructs simply represent the shared variance of various measures (e.g., Kline, 1998) – irrespective of the underlying source being a cognitive ability or a cognitive resource/process/mechanism, defined on the basis of whatever theoretical (e.g., neural or cognitive) model. Besides potentially increasing the validity of transfer measures, latent measures of change are also free of measurement error and may, therefore, overcome the problems associated with low reliability in the study of change (e.g., Cronbach & Furby, 1970; McArdle, 2009). Having these desirable characteristics in mind and acknowledging the call for analyses of transfer at the latent level (e.g., McArdle & Prindle, 2008; Morrison & Chein, 2011; Noack et al., 2009; Schmiedek et al., 2010; Shipstead et al., 2012), we investigated the prevalence of these methods among recent cognitive training studies that had some emphasis on transfer.

Specificity of Transfer

Are transfer effects specific? From a practical point of view this second question may be of minor relevance because the main criterion for training effectiveness is generality

(Barnett & Ceci, 2002). From a theoretical point of view, however, generality per se may not be as important as the precise pattern of transfer effects (Lövdén et al., 2010; see also Baltes et al., 1986). If this pattern is not theoretically plausible, that is, if the observed pattern contradicts the expected commonality between various transfer tasks, broad effects of transfer may even threaten rather than support the theoretical concept underlying the training intervention (Campell & Fiske, 1959). Changes at the level of non-cognitive, state-like, factors such as test motivation (e.g., Revelle, 1993) or test anxiety (Ashcraft & Kirk, 2001; Eysenck, Derakshan, Santos, & Calvo, 2007; Hopko, Crittendon, Grant, & Wilson, 2005), may offer more parsimonious explanations than changes at the level of general cognitive ability. The potentially broad effect of non-cognitive factors (e.g., Shipstead et al., 2010) is of major concern for latent analyses of training-related gains because latent constructs represent the common variance among multiple variables. To the extent that these variables are all sensitive to manipulations of non-cognitive factors, a proportion of the common variance in change will be attributable to these manipulations. As a consequence, researchers need to show that training-related change can be attributed to the targeted construct instead of any other confounding variables. As mentioned above, latent factors as well as latent measures of change are not tied to some theoretical construct of interest, which results in great flexibility and a broad range of possible applications but also requires careful interpretation.

Placebo-controlled designs have been proposed to minimize the effects of systematic biases like expectancy and motivation (Boot et al., 2011; 2013; Klingberg, 2010; Shipstead et al., 2012). Based on the classical 'Hawthorne Effect', Shipstead and colleagues (2012) argue, for example, that differences in expectations between trained and untrained participants may lead to differences in performance. Thus, according to these authors, challenging control conditions are needed to obscure group membership (experimental or control) to the participant. Rather than through direct manipulations of challenge or training intensity, this could be achieved through manipulations of the ability that is being targeted by treatment and control condition respectively (e.g., Shipstead et al., 2012; Redick et al., 2013, but see Boot et al., 2013). Following this logic, all participants would receive challenging interventions but one focusing on one ability (e.g., working memory) and the other focusing on another ability (e.g., processing speed). However, besides the obvious benefits of establishing challenging control conditions for minimizing non-cognitive biases, another complication arises from these designs: target and control training must have differential effects on the targeted transfer construct. Thus, these kinds of designs require a specific model of transfer that allows for the prediction of positive but also the prediction of absent transfer. That is, if researchers contrast the effects of improved working memory and perceptual speed on fluid intelligence (e.g., Colom et al., 2010), and argue for a positive relationship between working memory capacity and intelligence (Conway, Kane, & Engle, 2003; Kyllonen & Christal, 1990), they should also predict the absence of beneficial effects of increased perceptual speed on fluid intelligence. At least, they need to specify ways to distinguish differential impacts of the two abilities. Thus, the original idea to control for group differences in motivation and outcome expectancy (Shipstead et al., 2012; Redick et al., 2013) could then be extended to further validate the construct of training-related change. The training and assessment of multiple latent constructs (e.g., working memory and perceptual speed) as well as the theoretical model predicting the associations between the trained constructs and the targeted transfer construct (e.g., fluid intelligence) would allow for tests of discriminant validity of training-related change (Campbell & Fiske, 1959). In the present example, convergent validity could be assumed if working memory training leads to improvements in working memory and fluid intelligence, while discriminant validity could be assumed if working memory training would not lead to improvements in perceptual speed and perceptual speed training would not lead to improvements in fluid intelligence.

Transfer Models

We argued above that investigations of transfer require a theoretical model that includes both the definition of targeted transfer constructs and assumptions on the relationships between trained and transfer constructs. The model should be sufficiently specific to allow for predictions of positive and absent transfer. Beyond this requirement, however, the level of abstraction at which latent constructs may be defined is not of great importance for the argument presented here.

For example, such a model can be defined at the level of broad cognitive abilities. One such model draws on the observed association between working memory capacity and fluid intelligence (Conway et al., 2002; Engle, Tuholski, Laughlin, & Conway, 1999; Kyllonen & Christal, 1990). This model has been invoked frequently in recent training research. With only few exceptions, however, a discriminant part of the model was lacking. Being one of these exceptions, a study by Redick and colleagues (2013) implemented a visual search training based on the empirical finding that visual search performance is not related to working memory capacity or fluid intelligence respectively (Kane, Poole, Tuholski, & Engle, 2006).

Note, however, that Conway and Getz (2010) question the study of training-related changes at the level of cognitive abilities such as working memory and fluid intelligence, arguing that these concepts are too complex to allow for specific conclusions about where change is actually happening. According to these authors, mechanisms should be more at the focus of training research. We subscribe to this conclusion to the extent that the proposition of process models or mechanisms might allow for more specific predictions of expected patterns of transfer. However, theoretical constructs used to describe such mechanisms are typically defined at a latent level too and, as such, are not directly observable either. A small number (e.g., typically, one or two) of manifest variables, therefore, may not be sufficient to capture them well and the methodological considerations raised above apply to mechanisms as well as to cognitive abilities. For example, mental set switching as one component of executive control (Miyake et al., 2000) is typically operationalized using the task-set switching paradigm (e.g. Kray & Lindenberger, 2000). The task-set switching paradigm is only one possible instance of the latent construct of mental set switching, however, and, thus, change should be demonstrated at the latent level using multiple heterogeneous indicator variables to show that the latent construct of mental set switching had been altered with training. Recently, von Bastian and Oberauer (2013) provided an excellent example for how cognitive process models can be used to design specific training conditions and to derive specific predictions on transfer patterns. Very much inline with the present proposal, these authors also used multiple indicator variables and factor analytic methods to validate trained and transfer constructs.

Another possible transfer model can be derived from the neuronal overlap hypothesis, which posits that transfer can only be expected if trained task and transfer task share some overlap in neuronal activation (e.g., Dahlin et al. 2008; Jonides et al., 2004; Kuwajima & Sawaguchi, 2010; Lustig, et al., 2009; Thorell et al., 2009). As a test of the model, this overlap has to be shown to be sensitive and specific to the process or ability of interest. In a study of updating training, Dahlin and colleagues (2008) showed, for example, that trained (updating) and transfer (n-back) task invoked overlapping areas in the striatum. Based on earlier research showing that memory updating is related to striatal activity (e.g., Hazy, Frank, & O'Reilly, 2006), they argued that the common process shared among the trained and transfer tasks was updating. The finding that another transfer task (Stroop), which did not seem to involve memory updating, did not invoke striatal activity substantiated their hypothesis further. In fact, the transfer pattern conformed to the model showing transfer only to n-back but not to Stroop. Still, a note of caution should be voiced here because overlap in activation can be related to task-general targeted process (i.e., updating in this case) but it may also be related to task-specific skills transferring from the trained task to n-back but not to Stroop. The latter case would hold especially if the two tasks were rather similar at the surface, in which case overlapping areas of brain activity would be trivial. Thus, in principle, the same line of reasoning holds, no matter whether behavioral or neuroscientific approaches are used to investigate transfer: Single tasks are never process-pure and, therefore, it is difficult to know whether it is the targeted constructs (i.e., process, mechanism, or cognitive ability) rather than the specific skills and strategies that are being reflected in overlapping brain activation. One way to extract the targeted process more purely would, therefore, be to use multiple indicators and analyses at the latent factor level.

Interestingly, transfer models seem not only to exist in scientific theorizing but also in the heads of the participants. Boot and colleagues (2013) recently reported evidence for subjective transfer models in two survey studies with a total of 400 participants. They showed that participants expected differential training effects on multiple transfer tasks depending on the suspected training procedure. For example, after imagining playing *Unreal Tournament* for an extended period of time they expected greater improvements on measures of visuo-spatial functioning and processing speed than after imagining playing Tetris. On the other hand, for mental rotation they expected greater improvements after seeing a video of Tetris than after seeing a video of *Unreal Tournament*. Obviously, these findings add onto the notion of a broad effect of non-cognitive factors thereby challenging the suggested strategy of using the pattern of convergent and discriminant transfer to disentangle treatment from placebo effects. The effect of subjective transfer models may be difficult to anticipate in practice because these models may be in line with, independent of, or even opposed to the theoretical model underlying the study design. Thus, additional measures may be needed to account for the effect of subjective transfer models. Possible measures have been described recently (see Boot et al., 2013) and will not be further discussed here.

Measurement Invariance

When using latent-factor models to investigate training and transfer effects implementing, for example, latent change score models (McArdle, 2009), longitudinal measurement invariance becomes an important issue. The investigation of measurement invariance addresses the question of whether the meaning of the common factors is likely to have stayed the same over the course of the training intervention (McArdle & Nesselroade, 1994; Meredith & Horn, 2001). According to this concept, training-related changes at the latent level can only be interpreted at the latent factor level if the measurement model (i.e., factor loadings, intercepts, and residual variances) remained constant over time. Strict measurement invariance implies that all changes in task variables are represented by changes in the latent factor means and (co)variances.

Instead of, or in addition to, transfer at the latent factor level, however, there could also be task-specific effects. The presence of such task-specific effects is likely to disturb measurement invariance. A specific positive effect for one of the tasks of a common factor, for example, will lead to an increase of this variable's intercept at posttest. Likewise, individual differences in the strength of such a specific effect can increase the residual variance of this variable at posttest. As described above, only if the pattern of changes in means and (co)variances of specific effects mimicked the pattern of changes in means and (co)variances that would have to be expected if there was change at the latent factor level (e.g., through the action of general non-cognitive factors), specific effects could be mistaken for latent factor effects. As this is not a very likely scenario, task-specific effects will typically degrade measurement invariance. Importantly, however, the finding that measurement invariance does not hold needs not imply that transfer effects are only task-specific. It might well be that a training intervention leads to a combination of effects at the common factor level and task-specific effects, which in combination lead to a loss of measurement invariance.

In sum, these considerations lead to the conclusion that investigating measurement invariance in training research is important and informative, and that one should aim for the highest level of invariance possible. If strict measurement invariance holds and significant improvements are present for a latent factor that is operationalized with several heterogeneous tasks that all differ in content and/or paradigm from the trained tasks, this is strong evidence for improvements in the targeted construct. If only strong (invariant factor loadings and intercepts), weak (invariant factor loadings), or configural (same pattern of present and absent factor loadings) measurement invariance can be achieved, this might indicate the presence of task-specific effects, changes in the reliability of tasks, or changes in the nature of the latent construct itself (i.e., changes in validity). Whatever the exact reason, it is an informative finding that calls for further investigation. Just ignoring the whole issue of measurement invariance, for example, by creating composite scores at pretest and posttest is not a sensible option, in our view.

Statement of Problem

Above, we discussed the validity of transfer effects in training research. We showed that validity can be hampered by the acquisition of task-specific skill rather than improvement of task-general abilities, on the one hand, and by improvements of general non-cognitive factors rather than the specific targeted ability, on the other hand. We summarized possible strategies to strengthen the assumption that transfer effects actually represent effects at the level of the targeted construct. These calls for analyses of training-related gain at the latent level (McArdle & Prindle, 2008; Morrison & Chein, 2011; Schmiedek et al., 2010; Shipstead et al., 2010; 2012) and the corresponding investigations of measurement invariance, and for adequate placebo conditions (Boot et al., 2011; 2013; Redick et al., 2013; Shipstead et al., 2012), have been voiced before. Here, we review the recent training literature, investigating if these calls have been heard and if respective strategies have been implemented in practice.

Literature Review

We surveyed the prevalence of latent variable approaches in current training research doing a literature search in Thomson Reuters Web of Knowledge[SM]. Key words used in the searches included *training, transfer,* and *latent* but also the terms *intelligence, working memory, executive control,* and *executive functions,* which imply that latent constructs may have been targeted. We included studies of the last 5 years (i.e. 2007-2013 FOOTNOTE: Searches were conducted in March 2013), which were categorized in one or more of the research areas, *Psychology, Neuroscience, Geriatrics, Education,* or *Behavioral Sciences*. In the resulting list of articles, we scanned titles and abstracts for relevance and discarded those articles that featured certain exclusion criteria (for example, studying non-human animals or unhealthy populations). Following this procedure, we obtained a list of 133 original articles, 32 review articles, 8 commentaries, and 2 book sections. We evaluated the methods reported in original articles, categorizing the targeted cognitive ability or process (e.g., working memory, inductive reasoning, or executive functioning), the study design including sample size and training duration, the scope of training (single task, multiple tasks, or complex), scope of transfer assessment (none, single task, multiple tasks, complex), and the type of transfer evaluation (e.g., single, composite, latent). After evaluation, another 51 studies were classified irrelevant mostly because transfer was not assessed or control groups were not included and, thus, a final sample of 82 studies entered our evaluation. In the following sections, we give a short overview of the studies in the sample before discussing study design and analytic strategies.

Study Sample Overview

Training Targets

Working memory (n = 35) training was the most common approach in the current sample of studies. This strong focus can be attributed to the promising results reported by Jaeggi and colleages (2008) and the work by Klingberg and colleagues (2005) but also to the presence of a plausible transfer model (e.g., Morrison & Chein, 2011; Shipstead et al., 2012).

Similarly, training of executive control produced some promising results in the past (e.g., Bherer et al., 2005; Karbach & Kray, 2009; Kramer, Larish, & Strayer, 1995; see Noack et al., 2009 for reviews) motivating further investigation of the trainability of this ability (n = 16). Thirteen of the 16 studies reported data from older samples while, for example, only 9 of the 35 studies on working memory reported data from older samples. The focus on older participants may in part be due to the assumption that deficits in executive control and inhibition are central determinants of age-related cognitive decline in general (e.g., Hasher, Stoltzfus, Zacks, & Rypma, 1991; West, 1996).

Improvements in fluid abilities are a common aim of cognitive training procedures because of their strong association with scholastic success (e.g., Deary, Strand, Smith, & Fernandes, 2007), health, and longevity (Gottfredson & Deary, 2004). Tasks of inductive and deductive reasoning (n = 10) are central to fluid cognitive abilities (Cattell, 1972) and given "the difficulties of inducing transfer effects to reasoning ability following working memory training, an alternative approach is to train directly on tasks that load highly on reasoning ability." (Söderquist, Bergman Nutley, Ottersen, Grill, & Klingberg, 2012, p.2; see also Bergman Nutley et al., 2011; Mackey et al., 2011; Thorell, Lindquist, Bergman Nutley, Bohlin, & Klingberg, 2009). Training of general cognitive ability (n = 7) subsumed studies including various training tasks tapping into different mental abilities like processing speed, working memory, and episodic memory (e.g., Schmiedek et al., 2010) but also studies investigating the effects of broader commercial programs for brain fitness like Brain Age (Nouchi et al., 2013) or Brain Fitness (McDougall & House, 2012).

Other studies investigated the effects of video gaming (n = 5), memory strategy instruction (n = 4), physical fitness training (n = 3), processing speed training (n = 3), and musical ability (n = 3). Finally, some unique training approaches were also tested in single studies (n = 4). One study, for example, tested the effect of origami practice on imagination performance in younger women (Jausovec & Jausovec, 2012).

Sample Sizes

The 82 studies in the sample reported data on a total of 20826 participants ranging from 11 months to more than 85 years of age. On average, studies included 253 participants. This relatively large number is strongly biased, however, by three studies, which contributed almost 75% of all the participants (Irwing et al., 2008: n = 2492; McArdle & Prindle, 2008: n = 1397; Owen et al., 2010: n = 11430). Fifty percent of the studies reported data of 45 participants or fewer and 90% of the studies reported data of 100 participants or fewer. Dependent on study designs, these participants belonged to variable numbers of groups and, thus, the average group size may be a more sensible marker. Single groups comprised 93 participants on average, but again this estimate is strongly biased by the three large scale studies mentioned above (Irwing et al., 2008; McArdle & Prindle, 2008; Owen et al., 2010). Remarkably, 50% of the studies had an average group size of 20 participants or fewer and 90% of the studies had an average group size of 45 participants or fewer (see Fig. 1). Of note for the focus of the present review, sample sizes of fewer than 45 participants are generally insufficient to fit common factor models (MacCallum, Widaman, Zhang, & Hong, 1999). Thus, the vast majority of the reviewed studies are not well suited for latent transfer analyses because sample sizes are too small.

Training Durations

Studies varied also considerably with respect to training durations. Among those studies that provided sufficient information to estimate the average training duration at the level of hours (FOOTNOTE: Exact training duration could not be estimated in three studies (Lustig & Flegal, 2008; Mårtensson & Lövdén, 2011; McDougall & House, 2012)), three studies reported training durations of at least 100 hours (Degé, Wehrum, Stark, & Schwarzer, 2011; Schmiedek et al., 2010; Voelcker-Rehage, Godde, & Staudinger, 2011). Of these three, only one was designed as a cognitive intervention study (Schmiedek et al., 2010), one focused on the effect of physical fitness training (Voelcker-Rehage et al., 2011), and one evaluated the effect of extended music education in school (Degé et al., 2011), which included learning to play an instrument as well as lessons in music theory. The majority of the studies, however, implemented much less extensive training regimes (Fig. 2). Fifty percent of the studies reported training durations of 8 hours and 20 minutes or less and even ninety percent of the studies reported training durations of 32 hours or less. From a conceptual point of view, short training durations are troublesome, as training duration is likely to be a critical factor determining transfer magnitude (Lövdén et al., 2010).

Scope of Training

Training studies differed in scope of the training procedures. While 26 studies investigated the effects of practice on single tasks like the dual n-back task (e.g. Chooi & Thompson, 2012; Jaeggi et al., 2008; Rudbeck et al., 2012; Redick et al., 2013; Salminen, Strobach, & Schubert, 2012), the dual task (e.g. Bherer, Kramer, & Peterson, 2008; Lussier, Gagnon, & Bherer, 2012; Mackay-Brandt, 2011), or the task-switching paradigm (e.g. Karbach & Kray, 2009; Kray, Karbach, Haenig, & Freitag, 2012; Zinke et al., 2012), another 31 studies provided multiple task training that either tapped into one single cognitive ability (e.g. Chein & Morrison, 2011; Jackson, Hill, Payne, Roberts, & Stine-Morrow, 2012; McArdle & Prindle, 2008; Owen et al., 2010; Richmond, Morrison, Chein, & Olson, 2011; Söderquist et al., 2012; Thorell et al., 2009) or that spanned a broad set of different cognitive abilities (e.g. Penner et al., 2012; Schmiedek et al., 2010; Simpson, Camfield, Pipingas, Macpherson, & Stough, 2012). To train episodic memory, Schmiedek and colleagues (2010), for example, used three different training tasks pertaining to three different content domains: verbal, spatial, and numerical. The rationale for using multiple different training tasks is to reduce the relative importance of task-specific in favor of more task-general improvements. It is important to note, however, that multivariate training regimens may also have shortcomings compared the univariate task procedures when training-related improvements are not investigated at the latent level. Given that transfer effects are observed, it is much more difficult to evaluate where exactly such gains came from. Was it then the strengthened underlying ability, which caused the transfer, was it the broader set of acquired task-specific skills, or was it the task-specific skill of one of the various tasks alone?

Conceptual clarity is traded even more for ecological validity in complex training regimens (n = 25). Complex tasks, like action video gaming (e.g., Basak et al., 2008; Green, Sugarman, Medford, Klobusicky, & Bavellier, 2012; Maillot, Perrot, & Hartley, 2012; Sanchez, 2012), origami practice (Jausovec & Jausovec, 2012), painting (Tranter & Koutstaal, 2008), music education (Degé et al., 2011; Moreno et al., 2011), virtual breakfast cooking (Wang, Chang, & Su, 2011), or the participation in volunteer senior services (Carlson et al., 2008) were distinguished from the other tasks (in single and multiple-task trainings) because these activities were not designed to address one well-defined cognitive ability or to draw on one psychological process alone. It is important to note, however, that this distinction is somewhat arbitrary from an ecological perspective because all tasks, if designed for a specific purpose or not, are not process pure and usually draw on multiple cognitive abilities (e.g., Redick et al., 2013).

Evaluation of the Research Questions

Study Designs

We first analyzed study designs of the 82 studies in the sample to see whether the comparison of active control and treatment groups on the basis of trained abilities rather than training intensity was implemented in practice. This approach is preferable over the often used non-adaptive placebo design for three reasons (see also Boot et al., 2011; 2013; Shipstead et al., 2010; 2012; Sternberg, 2008): (1) Treatment and placebo conditions may be more comparable in terms of motivation because both conditions remain challenging over the entire training period; (2) conditions may differ less in outcome expectancy; and (3) theoretical implications of training gains can be clarified because more specific contrasts can be made with respect to the targeted cognitive ability. This holds particularly if transfer tasks are selected for construct validation of training gains (see below).

Among the 82 studies, only one study explicitly applied a study design implementing an active control condition selected based on expected relationships with the treatment condition and the transfer condition (Redick et al., 2013). Several other studies had similar designs, however. In contrast to the clear theoretical assignment of training and control group, however, these studies compared multiple training regimens to explore if one or more of them were effective or not (e.g., Bergman Nutley et al., 2011; Mackey et al., 2011; Söderquist et al., 2012; Thorell et al., 2009). For example, Thorell and colleagues (2009) investigated the effects of working memory and inhibition training in children between 4 and 5 years of age. Originating from the observation of a positive effect of working memory training on ADHD (Klingberg et al., 2005), the observation of a strong association between inhibition and working memory (Engle & Kane, 2004), and the observation of overlapping patterns of brain activity (McNab et al., 2009), the authors asked if inhibition training and working memory training would show transfer specific to the trained ability or general transfer to the respective other, non-trained, ability. The authors found within-ability transfer for the working memory training but not for inhibition training. These results suggest that – in contrast to inhibition – working memory can be improved by training. Obviously, the rationale behind this third approach strongly overlaps with the idea of construct validation mentioned above. However, in contrast to it, the separation between training and control condition seems less clear. This is reflected in the analytic strategy used in the example described above (Thorell et al., 2009). Rather than comparing the effects of inhibition and working memory directly, these authors contrasted each condition to the active control condition (i.e. computer games with emphasis on sensori-motor coordination).

It is often difficult to say if contrasts are being made between multiple treatments or between treatment and placebo. We still tried to classify the observed study designs, in order to give a short impression of current preferences in training research. We decided to subsume the two aforementioned approaches under the category of group comparisons (n = 26), which also included studies in which specific components of single tasks were tested – for example, when dual n-back training was compared to single n-back training (Jaeggi et al. 2010; Studer-Luethi, Jaeggi, Buschkuehl, & Perrig, 2012) or when full emphasis training was compared to variable priority training in the space fortress game (e.g., Boot et al., 2010; Lee et al., 2012; Stern et al., 2011). The defining criterion of this category of study designs was that training effects may be expected in multiple groups and that the focus of the analyses was on the differences between the respective outcomes. In contrast, the category of active control designs (n = 26) subsumed studies where control conditions mimicked the level of physical and social engagement related to coming to the lab or to doing certain tasks for a certain amount of time but where no specific training effect was expected. These control conditions included non-adaptive or non-demanding forms of the trained task or alternative activities like watching documentaries or answering trivia questions using the Internet. Finally, the largest portion of studies (n = 30), however, still contrasted training effects to passive control groups. It is important to note that these categories were mutually exclusive but not homogeneous. Studies implementing a passive control condition might, for example, implement an active control condition, and a group comparison as well (see e.g., Thorell et al., 2009; Wen, Butler, & Koutstaal, 2013). We ordered our categories hierarchically, starting with group comparisons at the top and ending at passive control designs at the bottom. Studies were classified according to their highest-ranking contrast on this continuum.

Patterns of Transfer

Although the focus of the present article was not on the effectiveness of training procedures, we evaluate the reported transfer effects in the study sample to exemplify the usefulness of using a transfer model with accompanying criteria for convergent and discriminant validity. Seventy-one of the 82 studies in the sample (87 %) reported reliable training-related differences in some transfer measure, suggesting that cognitive training is largely effective. However, we argued above that the observation of transfer might only be of limited value if the precise pattern of transfer effects is not considered. In the present sample, transfer was assessed using multiple transfer tasks in 74 of the 82 studies (90%) – with only 6 of these studies (8%) reporting transfer to all of the measures assessed. As described above, however, even training-related gains on all transfer measures would not necessarily imply training effectiveness and generality. To the contrary, general transfer could falsify the theoretical concept of training if theory predicted transfer to one construct but not to another (Lövdén et al., 2010). In one study, for example, Borella and colleagues (2010) reported reliable transfer effects of auditory working memory training in older people on measures of visuo-spatial working memory, performance in Cattell's Culture Fair, Color Stroop, and pattern comparison. Based on findings of parallel decline of working memory and processing speed in aging (e.g., Craik & Salthouse, 2000), the authors expected transfer of working-memory training gains to processing speed. The processing speed hypothesis of aging (Salthouse, 1996), however, posits that causality points from processing speed to other cognitive abilities. Therefore, it remains theoretically unclear how improved working memory performance could manifest in simple reaction time tasks. This example illustrates that statements on training effectiveness can only be made with appreciation of the specific study designs and hypotheses at hand. Such in-depth analyses would clearly go beyond the scope of the present review. Some examples may still be helpful to give a short overview of what we found.

First, 11 studies did not observe any transfer at all. These findings are rather clear with respect to the pattern of transfer but it may still be instructive to investigate the theoretical context within which they were obtained. Two studies reported the effects of long-term aerobic physical fitness training on cognition and brain activity (Voelcker-Rehage et al., 2011; Voss et al., 2010). The absence of transfer is surprising given the meta-analytic observations by Colcombe and Kramer (2003) who found that physical fitness training is effective in increasing cognitive performance especially if training persists over a long period of time. Depending on the underlying transfer model, however, the absence of transfer to other cognitive measures can still be theoretically plausible. If one assumes that beneficial effects of aerobic exercise are mediated by neurogenesis through up-regulation of neurotrophic factors (Perreira et al., 2007; van Praag, Kempermann, & Gage, 2000). If the integration of new neurons in the nervous system is experience-dependent (Fabel et al., 2009; Kempermann, 2008), it may follow that manipulations of physical exercise without additional manipulations of cognitive context may lead to altered patterns of brain activity that do not surface at the level of cognitive performance (Voelcker-Rehage et al., 2011; Voss et al., 2010). Whereas absent transfer in the studies just described may not be in conflict with the transfer model, there were two other illustrative studies where absent transfer actually indicated that training was ineffective and that the hypothesized transfer model did not provide an adequate description of the data (Chooi & Thompson, 2012; Redick et al., 2013). Both studies employed the dual n-back paradigm that had been used before to induce reliable transfer in measures of matrix reasoning (e.g., Jaeggi et al., 2008; 2010). Although both studies showed substantial improvements in the criterion task, they failed to demonstrate transfer to any of the transfer measures assessed, including measures of fluid intelligence, multitasking, working memory capacity, crystallized intelligence, and perceptual speed (Redick et al., 2013), as well as measures of verbal, perceptual, and mental rotation abilities (Chooi & Thompson, 2012).

The majority of the studies (n = 58) found transfer to some measures together with absent transfer to other measures. This pattern can be in accordance with the theory if transfer tasks are selected according to the logic of convergent and discriminant validity. In one study, Mackey and colleagues (2011) tested the effects of commercial games drawing either on reasoning abilities or processing speed in children at the age of 7 to 9 years. They expected transfer of the processing speed games to psychometric tests of processing speed and transfer of the reasoning games to psychometric tests of reasoning. In fact, their results supported their assumptions, showing that processing speed games led to improvements on a variant of digit-symbol substitution, and that reasoning games led to an improvement in matrix reasoning. Processing speed training, on the other hand, did not lead to improvements in matrix reasoning and reasoning training did not lead to improvements in one task of processing speed. Some other studies also found systematic transfer, which was not in full agreement with the transfer model, however. Implementing a working memory training with spatial n-back tasks, Li and colleagues (2008) found transfer of training gains to a numerical variant of the n-back task but no transfer to other measures of working memory capacity like operation span or reading span. Thus, it seems rather likely in this case that some specific task-related skill rather than working memory capacity *per se* was increased (see also Peng, Wen, Wang, & Gao, 2012, for similar results).

Different from this systematic absence of transfer, however, several studies also showed some unsystematic patterns of transfer. For example, Basak et al. (2008) examined the effect of playing strategy computer games on executive control and visuospatial attention in older participants. To test for improvements in executive control, operation span, task-switching performance, n-back speed and accuracy, and inhibition were compared between training and control group. Given that executive control is governed by three executive functions (shifting, updating, and inhibition; Miyake et al., 2000) and given that computer gaming enhanced executive control, improvements on all five measures can be expected. This expectation was not met by the data, however. Although the positive transfer to two of the five measures (task-switching, n-back speed) is certainly promising, the absence of transfer to the other three measures seems to cast some doubt on the validity of the underlying transfer model. Similarly, Anguera and colleagues (2013) recently showed transfer of gains in dual-task training to a measure of delayed recognition and a measure of visual attention. The authors interpret these results in terms of improvements in cognitive control abilities extending to working memory and visual attention. In addition to the finding of positive transfer, these authors also report that training gains did not transfer to other classical measures of dual-tasking, visual attention, or working memory, however. This inconsistency in the pattern of transfer results makes it difficult to support interpretations of training gains at the broad level of cognitive abilities or cognitive processes without further restraint.

Data Analysis and Methods

As outlined above, the central aim of training research is to identify improvements at the level of cognitive abilities rather than at the level of task-specific skills (e.g., McArdle & Prindle, 2008; Morrison & Chein, 2011; Schmiedek et al., 2010; Shipstead et al., 2010; 2012). This aim is reflected in the high number of studies that claim training effects and transfer at the level of latent constructs like working memory, executive functioning, or fluid intelligence. We argued that these claims would be strengthened considerably if theoretical constructs were analyzed at the latent level. Here, we evaluate if these kinds of analyses were common practice rather than the exception. The short answer to this question is clear: Only 6 of the 82 studies (7%) referred to the latent level in some way. Another 14 studies reported composite or sum scores, which tend to show better psychometric properties than individual task scores. Four studies used multivariate Analysis of Variance to establish an omnibus effect of training effectiveness including all transfer measures at the same time, and, finally, the vast majority of studies used univariate methods including ANOVA, ANCOVA, and simple t-tests of difference scores (n = 58). Unfortunately, univariate techniques are not only the most common but also the most problematic technique because single task scores show potentially low reliability together with high proportions of task-specific variance (i.e., low construct validity). A related problem is alpha-error accumulation due to multiple testing when several transfer measures are included in the study.

Similar to the univariate category, the latent category was not homogeneous with respect to the techniques used. Only three studies actually tested training gains and transfer effects at the latent level using latent growth curves (Jackson et al., 2012) or latent differences (McArdle & Prindle, 2008, Schmiedek et al., 2010). One study estimated factor scores at pretest and posttest separately and analyzed these estimates using univariate methods (Bergman Nutley et al., 2011). Two studies implemented the correlated vectors approach, introduced by Jensen (1998; Colom et al., 2010; Redick et al., 2013).

Correlated Vectors

Originally developed to identify the source of test differences between ethnic groups, this approach can also be used to determine the degree to which training-related differences can be attributed to some latent construct of interest (i.e., cognitive ability). In training research, correlated vectors simply represent the correlation between mean effect sizes on several tests and the respective loadings of the tests on the ability of interest. Factor loadings can either be estimated from the sample using factor analytic methods or, if available, can be taken from the standardization sample of the respective psychometric test (Jensen, 1998). Rather than analyzing training-related change (and individual differences therein) at the latent level, this approach is restricted to the estimation of the extent to which average gains can be attributed to some latent construct of interest. Construct validation is built-in in this approach because transfer is expected to occur only to variables that are related to the transfer construct.

Results of the correlated vectors analysis in the present sample of studies are not quite satisfactory, however. Colom and colleagues (2010) used the method post-hoc to explore the pattern of transfer onto four reasoning tests, where one of the four tests did not show transfer while the others did. The authors conducted exploratory factor analyses at pretest and posttest respectively interpreting the resulting factor in the sense of *g*. Correlation estimates between effect sizes and factor loadings were strongly negative, suggesting that those tasks showing the highest association with *g* were those that showed the smallest training effect. These results are difficult to interpret, however, at least for three reasons: (1) Factor loadings are obtained from exploratory analyses of pretest and posttest data separately, and there is, thus, no way of knowing whether these two factors actually capture the same construct (McArdle, 2007). (2) The reported correlations are based on four data points only and it must be suspected that their reliability is low. (3) The critical test is not specified. Should one simply expect higher correlations in the training group compared to the control group or should the control group show no or even negative correlations? It would also be possible that correlations were the same in the two groups but that mean changes were different. Unfortunately, Colom and colleagues do not discuss the theoretical implications of different patterns of correlations.

Redick and colleagues (2013) addressed the latter issue by assessing a set of 17 transfer measures before and after training. Unfortunately, these authors did not observe transfer in any of their transfer measures, and hence refrained from analyses of the latent level.

Univariate Analyses of Factor Score Estimates

In comparison to the correlated vectors approach, the procedure taken by Bergman Nutley and colleagues (2011) seems much more direct. These authors asked if improvements can be observed at the latent level. Rather than answering this question at the latent level directly, however, they estimated individual factor scores at pre- and post-test separately to analyze them at the manifest level. Despite of being practical, this step is somewhat at odds with the concept of latent factors, which implies that scores are not directly observable.

Consequently, estimation procedures are needed to make latent factor scores manifest. Estimation is never perfect, however, and the exact properties depend on the procedure that is being employed (Horn, 1965). Bergman Nutley and colleagues (2011) did not specify which procedure they used. Another critical aspect, which potentially hampers the interpretation of their results, is related to the fact that factor solutions were obtained separately at the two time points. Moreover, factor solutions were obtained for the whole sample at once, not separating between experimental conditions. Although this is straightforward at pretest, where no differences between groups should be expected, intervention might have influenced the factor structure differently in intervention and control groups, respectively, such that a formal demonstration of measurement invariance between groups at posttest appears in order. As noted above, the potential lack of measurement invariance, that is, the demonstration that factor solutions are comparable across time points and groups, weakens the interpretation of latent change scores in the sense of the underlying common factors (McArdle, 2007; 2009). In fact, factor loadings seem to have changed over time in the example of Bergman Nutley and colleagues (2011).

Analyses of Latent Change

In contrast to the studies discussed above, those three studies that investigated change at the latent level all established invariant measurement models first. The property of measurement invariance later allows for interpretation of latent change scores or latent slopes in the sense of actual change at the level of latent abilities.

Jackson and colleagues (2012) investigated potential changes in the personality trait “openness to experience” (Costa & McCrae, 1992) in response to inductive reasoning training. Assuming that personality traits and intellectual ability are mutually influencing each other, the authors expected to see increases in openness to experience in response to the training intervention. Personality measures were assessed in the training group and a passive control group four times over the course of the 16-week training program and cognitive measures were assessed before and after the intervention. Change in openness to experience was assessed using second-order latent growth curve models and change in cognitive measures was analyzed using latent change score models. These two methods overlap in many respects. Both build on metrically invariant common factors estimated at each time point. Change is then represented by latent change scores in the latent change score models (McArdle & Nesselroade, 1994) and by latent slopes in the latent growth curve models. Using these models, the authors found that – at the latent level – the training group gained more on reasoning ability than the passive control group, showing that training was effective. In agreement with their hypothesis, this improvement in reasoning ability was paralleled by increments in openness to experience. The latent slope for openness was positive in the training group but negative in the control group, leading to different levels of openness at posttest. Although these analyses may serve as a good example for the use of latent measures in training research, there is still more potential. The transfer model by Jackson and colleagues (2012) rests on the assumption of mutual influences between intellectual ability and personality traits, which seems to suggest correlations between baseline levels but also between baseline levels and change on the two dimensions. That is, one might expect, for example, that those who have high baseline levels of openness may also be those who profit most from the intervention. Similarly, it would have been interesting to see if training gains in reasoning performance were correlated to changes in the personality trait. Note that changes at the group level do not allow for inferences about individual differences in change within the trained group (Baltes, Reese Nesselroade, 1988; McArdle & Prindle, 2008; Noack, Lövdén, Schmiedek, & Lindenberger, 2013; Robinson, 2009). That is, there may be participants improving in reasoning performance and participants increasing in openness to experience, but these individuals need not to be the same. Correlations between changes in openness and changes in reasoning performance would be informative in this regard. Latent measures of change allow for these kinds of analyses because their estimates are free of measurement error and reliable correlations can, therefore, be obtained. Jackson and colleagues (2012) tested none of these correlations, however.

Schmiedek and colleagues (2010) went one step further. Their main focus was on demonstrating that improvements at the ability level can be achieved through broad, intensive, and extensive training. They constructed factors of the trained abilities (working memory, episodic memory, and reasoning) using indicator variables that were not part of the training. Training-related differences in change were then investigated at the latent level for each ability. In a second step, these authors compared correlations between latent measures of change derived from trained tasks and latent changes derived from transfer tasks. Now, looking at different abilities, the authors had the opportunity to cross-validate changes for each ability. In fact, the high correlation between changes in working memory trained and change in working memory transfer, and between changes in episodic memory trained and changes in episodic memory transfer point to the convergent validity of the observed transfer effects. On the other hand, lower correlations between working memory trained and episodic memory transfer as well as between episodic memory trained and working memory transfer may represent discriminant validity indicating that training-related improvements were specific for each ability.

Finally, McArdle & Prindle (2008) sought to test for relationships between baseline performance and change both within and between near and far measures of transfer. Conforming to theory, they assumed a directed link of change in near transfer to change in far transfer. They reanalyzed data from the reasoning training of the ACTIVE study (Ball et al., 2002; Jobe et al., 2001). In this study, trained participants received 10 sessions of reasoning training while control participants took part in the pre- and posttest assessment only. Measures of near transfer included the trained task but also two more non-trained tasks of the same ability, and measures of far transfer included tests of everyday functioning. Starting with a group-invariant bivariate latent change score model with common factors, these authors sequentially tested for group differences in latent change of near and far transfer, for lagged and cross-lagged regressions, and finally, for the crossed regression from near transfer change to far transfer change. This sequential model testing procedure revealed that latent change differed between groups in near transfer while all regression weights did not, suggesting that the task structure was invariant across groups and that training-related differences were restricted to differences in level. A positive standardized regression weight between change in near transfer and change in far transfer showed, however, that the observed mean level differences between the two measures can indeed be generalized to the individual to some degree. That is, participants who showed high gains in near transfer tended to show high gains in far transfer as well.

Conclusion

The recent review, together with others (e.g., Hertzog et al., 2009; Lövdén et al., 2010; Morrison & Chein, 2011; Noack et al., 2009; Klingberg, 2010; Shipstead et al., 2010; 2012), revealed that some transfer is typically observed with training procedures targeting working memory, executive functioning, general cognitive ability, and other constructs. However, the validity of positive transfer depends on both theoretical and methodological considerations. The threat of confusing task-specific effects with task-general effects, which are typically defined at a latent level, can be overcome using multiple heterogeneous transfer measures and analyzing them at a latent level. We emphasized that the analysis of latent measures applies to any theoretical construct that cannot be assessed with perfect reliability and validity at the manifest level. The call for latent methods in training research has been voiced before (Lövdén et al., 2010; McArdle & Prindle, 2008; Morrison & Chein, 2011; Noack et al., 2009; Schmiedek et al., 2010; Shipstead et al., 2010; 2012). Our review of the literature of the past five years revealed, however, that the theoretical ideal does not often transfer to the daily practice of training research.

Though latent analyses of transfer are well suited to arrive at valid conclusions about transfer, we noted that they may be sensitive to the operation of general but non-cognitive factors. We suggested implementing active control groups on the basis of a priori transfer models. Together with the assessment with multiple transfer tasks pertaining to different abilities, following the multitrait-multimethod logic (Campbell & Fiske, 1959), this would allow for estimates of convergent and discriminant validity. With the exception of one study (Redick et al., 2013; see also Mackey et al, 2011), placebo-controlled designs of this kind were not present among the studies in our sample. Our evaluation showed, to the contrary, that most studies implemented passive control conditions.

Finally, we highlighted the utility of latent change score models, showing that theoretical assumptions about associations between trained and transfer tasks can be modeled explicitly. These analyses are very helpful in verifying (or falsifying) hypothesized positive correlations between trained and transfer tasks within the trained group when group differences are present (Baltes, et al., 1988; McArdle & Prindle, 2008; Noack et al., 2013; Robinson, 2009).

Despite the potential conceptual benefits of the proposed methodological approaches, we acknowledge that these strategies are themselves tied to additional assumptions and preconditions, which might in part explain the low prevalence in research practice. For example, analysis of latent change requires the assessment of large samples using comprehensive test batteries. As described above, sample sizes in at least 90% of the studies reviewed here were likely too small to allow for this type of analyses, suggesting that investigation of latent measures of change is also a matter of costs. Besides these practical issues, there are some open conceptual questions as well. As mentioned earlier, testing for convergent and discriminant validity depends on the availability of a transfer model, which allows for predictions of the relationships between the trained and the transfer construct but also between a control construct and the earlier two. These predictions can only be made on the basis of good conceptual and empirical knowledge of the matter, which may not always be available from the start. Additionally, the presence of subjective transfer models may interfere with the theoretical assumptions and – if not properly addressed – may lead to false alarms as well as false negative results (Boot et al., 2013). Finally, latent factor models of cognitive constructs come with additional implications. For example, interpretation of latent factor models at the level of single individuals relies on the implicit assumption of ergodicity (i.e. the assumption that the data structure of multiple assessments of a single individual corresponds to the data structure of single assessments of multiple individuals; see Molenaar & Campbell, 2009). This means that measurement invariance does not only need to hold for the between-person structures across measurement occasions, but also for each individual, which could only be tested by measuring individuals multiple times to analyze their individual within-person structures. If within-person structures deviate from the between-person structures, findings on latent factor changes cannot be generalized to these individuals. Because the question of whether the efficiency of cognitive trainings differs across people and how such individual differences can be explained is highly relevant, these issues deserve careful investigation. Here again, latent factor approaches should be seen as opening opportunities rather than imposing constraints, in our view, because they provide the necessary means to investigate the structural equivalence of cognitive constructs.

To conclude, latent measures of change possess the potential to overcome the old issue of separating task-specific from task-general effects. The validity of these measures depends on the presence of a theoretical model of transfer and a study design that adheres to this model.

Acknowledgements

We would like to thank Ulman Lindenberger for contributing helpful comments on earlier versions of this manuscript.

References

Anguera, J. A., Boccanfuso, J., Rintoul, J. L., Al-Hashimi, O., Faraji, F., Janowich, J., . . . Gazzaley, A. (2013). Video game training enhances cognitive control in older adults. *Nature*, 501(7465), 97-101. doi: 10.1038/Nature12486

Ashcraft, M. H., & Kirk, E. P. (2001). The relationships among working memory, math anxiety, and performance. *Journal of Experimental Psychology-General, 130*(2), 224-237. doi: 10.1037/0096-3445.130.2.224

Ball, K. K., Berch, D. B., Helmers, K. F., Jobe, J. B., Leveck, M. D., Marsiske, M., . . . Grp, A. S. (2002). Effects of cognitive training interventions with older adults - A randomized controlled trial. *Jama-Journal of the American Medical Association, 288*(18), 2271-2281. doi: 10.1001/jama.288.18.2271

Baltes, P. B., Dittmann-Kohli, F., & Kliegl, R. (1986). Reserve-capacity of the elderly in aging-sensitive tests of fluid intelligence: Replication and extension. *Psychology and Aging, 1*(2), 172-177. doi: 10.1037/0882-7974.1.2.172

Baltes, P. B. & Lindenberger, U. (1988). On the range of cognitive plasticity in old age as a function of experience: 15 years of intervention research. *Behavior Therapy, 19,* 283-300. doi: 10.1016/S0005-7894(88)80003-0

Baltes, P. B., Reese Hayne W., & Nesselroade John R. (1988). *Life-span developmental psychology : introduction to research methods*. Hillsdale, NJ: Erlbaum.

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin, 128*(4), 612-637. doi: 10.1037//0033-2909.128.4.612

Basak, C., Boot, W. R., Voss, M. W., & Kramer, A. F. (2008). Can Training in a Real-Time Strategy Video Game Attenuate Cognitive Decline in Older Adults? *Psychology and Aging, 23*(4), 765-777. doi: 10.1037/a0013494

Bergman Nutley, S., Söderqvist, S., Bryde, S., Thorell, L. B., Humphreys, K., & Klingberg, T.

(2011). Gains in fluid intelligence after training non-verbal reasoning in 4-year-old children: a controlled, randomized study. *Developmental science, 14*(3), 591-601. doi: 10.1111/j.1467-7687.2010.01022.x

Bherer, L., Kramer, A. F., & Peterson, M. S. (2008). Transfer effects in task-set cost and dual-task cost after dual-task training in older and younger adults: Further evidence for cognitive plasticity in attentional control in late adulthood. *Experimental Aging Research, 34*(3), 188-219. doi: 10.1080/03610730802070068

Bherer, L., Kramer, A. F., Peterson, M. S., Colcombe, S., Erickson, K., & Becic, E. (2005). Training effects on dual-task performance: Are there age-related differences in plasticity of attentional control? *Psychology and Aging, 20*(4), 695-709. doi: 10.1037/0882-7974.20.4.695

Boot, W. R., Basak, C., Erickson, K. I., Neider, M., Simons, D. J., Fabiani, M., . . . Kramer, A. F. (2010). Transfer of skill engendered by complex task training under conditions of variable priority. *Acta Psychologica, 135*(3), 349-357. doi: 10.1016/j.actpsy.2010.09.005

Boot, W. R., Blakely, D. P., & Simons, D. J. (2011). Do action video games improve perception and cognition? *Frontiers in Psychology, 2*, 226. doi: 10.3389/fpsyg.2011.00226

Boot, W. R., Simons, D. J., Stothart, C., & Stutts, C. (2013). The Pervasive Problem With Placebos in Psychology: Why Active Control Groups Are Not Sufficient to Rule Out Placebo Effects. *Perspectives on Psychological Science*, 8(4), 445-454. doi: 10.1177/1745691613491271

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), 767-778. doi: 10.1037/a0020683

Brehmer, Y., Rieckmann, A., Bellander, M., Westerberg, H., Fischer, H., & Backman, L. (2011). Neural correlates of training-related working-memory gains in old age. *Neuroimage, 58*(4), 1110-1120. doi: 10.1016/j.neuroimage.2011.06.079

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological bulletin, 56*(2), 81-105. doi: 10.1037/h0046016

Carlson, M. C., Saczynski, J. S., Rebok, G. W., Seeman, T., Glass, T. A., McGill, S., . . . Fried, L. P. (2008). Exploring the Effects of an "Everyday" Activity Program on Executive Function and Memory in Older Adults: Experience Corps (R). *Gerontologist, 48*(6), 793-801.

Carroll, J. B. (1993). *Human cognitive abilities*. Cambridge: University Press.

Cattell, R. B. (1952). *Factor analysis: an introduction and manual for psychologists and social scientists*. New York: Harper.

Cattell, R. B. (1971). *Abilities: Their structure growth and action*. Boston: Houghton Mifflin.

Chein, J. M., & Morrison, A. B. (2010). Expanding the mind's workspace: Training and transfer effects with a complex working memory span task. *Psychonomic Bulletin & Review, 17*(2), 193-199. doi: 10.3758/pbr.17.2.193

Chooi, W.-T., & Thompson, L. A. (2012). Working memory training does not improve intelligence in healthy young adults. *Intelligence, 40*(6), 531-542. doi: 10.1016/j.intell.2012.07.004

Colcombe, S., & Kramer, A. F. (2003). Fitness effects on the cognitive function of older adults: A meta-analytic study. *Psychological Science, 14*(2), 125-130. doi: 10.1111/1467-9280.t01-1-01430

Colom, R., Martinez-Molina, A., Chun Shin, P., & Santacreu, J. (2010). Intelligence, working memory, and multitasking performance. *Intelligence, 38*(6), 543-551. doi: 10.1016/j.intell.2010.08.002

Conway, A. R. A., & Getz, S. J. (2010). Cognitive Ability: Does Working Memory Training Enhance Intelligence? *Current Biology, 20*(8), R362-R364. doi: 10.1016/j.cub.2010.03.001

Conway, A.R. A., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, 3, 163-183. doi:10.1016/S0160-2896(01)00096-4

Costa, P. T., & McCrae, R. R. (1992). Normal personality-assessment in clinical practice: The NEO personality inventory. *Psychological Assessment, 4*(1), 5-13. doi: 10.1037/1040-3590.4.1.5

Craik, F. I. M., & Salthouse, T. A. (2000). *Handbook of aging and cognition*. Hillsdale: Erlbaum.

Cronbach, L. J., & Furby, L. (1970). How we should measure "change": Or should we? *Psychological Bulletin, 74*(1), 68-80. doi: 10.1037/h0029382

Dahlin, E., Neely, A. S., Larsson, A., Backman, L., & Nyberg, L. (2008). Transfer of learning after updating training mediated by the striatum. *Science, 320*(5882), 1510-1512. doi: 10.1126/science.1155466

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence, 35*(1), 13-21. doi: 10.1016/j.intell.2006.02.001

Dege, F., Wehrum, S., Stark, R., & Schwarzer, G. (2011). The influence of two years of school music training in secondary school on visual and auditory memory. *European Journal of Developmental Psychology, 8*(5), 608-623. doi: 10.1080/17405629.2011.590668

Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. *Psychology of Learning and Motivation: Advances in Research and Theory, Vol 44, 44*, 145-199.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology-General, 128*(3), 309-331. doi: 10.1037/0096-3445.128.3.309

Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion, 7*(2), 336-353. doi: 10.1037/1528-3542.7.2.336

Fabel, K., Wolf, S. A., Ehninger, D., Babu, H., Leal-Galicia, P., & Kempermann, G. (2009). Additive effects of physical exercise and environmental enrichment on adult hippocampal neurogenesis in mice. *Frontiers in neuroscience, 3*, 50-50. doi: 10.3389/neuro.22.002.2009

Forte, R., Boreham, C. A. G., Leite, J. C., De Vito, G., Brennan, L., Gibney, E. R., & Pesce, C. (2013). Enhancing cognitive functioning in the elderly: multicomponent vs resistance training. *Clinical Interventions in Aging, 8*, 19-27. doi: 10.2147/cia.s36514

Gottfredson, L. S., & Deary, I. J. (2004). Intelligence predicts health and longevity, but why? *Current Directions in Psychological Science, 13*(1), 1-4. doi: 10.1111/j.0963-7214.2004.01301001.x

Green, C. S., Sugarman, M. A., Medford, K., Klobusicky, E., & Bavelier, D. (2012). The effect of action video game experience on task-switching. *Computers in Human Behavior, 28*(3), 984-994. doi: 10.1016/j.chb.2011.12.020

Hasher, L., Stolzfus, E. R., Zacks, R. T., & Rypma, B. (1991). Age and inhibition. *Journal of Experimental Psychology-Learning Memory and Cognition, 17*(1), 163-169. doi: 10.1037//0278-7393.17.1.163

Hazy, T. E., Frank, M. J., & O'Reilly, R. C. (2006). Banishing the homunculus: Making working memory work. *Neuroscience, 139*(1), 105-118. doi: 10.1016/j.neuroscience.2005.04.067

Hertzog, C., Kramer, A. F., Wilson, R. S., & Lindenberger, U. (2009). Enrichment effects on adult cognitive development: Can the functional capacity of older adults be preserved and enhanced? *Psychological Science in the Public Interest, 9*(1), 1-65. doi: 0.1111/j.1539-6053.2009.01034.x

Hopko, D. R., Crittendon, J. A., Grant, E., & Wilson, S. A. (2005). The impact of anxiety on performance IQ. *Anxiety Stress and Coping, 18*(1), 17-35. doi: 10.1080/10615800412336436

Horn, J. L. (1965). An empirical comparison of methods for estimating factor scores. *Educational and Psychological Measurement, 25*. doi: 10.1177/001316446502500202

Irwing, P., Hamza, A., Khaleefa, O., & Lynn, R. (2008). Effects of Abacus training on the intelligence of Sudanese children. *Personality and Individual Differences, 45*(7), 694-696. doi: 10.1016/j.paid.2008.06.011

Jackson, J. J., Hill, P. L., Payne, B. R., Roberts, B. W., & Stine-Morrow, E. A. L. (2012). Can An

Old Dog Learn (and Want To Experience) New Tricks? Cognitive Training Increases Openness to Experience in Older Adults. *Psychology and Aging, 27*(2), 286-292. doi: 10.1037/a0025918

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

Jaeggi, S. M., Studer-Luethi, B., Buschkuehl, M., Su, Y.-F., Jonides, J., & Perrig, W. J. (2010). The relationship between n-back performance and matrix reasoning - implications for training and transfer. *Intelligence, 38*(6), 625-635. doi: 10.1016/j.intell.2010.09.001

Jausovec, N., & Jausovec, K. (2012). Sex differences in mental rotation and cortical activation patterns: Can training change them? *Intelligence, 40*(2), 151-162. doi: 10.1016/j.intell.2012.01.005

Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport: Praeger.

Jobe, J. B., Smith, D. M., Ball, K., Tennstedt, S. L., Marsiske, M., Willis, S. L., . . . Kleinman, K. (2001). ACTIVE: A cognitive intervention trial to promote independence in older adults. *Controlled Clinical Trials, 22*(4), 453-479. doi: 10.1016/s0197-2456(01)00139-8

Jonides, J. (2004). How does practice makes perfect? *Nature Neuroscience, 7*(1), 10-11. doi: 10.1038/nn0104-10

Kane, M. J., Poole, B. J., Tuholski, S. W., & Engle, R. W. (2006). Working memory capacity and the top-down control of visual search: Exploring the boundaries of "executive attention". *Journal of Experimental Psychology-Learning Memory and Cognition, 32*(4), 749-777. doi: 10.1037/0278-7393.32.4.749

Karbach, J., & Kray, J. (2009). How useful is executive control training? Age differences in near and far transfer of task-switching training. *Developmental Science, 12*(6), 978-990. doi: 10.1111/j.1467-7687.2009.00846.x

Kempermann, G. (2008). The neurogenic reserve hypothesis: What is adult hippocampal neurogenesis good for? *Trends in Neurosciences, 31*(4), 163-169. doi: 10.1016/j.tins.2008.01.002

Kline, R. B. (1998). *Principles and practice of structural equation modeling*. New York: The Guilford Press.

Klingberg, T. (2010). Training and plasticity of working memory. *Trends in Cognitive Sciences, 14*(7), 317-324. doi: 10.1016/j.tics.2010.05.002

Klingberg, T., Fernell, E., Olesen, P. J., Johnson, M., Gustafsson, P., Dahlstrom, K., . . . Westerberg, H. (2005). Computerized training of working memory in children with ADHD - A randomized, controlled trial. *Journal of the American Academy of Child and Adolescent Psychiatry, 44*(2), 177-186. doi: 10.1097/00004583-200502000-00010

Kramer, A. F., Larish, J. F., & Strayer, D. L. (1995). Training for attentional control in dual-task settings - A comparison of young and old adults. *Journal of Experimental Psychology-Applied, 1*(1), 50-76. doi: 10.1037/1076-898X.1.1.50

Kray, J., Karbach, J., Haenig, S., & Freitag, C. (2012). Can task-switching training enhance executive control functioning in children with attention deficit/-hyperactivity disorder? *Frontiers in Human Neuroscience, 5*. doi: 10.3389/fnhum.2011.00180

Kray J. & Lindenberger, U. (2000). Adult age differences in task switching. *Psychology and Aging, 15*(1), 126-147. doi: 10.1037/0882-7974.15.1.126

Kuwajima, M., & Sawaguchi, T. (2010). Similar prefrontal cortical activities between general fluid intelligence and visuospatial working memory tasks in preschool children as revealed by optical topography. *Experimental Brain Research, 206*(4), 381-397. doi: 10.1007/s00221-010-2415-z

Kyllonen, P., & Christal, R. (1990). Reasoning ability is (little more than) working-memory capacity? *Intelligence, 14*(4), 389-433. doi: 10.1016/S0160-2896(05)80012-1

Kühn, S., Schmiedek, F., Noack, H., Wenger, E., Bodammer, N. C., Lindenberger, U., & Lövden, M.

(2013). The dynamics of change in striatal activity following updating training. *Hum Brain Mapp, 34*(7), 1530-1541. doi: 10.1002/hbm.22007

Lee, H., Boot, W. R., Basak, C., Voss, M. W., Prakash, R. S., Neider, M., . . . Kramer, A. F. (2012). Performance gains from directed training do not transfer to untrained tasks. *Acta Psychologica, 139*(1), 146-158. doi: 10.1016/j.actpsy.2011.11.003

Li, S.-C., Schmiedek, F., Huxhold, O., Roecke, C., Smith, J., & Lindenberger, U. (2008). Working Memory Plasticity in Old Age: Practice Gain, Transfer, and Maintenance. *Psychology and Aging, 23*(4), 731-742. doi: 10.1037/a0014343

Little, T. D., Lindenberger, U., & Nesselroade, J. R. (1999). On selecting indicators for multivariate measurement and modeling with latent variables: When "good" indicators are bad and "bad" indicators are good. *Psychological Methods, 4*(2), 192-211. doi: 10.1037//1082-989x.4.2.192

Lussier, M., Gagnon, C., & Bherer, L. (2012). An investigation of response and stimulus modality transfer effects after dual-task training in younger and older. *Frontiers in Human Neuroscience, 6*. doi: 10.3389/fnhum.2012.00129

Lustig, C., & Flegal, K. E. (2008). Targeting Latent Function: Encouraging Effective Encoding for Successful Memory Training and Transfer. *Psychology and Aging, 23*(4), 754-764. doi: 10.1037/a0014295

Lustig, C., Shah, P., Seidler, R., & Reuter-Lorenz, P. A. (2009). Aging, training, and the brain: A review and future directions. *Neuropsychology Review, 19*(4), 504-522. doi: 10.1007/s11065-009-9119-9

Lövdén, M., Bäckman, L., Lindenberger, U., Schaefer, S., & Schmiedek, F. (2010). A theoretical framework for the study of adult cognitive plasticity. *Psychological Bulletin, 136*(4), 659-676. doi: 10.1037/a0020080

Lövdén M, Schaefer S, Noack H, Bodammer NC, Kühn S, Heinze HJ, Düzel E, Bäckman L., & Lindenberger U, (2012). Spatial navigation training protects the hippocampus against age-related changes during early and late adulthood. *Neurobiology of Aging, 33*, 620.e9- 620.e22.

MacCallum, R. C., Widaman, K. F., Zhang, S. B., & Hong, S. H. (1999). Sample size in factor analysis. *Psychological Methods, 4*(1), 84-99. doi: 10.1037//1082-989x.4.1.84

Mackay-Brandt, A. (2011). Training attentional control in older adults. *Neuropsychology, development, and cognition. Section B, Aging, neuropsychology and cognition, 18*(4), 432-451. doi: 10.1080/13825585.2011.568046

Mackey, A. P., Hill, S. S., Stone, S. I., & Bunge, S. A. (2011). Differential effects of reasoning and speed training in children. *Developmental Science, 14*(3), 582-590. doi: 10.1111/j.1467-7687.2010.01005.x

Maillot, P., Perrot, A., & Hartley, A. (2012). Effects of Interactive Physical-Activity Video-Game Training on Physical and Cognitive Function in Older Adults. *Psychology and Aging, 27*(3), 589-600. doi: 10.1037/a0026268

Martensson, J., & Lövdén, M. (2011). Do intensive studies of a foreign language improve associative memory performance? *Frontiers in psychology, 2*, 12-12.

McArdle, J., J, & Nesselroade, J., R. (1994). Using multivariate data to structure developmental change. In S. Reese, H. & H. Cohen, W. (Eds.), *Life-span developmental psychology: Methodological contributions*. Hillsdale, NJ: Erlbaum.

McArdle, J. J. (2007). Five steps in the structural factor analysis of longitudinal data. In R. Cudeck & R. McCallum (Eds.), *Factor analysis at 100 years*. Mahwah, NJ: Erlbaum.

McArdle, J. J. (2009). Latent variable modelling of differences and changes with longitudinal data. *Annual Review of Psychology, 60*, 577-605. doi: 10.1146/annurev.psych.60.110707.163612

McArdle, J. J., & Prindle, J. J. (2008). A Latent Change Score Analysis of a Randomized Clinical Trial in Reasoning Training. *Psychology and Aging, 23*(4), 702-719. doi: 10.1037/a0014349

McDougall, S., & House, B. (2012). Brain training in older adults: Evidence of transfer to memory span performance and pseudo-Matthew effects. *Aging Neuropsychology and Cognition, 19*(1-2), 195-221. doi: 10.1080/13825585.2011.640656

McNab, F., Varrone, A., Farde, L., Jucaite, A., Bystritsky, P., Forssberg, H., & Klingberg, T. (2009). Changes in Cortical Dopamine D1 Receptor Binding Associated with Cognitive Training. *Science, 323*(5915), 800-802. doi: 10.1126/science.1166102

Melby-Lervag, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental psychology, 49*(2), 270-291. doi: 10.1037/a0028228

Meredith, W., & Horn, J. (2001). The role of factorial invariance in modeling growth and change. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change. Decade of behavior.* (pp. 203-240). Washington, DC: American Psychological Association.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*(1), 49-100. doi: 10.1006/cogp.1999.0734

Molenaar, P. C. M., & Campbell, C. G. (2009). The new person-specific paradigm in psychology. Current Directions in Psychological Science, 18, 112-117. doi: 10.1111/j.1467-8721.2009.01619.x

Moody, D. E. (2009). Can intelligence be increased by training on a task of working memory? *Intelligence, 37*(4), 327-328. doi: 10.1016/j.intell.2009.04.005

Moreno, S., Bialystok, E., Barac, R., Schellenberg, E. G., Cepeda, N. J., & Chau, T. (2011). Short-Term Music Training Enhances Verbal Intelligence and Executive Function. *Psychological Science, 22*(11), 1425-1433. doi: 10.1177/0956797611416999

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. doi: 10.3758/s13423-010-0034-0

Noack, H., Lövdén, M., Schmiedek, F., & Lindenberger, U. (2009). Cognitive plasticity in adulthood and old age: gauging the generality of cognitive intervention effects. *Restorative Neurology and Neuroscience, 27*(5), 435-453. doi: 10.3233/RNN-2009-0496

Noack, H., Lövdén, M., Schmiedek, F., & Lindenberger, U. (2013). Age-related differences in temporal and spatial dimensions of episodic memory performance before and after hundred days of practice. *Psychol Aging, 28*(2), 467-480. doi: 10.1037/a0031489

Nouchi, R., Taki, Y., Takeuchi, H., Hashizume, H., Nozawa, T., Kambara, T., . . . Kawashima, R. (2013). Brain training game boosts executive functions, working memory and processing speed in the young adults: a randomized controlled trial. *PloS one, 8*(2), e55518-e55518. doi: 10.1371/journal.pone.0055518

Olesen, P. J., Westerberg, H., & Klingberg, T. (2004). Increased prefrontal and parietal activity after training of working memory. *Nature Neuroscience, 7*(1), 75-79. doi: 10.1038/nn1165

Owen, A. M., Hampshire, A., Grahn, J. A., Stenton, R., Dajani, S., Burns, A. S., . . . Ballard, C. G. (2010). Putting brain training to the test. *Nature, 465*(7299), 775-U776. doi: 10.1038/nature09042

Peng, H., Wen, J., Wang, D., & Gao, Y. (2012). The Impact of Processing Speed Training on Working Memory in Old Adults. *Journal of Adult Development, 19*(3), 150-157. doi: 10.1007/s10804-012-9142-6

Penner, I.-K., Vogt, A., Stoecklin, M., Gschwind, L., Opwis, K., & Calabrese, P. (2012). Computerised working memory training in healthy adults: A comparison of two different training schedules. *Neuropsychological Rehabilitation, 22*(5), 716-733. doi: 10.1080/09602011.2012.686883

Pereira, A. C., Huddleston, D. E., Brickman, A. M., Sosunov, A. A., Hen, R., McKhann, G. M., . . . Small, S. A. (2007). An in vivo correlate of exercise-induced neurogenesis in the adult dentate gyrus. *Proceedings of the National Academy of Sciences of the United States of America, 104*(13), 5638-5643. doi: 10.1073/pnas.0611721104

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

Revelle, W. (1993). Individual differences in personality and motivation: 'Non-cognitive' determinants of cognitive performance. In A. Baddeley & L. Weiskrantz (Eds.), *Attention: Selection, awareness, and control: A tribute to Donald Broadbent*. Oxford: Oxford University Press.

Richmond, L. L., Morrison, A. B., Chein, J. M., & Olson, I. R. (2011). Working Memory Training and Transfer in Older Adults. *Psychology and Aging, 26*(4), 813-822. doi: 10.1037/a0023631

Robinson, W. S. (2009). Ecological Correlations and the Behavior of Individuals. *International Journal of Epidemiology, 38*(2), 337-341. doi: 10.1093/ije/dyn357

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*. doi: 10.3389/fnhum.2012.00166

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review, 103*(3), 403-428. doi: 10.1037/0033-295x.103.3.403

Sanchez, C. A. (2012). Enhancing visuospatial performance through video game training to increase learning in visuospatial science domains. *Psychonomic Bulletin & Review, 19*(1), 58-65. doi: 10.3758/s13423-011-0177-7

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

Shipstead, Z., Redick, T. S., & Engle, R. W. (2010). Does working memory training generalize? *Psychologica Belgica, 50*(3-4), 245-276.

Shipstead, Z., Redick, T. S., & Engle, R. W. (2012). Is working memory training effective?

Psychological Bulletin, 138(4), 628-654. doi: 10.1037/a0027473

Simpson, T., Camfield, D., Pipingas, A., Macpherson, H., & Stough, C. (2012). Improved processing speed: Online computer-based cognitive training in older adults. *Educational Gerontology, 38*(7), 445-458. doi: 10.1080/03601277.2011.559858

Soderqvist, S., Nutley, S. B., Ottersen, J., Grill, K. M., & Klingberg, T. (2012). Computerized training of non-verbal reasoning and working memory in children with intellectual disability. *Frontiers in Human Neuroscience, 6*. doi: 10.3389/fnhum.2012.00271

Stern, Y., Blumen, H. M., Rich, L. W., Richards, A., Herzberg, G., & Gopher, D. (2011). Space Fortress game training and executive control in older adults: a pilot intervention. *Neuropsychology, development, and cognition. Section B, Aging, neuropsychology and cognition, 18*(6), 653-677. doi: 10.1080/13825585.2011.613450

Sternberg, R. J. (2008). Increasing fluid intelligence is possible after all. *Proceedings of the National Academy of Sciences of the United States of America, 105*(19), 6791-6792. doi: 10.1073/pnas.0803396105

Studer-Luethi, B., Jaeggi, S. M., Buschkuehl, M., & Perrig, W. J. (2012). Influence of neuroticism and conscientiousness on working memory training outcome. *Personality and Individual Differences, 53*(1), 44-49. doi: 10.1016/j.paid.2012.02.012

Thorell, L. B., Lindqvist, S., Nutley, S. B., Bohlin, G., & Klingberg, T. (2009). Training and transfer effects of executive functions in preschool children. *Developmental Science, 12*(1), 106-113. doi: 10.1111/j.1467-7687.2008.00745.x

Thorndike, E. L. (1906). *Principles of teaching.* New York: Seiler.

Thorndike, E. L., & Woodworth, R. S. (1901). The influence of improvement in one mental function upon the efficiency of other functions. (I). *Psychological Review, 8*, 247-261. doi: 10.1037/h0074898

Tranter, L. J., & Koutstaal, W. (2008). Age and flexible thinking: An experimental demonstration of the beneficial effects of increased cognitively stimulating activity on fluid intelligence in healthy older adults. *Aging Neuropsychology and Cognition, 15*(2), 184-207. doi: 10.1080/13825580701322163

van Praag, H., Kempermann, G., & Gage, F. H. (2000). Neural consequences of environmental enrichment. *Nature Reviews Neuroscience, 1*(3), 191-198. doi: 10.1038/35044558

Voelcker-Rehage, C., Godde, B., & Staudinger, U. M. (2011). Cardiovascular and coordination training differentially improve cognitive performance and neural processing in older adults. *Frontiers in Human Neuroscience, 5*. doi: 10.3389/fnhum.2011.00026

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

Voss, M. W., Prakash, R. S., Erickson, K. I., Basak, C., Chaddock, L., Kim, J. S., . . . Kramer, A. F. (2010). Plasticity of brain networks in a randomized intervention trial of exercise training in older adults. *Frontiers in Aging Neuroscience, 2*. doi: 10.3389/fnagi.2010.00032

Wang, M.-Y., Chang, C.-Y., & Su, S.-Y. (2011). What's Cooking? - Cognitive Training of Executive Function in the Elderly. *Frontiers in psychology, 2*, 228-228.

Wen, M.-C., Butler, L. T., & Koutstaal, W. (2013). Improving insight and non-insight problem solving with brief interventions. *British Journal of Psychology, 104*, 97-118. doi: 10.1111/j.2044-8295.2012.02107.x

Wenger, E., Schaefer, S., Noack, H., Kühn, S., Mårtensson, J., Heinze, H. J., . . . Lövdén, M. (2012). Cortical thickness changes following spatial navigation training in adulthood and aging. *Neuroimage, 59*(4), 3389-3397. doi: 10.1016/j.neuroimage.2011.11.015

West, R. L. (1996). An application of prefrontal cortex function theory to cognitive aging. *Psychological Bulletin, 120*(2), 272-292. doi: 10.1037/0033-2909.120.2.272

Zelinski, E. M. (2009). Far transfer in cognitive training of older adults. *Restorative Neurology and Neuroscience, 27*(5), 455-471. doi: 10.3233/rnn-2009-0495

Zinke, K., Einert, M., Pfennig, L., & Kliegel, M. (2012). Plasticity of executive control through task switching training in adolescents. *Frontiers in Human Neuroscience*, 6. doi: 10.3389/fnhum.2012

Figure Captions

1. Histogram of average group sizes in the study sample. Three studies were omitted from the plot for better visualization (Irwing et al., 2008: n = 1246; McArdle & Prindle, 2008: n = 699; Owen et al., 2010: n = 3810).
2. Histogram of reported training durations. Three studies spanned more than 100h of training (Degé et al., 2011: ~400h; Schmiedek et al., 2010: ~ 116h; Voelcker-Rehage et al., 2011: 144h). These studies were omitted in the plot for better visualization.

Figure 1

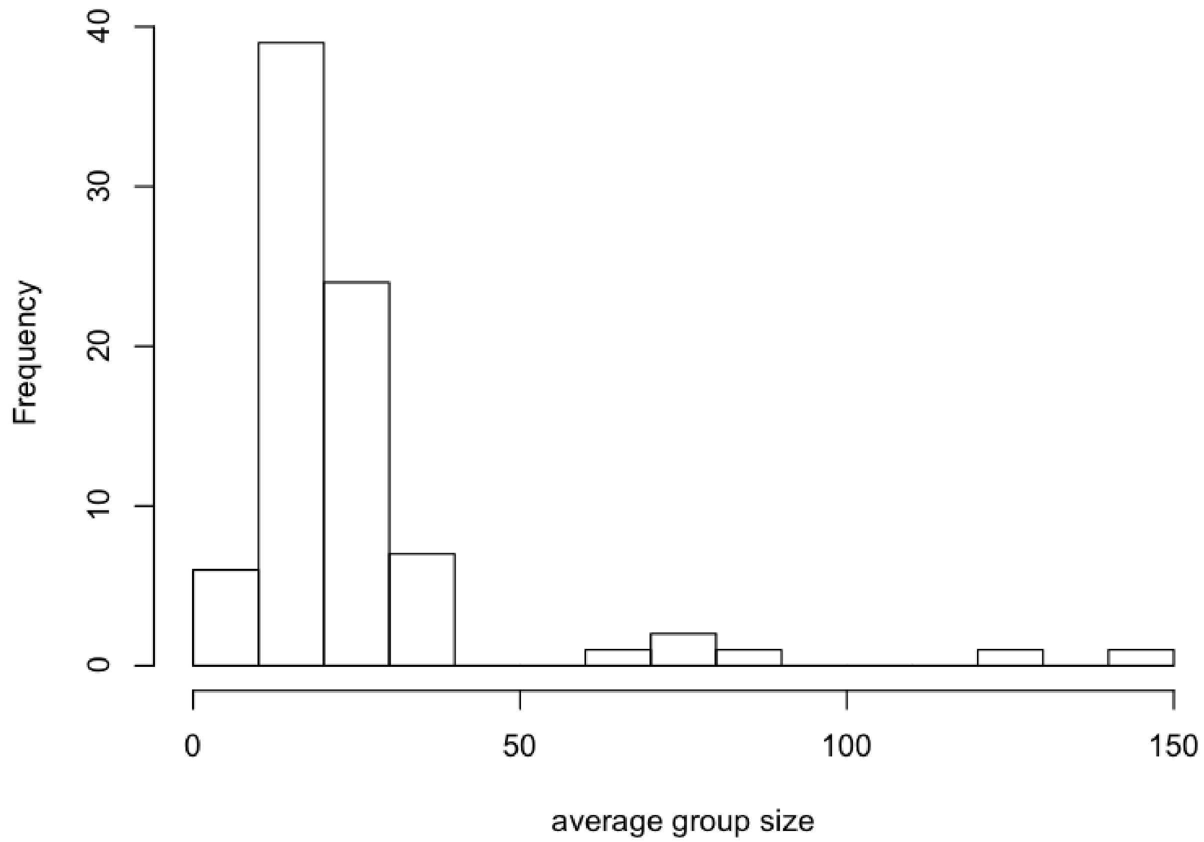

Figure 2