

MASARYK UNIVERSITY  
Faculty of Social Studies  
Department of psychology

Mgr. Vladimír Marček

**Effectiveness of n-back cognitive training:  
quantitative and qualitative aspects**

Dissertation

Supervisor: prof. PhDr. Tomáš Urbánek, Ph.D.

Brno 2014

## **Abstract**

Cognitive training has recently become a primary topic of interest in cognitive psychology. The discovery of a strong relationship between WM and *Gf* gave rise to new cognitive training methods (like dual n-back task), which challenged traditional views of intelligence as a fixed trait in healthy adults. Previous research has shown mixed results in the ability of cognitive training to improve fluid intelligence. Presented dissertation aims to first replicate such effects in a study with (N=142) participants, and then to explore the mediating role of personality systems interaction (PSI) personality factors. In addition, univariate and bivariate analyses of two n-back related, self-report questionnaires (N=258 and N=97) are presented. Experimental results showed improvements in one out of two IQ test scores, which reflects the ambivalent nature of previous research in this field. After examining the results in context of PSI theory, it was found that different training methods yielded different IQ gains in participants, depending on their personality styles. In addition, these correlations suggested a meaningful pattern, indicating that PSI theory may be able to account for the different outcomes of cognitive training studies. Analysis of self-report questionnaires suggests, among other things, that the use of mental strategies during n-back training does not influence prospective IQ gains, and neither does the motivation to participate in n-back study. Qualitative reports complement these findings by offering unique insights into the subjective experiences of people who trained n-back. The presented findings may facilitate tailor-made cognitive training interventions in the future, and can contribute to explaining the mechanisms underlying the far-transfer of working memory training to fluid intelligence.

## Acknowledgements

I feel honored to publicly thank those, who supported me throughout my doctoral studies and the writing of this dissertation. First and foremost, my gratitude goes to my professional mentors prof. PhDr. Tomáš Urbánek, Ph.D., and PhDr. Anton Polák. Without their diligent, professional and personal support, this endeavor (and many others of mine) would be impossible to accomplish.

The same goes to my family: my terrific parents Vierka and Vladimír, gorgeous sisters Vierka and Kristínka, and to my dearest partner, Jarka. Thank you for being there for me, believing in me and supporting me in countless ways to become who I am today.

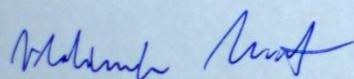
My gratitude extends towards all those, who contributed to our research projects: hundreds of dedicated participants, assistants Martin, Lenka, Clair, Jarka and Jake – without their help I would probably still be transcribing those 200,000 test items into a computer.

As we live in the age of internet, let me thank to all those anonymous individuals, who generously devoted hours of their precious time to responding to my questionnaires, and kept online discussions inspiring and respectful.

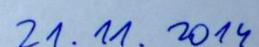
Lastly, in the spirit of science and humanism, let me express gratitude to all the people who are advancing our knowledge of ourselves and the world around us, and do so in a nice way.

## Disclaimer

I hereby declare that this dissertation is my own original work, and that I have acknowledged all sources used and have cited these in the reference section.



Vladimír Marček



Date

# Table of content

Introduction .....	1
Chapter I. Theoretical background .....	2
1.01 Working memory .....	2
1.02 Intelligence .....	3
1.03 Relationship between WM and fluid intelligence .....	4
1.04 WM training studies administering n-back task and Gf measures .....	6
(a) Present research.....	7
(b) The nature of cognitive improvements after CT .....	8
Chapter II. Experimental part .....	9
2.01 Method.....	9
(a) Participants .....	9
(b) Fluid intelligence measures .....	9
(c) Personality measures .....	10
(d) Training tasks .....	10
(e) Procedures .....	12
2.02 Results .....	12
2.03 Discussion .....	15
(a) Mediating role of PSI personality factors in CT outcomes .....	15
(b) Mediating role of anxiety in CT outcomes.....	18
(c) Limitations.....	19
2.04 Conclusions .....	19
Chapter III. BRAWO questionnaire.....	20
3.01 About.....	20
3.02 Preparation and limitations.....	20
3.03 Univariate analysis .....	22
(a) Timestamps .....	22
(b) Age .....	22
(c) The amount of n-back training .....	23
(d) Version of n-back exercise .....	27
(e) Level of n-back exercise.....	28
(f) Intelligence quotient .....	31
(g) IQ mindset (Do you think IQ can be improved?) .....	33
(h) Drug use (and abuse) .....	34
(i) Self-reported effects of n-back training (categories) .....	36
(j) Self-reported effects of n-back training (qualitative aspects) .....	39

3.04	Bivariate analysis .....	45
(a)	Multiple comparison problem .....	45
(b)	H1: There is a directly proportional relationship between <i>Total training time</i> and <i>Self-reported improvement of IQ</i> .....	48
(c)	H2: There is a directly proportional relationship between <i>Total training time</i> and <i>Self-reported improvement of mood</i> .....	48
(d)	H3: There is a directly proportional relationship between Self-reported IQ and <i>N-back level</i> .....	49
(e)	H4: There is a directly proportional relationship between <i>N-back level</i> and <i>Self-reported improvement of IQ</i> .....	50
(f)	H5: There is not a proportional relationship between Self-reported IQ and IQ Mindset .....	51
(g)	Discussion .....	51
Chapter IV.	BIQ questionnaire.....	53
4.01	About .....	53
4.02	Preparation and limitations.....	53
4.03	Univariate analysis .....	54
(a)	Motivation to participate .....	55
(b)	Self-reported positive effects of n-back training (quantitative summary).....	55
(c)	Self-reported negative effects of n-back training (quantitative summary) .....	57
(d)	Self-reported difficulty of training .....	59
(e)	Use of mental strategies in n-back training .....	61
(f)	Self-reported changes after n-back training (qualitative content) .....	63
4.04	Bivariate analysis .....	65
(a)	Multiple comparison problem .....	65
(b)	H6: <i>Intrinsic motivation</i> to participate in n-back training is in directly proportional relationship with <i>IQ gains</i> .....	66
(c)	H7: The use of different <i>Mental strategies</i> in n-back training yields different <i>IQ gains</i> in both RAPM and BOMAT .....	67
(d)	H8: Reports of <i>Subjectively improved thinking</i> are in directly proportional relationship to psychometric <i>IQ gains</i> in both RAPM and BOMAT .....	68
(e)	Discussion .....	69
Chapter V.	Summary .....	71
5.01	Conclusion.....	73
Chapter VI.	References .....	74

## List of attachments (tables / figures)

Table 1. Sample characteristics .....	9
Table 2. Correlations of intelligence scores and personality scales in pretest .....	13
Table 3. Correlational analyses of PSDI scales and IQ score changes (BOMAT) for each of the possible cognitive training methods (see legend for description). ....	14
Table 4. Descriptive statistics - Age variable .....	22
Table 5. Descriptive statistics - regular training schedule.....	24
Table 6. Descriptive statistics - Minutes per session.....	24
Table 7. Descriptive statistics - Weekly training frequency .....	25
Table 8. Descriptive statistics - Partial aggregate time .....	26
Table 9. Descriptive statistics - N-back version .....	27
Table 10. Descriptive statistics - Usual training level.....	29
Table 11. Descriptive Statistics - aggregateBackLevel.....	30
Table 12. Descriptive statistics - IQ estimate .....	32
Table 13. Descriptive statistics - IQ malleability mindset.....	34
Table 14. Descriptive statistics - Drug use (and abuse).....	35
Table 15. Descriptive statistics - Self-reported positive effects of n-back training.....	55
Table 16. Descriptive statistics - Self-reported negative effects of n-back training. ....	58
Table 17. Descriptive statistics - How hard is it to start n-back training daily? .....	60
Table 18. Descriptive statistics - How hard was actual training? .....	60
Table 19. Descriptive statistics - Use of mental strategies during n-back training .....	61
Table 20. Correlations of IQ gain and Strategy use during n-back.....	68
Figure 1. An example of the instructions for the adaptive mental rotation task.....	11
Figure 2. An example of possible response options in the adaptive mental rotation task.....	12
Figure 3. The STAR model of personality dimensions by Kuhl & Kazén (1997).....	17
Figure 4. Histogram and density trace - Age variable .....	22
Figure 5. Bar of counts – regular training schedule .....	24
Figure 6. Bar of counts - Minutes per session .....	25
Figure 7. Histogram and density trace - Partial aggregate time .....	26
Figure 8. Bar of counts - Usual training level.....	28
Figure 9. Histogram - AggregateBackLevel.....	31
Figure 10. Histogram and density trace - IQ Estimate.....	33
Figure 11. Descriptive statistics - IQ malleability mindset .....	34
Figure 12. Subjectively felt improvements after n-back training.....	38
Figure 13. Comparison of BIQ and BRAWO self-reported improvements .....	56
Figure 14. BIQ-only, self-reported improvements .....	57
Figure 15. Self-reported negative effects of n-back training .....	58
Figure 16. How hard is it to start n-back training daily? .....	60
Figure 17. Descriptive statistics - How hard was actual training?.....	60
Figure 18. Use of mental strategies during n-back training .....	62

## Glossary of Abbreviations

BIQ	Brno inquiries questionnaire
BOMAT	Bochumer Matrizentest
BRAWO	Brain workshop questionnaire
e.g.	For example
fMRI	Functional magnetic resonance imaging
<i>Gc</i>	Crystallized intelligence
<i>Gf</i>	Fluid intelligence
GAČR	Czech Science Foundation
GRE	Graduate Record Examinations
IQ	Intelligence quotient
i.e.	That is
NEO-FFI	NEO Five-Factor Inventory
<i>p</i>	P-value
PCE	Pharmacological cognitive enhancers
P-FIT	Parieto-Frontal Integration Theory
PSI	Personality systems interaction theory
PSSI	The Personality Styles and Disorder Inventory
<i>r<sub>s</sub></i>	Spearman's rank correlation coefficient
RAPM	Raven's Progressive Matrices
RM-ANOVA	Repeated measures analysis of variance
SAT	Scholastic Aptitude Test
STM	Short-term memory
tCDS	Transcranial direct current stimulation (tDCS)
TMS	Transcranial magnetic stimulation
VPR	Verbal, perceptual and image rotation
WAIS	Wechsler Adult Intelligence Scale
WM	Working Memory

# Introduction

*Almost in the beginning was curiosity.*

Isaac Asimov

**H**uman intelligence is a fascinating phenomenon – both subjectively as a part of our everyday experience, and more objectively, as a topic of scientific interest. From a scientific perspective, complexity has its downsides – it is not easy to define intelligence, to measure and explore it, and to interpret our discoveries.

A few years ago, Suzan Jaeggi and her colleagues published a landmark study on improving intelligence in a high-impact scientific journal (Jaeggi et al., 2008). Despite its shortcomings, this study jump started the research on improving intelligence with cognitive training. Within six years from its publication, it has been cited more than 800 times, and more than 30 studies explored the effect in question, using different methodologies, but the same training paradigm.

This time period corresponded well with my doctoral studies – I was able to watch all the research come out in real time, and by successfully co-applying for two research grants, with my mentor prof. Tomáš Urbánek, we also had the chance to conduct two large studies on improving intelligence ourselves. One of the outcomes of this endeavor is this dissertation. As a dissertation should substantiate one's ability to do independent research, I denote the section which was co-authored (Chapter 2) by the use of plural (pronoun “we”).

The first chapter deals with theoretical aspects of working memory and intelligence, which are relevant to cognitive training. The second chapter presents our study on improving fluid intelligence with n-back training ( $N=142$ ), interpreted in the context of PSI personality theory. In third and fourth chapter I summarize quantitative and qualitative self-reports of persons with experience in n-back training ( $N=258$  and  $N=97$ ), and explore relationships between selected variables. The fifth chapter is dedicated to a summary and conclusion.

# Chapter I. Theoretical background

## 1.01 Working memory

Working memory (WM) was firstly conceptualized by Atkinson & Shiffrin (1968) as a part of "multi-store model", which included sensory memory, short-term memory and long-term memory. Sensory (or iconic) memory was proved to store up to 16 bits of information for up to 1000 ms, Sperling (1960). Related, but different concept of short-term memory (STM) was coined by Miller (1956) in his popular article "The Magical Number Seven, Plus or Minus Two". Number of possible chunks, which can be held in it for up to 30 s was actually decreased to 4 by Cowan (2001). This number was further challenged by Gobet & Clarkson (2004), who stated that only 2 representations can be actively maintained, while McElree (2006) reduced it to only one. Oberauer (2002) proposes viable solution here – according to him, there is a "focus of attention" component, which can process only one chunk of information at a time. Nevertheless, there are several chunks in WM, and even several partly activated representations from long-term memory. Probably the most popular model of WM was introduced by Baddeley & Hitch (1974). It consists of executing component with two slave systems: phonological loop and visual-spatial sketchpad. Later, Baddeley added episodic buffer into his theory of WM, Baddeley (2000). This turned out to be a bit elusive component, responsible for episodic integration of information into multidimensional representations, with a capacity of four elements (Baddeley, 2012).

The main focus of research in this area seems to be the structure of WM, including the WM - STM relationship, and domain-specificity of WM. Current results suggest that STM is more of a storage simply maintaining information, while WM is more related to attention and is involved in possible transformations of maintained information Kane et al. (2004). Regarding domain specificity of WM, today's prevalent view is, that WM is domain specific (Shah & Miyake, 1996; Friedman & Miyake, 2000; Mackintosh & Bennett, 2003; Tillman, Nyberg & Bohlin, 2007), while verbal and visuospatial components share some common resources too – both in terms of capacity (Kane et al., 2004; Colom et al., 2006), and of executive and attentional processes (Baddeley, 2012; Nee & Jonides, 2013; Morey & Bieler, 2013).

Not contradicting mentioned WM theories, we can settle for general definition of WM: it is a cognitive system responsible for actively maintaining and updating task-relevant information, it is related to attention, influenced by interference, and has a limited overall capacity.

## 1.02 Intelligence

Human intelligence is one of the main subjects of current research in cognitive psychology. Origins of this research actually overlap with origins of psychology as a scientific discipline. For a brief summary of research in psychometric intelligence, let us start more than a century ago with Spearman (1904), saying that there is a common „g-factor“ behind different measures of intellectual performance. Thurstone (1938) stated an alternative hypothesis about seven „primary mental abilities“. Guilford (1967) extended the number of factors composing human intellect to more than 120 (three dimensions per several components, interacting with each other). In this work, we are going to use Cattell's (1971), now broadly accepted theory of intelligence. He was the first to propose a hierarchical model, which divided g-factor into *fluid (Gf) and crystallized (Gc) intelligence*. The former

- accounts for speed and precision of abstract thinking particularly in novel situations
- can be measured with tests like analogies or series completion
- its measures were shown to be quite resistant to environmental influences except aging (and pathology), see Salthouse (2006) for comprehensive summary.

*Gc* derives from *Gf* over time, and can be described as an ability to use already acquired, long-term knowledge, experiences and vocabulary. Cattell's dichotomy was further developed in the „*three-stratum*“ model of intelligence (Carroll, 1993), which preserves the *Gf - Gc* discrimination, and adds another layer with more specialized components.

Besides CHC (*Cattell-Horn-Carrol*) theory, there are additional contemporary theories of intelligence, relevant to cognitive training research. E.g., Jung & Haier (2007) authored a neuroanatomical theory of intelligence – *Parieto-Frontal Integration Theory (P-FIT)*. Based on analysis of tens of neuroimaging studies, it proposes that several brain regions are responsible for processing cognitive information, which takes place in four distinct stages: processing of sensory information; integration and abstraction; problem solving, evaluation, hypothesis testing; response inhibition / selection. P-FIT theory is useful not only as a framework for examining intelligence, but WM too. E.g., Haier et al. (2003) states, that according to fMRI data it is precisely fronto-parietal network that mediates the correlation between *Gf* and 3-back task. P-FIT theory was further supported by studies examining brain networks, which are shared between intelligence and WM (Colom et al., 2009), sometimes even using n-back task Barbey et al. (2014).

Considering more psychometrically oriented theories of intelligence, interesting example can be *VPR* (*Verbal, perceptual and image rotation*) model, authored by Johnson and Bouchard (Johnson & Bouchard, 2005a, 2005b; Johnson, te Nijenhuis, & Bouchard, 2007). They reanalyzed data from 436 adults who completed 42 mental ability tests, and compared statistical performance of their VPR model with CHC and Carroll's three-stratum theory. VPR model was the best theoretical fit for given data set. The interesting thing is, that this not-so-famous ability of mental image rotation came out to be very important in the model. Authors elaborate further on this ability, and conclude that "mental image rotation tasks have not been given the attention they deserve as important and relatively independent contributors to the manifestation of human intelligence." (p. 409), and that "spatial image rotation ability is highly relevant to the overall structure of human intellect" (p. 414).

And there are still more theories of intelligence, which can be of interest for cognitive training researchers. E.g. *Dual-Process theory* (Davidson & Kemp, 2011; Kaufman, 2009, 2011, 2013), which builds on differentiation between implicit and explicit cognition, eventually recognizes two types of cognitive processes. One of them is a *goal-oriented cognition*, which consists of intentional thinking and reflecting. All the processes in this category (e.g. long-term planning, inhibition, self-regulation, perseverance and other factors) compete for a limited pool of attentional resources, and require the use of WM. On the other hand, there is *spontaneous cognition*, which is characterized by automatic acquiring and processing of information (e.g. in form of implicit learning, mind-wandering or expert intuition) – this processes are not dependent on input from higher-level cognitive processes like attention or WM, Stanovich & Toplak (2012). For up-to-date discussion on Dual-Process theory see Evans & Stanovich (2013).

### 1.03 Relationship between WM and fluid intelligence

Basic idea here is, that when Gc is by definition related to long-term memory, could Gf be significantly correlated to WM/STM? One of the first studies exploring this relationship was Larson & Saccuzzo (1989). They found a 0.5 correlation of WM performance task with Raven's Advanced Progressive Matrices (RAPM) scores. Kyllonen & Christal (1990) in their large study with two experiments (nearly 400 participants each) found several correlations up to 0.5 between WM and reasoning tasks. Thirdly, there is Stankov & Crawford (1993) who did find correlations with Gf measurements starting at 0.2 after easier version of WM tasks, up to 0.46 when participants did three of four exchanges in sequence of letters mentally in a

row. These studies were the first to suggest, that the ability to keep the results of one task accessible while performing another task, brings significant correlations with Gf.

In current decade of research, Engle et al. (1999) in a study with 133 participants finds controlled correlation WM - Gf of 0.49 and argue, that WM shows a strong connection to Gf. Süss et al. (2002) administered a battery of 17 different WM tasks to 128 young adults, together with an intelligence test. They conclude that „Working-memory capacity is highly related to intelligence. The strongest relationship was found to reasoning ability. ...The common variance between the WM and the reasoning tests... can hardly be attributed to something other than WM capacity...“ (p. 261). Colom et al. (2003) did study with 187 participants, eight computerized WM tasks and two types of Gf measures: „The results show that WM can be considered as one general cognitive resource, and that this resource is strongly related with intelligence ( $r = +0.7$ ).“ (p. 33)

These quite explicit claims met an explicit response, e.g. Conway et al. (2003) state in their meta-analysis that „...WMC (WM capacity) and g are indeed highly related, but not identical.“ (p. 547) Ackerman et al. (2005) were even more critical: „The authors conducted a meta-analysis of 86 samples that relate WM to intelligence. The average correlation between true-score estimates of WM and g is substantially less than unity ( $\rho = .479$ ).“ (p. 30). There were two comments on this: Oberauer et al. (2005) stated that „reanalysis of the data ...using the correct statistical procedures demonstrates that g and WM capacity are very highly correlated.“ (p. 61). Second comment came from Kane et al. (2005), who agreed, that WM capacity is not isomorphic with Gf, but „reanalysis of 14 data sets from 10 published studies, representing more than 3,100 young-adult subjects, suggests a strong correlation between WM capacity and Gf (median  $r = .72$ ), indicating that the WM capacity and Gf constructs share approximately 50% of their variance.“ Story continued with Buehner et al. (2006), who administered 20 working memory tests, 2 attention tests, and 18 intelligence subtests to 121 students. They found that WM and sustained attention together account for about 83% of variance in reasoning abilities. But after detailed discussion of their models and results they conclude: „We believe that Oberauer et al. (2005) made a reasonable estimate of how strongly WM and reasoning overlap, namely, about 70%.“ (p. 57). Colom et al. (2008) in large study with 661 participants confirm one more time that „WM and the general factor of intelligence (g) are highly related constructs.“ (p. 584).

## 1.04 WM training studies administering n-back task and Gf measures

Research on cognitive training is a swamp.

*Scott Barry Kaufman*

Cognitive training (CT) has recently become a primary topic of interest in cognitive psychology, and several lines of research have been producing consistent and encouraging results. CT has been shown to compensate for the natural aging process in healthy adults (Caretti, Borella, Zavagnin, & de Beni, 2012; Gates & Valenzuela, 2010; Kueider, Parisi, Gross, & Rebok, 2012; Reijnders, van Heugten, & van Boxtel, 2012) and to benefit individuals with schizophrenia (Subramaniam et al., 2012; Wykes et al., 2011). There are, however, some applications of CT, such as the improvement of fluid intelligence, which have regularly yielded mixed results (discussed below). One of the most striking instances of such discrepancy is for the effects of working memory (WM) training on fluid intelligence (Gf).

Jaeggi, Buschkuhl, Jonides, and Perrig (2008) claimed to have significantly improved Gf in healthy adults in only one month, by administering a WM-taxing exercise (n-back) to participants for 20 minutes per day. Their findings promptly received both positive (Sternberg, 2008) and negative (Moody, 2009) reviews, and have been scrutinized by the scientific community to this day. Such scrutiny seems justified because, remarkably, six years later there have been a comparable number of methodologically improved studies that have replicated and sometimes extended the original results (Bauernschmidt, Conway, & Pisoni, 2009; Colom et al., 2010; Jaeggi, Buschkuhl, Jonides, & Shah, 2011; Jaeggi, Buschkuhl, Shah, & Jonides, 2014; Jausovec & Jausovec, 2012b; Karbach & Kray, 2009; Schmiedek, Lövdén, & Lindenberger, 2010; Stephenson, 2010; Wang, Zhou, & Shah, 2014), and that have found no Gf improvements in healthy adults after CT (Baniqued et al., 2013; Brehmer, Westerberg, & Bäckman, 2012; Chein & Morrison, 2010; Chooi and Thompson, 2012; Oelhafen et al., 2013; Owen et al., 2010; Pugin et al., 2014; Redick et al., 2012; Salminen, Strobach, & Schubert, 2012; Thompson et al., 2013;).

Reviews of literature have so far emphasized the inconclusiveness of results and the heterogeneity and shortcomings of different methodologies, especially regarding control groups and measurement processes, and have called for further research to uncover the mechanisms underlying far-transfer of learning in CT to Gf (Jaeggi et al., 2010; Morrison & Chein, 2011; Shipstead, Redick, & Engle, 2012). In two recent methodologically rigorous studies, Stephenson and Halpern (2013) and Colom et al. (2013) have found improvements to

several Gf measures following CT. Despite this, however, these studies have gone on to consider latent factors of Gf or to consider Gf as a construct of the tests administered, respectively, and have led to the interpretation that CT could improve visual performance rather than Gf.

(a) Present research

Lately, first metaanalyses regarding n-back training became available. Au et al. (2014) analysed 20 relevant studies and concluded, that there is “small but significant positive effect of n-back training on improving Gf.” Authors elaborate on several mediating factors as well, e.g. they found no difference in using different versions of n-back. While this may seem to contradict results described in Chapter 2 (see below), actually it does not, because personality traits are supposed to be distributed evenly in population, therefore any version of n-back used on a random sample should have comparable effect.

Although topic of this dissertation lies with population of healthy adults, let us briefly mention three metaanalyses that came out lately, all of them discovering small far-transfer effects of short-term cognitive training to general cognitive abilities in older adults. Karbach & Verhaegen (2014) reanalysed 61 independent samples and considered several mediating factors as well. They corroborated Au et al. (2014) e.g. by finding no difference regarding the use of active or passive control groups. Kelly et al. (2014) included 31 groups of healthy older adults and emphasized that more research is needed to examine the transfer of training gains to everyday functioning. Third metaanalysis was published by Lampit, Hallock & Valenzuela in an online-only, pay-to-publish model journal with not so rigorous peer review (PLOS ONE, “we will publish every technically sound paper”), and found that transfer effects in healthy older adults vary across cognitive domains, but are significant in many of them.

These studies were unfortunately covered by media in an exaggerated and misleading way, often in context of business interests. Therefore renowned authors in the field (e.g. Jaeggi, Melby-Lervåg, Oberauer, Salthouse, among many others) signed a public letter (“A Consensus on the Brain Training Industry from the Scientific Community”) on October 20<sup>th</sup>, 2014. Their main point is, that there is no scientific evidence that so-called “brain training games” could reduce or reverse cognitive decline in older adults. Although this letter represents considerable scientific consensus and is aimed towards population of older adults, it is also a message toning down exaggerated claims about general cognitive training in healthy adults.

(b) The nature of cognitive improvements after CT

Improvements in tests of fluid intelligence after CT have recently been interpreted as improvements not of Gf, but rather of “visual performance” (Colom et al., 2013; Stephenson & Halpern, 2013). I have several conceptual objections to this interpretation. Firstly, this general kind of “visual performance” may very well be what one actually aims to improve with CT, considering that several theories of intelligence recognize visuospatial abilities as an essential component of general intelligence (Wechsler, 2008; Johnson & Bouchard, 2005b). More importantly, visual performance (see e.g., the Gv factor of CHC theory; Carroll, 1993) is measured by administering visually complex stimuli (e.g., noise-obsured objects, complex patterns, fields of letters), and one task operation (e.g., change size, rotate, search for something visually). In contrast, matrix tests (such as the RAPM and BOMAT used in this study) use items consisting of visually simple stimuli, meant only to be carriers of an underlying reasoning problem containing complex transformations that must be discovered and interpreted during the task. Finally, tests of visual performance increase their difficulty by increasing the visual complexity of stimuli, whereas matrix tests increase their difficulty by increasing the number of possible and necessary logical relationships, regardless of visual complexity (indeed, some of the hardest RAPM or BOMAT items look visually simpler than the easiest ones – it is the number of potential logical relationships one has to discover and test, which makes them difficult). Taken together, visually performing a search for an object partially obscured by visual noise is substantially different from the inductive reasoning required to find the logical relationships of the RAPM and BOMAT. In addition, Jaeggi et al. (2014) documented far-transfer from an auditory n-back task to visuospatial reasoning, indicating improvements to modality-independent cognitive abilities.

## Chapter II. Experimental part

### 2.01 Method

#### (a) Participants

By using printed media ads and public social networks, we recruited 142 participants (62 males, 43.7%). Their mean age was 25.2 (SD = 6.47, range: 18–50; for details see Table 1). All participants reported no present or previous history of psychiatric illness or medication. Nine participants did not finish the study and were excluded from the final analysis (they either did not meet the required training time, or did not attend the posttest session). Participants were required to undergo nearly 3 hours of pretesting, 8 hours of training in total spread across 25 working days, and 3 hours of post-testing. For participation, each received a small financial compensation (ca. €25) and their own test results including a short interpretation.

*Table 1. Sample characteristics*

Group	N	% of males	Mean age (SD)
Single n-back (SNB)	37	43	25.7 (6.2)
Triple n-back (TNB)	31	48	24.8 (5.8)
Mental rotations 3D	31	32	24.7 (5.8)
Sudoku (control)	34	47	24.6 (7.1)

#### (b) Fluid intelligence measures

We administered Raven's Advanced Progressive Matrices, set II (RAPM; Raven, 1990) with a 40-minute time limit. RAPM contains 36 items, each composed of a  $3 \times 3$  grid of pattern elements. One element is always missing and the participant's task is to choose an option that appropriately fills in the pattern. As a second measure of fluid reasoning, we used the Bochum Matrices Test (BOMAT) devised by Hossiep, Turck, and Hasella (1999), with an 80-minute time limit. The BOMAT contains 40 items, each of which is composed of  $5 \times 3$  elements. This allows for more pattern relationships and makes the test significantly more difficult than the RAPM, on which some participants performed at ceiling. No participants reached the maximum performance ceiling on the BOMAT.

(c) Personality measures

(i) *Personality Styles and Disorders Inventory (PSDI)*

This personality inventory was developed by Kuhl and Kazén in 1997. It has 140 items and measures 14 scales of nonpathological personality styles. Each scale ranges from a minimal level and extent of the trait to a greater level of severity or intensity (e.g., “confident-dissocial,” “distrustful-paranoid,” or “conscientious-obsessive”). This is based on Personality Systems Interaction Theory (PSI; Kuhl, 2000b), which describes the relationship between cognition and personality macrosystems. Therefore, we assumed this scale would be useful to explore the interplay of cognition and personality during and after CT.

(ii) *State-Trait Anxiety Inventory (STAI)*

To measure anxiety, we administered the State-Trait Anxiety Inventory (STAI; Spielberger & Gorsuch, 1983). This is a widely accepted clinical psychometric instrument comprising 40 questions. It measures anxiety both as a current state and as a long-term personality trait.

(d) Training tasks

(i) *Adaptive single n-back task (SNB)*

The goal of the n-back task is for participants to keep track of regularly appearing stimuli and to respond when a stimulus matches one presented n number of stimuli prior. For the task to be adaptive, n has to change periodically to match the participant’s actual level of performance. In our “single” version of n-back task, participants had to keep track of one stream of stimuli: positions of squares appearing one at a time on a computer screen. For this, we used an open source software (Brain Workshop, v. 4.8.1, <http://brainworkshop.sourceforge.net/>), configured to match the parameters used by Jaeggi et al. (2008), except we nearly doubled the number of high-interference (“lure”) trials to 20% of cases, because one’s ability to cope with such trials has been shown to correlate significantly with Gf (Burgess, Gray, Conway, & Braver, 2011).

(ii) *Adaptive triple n-back task (TNB)*

We also included a “triple” version of the n-back task, using the same software and configuration as the “single” version, except that in this task participants were required to keep track of three streams of stimuli: positions of squares, colors of squares, and letters presented auditorily. This task taxes multiple sensory modalities at once: spatial (positions) and color visual abilities as well as symbolic (letters) auditory processing.

### *(iii) Adaptive mental rotations task*

Efforts to improve Gf with cognitive training have so far focused on training WM.

This has been justified because WM has been reliably linked to Gf (Colom, Abad, Quiroga, Shih, & Flores-Mendoza, 2008; Kane, Hambrick, & Conway, 2005; Oberauer, Süs, Wilhelm, & Wittmann, 2008) and has been considered to be improvable with training (Morrison & Chein, 2011; Rafi & Samsudin, 2009). Mental rotation ability (MRA), however, is another cognitive process that meets these criteria and may be an additional pathway to improved Gf. First, MRA strongly correlates with Gf scores (Kaufmann, DeYoung, Gray, Brown, & Mackintosh, 2009), and Johnson and Bouchard (2005a; 2005b) have even suggested that it is a core component of general intelligence. Second, the outcomes of MRA practice have been considered “dramatic” (Peters et al., 1995), and its trainability has recently been supported by additional study (Stransky, Wilcox, & Dubrowski, 2010; Jausovec & Jausovec, 2012a). As the transfer of MRA training to Gf seems reasonable from this perspective, we developed an adaptive 3D mental rotation task. We presented users with a random 3D object and a series of instructions for mentally rotating it along X, Y, and Z axes (Figure 1). Participants were then required to select the correct perspective of the object out of six possible alternatives (Figure 2). The number of instructions was adaptively changed based on the previous success rate.

*Figure 1. An example of the instructions for the adaptive mental rotation task*

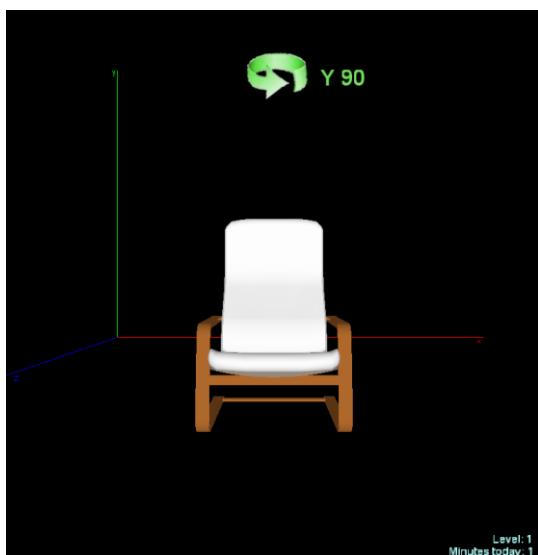
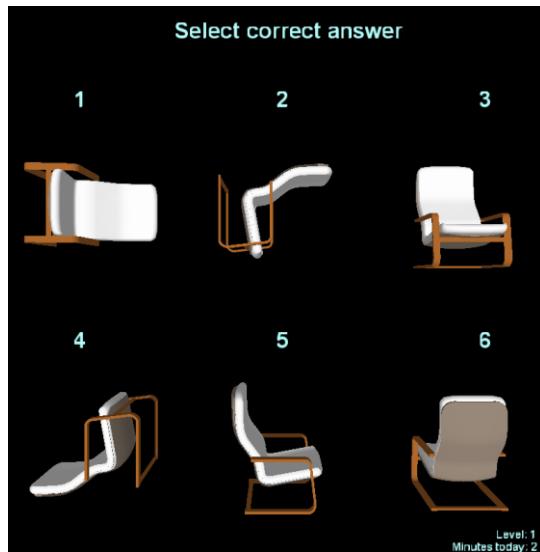


Figure 2. An example of possible response options in the adaptive mental rotation task.



(iv) *Control group task*

We used a non-adaptive, beginner-level Sudoku game as our active-contact control group training task. The goal of Sudoku is to fill each nine-square row, each nine-square column, and each nine-square box with the numbers 1 through 9, by using each number only once in each. We consider this simple version of Sudoku to be an appropriate control training task for WM training, because at this level it does not require any significant WM processing, but still involves attention, short-term memory, and some comparing and matching of stimuli.

(e) *Procedures*

After informing participants on testing and training procedures, each signed a research contract and consent form. Pretest sessions were conducted in small groups by administering either the odd- or even-numbered items of the RAPM and BOMAT in half of the standard time limit. Next, training and control activities took place on auto-timed software that had been distributed to participants (Penner et al., 2012) for 20 minutes per day, across 25 days in total. Data were automatically transmitted to us daily. Posttest sessions were conducted in the same way as the pretests, except that the complementary odd- or even-numbered versions of the RAPM and BOMAT were administered. The version of test (odd or even) was chosen randomly for each participant before the pretest.

## 2.02 Results

Our primary aim was to check for differences in RAPM and BOMAT performance scores from pretest to posttest (comparing odd- and even-numbered item sets). We standardized the scores from these different sets by transforming them to z-scores (the

distance of each user's score from the mean of odd- or even-numbered items of all participants). Consequently, paired t-test analyses revealed statistically significant improvement from pretest to posttest of RAPM scores in the TNB training group ( $t = 2.67$ ,  $df = 30$ ,  $p = 0.012$ ). There was no significant difference in RAPM scores in the SNB or mental rotations training groups or in the control group (all  $p > 0.13$ ), nor was there any difference in BOMAT scores for any training or control group (all  $p > 0.20$ ).

To further investigate possible interactions between time (pretest and posttest), test versions (odd and even items), and training group (Sudoku, mental rotations, SNB, and TNB), we conducted a repeated-measures analysis of variance (RM-ANOVA). As the sample sizes in our subgroups were rather small, this omnibus test revealed no interactions reaching statistical significance for any of the intelligence measures.

Correlations of pretest IQ levels and personality traits are summarized in Table 2. From 14 personality styles, only one was related to both RAPM and BOMAT IQ tests: a negative correlation between test performance and intuitive/schizotypal style (e.g., "I believe telepathy is possible."). Additionally, RAPM score correlated negatively with situational anxiety, while BOMAT score correlated negatively with anxiety as a trait.

*Table 2. Correlations of intelligence scores and personality scales in pretest*

	RAPM	BOMAT
RAPM	1	0.49
BOMAT	0.49	1
URBAN	-0.13	-0.02
Anxiety state	-0.21*	-0.1
Anxiety trait	-0.08	-0.17*
Self-determined (dissocial)	-0.12	0
Cautious (paranoid)	-0.05	0.02
Reserved (schizoid)	0.15†	0.04
Self-critical (self-insecure)	-0.15†	-0.11
Conscientious (compulsive)	0.2*	0.05
Intuitive (schizotypal)	-0.24*	-0.24**
Optimistic (rhapsodic)	-0.04	0.04
Ambitious (narcissistic)	-0.13	0.01
Critical (negativistic)	-0.05	-0.02
Loyal (dependent)	-0.05	-0.01
Spontaneous (borderline)	-0.16†	-0.15†
Charming (histrionic)	-0.2*	-0.14
Calm (depressive)	-0.06	-0.14
Helpful (self-sacrificing)	0.02	-0.02

$N = 132$ ; \*\* $p < 0.01$ , \* $p < 0.05$ , † $p < 0.10$

Since RAPM scores were sometimes at ceiling, we focused further analyses only on BOMAT scores. Correlations of changes to IQ as measured by BOMAT score and the different PSI personality styles under different training methods are shown in Table 3. Primarily, significant correlations occurred for both SNB and TNB training, sometimes mirroring each other (e.g., changes in BOMAT scores due to SNB training and pretest cautious personality style:  $r(36) = 0.51$ ,  $p < 0.01$ ; changes in BOMAT scores due to TNB training and pretest cautious personality style:  $r(31) = -0.31$ ,  $p = 0.09$ ). Posttest personality traits correlated with the different versions of n-back in a very similar fashion, and sometimes the mirroring was even more pronounced. For example, the correlation between the changes in BOMAT scores due to SNB training and posttest optimistic personality style ( $r(36) = -0.36$ ,  $p = 0.03$ ) was nearly the opposite of the correlation due to TNB training ( $r(31) = 0.40$ ,  $p = 0.03$ ). Similarly, the correlations between BOMAT score changes and posttest depressive personality style scores were nearly opposite for SNB training ( $r(36) = 0.40$ ,  $p = 0.02$ ) and TNB training ( $r(31) = -0.36$ ,  $p = 0.05$ ).

*Table 3. Correlational analyses of PSDI scales and IQ score changes (BOMAT) for each of the possible cognitive training methods (see legend for description).*

PSDI and STAI scores	All training groups		Control (Sudoku)		MR		SNB		TNB	
	$r_G$	$r_{part}$	$r_G$	$r_{part}$	$r_G$	$r_{part}$	$r_G$	$r_{part}$	$r_G$	$r_{part}$
Self-determined (dissocial)	0.12	0.13	-0.22	-0.20	0.25	0.33 <sup>†</sup>	0.19	0.19	-0.17	-0.19
Cautious (paranoid)	0.13	0.16	0.01	-0.02	0.17	0.20	0.51**	0.50**	-0.31 <sup>†</sup>	-0.22
Reserved (schizoid)	0.08	0.08	-0.09	-0.02	-0.2	-0.17	0.52**	0.49**	-0.23	-0.12
Self-critical (self-insecure)	-0.14	-0.17 <sup>†</sup>	0.25	0.14	-0.17	-0.20	-0.01	-0.02	-0.32 <sup>†</sup>	-0.39*
Conscientious (compulsive)	0.13	0.17 <sup>†</sup>	0.09	0.12	-0.05	-0.06	0.17	0.23	0.32 <sup>†</sup>	0.34 <sup>†</sup>
Intuitive (schizotypal)	0.03	-0.08	0	-0.12	0.07	-0.02	-0.30 <sup>†</sup>	-0.37*	0.28	0.16
Optimistic (rhapsodic)	-0.01	0.02	-0.11	-0.10	0.18	0.25	-0.37*	-0.33 <sup>†</sup>	0.19	0.11
Ambitious (narcissistic)	-0.07	-0.01	0.06	-0.07	-0.12	0.04	-0.09	-0.05	0	0.03
Critical (negativistic)	0.18 <sup>†</sup>	0.19 <sup>†</sup>	-0.1	-0.14	0.13	0.28	0.41*	0.40*	-0.11	-0.12
Loyal (dependent)	-0.09	-0.04	0.30 <sup>†</sup>	0.14	-0.21	-0.04	-0.06	0.00	0.01	-0.07
Spontaneous (borderline)	-0.06	-0.13	0.12	-0.02	0.02	-0.04	0.06	-0.02	-0.32 <sup>†</sup>	-0.29
Charming	0.12	0.06	-0.16	-0.22	0.18	0.18	-0.15	-0.16	0.29	0.16

(histrionic)										
Calm (depressive)	0.01	-0.06	0.22	0.15	-0.14	-0.21	0.30 <sup>†</sup>	0.23	-0.22	-0.24
Helpful (self-sacrificing)	-0.08	-0.07	0.12	0.06	-0.27	-0.14	0.03	0.02	0.02	-0.05
State anxiety	-0.05	-0.09	0.01	-0.09	-0.19	-0.21	0.07	0.01	-0.02	0.02
Trait anxiety	0.09	0.00	0.08	0.03	-0.13	-0.10	0.38*	0.33 <sup>†</sup>	0	-0.20

$N_{\text{allTraining}} = 97$ ,  $N_{\text{sudoku}} = 34$ ,  $N_{\text{MR}} = 31$ ,  $N_{\text{SNB}} = 36$ ,  $N_{\text{TNB}} = 31$ ; \*\* $p < 0.01$ , \* $p < 0.05$ , <sup>†</sup> $p < 0.10$ ; SNB = single n-back, TNB = triple n-back, rg = correlation (zero-order) of the personality scale with IQ gain, r<sub>part</sub> = partial correlation of the pretest personality scale with the posttest IQ (pretest IQ partialled out)

## 2.03 Discussion

In this study we investigated whether CT would cause any statistically significant changes in intellectual performance from pretest to posttest (as measured by RAPM and BOMAT), and whether these potential changes would correlate with different personality traits. The TNB training method significantly improved performance on the RAPM, but no training method improved scores on the BOMAT. These mixed results in our study with fully timed IQ tests and an active-contact control group reflect the ambivalent outcomes of WM training research in general. While failures to replicate and null results are certainly scientific contributions generally desirable in psychology research (Makel, Plucker, & Hegarty, 2012) and publishing them is important to combat the “file-drawer” problem (Redick et al., 2012; Shipstead et al., 2012), the ambiguity of cognitive training outcomes has become an important issue requiring further explanation.

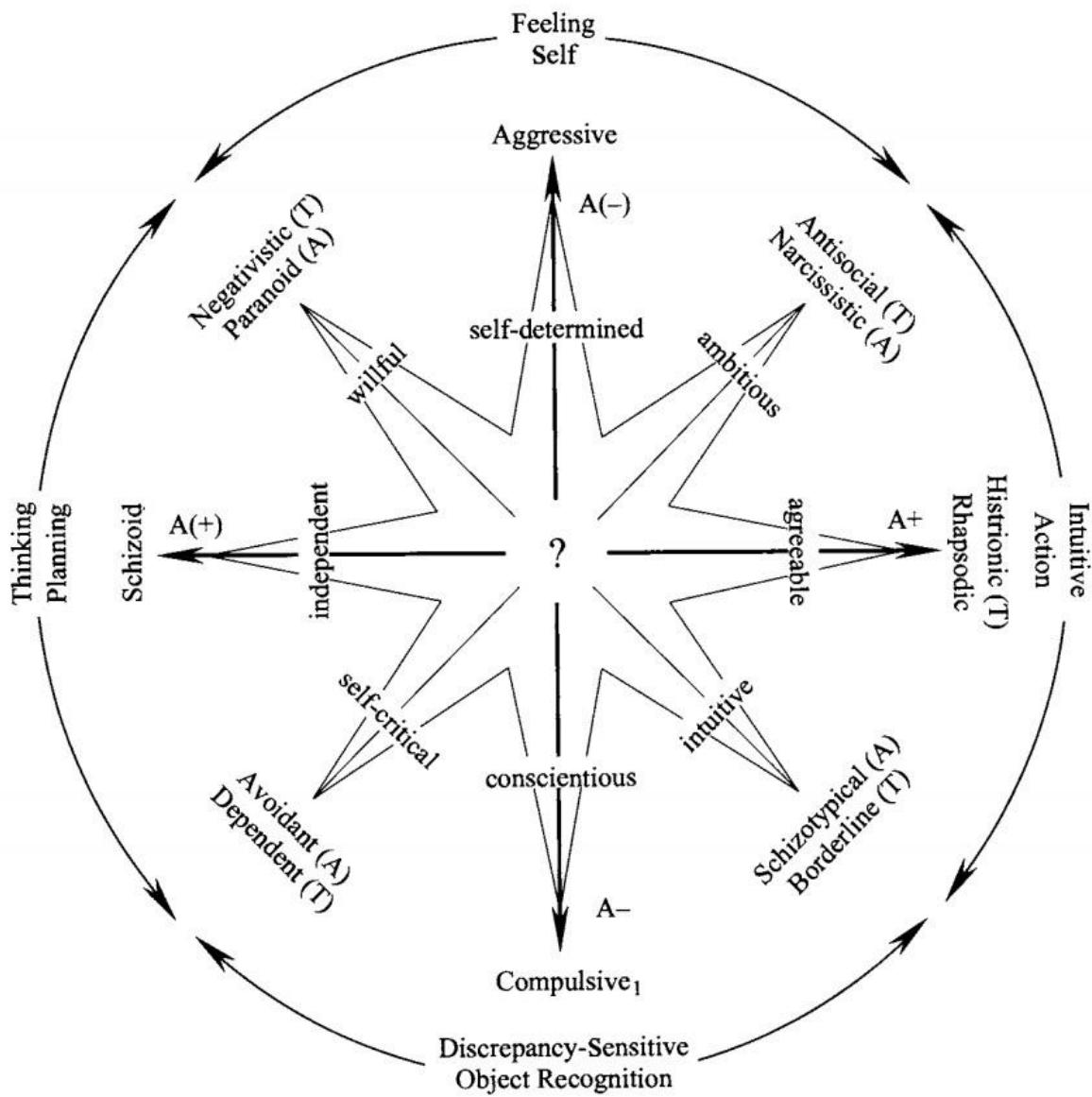
### (a) Mediating role of PSI personality factors in CT outcomes

We started by correlating baseline IQ scores of participants with their scores on 14 PSDI personality trait scales. Some weak correlations appeared (see Table 2), which are consistent with the notion that ability traits (e.g., IQ) and non-ability traits (e.g., the “big five” personality dimensions) are relatively independent (Demetriou, Kyriakides, & Avraamidou, 2003; Farsides & Woodfield, 2003; Soubelet & Salthouse, 2011). We then correlated PSI personality profiles with the IQ gains from the multiple CT methods, and with the pooled sample of training as a whole (not differentiating between training methods; see Table 3). There was no significant relationship with any of the PSDI scales, which suggests that there is no effective way to predict if and how much a person will gain from CT based on their personality scores. However, when we split our sample according to the type of training method, substantially different results appeared: in the SNB training group, gains in BOMAT scores correlated positively with negativistic, paranoid, and schizoid PSDI scales, while at the

same time, correlated negatively with schizotypal and rhapsodic PSDI scales (see Table 3). Looking at the STAR model proposed by Kuhl (2000a; Figure 3), the PSDI scales that are positively associated with BOMAT scores are situated exactly opposite those PSDI scales that are negatively associated with BOMAT scores. PSI theory, which integrates cognitive and personality styles into a unifying framework, therefore may provide an explanation of the underlying factors that affect the gain in IQ after SNB cognitive training.

According to PSI theory, the two pairs of styles correlating with IQ gain positively (distrustful and reserved) and those correlating with IQ gain negatively (intuitive and optimistic) are fundamentally different from each other. Namely, they differ in terms of their sensitivity to rewards and punishments and in the dominance of cognitive macrosystems as described by PSI theory. While both distrustful and reserved styles are dominated by analytical thinking and planning (and include low sensitivity to both rewards and punishments), intuitive and optimistic PSI styles are dominated by intuitive behavior control and are characterized by high sensitivity to rewards and punishments (see Kuhl, 2000a). This perspective therefore suggests that the effects of SNB exercise (utilizing only one sensory modality, presenting only one stimulus every 3 seconds, and requiring greater delays in WM processing as higher levels of n are reached) correlates positively with personality styles that typically engage in time-planning and that are less dependent on rewards or punishments. On the other hand, the substantially different training method of TNB (which requires more immediate processing, and presents participants with diverse, multimodal stimuli) favored persons with intuitive and action oriented personality styles (although evidence for this is based on our small sample of N = 31 and correlations with IQ gain that were marginally significant at  $p < 0.1$ , see Table 3).

Figure 3. The STAR model of personality dimensions by Kuhl & Kazén (1997).



PSI theory provides quite complex explanations and testable predictions about human behavior (Kuhl, 2000b). For example, while Eysenck's theory (Eysenck, 1950) operates with static traits and their content (i.e., traits as a tendency to experience certain affects), PSI theory is more dynamic and process oriented, with traits representing a tendency and capacity to up- or down-regulate certain affects. In addition to the 14 personality traits, PSI theory also introduces four interacting personality macrosystems (intention memory, extension memory, intuitive behavior control, and recognition of objects; Kuhl, 2000a). This complexity has allowed us to elaborate further on our results. It is evident that opposing PSI personality styles influence the effectiveness of cognitive training differently, and the fact that opposing correlations in our data both reached statistical significance further supports the emerging

correlation pattern that fits the PSI model. In addition, this pattern repeats itself in retest sessions and seems to be mirrored in independent groups training on either of the considerably different TNB and SNB training methods. The pattern of correlations present in our data fits well with PSI theory, and we suggest that this explanation of ambiguous CT results is the most important finding of our study.

We did not find this pattern in the RAPM. One possible reason for this could be that RAPM was much more influenced by a ceiling effect. No participants scored perfectly in the BOMAT, but 23% of participants did in the RAPM.

To explore the directionality of significant correlations between pre- and posttest personality and cognition, we conducted Cross-Lagged Panel Analyses. In several cases, pretest personality traits significantly added to the prediction of the posttest IQ score (but not vice versa, see Table 3). Caution is always advised when interpreting causality, but there is some evidence suggesting that posttest IQ score changes are influenced by pretest personality traits, rather than posttest personality traits being influenced by pretest IQ scores (Granger, 1969).

#### (b) Mediating role of anxiety in CT outcomes

Studer-Luethi, Jaeggi, Buschkuhl, and Perrig (2012) have also documented the influence of personality traits on CT outcomes. In their study, individuals scoring high in neuroticism benefitted more from SNB and individuals scoring low in neuroticism benefitted more from dual n-back training. The authors hypothesized that anxiety could be a mediating factor: individuals high in neuroticism are overwhelmed sooner by more complex, multimodal n-back exercises (i.e., the dual n-back), and this anxiety in turn blocks the process of cognitive gain. In our experiment, we tested this hypothesis, by measuring both state and trait anxiety using the State-Trait Anxiety Inventory (STAI; Spielberger & Gorsuch, 1983). We found a positive correlation of anxiety as a long-term personality trait (which presumably influenced participants CT continuously, as opposed to the transient state of anxiety during the test-taking situation), with BOMAT gain for the SNB training group (see Table 3). This “high-anxiety – gain after SNB” correlation is in concordance with the hypothesis proposed by Studer-Luethi, Jaeggi, Buschkuhl, and Perrig (2012). In addition, there seems to be certain phenomenological closeness in all of the personality traits thus far found to correlate positively with cognitive gain after SNB (anxiety, neuroticism, and critical, reserved and cautious personality style).

### (c) Limitations

Some inherent limitations of our study include the problem of psychometric precision and difficulty identifying small inter-group differences (one or two points of raw score from pretest to posttest), against noticeable intra-group differences caused by unavoidable differences in level of difficulty between the odd- and even-numbered IQ subtests (RAPM:  $t = 7.50$ ,  $p < 0.001$ ; BOMAT:  $t = 3.13$ ,  $p = 0.002$ ). We consider total training time (approximately 8 hours) to be another constraint to examining the effects of CT to Gf thoroughly. Stronger effects may have been observed with additional training. Another limitation may be the relatively high pretest scores of our sample (RAPM mean = 14.78 out of 18, SD = 2.6; BOMAT mean = 13.38 out of 20, SD = 2.34). In addition, our control group task (basic-level Sudoku) was perceived as quite demanding by several training subjects retrospectively, which could result in type-II errors of negative bias in favor of the null hypothesis. We are also aware of the potential risk of some statistical tests being significant by chance due to our multiple-comparisons (analyzing several personality factors in each group). Nevertheless, the emerging pattern of correlations seems to fit logically into the PSI personality model rather than be random.

## 2.04 Conclusions

Our study echoes the findings of several previous studies suggesting that ability and non-ability traits can emerge in any combination in a person, with a loose relationship between the two at best. Nevertheless, we presented some evidence that changing or training an already established, complex ability trait (e.g., fluid intelligence) can require an understanding of non-ability traits (e.g., PSI personality styles). Intuitively, the opposite seems to be plausible as well: if one seeks to modify personality traits, for example in psychotherapy, this may depend in part on levels of ability traits such as intelligence. Of course, any investigation into the interplay of two big areas in psychology (intelligence and personality) is bound to be incomplete, and we have discussed some potential limitations to our study above. Still, we believe our findings can contribute to the efficacy of tailor-made cognitive training interventions and inspire further research in this exciting area.

## Chapter III. BRAWO questionnaire

### 3.01 About

To further examine the n-back phenomenon, during my doctoral studies I became interested in experiences of people who used n-back to exercise their cognition (let's call them „n-backers“). The best place to meet n-backers at that time was a Google discussion group called „Dual N-back, Brain Training & Intelligence“. Founded in 2008, this group discussed over 2500 topics in six years, ranging from n-back to nootropics to TMS. Today it consists of thousands of mostly inactive members, but still averages 1-5 short posts daily from around 20-50 members per month.

The reason this group attracts n-backers (besides containing huge amount of keywords) is that it is adjacent to a very popular, open-source n-back training software, called „Brain workshop“. It was released by Paul Hoskinson in 2008, and downloaded over 750,000 times over the next six years. From its earliest versions, it supported the configuration which was used in Jaeggi et al. (2008). In its latest version (4.8.4) it is widely customizable, and features many of original n-back modifications (like arithmetic mode, variable N mode, or „crab back“ mode - reverse order of sets of N stimuli). Today there are several other ways one can use to n-back (including websites and mobile platform), but Brain workshop still offers the most customizable experience of n-back training. This is the reason we used its original or proprietary-modified versions in our experiments.

In 2010, I created the BRAWO (BRAin WOrkshop) questionnaire, to collect basic quantitative and qualitative aspects of the n-back training experience. For four years, it recorded responses from members of the group regarding their lifestyle, n-back training and its outcomes. This chapter is dedicated to its analysis.

### 3.02 Preparation and limitations

There were 291 entries recorded altogether. Unfortunately, 33 of them (11%) had to be eliminated from final analysis due to following:

- Empty entries
- Entirely duplicated entries (same time stamp second or minute, identical responses – newer out of two entries maintained)
- Almost duplicated entries (during a few minutes nearly identical entry appeared, with one or two changed bits of information out of 50 – person supposedly wanted to correct himself – newer entry maintained)

- Sham entries (every few months, entry with all the default answers appeared – one or several individuals supposedly just wanted to see updated result section, which was available in read-only mode after one completed the questionnaire)

In some entries, some answers had to be invalidated, for the following reasons

- In an effort to simplify the questionnaire and improve the user-experience and involvement, I always tried to offer options (i.e. you could choose your age from a list numbered from 13 to 50, rather than writing it down). The issue was, that sometimes I made the cut-offs too narrow („less than 13“ and „more than 50“), which were actually chosen ca. 10 times. To preserve the type of variable (interval), I had to claim these 10 answers to „age“ question as missing.
- Particular answers to some questions make other questions of the same entry irrelevant (i.e. when a person states, that s/he tried n-back only few times or not at all, whole section on results from training is not applicable)
- Contradictory answers – In qualitative section of some entries, users left a comment that invalidated one of their answers (i.e. they left a comment „I have no idea what my IQ is, so I left the default option for that question“)

In addition, due to its methodological nature, BRAWO as a self-report, anonymous, online questionnaire is prone to inaccuracies and biases (Podsakoff et al., 2003). The following biases were especially considered:

- Repeated entries by the same person, and / or sham entries
- Self-recruitment of participants to answer the questionnaire
- Social desirability (tendency of some people to respond to items more as a result of their social acceptability than their true feelings)
- Implicit theories and illusory correlations (beliefs about the covariation among particular traits, behaviors, and/or outcome)
- Mood, especially as an actual state (relatively recent mood-inducing events to influence the manner in which respondents view themselves and the world around them)
- Item ambiguity / misinterpretation of questions, without the possibility to ask for explanation

On the other hand, the online framework and anonymity could potentially lead to improved ecological validity, because participants

- Are able to choose the place they want to respond from (e.g. privacy of their home)

- Are able to choose the time and pace of their responses

In conclusion, although there are over 250 entries in BRAWO, we need to be very modest in interpretations of this questionnaire. As one of the reminders of this need, I refrain from using decimal places in statistical analysis, except when inevitable.

### 3.03 Univariate analysis

#### (a) Timestamps

Usable responses ( $N=258$ ) were recorded during the time period from *2-Aug-2010* to *26-Aug-2014*, for a range of 4 years (1484 days precisely). So-called median was on *22-Jan-2012*.

*Time-of-the-day mean was 12:59:13.* This probably reflects the fact that entries were made from many different time zones at many different times of the day, amounting to random times. This would have averaged in the middle of 0-24 hours format – at 12:00, but the questionnaire itself was created and recorded the responses in GMT+1 time zone.

#### (b) Age

Figure 4. Histogram and density trace - Age variable

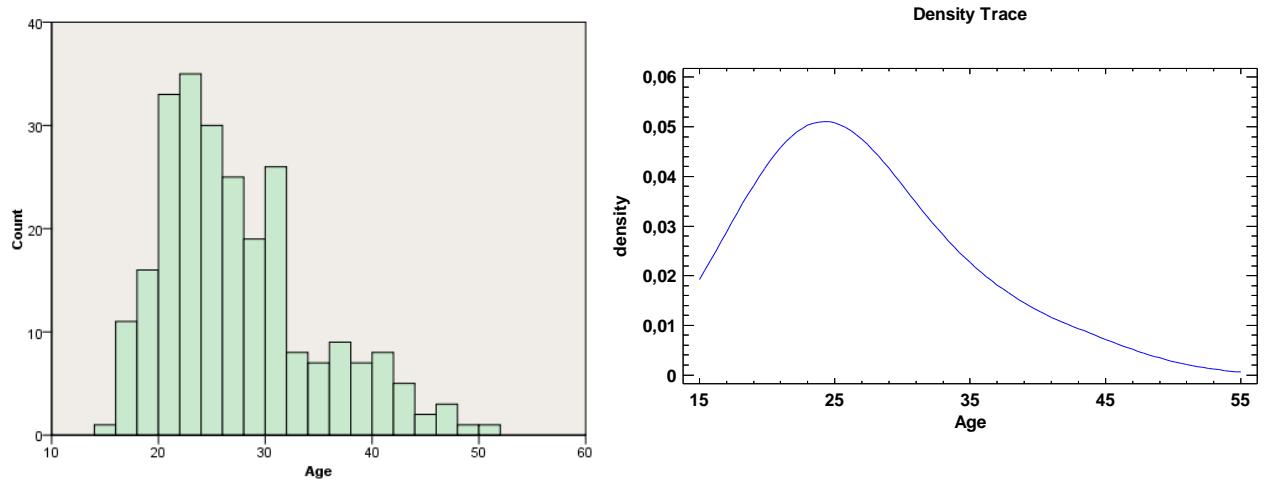


Table 4. Descriptive statistics - Age variable

	N	Range	Minimum	Maximum	Mean	SD	Skewness		Kurtosis	
							Statistic	Std. E	Statistic	Std. E
Age	247	35	15	50	27	7	,893	,155	,328	,309
Valid N	247									

There were 3 entries with „less than 13 years old“, and 8 entries with „more than 50 years old“. Because no precise value was known, for the means and purposes of quantitative analyses I consider these pieces of information „missing“. Skewness of *Age* significantly departs from normal distribution, which renders parametric tests on this variable inappropriate.

### (c) The amount of n-back training

In an effort to quantify the total amount of n-back training for each person, I had to determine several aspects of training:

1. How long did they follow regular training schedule (in weeks)
2. How often do they usually train per week (in sessions)
3. How long does their usual training session take (in minutes)

Similarly to *Membership Time*, these figures probably changed for many users after the completion of BRAWO questionnaire, as many of them continued training. Fortunately, we do not need to know the lifetime amount of training time to investigate potential training effects – the questionnaire „snapshot“ at any time is enough, because it includes other variables recorded at the same (and quantified) amount of training time. By multiplication of these variables, we get the amount of n-back training:

$$\text{Total training time} = \text{weeksOfRegularTrainingSchedule} \times \text{sessionsPerWeek} \times \\ \text{minutesInSession}$$

Understandably, all the factors are based on self-report and sometimes quite distant memories, so the final product is a highly speculative number. Even more so because, as we will see, I needed to do some extrapolation. Nevertheless, the first figure (the number of weeks, during which participants followed some regular n-back training schedule), is probably the most reliable in determining the general amount of training at the time of BRAWO completion.

#### (i) Duration of regular training schedule

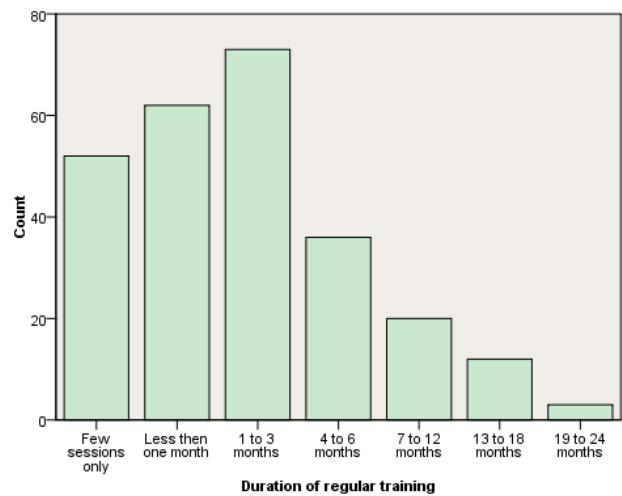
Nearly a half of BRAWO respondents trained less than one month, but 80% trained more than just a few sessions (which still leaves us with considerable N=206 for analyzing the effects of N-back training later). Out of these, N=71 trained regularly for four months or more.

*Weighted mean* (the mean of middle values from each category, weighted by their frequencies in the sample) is  $\bar{X} = 7$  weeks.

Table 5. Descriptive statistics - regular training schedule

	Frequency	Percent	Cumulative Percent
Few sessions only	52	20	20
Less than one month	62	24	44
1 to 3 months	73	28	72
4 to 6 months	36	14	86
7 to 12 months	20	8	94
13 to 18 months	12	5	99
19 to 24 months	3	1	100
Total	258	100	

Figure 5. Bar of counts – regular training schedule



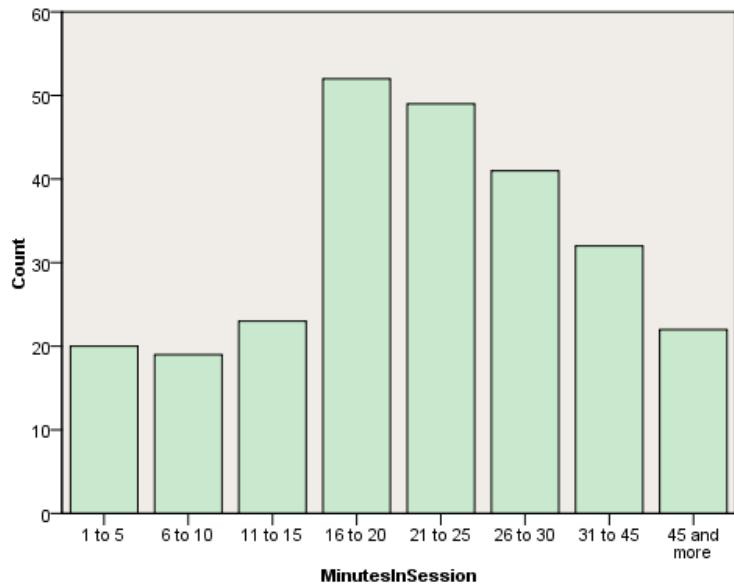
### (ii) Minutes per session

Next, we need to determine how long each session lasted. To maximize the response rate to BRAWO (its fluency and user experience of respondents), I again prepared time categories and let the user chose one of them. Most popular session duration apparently varies around the time-limit that Jaeggi et al. (2008) used in her landmark study (25 minutes per session). And indeed, this is supported with weighted mean value:  $\bar{X} = 23.5 \text{ minutes}$ . But even longer times, up to 45 minutes and more, are no exception.

Table 6. Descriptive statistics - Minutes per session

	Frequency	Percent	Cumulative Percent
1 to 5	20	8	8
6 to 10	19	7	15
11 to 15	23	9	24
16 to 20	52	20	44
21 to 25	49	19	63
26 to 30	41	16	79
31 to 45	32	12	91
45 and more	22	9	100,0
Total	258	100,0	

Figure 6. Bar of counts - Minutes per session



### (iii) Weekly Training Frequency (WF)

The last parameter required to estimate the total amount of n-back training time is the number of n-back training sessions per week (or weekly training frequency):

Table 7. Descriptive statistics - Weekly training frequency

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Less than once	97	37,6	37,6	37,6
1 to 3 times	56	21,7	21,7	59,3
4 to 7 times	105	40,7	40,7	100,0
Total	258	100,0	100,0	

As I mentioned before, here we have to deal with the biggest drawback in BRAWO design. In 2010, I just created a question investigating the *present* WF – but in many cases, respondents became acquainted with questionnaire *after* they already finished the training. So although at the time of BRAWO some respondents claimed their WF was close to zero, it often was quite different from zero in the past (except for those of course, who selected „I did only few sessions altogether“ option in previous question). Therefore, this figure has little practical use.

#### (iv) Total training time

Because we miss the “sessionsPerWeekInThePast” number, we will control for it by keeping it equal to “1” (assuming just one session per each week of regular training, although we are losing statistical power here). Total training time will be then computed only out of two remaining time-parameters:

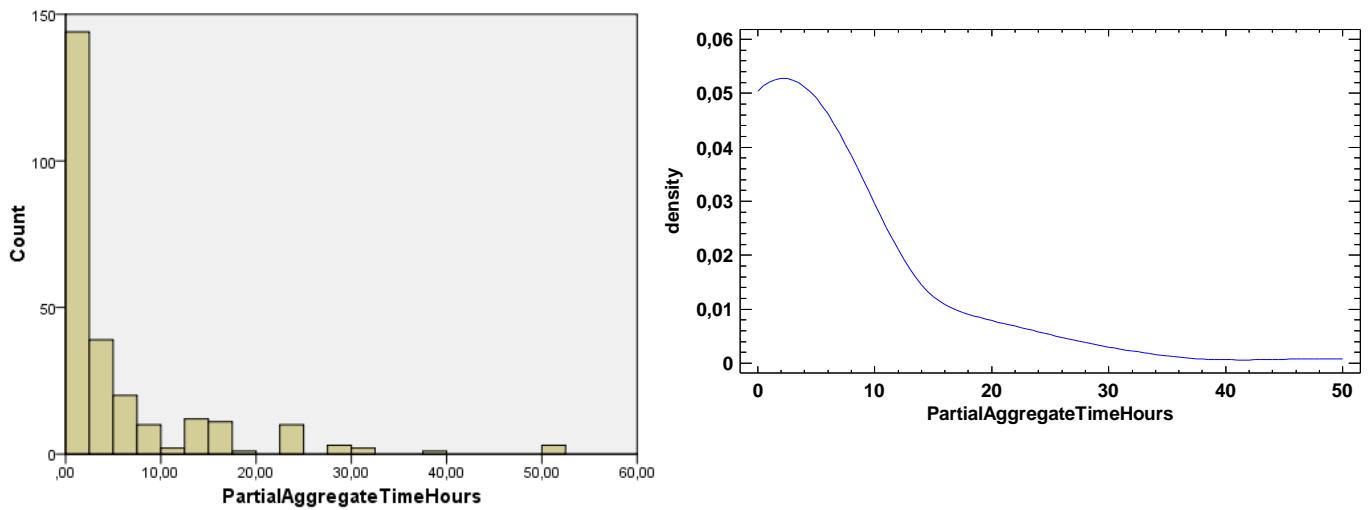
$$\text{PartialAggregateTime} = \text{weeksOfRegularTrainingSchedule} \times (\text{minutesInSession} / 60)$$

Although an arbitrary number, based on estimates and ordinal in nature, *PartialAggregateTime* relates directly to BRAWO members experiences, and it can sketch an interesting picture of n-back training around the world. Let's start with raw descriptives:

Table 8. Descriptive statistics - Partial aggregate time

	N	Minimu	Maximu	Mean	Std. Deviation	Skewness		Kurtosis	
		m	m			Std. Err	Std. Err		
PartialAggregateTime Hours	25 8	,00	50,00	5	8,5	2,702 ,152		8,767 ,302	

Figure 7. Histogram and density trace - Partial aggregate time



So if we pretend, that participants trained only once per week of training, they trained n-back for an estimated average amount of 4 to 6.7 hours (99% confidence interval), with 8.5 SD. Nevertheless, as we do know weekly training frequency of those who train in present

(weighted mean  $\bar{X} = 4$ ), actual mean of total training time can be up to 4 times greater, i.e. around 20 hours. These numbers continued to change with passing time – nearly two thirds of respondents continued to train at the time of BRAWO completion, one third already finished training, and 5% hardly ever gave it a try).

#### (d) Version of n-back exercise

N-back comes in many shapes and forms (different sounds, images, arithmetic, variable n, rotation of matrix and so on), but the most popular classification comes from the number of involved sensory modalities, or streams of stimuli:

- Single n-back only involves updating of visual position of the squares, or audio letters
- Dual n-back requires updating both visual and audio channel
- Triple n-back should actually involve additional sensory modality (for example touch, Klatzky et al., 2008). Nevertheless, triple n-back commonly stands for a stream of audio letters, visual positions, and different colors of visual stimuli
- Quad n-back adds shape of visual stimuli
- Penta n-back adds another visual or audio stimuli

In our sample, respondents who actually trained ( $N=206$ ), strongly preferred dual n-back (80 %). Single n-back was second (10%), and next three versions were represented marginally:

*Table 9. Descriptive statistics - N-back version*

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Single n-back	14	6,8	6,8	6,8
Dual n-back	166	80,6	80,6	87,4
Triple n-back	11	5,3	5,3	92,7
Quad n-back	13	6,3	6,3	99,0
Penta n-back	2	1,0	1,0	100,0
Total	206	100,0	100,0	

### (e) Level of n-back exercise

Independently of the *version* of n-back, one can only exercise it on a specific *level*: 1-back, 2-back, 3-back and so on. This level changes every few minutes, in order to constantly match (and challenge) user abilities, yet not overwhelm him.

Now, it is definitely easier to reach single 3-back than dual 3-back (or to reach dual 4-back then triple 4-back). But how about dual 4-back and quad 2-back? Both seem to deal with the same number of items in WM but are they equally demanding of cognitive resources? To my knowledge, this question remains unanswered thus far, although the answer could potentially improve our knowledge of WM structure (how much do different modalities draw from domain general / domain specific resource pool? See Cowan, Saults & Blume, 2014). In any case, the difficulty of n-back is strongly influenced by both its level *and* version. So what do our respondents tell us about the former?

#### (i) Usual training level

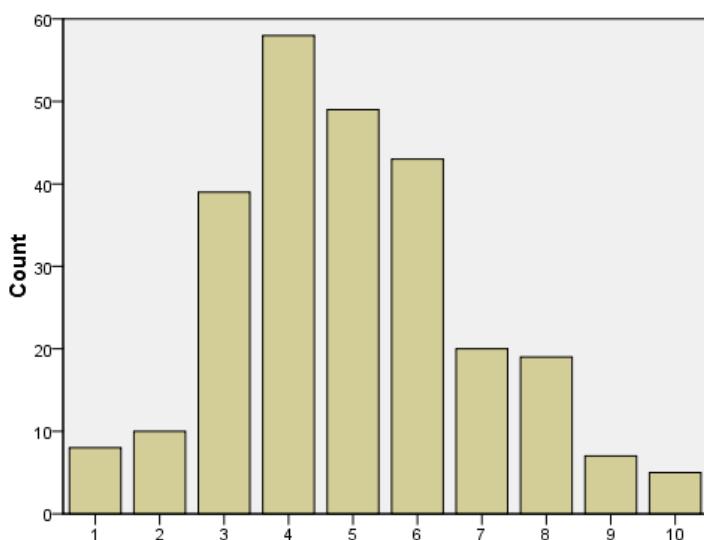


Figure 8. Bar of counts - Usual training level

For the whole sample ( $N=258$ ), weighted mean of level  $\bar{X} = 4,7$  with  $SD = 1.4$ , and 99% CI [4,68; 4,75]. But this figure actually doesn't tell us much, because it is a mix of 5 different versions of n-back (single, dual, triple, quad, penta). So let's take a look only at those who actually trained more than few times ( $N=206$ ), and pool them into groups by n-back version (without penta n-back, because its  $N=2$ ):

Table 10. Descriptive statistics - Usual training level

	N	Minimu m	Maximu m	Mean	Std. Deviation
Single n-back group	14	1	8	5,57	2,138
Dual n-back group	166	1	10	5,50	1,845
Triple n-back group	11	1	7	4,27	1,618
Quad n-back group	13	3	6	4,15	1,214

As one would expect, there's an inversely proportional relationship between the complexity of n-back and achieved level (that is, the more complex version, the lower average level users achieved). But shouldn't it be much more pronounced? After all, triple n-back requires user to keep in mind 3 times as many items as single 2-back. The most probable reasons for these unexpectedly small differences in means across versions could be:

- Huge differences in sample size
- Maybe going from single 3-back to dual 3-back is less cognitively taxing, then going from single 3-back to single 6-back (although the number of items to keep in WM is the same)<sup>1</sup>
- The way people choose particular versions to train. For example, single n-back could be chosen by respondents who consider themselves incapable of dealing with dual n-back, or they just want to „try it out“ - which in addition reflects on total training time.

Analogically, triple and quad n-back, although significantly harder, can be chosen by a high-ability person as a challenge. In this way, the means could harmonize across different versions of n-back.

In any case, from present data it is impossible to quantify the real differences in difficulty of different n-back versions (mean differences are unrealistically small). But we will at least check for explaining factors for this (like IQ or total training time) in the bivariate analyses section. Interestingly though, there seems to be some regularity in standard deviations of means – more complex versions of n-back show smaller variance around mean, which lends some support to our previous reasoning.

---

<sup>1</sup> This hypothesis is definitely worth examining.

### (ii) Aggregate n-back level

Although in our sample, the average means for different versions of n-back didn't differ much, there seems to be no doubt, that difficulty of n-back exercise increases from „single“ to „penta“. We can actually determine, how many items one has to hold in WM to meet the demands of certain version and level of n-back. For example, dual 2-back requires the user to continually update the last two items from two streams of stimuli – that is, to hold 4 items in WM all the time. Triple 4-back deals with last four items from three streams, that is 12 items, and so on.

As I hypothesized earlier, training with the same number of items in WM can put different load on user depending on configuration (dual 4-back versus quad 2-back). In spite of these differences, it certainly makes sense to know the number of items in a configuration, which people used to train. Additionally, this figure will serve as a „normalization of training difficulty“, and allow us to include all the versions of n-back in bivariate analyses later (not only dual n-back responses, but N=50 from other 4 groups too). The equation, followed by basic descriptives of *aggregateBackLevel* (of those who trained more than just few times, N=206):

$$\text{aggregateBackLevel} = \text{nBackVersion} \times \text{achievedLevel}$$

Table 11. Descriptive Statistics - aggregateBackLevel

	N	Minimum	Maximum	Mean	Std. Deviation
aggregateBackLevel	206	1	24	11,08	4,245
Valid N (listwise)	206				

This means that users (at the time they completed BRAWO) trained with approximately 11 items in WM on average (which translates to single 11-back, or circa dual 5-back or triple 4-back). On the histogram, we can see the disproportionately high counts of dual n-back subgroup (N=166), intertwined with smaller counts of other n-back versions.

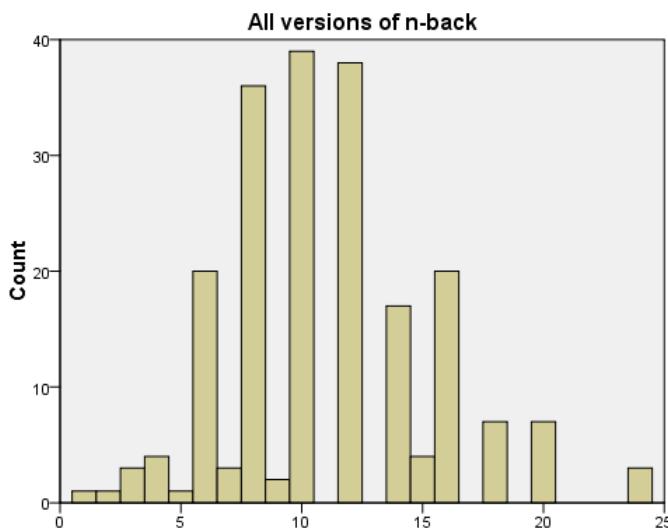


Figure 9. Histogram - AggregateBackLevel

#### (f) Intelligence quotient

As I mentioned the theoretical part, conceptualization and measurement of general human cognitive ability have a long and difficult scientific history. Today, psychologists have considerably valid and reliable intelligence tests at their disposal, but most people seem to be pretty happy without knowing their precise IQ. There even seems to be something about intelligence tests that scares people off. It seems to me that being „intelligent“ is far too important in our culture – having a perfectly average IQ (say 100) is often considered not good enough, and even kind of embarrassing. We mystify intelligence and overemphasize its importance and, as a result, are afraid to take the test.

To deal with these attitudes and lack of knowledge regarding one's IQ, I took several measures when constructing items that dealt with IQ in BRAWO questionnaire:

- I briefly stated that concept of intelligence and ways to measure it are open to dispute, and that intelligence is widely independent from personality traits, ethics, or happiness and the like
- I asked respondents to indicate a range, to which they think their *lower* IQ limit ( $SD = 15$ ) realistically belongs to
- Then I asked them to indicate a range, to which they think their *upper* IQ limit ( $SD = 15$ ) realistically belongs to
- Then I asked, if they have any references for these estimates

References for these estimates generally belong into one of following categories:

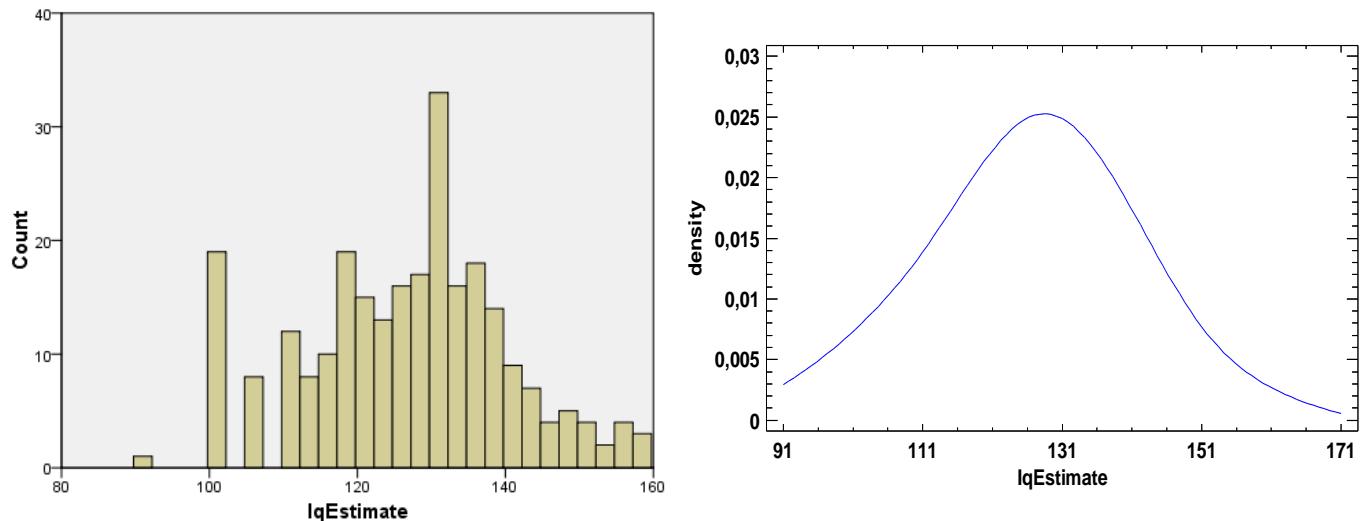
- Not stated or unknown (N=123, 48%)
- Random internet IQ tests (mostly iqtest.dk)
- High-range IQ tests (Concep-T, Algebrica)
- SAT / GRE scores
- Professional IQ test (WAIS, RAPM, Mensa tests, “psychiatrist’s IQ test”, and the like)

After computing means for *lower limit range* and *upper limit range*, I averaged these two means, and ended up with as precise an estimate of one’s own intelligence as possible. Nevertheless, I didn’t go as far as weighting each IQ score depending on credibility of the source. In my opinion, there’s no useful factor, or even reliable order of credibility of references – RAPM interpreted by expired norms, or taken 10 years ago, can be similarly (in)accurate as an unknown, but up-to-date online IQ test with some scientific background. In addition, although people tend to have imperfect information about their intelligence (and men overestimate their intelligence more than women, Furnham, 2001; Syzmanowicz & Furnham, 2011), there is evidence for a considerable correlation of self-assessed intelligence (SAI) and psychometrically measured IQ: from  $r = .30$  (Furnham, 2001; Freund and Kasten, 2012) up to  $r = .49$  (Stumm, 2014). And magnitude of this correlation is comparable to the one between different factors of intelligence (fluid and verbal scores), see e.g. Soh, Jacobs (2013) or Kornilova and Novikova (2012). Taken together, in online self-reported questionnaires, self-estimates of “intelligence” with and without layman references can be considered comparably inaccurate.

Table 12. Descriptive statistics - IQ estimate

	N	Minimum	Maximum	Mean	Std. Dev.	Skewness		Kurtosis	
						Std. Err	Std. Err	Std. Err	Std. Err
IqEstimate	257	91	158	126,54	13,739	,127	,152	-,131	,303

Figure 10. Histogram and density trace - IQ Estimate



Looking at the statistics, respondents of BRAWO tend to report IQ's well above average.

There were no outliers, data nicely fit normal distribution, the mean being

$$\bar{X} = 126 \quad SD = 14 \quad 99\% CI [124, 129]$$

#### (g) IQ mindset (Do you think IQ can be improved?)

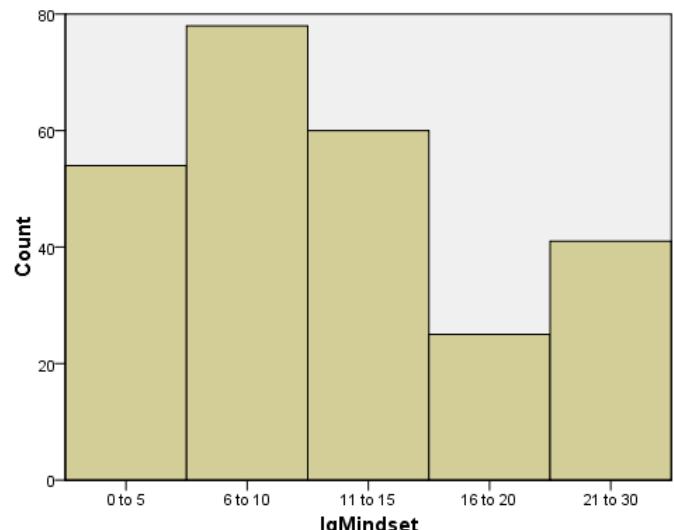
Some researchers differentiate between *fixed* and *growth* mindset (Dweck, 2000). So-called „fixed mindset“ is represented by the belief, that human cognitive abilities are inborn, and not significantly malleable. On the other hand the „growth mindset“ advocates the opinion that intelligence can be developed, which seems to lead to higher motivation and persistence in face of real-world challenges (Dweck, 2012). Mangels (2006) even identified some neurological processes (differences in top-down control of attention), which could be related to different mindsets.

Cognitive training, and the n-back exercise in particular became a synonym for this growth/fixed debate in recent years. I was curious what people who use n-back think of this dilemma, so I added the appropriate questions to BRAWO questionnaire (and offered some categories):

Table 13. Descriptive statistics - IQ malleability mindset

	Frequency	Percent	Valid Percent	Cumulative Percent
0 to 5	54	20,9	20,9	20,9
6 to 10	78	30,2	30,2	51,2
11 to 15	60	23,3	23,3	74,4
16 to 20	25	9,7	9,7	84,1
21 to 30	41	15,9	15,9	100,0
Total	258	100,0	100,0	

Figure 11. Descriptive statistics - IQ malleability mindset



As one could expect from group of people who are presumably trying to improve their cognition, 80% of respondents think that intelligence can be significantly improved. Weighted mean:  $\bar{X} = 10,5$   $SD = 6,5$

#### (h) Drug use (and abuse)

Participants of the *Dual N-back*, *Brain Training & Intelligence* Google group regularly discuss (and sometimes ruminate) about everything that seems to be even remotely related to improving one's cognition. There are many threads about trans-cranial magnetic stimulation (TMS), trans-cranial direct current stimulation (tDCS), nootropics, image streaming, food supplements, psychiatric medication or even illegal drugs. Some members reported trying home-made tCDS devices on themselves, or exercising n-back for 4 hours straight (...). It seemed to me that question about lifestyle (food supplements, medication, or dependent behaviors) is in place. Overview follows (N=258):

Table 14. Descriptive statistics - Drug use (and abuse)

Variable	YES		NO	
	Frequency	Percent	Frequency	Percent
Vitamin supplements	128	<b>50</b>	128	50
Minerals supplements	82	<b>32</b>	176	68
Natural "nootropics" (omega 3, caffeine...)	125	<b>48</b>	133	52
Artificial nootropics (racetams, melatonin...)	65	<b>25</b>	193	75
Prescription stimulants (modafinil...)	26	<b>10</b>	232	90
Prescription antidepressants	22	<b>9</b>	236	91
Tried illegal drugs (stimulants, opiates, psychedelics...)	19	<b>7</b>	239	93
Drinking on weekends	49	<b>19</b>	209	81
Smoking marijuana	22	<b>9</b>	236	91
Smoking tobacco	22	<b>9</b>	236	91

First three variables (vitamins, minerals and natural food supplements) do not seem to deviate much from US population, according to the report by National Center for Health Statistics (2012). Circa half of the American population in 2012 used vitamin supplements, and this didn't change much throughout the whole decade.

On the other hand, artificial nootropics, also referred to as smart drugs, or pharmacological cognitive enhancers (PCE), seem to be quite popular in BRAWO sample – every fourth respondent is using them. In reality, there is scarce evidence of PCE causing any cognitive improvements in healthy adults. Even up-to-date reviews of literature, speculating about benefits of low doses of psycho-stimulants (Wood et al, 2014) sound unconvincing in terms of both psychologically safe, and non-zero-sum<sup>2</sup> cognitive improvements in healthy adults (Urban & Gao, 2014). Similar results were presented by (Mohamed, 2012 and 2014), who discovered that Modafinil has potential for addiction and in healthy adults it decreases creative thinking and reaction time.

---

<sup>2</sup> By zero-sum cognitive improvement I mean e.g. „increasing attention“ in terms of its tenacity, but decreasing its flexibility, capacity or stability.

In addition, should effective and safe PCE be available one day, essential ethical issues arise (Maslen, Faulmüller & Savulescu, 2014), such as whether

- the medical safety-profile of PCEs justifies restricting or permitting their optional or required use (e.g. surgeons, pilots etc.)
- the enhanced mind can be an “authentic” mind
- individuals might be coerced into using PCEs
- there is a meaningful distinction to be made between the treatment vs. enhancement effect of the same PCE
- unequal access to PCEs would have implications for distributive justice
- PCE use constitutes cheating in competitive contexts

Moving on to prescription antidepressant, the percentage among members of the group (9%) matches the average of US population in 2012.

Regarding illicit drug use in the last month, the US population prevalence is 9% (as reported in 2014 National Survey on Drug Use and Health). Our respondents reported only lifetime use (“I tried it”), in a below-average amount of 7%.

Alcohol binge drinking (5 or more drinks in 24h) of adult US population is 25%, while 19% of BRAWO respondents report they “drink on weekends”.

In the US population, 8% used marijuana last month (2012 data), while 9% of examined sample did. Regarding tobacco smoking, it is 22% (US population 2012) to 9% (sample).

Taken together, BRAWO respondents seem to score lower on both drug use and abuse than average adult US citizen, with the exception of smart drugs experiments.

#### (i) Self-reported effects of n-back training (categories)

Throughout the last century, intelligence research was often controversial, stirred lots of emotions, and scientists often overemphasized the importance of intelligence (or more precisely, IQ). Today, EBSCO database alone indexes more than 60,000 peer-reviewed, scientific articles from the area of psychology with “intelligence” as a keyword. Not surprisingly, this contributed to less overinterpretation, calmer discussion and better understanding of intelligence by *scientists*. Nevertheless, the public is still largely misinformed about

- How is IQ computed, and how it is not (no longer a ratio of mental / physical age, but a percentile of cognitive performance, standardized on certain population, mostly considered an interval variable, which allows adding and subtraction at best)
- What IQ relates to/correlates with, and what not (e.g. no reliable correlation to psychopathology, personality traits and so on). This means, gifted and/or motivated children may have special *learning* needs (like higher need for cognition, or more autonomous learning style), but as a group they are not different in their personality traits, psychological needs or mental health (and neither are gifted adults). Unfortunately, there seems to be demand for counselors who base their careers on the misconception that “gifted adults have special kind of psychological problems”.
- What is an appropriate interpretation of psychometric IQ (e.g. full score in RAPM, which is supposed to indicate an IQ of 155+, can go an infinitely long way to a real-world genius)

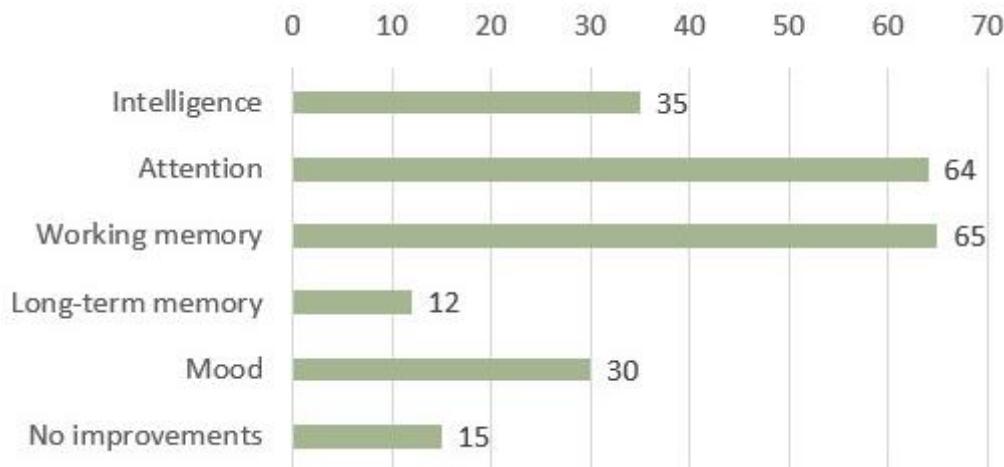
Now, considering this lack of clarity about *IQ itself*, how can the public make sense of *improving IQ* studies? Even more so, when research on the topic is considered a “swamp” even by experts in the field? The simple answer is it can’t. As a layman in psychology, hopefully you can tolerate the feeling of not knowing or acknowledging that we do not have enough evidence and wait and learn on the go. If you can’t tolerate the unknown you’ll probably pick your favorite study and fall prey to biased media coverage, or your friendly neighborhood intellectual, who has all the knowledge, intelligence, and especially confidence and time to argue about his opinions.

This happened repeatedly in the “Dual N-back, Brain Training & Intelligence” discussion group during my participant observation. This group, according to its owner Paul Hoskinson, has around 2500 members, from which 900 subscribed to the regular newsletter. There happened to be members who made up their minds about n-back very early on, ignored or belittled studies with contradictory evidence, loathed their authors and even systematically served their pseudo-neutral reviews to new members (ironically, these texts never went on to become “just another deficient, single peer-reviewed article”).

Fortunately, very few individuals behaved in such an extreme way. Eventually it seemed that this compulsive negativistic writing manifested itself in a negative narrow-mindedness of the group, and lowered the motivation of new members to give n-back a

serious try.<sup>3</sup> Sometimes, there even seemed to be a relationship to PSSI personality traits. Nevertheless, here are the BRAWO self-reports of n-back effects from those who *did* train more than a few times (N=206, less those who didn't provide answers on this particular set of questions: N=8, or 4%):

Figure 12. Subjectively felt improvements after n-back training



Redick et al. (2012) in their rigorous study also surveyed their participants on subjective improvements after n-back training. Self-reports in their study (even contrasted to self-reports against an active control group) are in concordance with BRAWO reports regarding *memory* and *attention*. But because Redick et al. (2012) in their study didn't find matching improvements in more objective, psychometric measures of these constructs, they hypothesize these reports to be an „illusory placebo effect“ (Pratkanis, Eskenazi and Greenwald, 1994). While this is perfectly possible, the question is why would these hypothetic self-suggestions appear significant only in self-reports by training group, and not in self-reports by active controls? Moreover, let's not forget the possibility, that although they claim their study to be of „high statistical power“ (and training sessions really had a high duration 30 to 40 minutes), size of their training group (N=24) is not so big, and dividing each performance test into thirds takes away a lot of ANOVA statistical power (by both increasing between-group differences and decreasing test sensitivity, as elaborated by Jaeggi et al., 2013).

---

<sup>3</sup> Based on the rationale from the experimental part of this dissertation (critical persons gain most from *single* n-back exercise), and the fact, that *dual* n-back is by far the most frequent choice of training, it is possible that an unfortunate mismatch happened: the most critical members of the group, experiencing smallest gains from dual n-back, had especially good reasons to be critical about their experience.

In BRAWO self-reports, we have only estimates of training duration, and no control group to contrast with, but still the results are interesting: only 15% of respondents chose the *No improvements* option. Even if we add to them the 4% members who didn't provide answers, it means that only one in five respondents was not aware of any improvements that they'd ascribe to n-back training.

Approximately one in ten respondents reports improvements in *Long-term memory*, but circa two thirds of sample felt improvements in *Working memory*, as well as in *Attention*. In addition one third of participants felt their *Mood* improved, and approximately the same amount noticed improvements of *Intelligence* (or more precisely, their implicit theory thereof).

#### (j) Self-reported effects of n-back training (qualitative aspects)

Positive experiences reported in previous item, were quite often (in 35% of all cases, N=73) supplemented by qualitative answers. After a brief qualitative content categorization, these points are the most frequent and / or noteworthy to mention:

- Surprisingly, the most frequently mentioned qualitative phenomenon (N=23) was *change in sleep quality and quantity*. Relevance of this discovery is supported by the fact that sleep was not mentioned anywhere in the questionnaire, yet every third respondent who reported some qualitative aspects mentioned it. Some wrote about vivid, sometimes even lucid dreams, better ability to recall dreams, and several respondents rated their quality of sleep as higher. On the other hand, two respondents said n-back made them sleepy, although they had less problems waking up in the morning. Examples:
  - vivid dreams / I dream more vividly while training DNB. / strange vivid dreams
  - I've been experiencing some strange effect of DnB on my dreams, they've been more vivid and I could remember them more easily, it has happened almost every time I did a 20 minute session, so I think that rationalization is valid.
  - ...and in my dreams, which I recall much more readily than before, I seem to enjoy much more complex narratives, to say the least.
  - I sometimes have very detailed dreams - it didn't happen before I took up training.
  - I sleep more deeply and dream more vividly the nights that I train DNB.

- The second most frequent qualitative report was *changed quality of thinking* (N=18). Respondents described that their thinking is somehow of higher quality (clearer / more organized / quicker), and that it felt natural. On the other hand, some of them explicitly stated they didn't notice a higher quality of thinking. Examples:
  - Mental clarity / clarity of mind / clarity after doing the task / my thoughts seem clearer and more organized
  - Understanding complex things faster / Able to handle more complicated thoughts at once / Thought process is generally faster / just the same but faster in my mind
  - It has taught me a few, very hard to describe, ways of thinking similar to new qualia of thought / easier to follow others arguments in discussions and see gaps
  - However, I think it is important to emphasize, since I have not experienced any increases in this area, that depth of thought has not improved as much as I would have hoped. / Not sure about improvement in problem-solving ability. / Easier to visual ideas (however, the quality of ideas has not improved)
  - I can easily create new concepts based on the knowledge already acquired.
  - I'm finding it easy to solve mathematical problems and questions involving logic and comprehensive interpretation
  - The most noticeable effect I could mention regards my ability to deal with bigger amounts of information and organize them in much more coherent, comprehensive and still complex picture. ... I feel that when I am training consistently, the feedback I give to the students at the end of the sessions is, as I said, much more comprehensive and, even though, more complex than when I am off training
- Thinking was closely followed by improvements in *memory*. Examples:
  - Memory / working memory / my memory improved a lot / noticeable changes in short term memory
  - I attribute this through increases in working memory (leading to easier understanding of conceptual material) and increased attention span.
  - I got more conscious about how I use my working-memory and where my deficits might be / developing an understanding how interference kills working memory
  - It has helped enormously with memory-intensive (eg. most) college courses

- my memory being so much better, I get easily frustrated by having to repeat things to people multiple times
- Similarly pronounced change was *increased quality of attention*. Examples:
  - I notice that I actually feel very focused after n-back / 3. attention gains / increased attention span / ability to concentrate/focus / I am a lot more focused now. / better and more control over attention
  - In general, attention and focus improve whenever I have the chance to train regularly.
- *Improved language skills* were mentioned often too – like reading comprehension, writing skills and speech fluidity / eloquence. Examples:
  - Better at writing. / Better verbal fluency (vocabulary, speed). / I could talk faster and more fluently
  - Writing and reading becomes easier as one can keep track of long trains of reasoning and remember key points.
  - I've noticed an improvement in my reading comprehension. I've only trained for about a month yet feel a significant effect on my daily life
  - I can defend myself verbally much better, I'd say.
  - I've had sudden spikes of heightened verbal fluency... The little bit of Spanish that I have has become more useful with no additional practice.
  - Ability to read faster and with better efficiency. Greatly improved verbal fluency.
  - More articulate, explaining difficult concepts and debating becomes easier. Also vocabulary access and recall improves which is noticeable in verbal and written communication
- Many times, BRAWO respondents reported an *increase in energy, motivation, a need for new information and experiences, sometimes creativity*. Examples:
  - More motivation / It serves the purpose of a caffeine substitute / I became more energized after. / Better self-discipline.
  - I always thought being alert would be strenuous, now that I myself am alert, too, I notice it isn't strenuous at all.
  - Creativity / My creativity has increased / increased creativity / I definitely feel more creative / Made me more creative.

- does not impede creativity in my case. I thought it would but it doesn't. I'd say without doubt that I'm very creative and even ADD--I'm just as creative after many months of dual-n-back
  - A better understanding of mathematics and a stronger desire for learning it. / My mathematical ability (and inclination to pursue difficult intellectual work) also improve. / Became more interested in things I wouldn't normally be interested in, like advanced math. This could be placebo. /
  - More enthusiastic about learning. / I act on my curiosity and read a lot more.
  - Didn't see any noticeable improvement in reasoning ability (my primary motive for n-backing), information processing speed and focus/concentration. I have also occasionally reached D11B although I can sustain D8B for long periods. Significant changes in personality traits- conscientiousness, openness to new experiences, appreciation of art, feeling very very intense 'chills' listening to classical music
- *Changes in mood* were not an exception among those who responded to qualitative section. Examples:
  - Mood / General mood lift, maybe because of the other effects.
  - Another experience is that I felt an improvement in my mood too, I feel more confident and happy in many ways.
  - It may have actually worsened my mood -- having better memory comes at the cost of remembering things you might have preferred to forget.
  - Clear my mind = less stress and anxiety. / lowered my depressive moods / It is great for Anxiety Patients I feel, as I am one.
  - Sleep, dreams, and after the most intense sessions, humor
  - The quality of my sarcasm! Strange, but a noticeable change.
  - I make more jokes and it lowers the social anxiety I can occasionally have.
- *Negative effects*. These were mostly present as tiredness and unpleasant sensations after training. Complete enumeration of negative reports follows:
  - Sometimes feels like my mind gets fatigued after too much training which leads to worsened cognitive ability
  - I got headaches from doing dual n-back then and they seem to have returned recently.
  - One doesn't need to sleep as much after intense training
  - Increased need for sleep after days in which I dual-n-back.

- zoning. gluing to something - sometimes a bit annoying.
- Music sticks in my mind much more now than it used to -- the 'getting a song stuck in your head' problem occurs much more now as a result.
- If I n-back too many days in a row I am very mentally fatigued, but taking time off makes me extremely energetic.
- Tingling in my head.
- It's given me some clues about psychological issues (distracted attention because I'm too invested in getting things right), but no solutions
- Higher dream recall (might be a good or a bad thing, depending on what you dream of).
- Sometimes I might be a little bit too "attentive" or "lucid" and that's a bad consequence when you're in a discussion, as you get bored easily, I can't talk about the weather usually but when I do DNB i tend to speak only of "important" subject.
- Less patience with stupidity and banality
- Feel that the effects decrease after a long hiatus, however, they still remain above the baseline set before any training was done

All of these reported effects are obviously just some (repeated) anecdotal evidence. Nevertheless, they echo and extend the answers to previous, quantitatively stated questions in BRAWO questionnaire. In addition, I saw similar experiences being reported right in the group, and had a similar one myself:

I did n-back quite intensively for approximately 6 months (about 30 minutes a day, triple-n-back, high interference trials, no feedback about errors, no breaks after each minute, but once per 10 minutes). Subjectively, I'm not sure after what time I felt changes I'm going to describe, but eventually they became clearly pronounced, and I'm very sure it wasn't due to a placebo effect (or not just because of it):

- I felt more „fresh“ cognitively – same cognitive challenges required less energy to complete
- I felt kind of „intellectual hunger“ – a need to improve my *general* knowledge / education. I started to “explore things” more, read overviews and enjoy topics I considered important/valuable, but never really interested me before (e.g. history, classical music and classical art)
- My academic writing definitely improved

- The time I did n-back, was the only time so far I could beat my father in scrabble 😊
- My mood was slightly, but noticeably more positive and stable throughout the day
- I started to run every other day, and I've kept this habit until this day

There were some other effects that were not that pronounced and / or they could be attributed purely to the placebo effect (social interactions felt more “smooth”, and I think my creativity improved). I did not measure my fluid intelligence before I started n-back (didn't really think it would make much of a difference), but when I tested myself after two months, I was definitely satisfied with my performance.

The only negative effect I remember, was that once or twice a day I felt clearly more anxious than the rest of the day. It was sometimes before training, or just at random times, and it lasted for few minutes. It was not too strong, or limiting in any way, but definitely noticeable and mildly unpleasant.

Why did I stop doing n-back? After 6 months of training, I got sick for two weeks, and had to abruptly stop both running and doing n-back. After these two weeks, I got a bit lazy and started to enjoy “vacation” from both mental and physical exercise. After another two weeks, I started to experience weird negative psychological effects, which I attributed to “runners blue” – to coming off of healthy addiction to both intensive exercises. For about one week, I became frustrated very easily, I experienced mood swings and general discontent, and I was aware the entire time that I had no appropriate reason for this. After few days it went away as if nothing happened, but it was quite an unpleasant experience. And although I returned to running in few days, I never really returned to proper n-back training – partly out of laziness, partly out of respect for these abrupt-ending after-effects. Nevertheless, a year ago (three years after my n-back experience), I adopted daily mental training (15 minutes) at lumosity.com. It is of course much easier than n-back, but games are more diverse, enjoyable, and I experienced some cognitive improvements too (as measured by luminosity, after 9 months of training circa 10% improvement, although the slope of improvements is becoming continually less steep).

I'm absolutely aware this is just my experience, surely depending on my genetics, personality and environment. But considering all the qualitative evidence, the general notion here is that n-back definitely *can* make a difference – even a considerable one. And if it does, it is change for the better, in great concordance with research, and with minimum side effects.

### 3.04 Bivariate analysis

#### (a) Multiple comparison problem

In the BRAWO questionnaire, I asked each respondent to provide 38 pieces of information. This means that the raw data was differentiated into 38 variables, including 3 variables that contained qualitative reports (*IQ level justification, Other subjective effects of training* and *General questionnaire feedback*). During univariate analyses, the number of variables increased to over 60, because of transformed variables, weighting, filter and aggregate variables. Among 60 variables, there are

$$N = \frac{60(60-1)}{2} = 1770$$

relationships. Even among 35 basic quantitative variables, there are still 595 relationships. If we performed 595 tests at  $\bar{\alpha} = 0.05$  level, it would increase our chances of getting a false positive to

$$1 - (1 - 0.05)^{595} = \sim 99, \bar{9}\%$$

In other words, if we checked for all the relationships between 38 variables (even if they were filled with random data), we are also bound to find some rare events (some statistically significant relationships), which are nevertheless caused by chance. If we want to keep the *overall* probability of Type I errors (probability of incorrectly rejecting the null hypothesis for the whole experiment) at  $\bar{\alpha} = 0.05$  level, then we have to lower partial  $\bar{\alpha}$  for each of the 595 individual tests accordingly. This is the problem of multiple comparisons, and it can be dealt with in several ways:

- Bayesian statistics
- Bonferroni correction
- Other multiplicity corrections
- Pre-select the most important variables / prioritize hypotheses, and report the results in context and in priority order

Unfortunately, I'm a stranger to Bayesian statistics. The second option, Bonferroni correction, is the simplest one available: it means lowering all the partials in such a way, that their product doesn't exceed the original confidence level (which is sometimes called the *Familywise error rate*). This means dividing the original alpha by the number of individual tests:

$$\bar{\alpha} = \frac{0.05}{595} = 0.000084$$

The problem with Bonferroni correction is that it is very conservative – in our case, for a result to be significant we would need it to meet a specific level  $\bar{\alpha} = 0.000084$ .

Additionally, Perneger (1998) importantly states, that having a wide “family” of relationships in an experiment means that the probability of type II errors (failing to reject a false null hypothesis – i.e. that truly existing differences will be deemed non-significant) increases as well. Nevertheless, applying the Bonferroni correction lowers the statistical power of a test, thus further increasing the probability of type II errors. Bonferroni is quite counterintuitive too – with 10 tests each with  $\bar{\alpha} = 0.04$ , none of them is deemed significant when testing with Bonferroni correction at the overall  $\bar{\alpha} = 0.05$  level.

There are other popular multiplicity correction methods, like Holm–Bonferroni, Fisher LSD, Tukey's HSD, Duncan's New Multiple Range Test and Newmann-Keulls, which do not deprive tests of so much power. The most advanced method (False discovery rate, FDR) deals with the ranking of all individual tests in a given “family” by their  $\bar{\alpha}$  level, and computing a final cut-off level based on certain position in this rank.

Nevertheless, I do not consider these (generally undiscriminating) multiplicity corrections to be the best method of dealing with multiple comparison problems in the case of the BRAWO questionnaire (although I still consider them relevant as supplementary information). The main reasons for this are:

- These corrections actually do not follow the basic statistical principles and are not thoroughly justified
- Number of variables in BRAWO questionnaire is really high, which means any corrections would probably lower  $\bar{\alpha}$  to an unrealistic level
- Not all of the variables are equally loaded with data relevant to the topic of this dissertation. E.g.
  - Variable *Time of group membership* is of peripheral importance to actual cognitive training
  - Variable *Minutes per one session* is secondary to *Total training time*
  - Higher and lower IQ estimates are both secondary to *Total IQ estimate*
- Not all hypotheses are equally meaningful regarding our research purposes (even when we use only the most “important” variables). E.g.:
  - Relationship (*Total training time, Age*) is less relevant to our research topic, than the relationship (*Total training time, Subjectively improved working memory*)

From this perspective, it is necessary to single out the variables that are the most loaded with information relevant to our research topic, and to enumerate and prioritize the hypotheses that are the most meaningful in regard to our research purpose. So here is the list of the most research-relevant variables (in no particular order):

- *N-back aggregate level* (ordinal)
- *Total training time* (ordinal)
- *IQ estimate* (interval)
- *IQ Mindset* (ordinal)
- *Self-reported improvement of IQ* (binomial)
- *Self-reported improvement of mood* (binomial)

We are down to 6 variables, which still yields 15 relationships (or 30 unidirectional hypotheses). After rigorously weeding them out and sorting them by research relevance, I ended up with this prioritized list of 6 hypotheses:

1. H1: There is a directly proportional relationship between *Total training time* and *Self-reported improvement of IQ*
  - This is prevalent view of studies which propose far-transfer
2. H2: There is a directly proportional relationship between *Total training time* and *Self-reported improvement of mood*
  - Rationale analogical to previous hypothesis
3. H3: There is a directly proportional relationship between *IQ Estimate* and *N-back aggregate level*
  - As elaborated by Jaeggi et al. (2010)
4. H4: There is a directly proportional relationship between *N-back aggregate level* and *Self-reported improvement of IQ*
  - As elaborated by Jaeggi et al. (2011)
5. H5: There is not a proportional relationship between *IQ Estimate* and *IQ Mindset*
  - *IQ mindset* can be considered a personality trait, therefore probably independent from general IQ (Demetriou, Kyriakides, & Avraamidou, 2003; Farsides & Woodfield, 2003; Sobelet & Salthouse, 2011)

(b) H1: There is a directly proportional relationship between *Total training time* and *Self-reported improvement of IQ*

- Variable 1
  - *PartialAggregateTime* – ordinal variable. Arbitrary number, computed as “number of weeks of training \* minutes per session”. Unfortunately, we do not have weekly frequency to add to the equation (although presumably most of the respondents trained more than once per week), which decreases statistical power. Still, it is the most reliable piece of information regarding total training time at the time of BRAWO completion.
- Variable 2
  - *SubjectivelyImprovedIntelligence* – Binomial variable. Self-report regarding one's increased IQ due to n-back training.

I. Statistical method

- Nonparametric, 2-tailed. Spearman's rank correlation coefficient,

II. Assumptions, sample pooling, outliers

- Assumptions met. Pooling is necessary – for checking effects of training, we need to select only respondents who trained at least a few times. No need to filter outliers, they are limited to the value of their rank by the test itself.

III. Result & interpretation

- $r_s(196) = .193, p < .001$
- H1 confirmed.
- Based on the results of BRAWO self-report questionnaire, respondents who've trained with n-back for a longer time report more often that their intelligence has improved. *Longer n-back training (any kind) is linked to self-reported IQ gain.*
- See interpretation of disproved H8 (p. 69) for further context

(c) H2: There is a directly proportional relationship between *Total training time* and *Self-reported improvement of mood*

- Variable 1
  - *PartialAggregateTime* – ordinal variable. Arbitrary number, computed as “number of weeks of training \* minutes per session”. Unfortunately, we do not have a weekly frequency to add to the equation (although presumably most of the respondents trained more than once per week), which decreases statistical

power. Still, it is the most reliable piece of information regarding total training time at the time of BRAWO completion.

- Variable 2
  - *SubjectivelyImprovedMood* – binomial variable. Self-report regarding one's increased mood due to n-back training.
- Statistical method
  - Nonparametric, 2-tailed. Spearman's rank correlation coefficient.
- Assumptions, sample pooling, outliers
  - Assumptions met. Pooling is necessary – to check the effects of training we need to select only the respondents who trained at least a few times. No need to filter outliers, they are limited to the value of their rank by the test itself.
- Result & interpretation
  - $r_s(196) = .028, p < .699$
  - H2 disproved.
  - Based on the results of the BRAWO self-report questionnaire, it is not true that respondents who have trained n-back for a longer time (more weeks), report more often that their mood has improved. *Longer n-back training (any kind) is not linked to self-reported mood improvement.*

(d) H3: There is a directly proportional relationship between Self-reported IQ and *N-back level*

- Variable 1
  - *IQ Estimate* – interval variable. Self-report regarding one's IQ.
- Variable 2
  - *N-back aggregate level* – ordinal variable. User's usual n-back training level (taking into consideration version of n-back, which is important). *N-back aggregate level* is a relatively diverse number, which is nevertheless arbitrary, therefore the variable has to be ordinal.<sup>4</sup>
- Statistical method
  - Nonparametric, 2-tailed. Spearman's rank correlation coefficient.
- Assumptions, sample pooling, outliers

---

<sup>4</sup> Another option would be to limit the version of n-back to the most frequent one (dual), and use directly raw *User training level* – but this would “cost” us N=50, the variable would still need to be ordinal, and we would get very similar results.

- Assumptions met, no pooling, no need to filter outliers, they are limited to the value of their rank by the test itself.
- Result & interpretation
  - $r_s(255) = .294, p < .001$
  - H3 confirmed.
  - Based on the results of BRAWO self-report questionnaire, respondents who report higher IQ's, report training on more difficult level of n-back (taking into account the version of n-back). *Higher self-reported IQ is linked to training on more difficult levels of n-back*. This reflects the notion (in a limited scope), that the n-back task has some psychometric properties, compare Jaeggi et al. (2010) and Redick & Lindsey (2013).

(e) H4: There is a directly proportional relationship between *N-back level* and *Self-reported improvement of IQ*

- Variable 1
  - *N-back aggregate level* – ordinal variable. User's usual n-back training level (taking into consideration version of n-back, which is important). *N-back aggregate level* is a relatively diverse number, which is nevertheless arbitrary, therefore the variable has to be ordinal.<sup>5</sup>
- Variable 2
  - *Subjectively Improved Intelligence* – Binomial variable. Self-report regarding one's increased IQ due to n-back training.
- Statistical method
  - Nonparametric, 2-tailed. Spearman's rank correlation coefficient,
- Assumptions, sample pooling, outliers
  - Assumptions met, no pooling, no need to filter outliers, they are limited to the value of their rank by the test itself.
- Result & interpretation
  - $r_s(239) = .292, p < .001$
  - H4 confirmed.

---

<sup>5</sup> Again, another option would be to limit version of n-back to the most frequent one (dual), and use directly raw *User training level* – but this would “cost” us N=50, variable would still need to be ordinal, and we would get very similar results.

- Based on the results of BRAWO self-report questionnaire, respondents who report training on more difficult levels of n-back (taking into account version of n-back), report more often that their intelligence has improved. *Training on more difficult levels of n-back is linked to higher IQ gains.*

(f) H5: There is not a proportional relationship between Self-reported IQ and IQ Mindset

- Variable 1
  - *IQ Estimate* – interval variable. Self-report regarding one's IQ.
- Variable 2
  - *IQ Mindset* – ordinal variable. Self-report regarding one's beliefs about possibility of IQ improvement.
- Statistical method
  - Nonparametric, 2-tailed. Spearman's rank correlation coefficient.
- Assumptions, sample pooling, outliers
  - Assumptions met, no pooling, no need to filter outliers, they are limited to the value of their rank by test itself.
- Result & interpretation
  - $r_s(255) = .218, p < .001$
  - H5 confirmed.
  - Based on the results of BRAWO self-report questionnaire, respondents with higher self-reported IQ believe in higher possibility of IQ improvement.  
*People with higher IQ's believe in greater malleability of intelligence.*

(g) Discussion

When interpreting these findings, it is important to be aware of the following limitations:

- BRAWO is an anonymous, self-reported online questionnaire, with self-selected sample from an “IQ training” group – all the biases mentioned at the beginning of analysis can be expected, plus a bias towards a liberal IQ mindset.
- There is some chance that people with negative training experiences tend not to be members of the group (although this was not what I observed as a participant).
- Confirmed correlations (H1, H3, H4, and H6) say nothing about causality or directionality of the link. Therefore it is e.g. possible, that

- people with higher IQ's tend to train on higher N-back levels not because of their IQ, but because e.g. their persistence in training. Or maybe there is some other reason largely independent from variables examined in this study.
- *Total training time* does not lead to *self-reported IQ gains*, but vice-versa: people with a tendency to believe that their IQ has improved, tend to train longer on n-back.

Nevertheless, there are several reasons why not to discard these findings in advance:

- Design of BRAWO aimed for less precise, but more reliable data gathering (category questions)
- All the entries were carefully checked for erroneous or nonsensical records, while still leaving sample size of more than N=250
- Univariate statistics show normal (or US population comparable) distributions of several personal and training statements, which support the notion that the sample as a whole is quite representative.
- Multiple comparison problems were addressed and dealt with in a strict way (only the 6 most important hypotheses were tested, from all the wide possibilities)
- Modest statistical procedures were used (nonparametric, rank-based Spearman's Rho), with transparent pooling in inevitable cases only
- Findings from BRAWO bivariate analyses fit well with
  - BRAWO qualitative reports
  - presented experimental study
  - previous research in the field (see e.g. metaanalysis Au et al., 2014)

## Chapter IV. BIQ questionnaire

### 4.01 About

During my doctoral studies, with my mentor prof. Urbánek we were awarded two research grants:

- “Different versions of N-back task: comparing efficiency in improving fluid intelligence scores”, 2011-2012 (GAČR P407/11/2130)
- “Personality traits as mediators of the cognitive training effectiveness”, 2013-2014 (GAČR13-36836S)

All related experiments took place under the The Institute of Psychology of Academy of Science of Czech Republic, in casual cooperation with Department of psychology, Faculty of Social Studies in Masaryk University. As both institutions are located in the city of Brno and all experiments included self-report questionnaires, these were joined into one BIQ questionnaire (Brno InQuiries).

### 4.02 Preparation and limitations

In our early experiments, we made self-reports voluntary and anonymous. Because of this, we gathered  $N=97$  entries altogether, but approximately two fifths of them ( $N=38$ ) answered anonymously, so we could not attribute psychometric results to these participants (i.e. age, gender, IQ tests and personality tests). Nevertheless, for  $N=59$  participants it was possible to connect all the data, and self-reports from all ( $N=97$ ) participants include

- motivation to participate in experiment
- improvements after n-back training
- impairments after n-back training
- qualitative reports
- use of strategies during n-back training

and some other details. All  $N=97$  entries are valid and will be included in analysis (although not every variable will be reported, see below). In contrast to BRAWO questionnaire, fewer biases fall into consideration in BIQ self-reports:

- Social desirability (tendency of some people to respond to items more as a result of their social acceptability than their true feelings)

- Implicit theories and illusory correlations (beliefs about the covariation among particular traits, behaviors, and/or outcome)
- Mood, especially as an actual state (relatively recent mood-inducing events to influence the manner in which respondents view themselves and the world around them)

In addition, these biases are suspected to be less pronounced, because

- We recruited the participants through several channels (several social networks, university information system, e-mail, paper ads), so the sample is quite diverse regarding age, gender, and basic population characteristics
- Some participants were paid, some were not, and each knew and agreed with his/her payment status before they enrolled in study. This allows for diverse motivation in BIQ sample
- Participants signed an informed consent contract, self-reports were gathered after the study, and we emphasized that we need and appreciate their honest answers to all questions

#### 4.03 Univariate analysis

There were 87 variables in BIQ data sample. These were excluded from *univariate* analysis:

- Variables related to organization of the study (ID number, version of subtests, recruitment feedback, evaluation of organization, and so on)
- Most of the variables which were known only for 3/5 of the sample:
  - Gender, age
  - Both pretest and posttest: RAPM IQ, BOMAT IQ
  - Both pretest and posttest: 14 PSSI traits, 5 NEO-PI traits

We think that IQ's are interesting despite the fact that they are available only for 3/5 of the sample, but unfortunately both RAPM and BOMAT do not have reliable norms to compute IQ for our population, and in addition we used only half-split versions of these tests.

Nevertheless, we will use raw scores from both tests as ordinal variables in bivariate analyses, as well as normalized z-score of IQ gains from both tests too. For now, we are left with 37 variables – 28 binomial, 4 interval, 3 qualitative and 2 ordinal.

### (a) Motivation to participate

Participants could check none, one or both answers to a question “I wanted to participate because of”

- *Financial compensation*: chosen by 82% of participants
- *Curiosity / chance for cognitive improvement*: chosen by 25% of participants

Internally motivated participants are prevalent in our sample.

### (b) Self-reported positive effects of n-back training (quantitative summary)

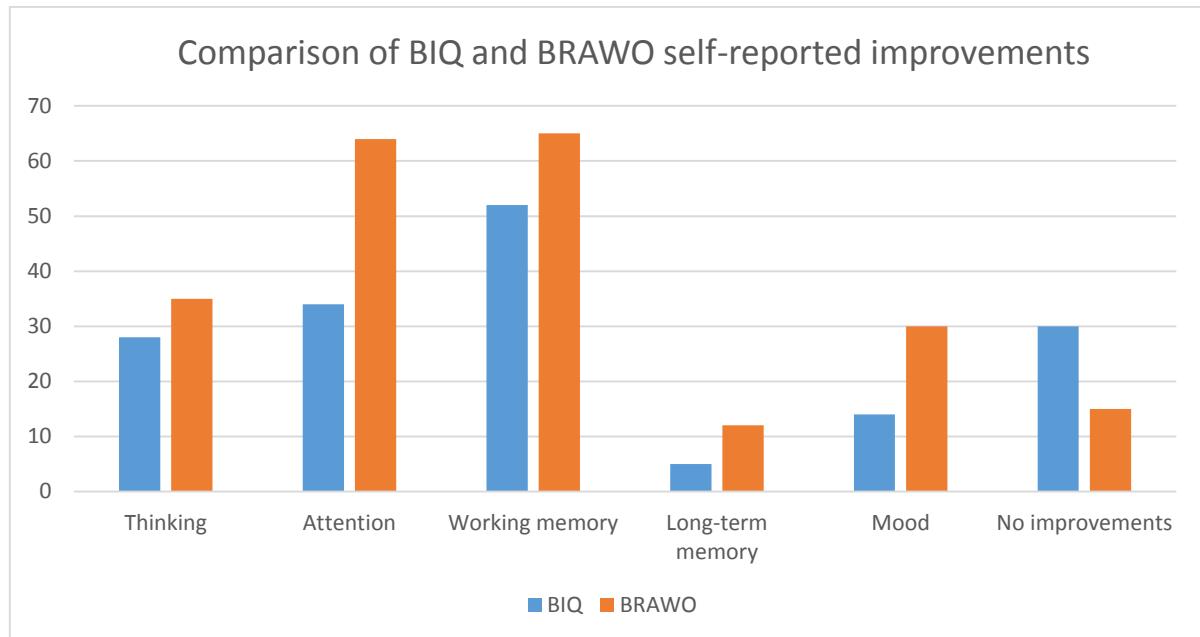
Having learned from BRAWO questionnaire, I increased the number of options in “Improvements” question:

*Table 15. Descriptive statistics - Self-reported positive effects of n-back training*

Variable	N	YES		NO	
		Frequency	Percent	Frequency	Percent
Subjectively improved thinking	97	27	<b>28</b>	70	72
Subjectively improved attention	97	33	<b>34</b>	64	66
Subjectively improved working memory	97	50	<b>52</b>	40	48
Subjectively improved long-term memory	97	5	<b>5</b>	92	95
Subjectively improved mood	97	14	<b>14</b>	83	86
Subjectively improved curiosity	97	7	<b>7</b>	90	93
Subjectively improved creativity	97	9	<b>9</b>	88	91
Subjectively improved impulse control	97	2	<b>2</b>	95	98
Subjectively improved will to perform	97	18	<b>19</b>	79	81
Subjectively improved lifestyle	97	2	<b>2</b>	95	98
Nothing subjectively improved	97	30	<b>31</b>	67	69

Each option was thoroughly specified in the BIQ questionnaire (in Czech language). Here we can see the comparison of relevant percentages in BIQ versus BRAWO sample:

Figure 13. Comparison of BIQ and BRAWO self-reported improvements



We can see, that:

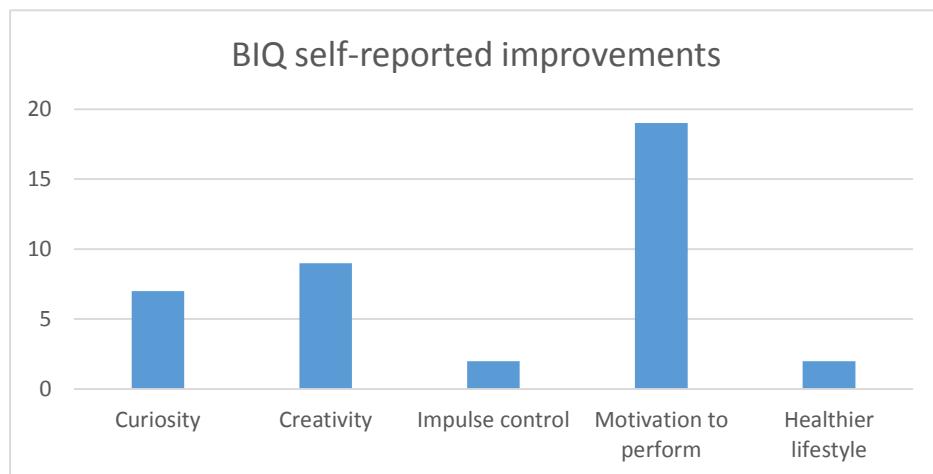
- The order of reported improvements is exactly the same in both questionnaires, namely
  - Working memory
  - Attention
  - Intelligence / thinking
  - Mood
  - Long-term memory
- In each area, BRAWO respondents report improvements more frequently than BIQ participants. The most likely explanation:
  - While it is complicated to estimate the average *Total training time* in BRAWO sample, it is very probably higher than the *Total training time* of BIQ sample (which is at maximum 25 minutes x 25 days, approximately 10 hours). If n-back training improvements are dosage dependent (which is the prevalent view), it is perfectly appropriate for BRAWO respondents to report improvements more often.
- Differences in improvements of *Attention*, *Mood* and *Long-term memory* seem to be disproportionate. The most likely explanation:
  - While BIQ participants without mental health issues and who weren't on medication were chosen (which means their attention and mood is near their

optimum), 30% of BRAWO respondents take either prescription antidepressants (suspected *Mood* issues), artificial nootropics, or prescription stimulants (suspected *Attention* issues). Because it is presumably harder to improve abilities of healthy adults, BIQ reports are more modest (and realistic, for that matter).

- *Long-term memory* is the only area in question, in which improvements take longer time to become aware of. Therefore longer training times could strongly favor improvements in this area (i.e. the self-reports of improvements could be non-linear in time)

Here is how often respondents reported improvements in BIQ-specific variables (in percent<sup>6</sup>):

*Figure 14. BIQ-only, self-reported improvements*



We can see that appreciable 19% of respondents reported improvements in *Motivation to perform* (which was further specified in questionnaire as „I have a stronger will, I procrastinate less“). Improvements in *Impulse control* and *Healthier lifestyle* were both negligible, and improvements in *Creativity* and *Curiosity* seldom (one person in ten or fifteen).

### (c) Self-reported negative effects of n-back training (quantitative summary)

To get a more complete picture of training effects, respondents were asked if they noticed any impairments as a result of n-back training (in the same areas):

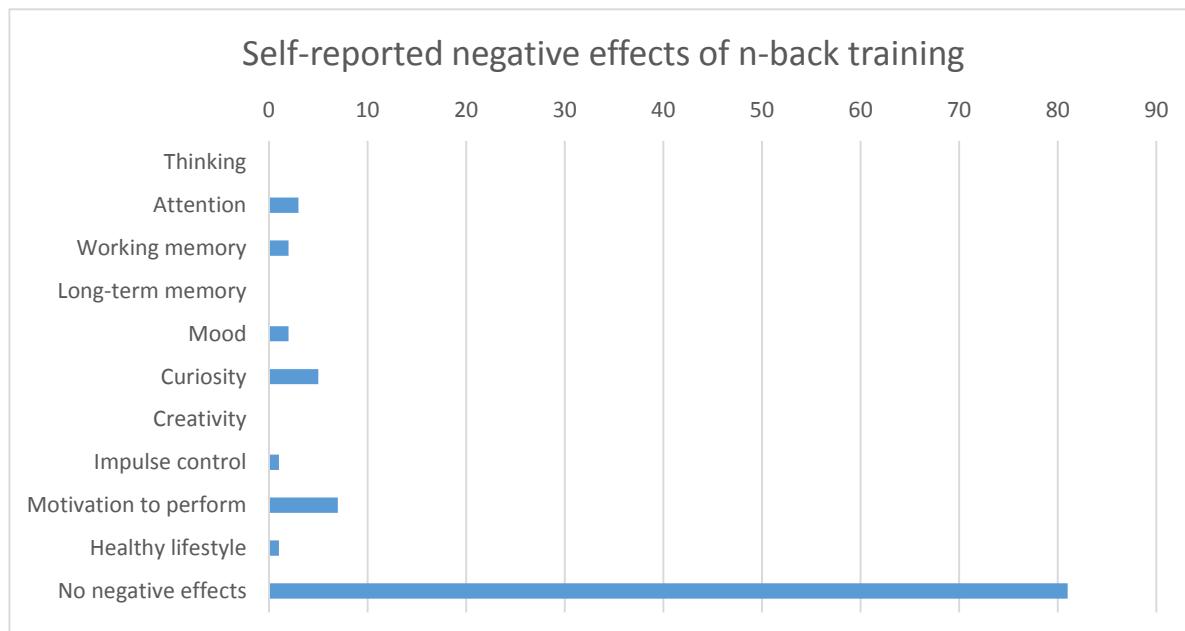
---

<sup>6</sup> Different measure of vertical axis used for visual reasons.

Table 16. Descriptive statistics - Self-reported negative effects of n-back training.

Variable	N	YES		NO	
		Frequenc y	Percent	Frequen cy	Percent
Subjectively declined thinking	97	0	0	97	100
Subjectively declined attention	97	3	3	94	97
Subjectively declined working memory	97	2	2	95	98
Subjectively declined long-term memory	97	0	0	97	100
Subjectively declined mood	97	2	2	95	98
Subjectively declined curiosity	97	5	5	92	95
Subjectively declined creativity	97	0	0	97	100
Subjectively declined impulse control	97	1	1	96	99
Subjectively declined motivation to perform	97	7	7	90	92
Subjectively declined lifestyle	97	1	1	96	99

Figure 15. Self-reported negative effects of n-back training



Strong majority of BIQ participants explicitly stated, that they didn't experience any decline in their life. A few of them reported decreased *Motivation to perform*<sup>7</sup> and *Curiosity*, other complaints were relatively negligible.

#### (d) Self-reported difficulty of training

N-back is considered to be one of the most WM taxing cognitive exercises available (if not the most). The person doing it needs to constantly update several items in WM, while regularly retrieving the one which lies at the distant end of one's WM capacity. This process requires a great deal of attention, inhibition (of items which are no more valid to the task), and dealing with significant interference (of still valid items).

When explaining the principles of n-back to participants, most of them are usually worried about the difficulty of exercise and/or their own performance. And after they give n-back a try, they use metaphors like „mental sprint“ or „insanely hard“, to express the opinion that they can't train for more than few minutes, or even think they are „dumb“ because of having problems with dual-2 back.

Nevertheless, having problems with dual 2-back in the beginning is perfectly normal – according to studies which reported n-back specific improvements (in n-back naive participants), it seems that even the most competent individuals have to start at dual 2-back or dual-3 back. It is actually striking that after few hours of training, participants were able to double or triple their baseline performance in such an extremely demanding mental exercise (despite developing and using strategies). If one would indeed compare n-back to running, it could probably be best compared to long-distance training, not sprinting (you can exponentially increase your running distance, but not your sprinting speed). Then again, it is not clear exactly how differences in the cognitive load of e.g. *dual 6-back* and *quad 3-back*, although they theoretically require the updating of the same number of items in WM.

To examine the perceived level of difficulty of n-back training, we asked BIQ participants:

- How hard was it for you to motivate yourself to daily n-back training (on average)?
  - 1 = absolutely undemanding, 7 = extremely demanding
- How demanding was training for you?
  - 1 = absolutely undemanding, 7 = extremely demanding

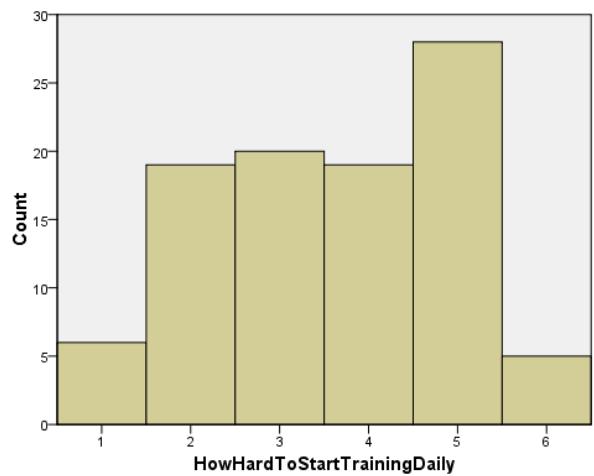
---

<sup>7</sup> This small figure can be contaminated by procrastination of training itself and aversion to it – some participants perceived the training as quite difficult and dull, as reported below.

Table 17. Descriptive statistics - How hard is it to start n-back training daily?

Figure 16. How hard is it to start n-back training daily?

	Frequency	Percent	Cumulative Percent
Valid 1	6	6,2	6,2
2	19	19,6	25,8
3	20	20,6	46,4
4	19	19,6	66,0
5	28	28,9	94,8
6	5	5,2	100,0
Total	97	100,0	



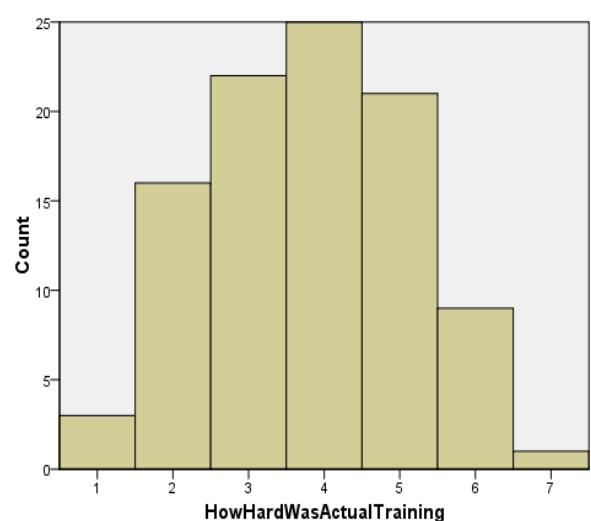
No person chose the most intense answer to first question (weighed mean  $\bar{X} = 3.8$  with SD = 1.2) – it seems, that while it is not easy to motivate oneself to train, it is not extremely hard too.

More surprisingly, answers to second question were similarly modest (weighed mean  $\bar{X} = 3.9$  with SD = 1.3). One of the possible explanations is, that in retrospect (BIQ questionnaire were gathered after the whole training was completed) participants were adapted to the difficulty level.

Table 18. Descriptive statistics - How hard was actual training?

Figure 17. Descriptive statistics - How hard was actual training?

	Frequency	Percent	Cumulative Percent
Valid 1	3	3,1	3,1
2	16	16,5	19,6
3	22	22,7	42,3
4	25	25,8	68,0
5	21	21,6	89,7
6	9	9,3	99,0
7	1	1,0	100,0
Total	97	100,0	



### (e) Use of mental strategies in n-back training

One of important topics regarding the effectiveness of n-back training (discussed especially in its early days), was the use of mental strategies (e.g. using inner voice to repeat auditorially presented letters). Jaeggi et al. (2008) claimed, that an important feature of dual n-back is, that it makes the reliance of mental strategies quite hard, therefore forcing users to exercise their fluid abilities. Morrison & Chein (2011) approached this topic by differentiating between *strategy training* and *core training*, n-back falling into the latter category. Nevertheless, based on observant participation in the discussion group, it seemed to me that the use of strategies is quite common. Unfortunately, there was no data on whether this use was beneficial or impeding to potential far-transfer.

In BIQ sample, we examined the use of strategies regarding n-back training, offering several possible answers, and a text-box for qualitative addendums:

- During the training, did you use any mental strategies to remember the positions of squares / sounds / colors?
  - Yes, I counted items in my mind, or otherwise made use of my inner voice
  - Yes, I did connect the positions of squares into triangles
  - Yes, I tried to imagine series of successive squares positions as a „snake“ with a length of „N“ items
  - Yes, I tried to remembered previous „N“ items in a batch, then compare it to the new batch of „N“ items, while trying to remember this new batch
  - No, I was just trying to be aware of previous „N“ items
  - Other (please specify)

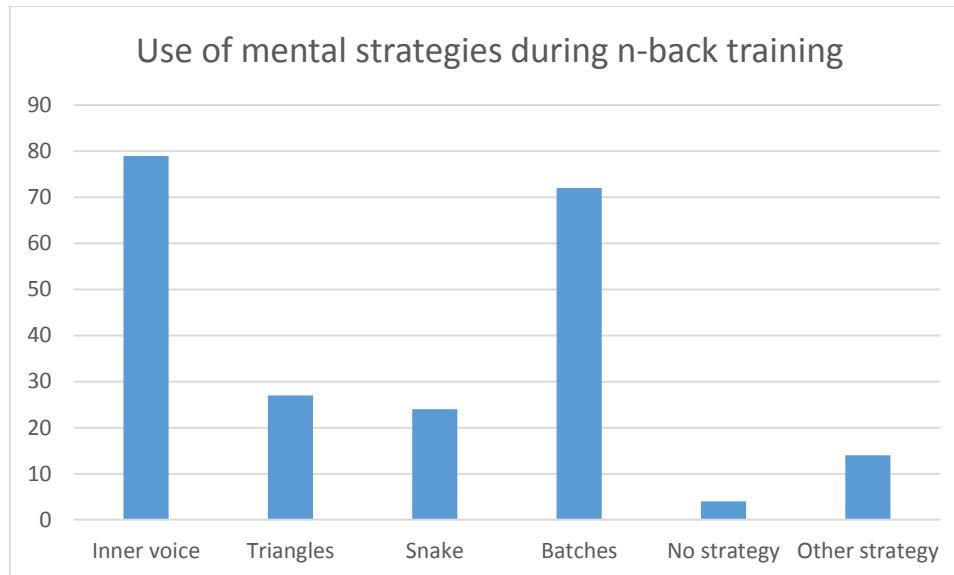
Here are descriptions of recorded answers:

*Table 19. Descriptive statistics - Use of mental strategies during n-back training*

Variable	N	YES		NO	
		Frequenc y	Percent	Frequency	Percent
Inner voice	97	77	<b>79</b>	20	21
Triangles	97	26	<b>27</b>	71	73
Snake	97	23	<b>24</b>	74	76
Batches	97	70	<b>72</b>	27	28

No strategy	97	4	<b>4</b>	93	96
Other strategy	97	14	<b>14</b>	83	86

Figure 18. Use of mental strategies during n-back training



The use of strategies was apparently quite common among BIQ participants. There's some small possibility that the question itself rendered the usage of strategy more desirable in the eyes of participants. But considering some of the creative answers in the qualitative addendum, it seems that strategy use definitely looms largely among n-backers (as translated from Czech):

- To memorize positions, I moved my head accordingly, so as to point to the squares
- I just watched the squares, as I pretended to move my head in their direction
- As „N“ increased to level 4 and 5, my strategy changed from „repeating“ to more intuitive
- I repeated the whole batch aloud, and moved my body according to square positions
- I remembered first square in the batch by looking at it, and followed the rest with my peripheral vision
- I divided batches of „N“ to pairs and triplets
- Clusters of letters remind me of words, and the rest I just figured out somehow
- I tried to teach myself the ability to assign color to the sound, and with my eyes I anticipated the position in which then next square should appear

- At the 3-back and 4-back, I successfully applied the „cache memory“ strategy – I memorized the first batch, did not actively memorize the second, but immediately after second batch ended I just „made it up“ – though this was not possible with higher levels of „N“
- At more complex levels of n-back (7 and 8) I remembered first set, half of the second set, and the rest I just subtracted from the end of a session (as it was already near)
- At higher „N“ levels, I alternated remembering triplets and quartets (set of 7 as a set of 4 + set of 3, then I alternate 4, then I alternate 3...)
- I pointed at the squares with my finger

Some of these creative strategies even preconceive the idea of „haptic n-back“<sup>8</sup>, which could offer an even more immersive experience – and a higher cognitive load. In the bivariate analyses section, we will examine how different strategies interacted with other variables of BIQ questionnaire.

#### (f) Self-reported changes after n-back training (qualitative content)

There were a few (N=10) qualitative comments on the effects of n-back:

- Every day started to seem more important and interesting, because I had the opportunity to be challenged by something, to get better in it (to succeed, to advance to another level in the exercise)
- I found the training extremely dull and I did not enjoy it. If I had enjoyed it, I think I would achieve better results in it.
- I'm sleeping harder at night
- Some of the days, I dreamt about the training
- I think that my imagination, self-esteem, and my ability to anticipate things improved, after I experienced what I'm able to memorize. My relationships with people improved thanks to better self-esteem. Although with some people, they got worse, because they started to get on my nerves with their limited view on certain things (I mean, I'm not rude to them, but I think these things to myself)
- Sometimes I was exhausted during the training (especially at the end) desperate, when I reached level 8 (I was just staring dully :D ). When I did the n-back session after

---

<sup>8</sup> As mentioned earlier, real “triple n-back” would actually need to involve third sensory modality – most probably touch, Klatzky et al. (2008). With todays vibrating game controllers and smartphones, this should be a very attainable goal.

running, I got a better score. If I didn't repeat the series aloud during the session, my score got considerably worse.

- I was sick during most of the training
- I had to organize my day to make time for training. As I have two temporary jobs (tutoring –mentally exhausting; nighttime inventorying – physically exhausting), I mostly did the training after the work. And I got the feeling, that in the night or in early morning (without sleep) I got better scores than throughout the day.
- I enjoyed the training – seeing an ascending curve of my results increased my self-esteem
- More dreams, and I better remembered them in the morning

These reports are as similarly diverse as BRAWO reports, with similar ratio of sleep/dreams references.

## 4.04 Bivariate analysis

### (a) Multiple comparison problem

As I mentioned earlier, there's 37 variables of interest in BIQ questionnaire.

Additionally, we decided to include IQ gain variables in bivariate analysis (i.e. RAMP z-score gain, BOMAT z-score gain). This leaves us with 741 relationships between all the variables, so we have to deal with multiple comparison problem again. I chose to deal with the same way as in BRAWO analysis: pre-select the most important variables, prioritize hypotheses, and report the results in context and in priority order (for rationale of this decision see section 3.04, p. 44).

Here's list of variables that are most loaded with information relevant to our research topic (in no particular order):

- *IQ gain in RAPM (z-score, scale)*
- *IQ gain in BOMAT (z-score, scale)*
- *Intrinsic Motivation* (dichotomous)
- *Extrinsic Motivation* (dichotomous)
- *Self-reported improvement of thinking* (dichotomous)
- *Strategy use – inner voice* (dichotomous)
- *Strategy use – triangles* (dichotomous)
- *Strategy use – snake* (dichotomous)
- *Strategy use – batches* (dichotomous)
- *Strategy use – none* (dichotomous)

We are down to 9 variables, which still yields 36 relationships (or 72 unidirectional hypotheses). After weeding them out, and sorting them by research relevance, I ended up with this prioritized list of 3 hypotheses:

- H6: *Intrinsic motivation* to participate in n-back training is in directly proportional relationship with *IQ gains* in both RAPM and BOMAT.
  - Au et al. (2014)
- H7: The use of different *Mental strategies* in n-back training yields different *IQ gains* in both RAPM and BOMAT.
  - Morrison & Chein (2011)
- H8: Reports of *Subjectively improved thinking* are in directly proportional relationship to psychometric *IQ gains* in both RAPM and BOMAT.

(b) H6: *Intrinsic motivation* to participate in n-back training is in directly proportional relationship with *IQ gains*.

- Variable 1
  - *IntrinsicMotivation* – dichotomous variable.
- Variable 2
  - *RapmGainZ, BomatGainZ* – Both variables are scale. Because even- and odd-numbered IQ subtests were of different statistical power, and we needed to standardize them by computing z-scores (the distance of each user's score from the mean of odd- or even-numbered items of all participants). Method of using z-scores is clearly inferior to using pre/post IQ tests with the same psychometrical properties (i.e. of the same validity, which would lead to the same level of difficulty). Although it can be hard to obtain IQ tests of identical difficulty, it is essential for any research of small pretest/posttest differences.
- Statistical method
  - The point-biserial correlation coefficient  $r_{pbi}$  (actually Pearson's  $r$ , slightly modified for determining the level of association between dichotomous and interval / scale variable)
  - Potentially logistic regression
- Assumptions, sample pooling, outliers
  - Assumptions met (scale variables come from normal population, as assessed by Shapiro-Wilk's test of *RapmGainZ*,  $p < .918$ , and *BomatGainZ*  $p < .313$ ). IQ gains are missing for 2/5 of sample ( $N=38$ ). Pooling is necessary, because some participants ( $N=7$ ) reported both extrinsic and intrinsic motivation. And because we're using a parametric test, there's need to filter 1 outlier ( $rapmGainZ > 3.1$ ), pairwise.
- Result & interpretation
  - *IntrinsicMotivation* and *RapmGainZ*       $r_{pbi}(48) = -.136$ ,  $p < .347$
  - *IntrinsicMotivation* and *BomatGainZ*       $r_{pbi}(49) = .092$ ,  $p < .523$
  - H6 disproved.
  - Based on the BIQ data, participants who entered the research project with intrinsic motivation, didn't improve their psychometric IQ more than others. And because double motivation participants were discarded from this test (i.e. every participant now has either intrinsic, or extrinsic motivation), this means

extrinsic motivation does not play role in IQ gains as well. *Motivation to participate is not linked to lower or higher IQ gains.*

(c) H7: The use of different *Mental strategies* in n-back training yields different *IQ gains* in both RAPM and BOMAT.

- Variable 1
  - *StrategyInnerVoice, StrategyTriangles, StrategySnake, StrategyBatches, StrategyNone* – All 5 variables denoting the use of mental strategies are dichotomous.
- Variable 2
  - *RapmGainZ, BomatGainZ* – Both variables are scale. Because even- and odd-numbered IQ subtests were of different statistical power, and we needed to standardize them by computing z-scores (the distance of each user's score from the mean of odd- or even-numbered items of all participants). Method of using z-scores is clearly inferior to using pre/post IQ tests with the same psychometrical properties (i.e. of the same validity, which would lead to the same level of difficulty). Although it can be hard to obtain IQ tests of identical difficulty, it is essential for any research of small pretest/posttest differences.
- Statistical method
  - The point-biserial correlation coefficient  $r_{pbi}$  (actually Pearson's  $r$ , slightly modified for determining the level of association between dichotomous and interval / scale variable)
  - Potentially logistic regression
- Assumptions, sample pooling, outliers
  - Assumptions met (scale variables come from normal population, as assessed by Shapiro-Wilk's test of *RapmGainZ*,  $p < .918$ , and *BomatGainZ*  $p < .313$ ). IQ gains are missing for 2/5 of sample ( $N=38$ ). And because we're using a parametric test, there's need to filter 1 outlier ( $\text{rapmGainZ} > 3.1$ ), pairwise.
  - Result & interpretation

Table 20. Correlations of IQ gain and Strategy use during n-back

		Strategy InnerVoic e	Strategy Triangle s	Strategy Snake	Strategy Batches	Strategy None
rapmGainZ	Pearson Correlation	-,137	,233	-,125	-,032	-,064
	Sig. (2-tailed)	,311	,082	,353	,810	,635
	N	57	57	57	57	57
BomatGainZ	Pearson Correlation	,107	,117	-,140	-,012	,352**
	Sig. (2-tailed)	,424	,382	,296	,929	,007
	N	58	58	58	58	58

- This means H7 is disproved.
- Based on the BIQ data, use of different mental strategies when training n-back does not yield different IQ gains. There is one correlation significant at  $p < .007$ , but there are only N=4 members of that group (*StrategyNone*), there's no corresponding relationship to another IQ test, and multiple comparison problem here is too pronounced. *There is no link between using certain, or any mental strategies, and IQ gain.*

(d) H8: Reports of *Subjectively improved thinking* are in directly proportional relationship to psychometric *IQ gains* in both RAPM and BOMAT.

- Variable 1
  - *SubjectivelyImprovedThinking* – dichotomous variable.
- Variable 2
  - *RapmGainZ, BomatGainZ* – Both variables are scale. Because even- and odd-numbered IQ subtests were of different statistical power, and we needed to standardize them by computing z-scores (the distance of each user's score from the mean of odd- or even-numbered items of all participants). Method of using z-scores is clearly inferior to using pre/post IQ tests with the same psychometrical properties (i.e. of the same validity, which would lead to the

same level of difficulty). Although it can be hard to obtain IQ tests of identical difficulty, it is essential for any research of small pretest/posttest differences.

- Statistical method
  - The point-biserial correlation coefficient  $r_{pbi}$  (actually Pearson's  $r$ , slightly modified for determining the level of association between dichotomous and interval / scale variable)
  - Potentially logistic regression
- Assumptions, sample pooling, outliers
  - Assumptions met (scale variables come from normal population, as assessed by Shapiro-Wilk's test of *RapmGainZ*,  $p < .918$ , and *BomatGainZ*  $p < .313$ ). IQ gains are missing for 2/5 of sample ( $N=38$ ). Pooling is not necessary. Because we're using a parametric test, there's need to filter 1 outlier ( $rapmGainZ > 3.1$ ), pairwise.
- Result & interpretation
  - *SubjectivelyImprovedThinking* and *RapmGainZ*  
 $r_{pbi}(55) = .066$ ,  $p < .628$
  - *SubjectivelyImprovedThinking* and *BomatGainZ*  
 $r_{pbi}(56) = -.167$ ,  $p < .210$
  - H8 disproved.
  - Based on the BIQ data, there's no statistically significant relationship between self-reported improvements in thinking, and psychometric IQ gains. *Self-reported improvements of thinking are not linked to IQ gains.*

#### (e) Discussion

When interpreting these findings, it is important to be aware of following limitations (in order of importance):

- Z-scores of even- and odd- IQ subtests were used to measure IQ gain, which deprives the tests of statistical power
- BIQ participants (who submitted self-reports) were gathered across several years of experiments, therefore sample as a whole underwent considerably heterogeneous n-back training (unknown ratio of single, dual and triple n-back)

- We just tested 14 hypotheses on a medium sized sample (ca. 50 subjects), therefore multiple comparison problem is hard to avoid. There's a  $1 - (1 - 0.05)^{14} = \sim 51\%$  chance of making type II error (false positive), and unknown chance of type I (false negative) error.
- Just for the record, versions of n-back were not aligned with personality traits (see experimental part)

On the other hand, these reasons contribute to credibility of these findings:

- Although training methods were quite heterogeneous, we can be sure that each participant included in sample underwent 6 to 8 hours of n-back training
- All the psychometric methods and questionnaires were part of controlled environment

Taken together, availability of both psychometric and subjective data yielded interesting insights into n-back training phenomenon. Of course these have yet to be supported by more research in the future, but they allow us to consider possible consequences:

- Disproving H6 is a bit counterintuitive, nevertheless according to this finding, missing evidence of far-transfer can't be attributed to missing intrinsic motivation (and vice versa).
- Disproving H7, meaning that no particular strategy is linked to higher or lower IQ gain, is probably less surprising, than the fact that 93% of participants was using one or more strategies. This can mean that even certain n-back-task-specific strategies do not have to diminish potential far-transfer.
- Disproving H8 (meaning that the feeling of cognitive improvement is independent from psychometric improvement) doesn't influence any previously published research (which reported only psychometric data), but is substantially relevant to first hypothesis of BRAWO questionnaire (H1), and interpretations of its results. While it does not disprove potential psychometric improvements of BRAWO respondents, it is a step back into uncertainty. We do not know if participants improved their psychometric IQ – their self-reports (both positive and negative) seem equally irrelevant in this regard. In any case, as it is so often repeated in field of n-back research – we are talking about first occurrence of this finding, and more research is necessary to support it.

## Chapter V. Summary and conclusion

In this dissertation, I first elaborated on the relationship between WM and *Gf*, which was evaluated as substantial by a majority of experts in the field. Then I attempted to construct a bridge over the swamp of previous cognitive training research, which has consistently shown mixed results for the ability of n-back training to improve fluid intelligence.

In the experimental part of the dissertation, I presented a co-authored study, which aimed to replicate far-transfer effects, and to explore the mediating role of personality systems interaction (PSI) with personality factors. By using three different training methods and an active-contact control group, we examined the effects of 25 days of cognitive training on 142 participants. We observed improvements in one out of two IQ test scores, which reflects the ambivalent nature of previous research in this field. After examining the results within the context of PSI theory, we found that different training methods yielded different IQ gains in participants, depending on their personality styles. In addition, these correlations suggested a meaningful pattern, indicating that PSI theory may be able to account for the different outcomes of cognitive training studies. These findings may facilitate tailor-made cognitive training interventions in the future, and can contribute to explaining the mechanisms underlying the far-transfer of working memory training to fluid intelligence.

In chapter 3, I conducted mostly univariate analysis of BRAWO, an anonymous online questionnaire for members of the n-back discussion group (N=258). This analysis offered insights into the training habits of n-backers, including performance parameters. It pondered the concept of *Aggregate n-back level*, public attitudes and misconceptions regarding IQ and its improvement. Next, it seems that respondents of BRAWO sample lead healthier lifestyle in comparison average US population, except they are fond of nootropics (which is a bit odd, considering their effectiveness and safety is less scientifically justified than that of n-back training). I acknowledged both overly optimistic and pessimistic media coverage, pseudo-neutral reviews and occasionally offensive atmosphere of online n-back discussions. Majority of BRAWO respondents reported subjectively felt improvements after n-back training in these particular cognitive areas (in order of frequency): 1. working memory, 2. attention, 3. intelligence/thinking. Nevertheless, later analysis suggests, that self-reported and psychometric IQ gains are independent from each other – i.e. placebo effect (false impression of one's IQ gain) is as possible as „anti-placebo“ (false impression of no gain). It hardly comes as surprise, that individuals can't reliably determine *changes* in their cognitive

performance. But it begs a question too – is our prevalent research methodology (measuring the effects of *few hours* of cognitive intervention on *general intelligence*, with *IQ tests split into halves or thirds*) able to reliably detect improvements after cognitive training, however good it may be? I'm not sure. But I'm positive there's both room and need for improvements of our methodology.

Delving deeper into qualitative aspects of n-back, many BRAWO respondents extended their previous questionnaire answers with qualitative addenda. Beside statements about cognition, appreciable part of respondents (about 20%) declared improved mood and motivation to perform. These areas are subjective in nature, so placebo effect is not possible (there's no „more objective“ mood to compare to), and we have to take them at face value. BRAWO qualitative reports include quite interesting and even surprising statements, e.g. most frequent change was related to sleep quality and recall of dreams. I conclude the section with description of my subjective n-back training experience.

In bivariate section of BRAWO analysis, multiple comparison problem was addressed, and selected hypotheses were tested. Among them, hypothesis about psychometric properties of n-back was confirmed (people with higher IQ's train on higher levels of “n”), with the disclaimer that IQ's were self-reported (self-reports of IQ, not its change, actually correlate with psychometric IQ considerably). Another confirmed hypothesis was, that people with self-reported higher IQ's believe in higher malleability of intelligence.

Chapter 4 was dedicated to analysis of BIQ questionnaire (N=97), in which self-reports were often connected to psychometric research data, which allowed for further insights into inner workings of n-back cognitive training. Interestingly, participants reported the same order of subjectively improved areas (1. working memory, 2. attention, 3. thinking), and similar ratios of mood and motivation improvements. Similarly, negative effects were marginal. After the whole experiment, participants perceived n-back training as moderately demanding, although initial reactions to n-back difficulty are usually significantly more pronounced. Bivariate analysis of BIQ questionnaire first addressed the multiple comparison problem, then advanced to the several findings: motivation to participate in n-back studies (intrinsic or extrinsic) does not influences the prospective IQ gain. Next, mental strategies used when training (like the use of inner voice) doesn't seem to exercise influence on IQ gains either. The last finding (subjective and psychometric cognitive gains are independent of each other) was already discussed in this section.

I tried to rigorously consider limitations and consequences of each reported finding in appropriate sections, both in the scope of this study, and in the scope of what I consider to be

relevant research context. I'd be pleased to respond to any questions, comments or correspondence regarding this work. You can reach me at [vlad.marcek@gmail.com](mailto:vlad.marcek@gmail.com).

## 5.01 Conclusion

Cognitive training became one of the most discussed topics in psychology during the last few years, and even in the general public it is more popular than ever before. The discovery of a strong relationship between WM and *Gf* gave rise to elaborate training methods, which dare to challenge traditional views of intelligence as a fixed trait in healthy adults. On one hand, revolutions in social sciences became quite cheap in 21<sup>st</sup> century – especially in psychology, new paradigms are allegedly born each month (although nobody seems to know what was the previous one, and maybe there is not one dominant paradigm at all, see Marček & Urbánek, 2011). On the other hand, scientific evidence assembled between 2008 and 2014 now seems to fully justify current multidisciplinary interest in n-back, as well as its public popularity.

There's a saying which I think often applies to attitudes in n-back research: "A pessimist is a person who has had to listen to too many optimists." Although optimism is what inspires one to explore the unknown (and unfortunately, sells products too), its role is secondary in contemporary n-back research. Considering today's evidence, n-back research interest is based on increasingly compelling evidence, and responsibility towards utilizing scientific knowledge for public well-being.

## Chapter VI. References

- Atkinson, R. C. & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. *The Psychology of Learning and Motivation*, 2, 89–195.
- Au, J., Sheehan, E., Tsai, N., Duncan, G. J., Buschkuhl, M., & Jaeggi, S. M. (2014). Improving fluid intelligence with training on working memory: a meta-analysis. *Psychonomic bulletin & review*, 1-12.
- Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*, 4(11), 417–423. Chicago.
- Baddeley, A. (2012). Working memory: theories, models, and controversies. *Annual Review of Psychology*, 63, 1–29.
- Baddeley, A. D. & Hitch, G. J. (1974). Working memory. *The Psychology of Learning and Motivation*, 8, 47–89.
- Bauernschmidt, A., Conway, C. M., & Pisoni, D. B. (2009). Working memory training and implicit learning. *Research on Spoken Language Processing Progress Report*, 29.
- Baniquid, P. L., Kranz, M. B., Voss, M. W., Lee, H., Cosman, J. D., Severson, J., & Kramer, A. F. (2013). Cognitive training with casual video games: points to consider. *Frontiers in Psychology*, 4.
- Brehmer, Y., Westerberg, H., & Bäckman, L. (2012). Working-memory training in younger and older adults: training gains, transfer, and maintenance. *Frontiers in Human Neuroscience*, 6.
- Buehner, M., Krumm, S., Ziegler, M., & Pluecken, T. (2006). Cognitive abilities and their interplay. *Journal of Individual Differences*, 27(2), 57-72.
- Burgess, G. C., Gray, J. R., Conway, A. R. A., & Braver, T. S. (2011). Neural mechanisms of interference control underlie the relationship between fluid intelligence and working memory span. *Journal of Experimental Psychology: General*, 140(4), 674–92. doi:10.1037/a0024695
- Carretti, B., Borella, E., Zavagnin, M., & de Beni, R. (2012). Gains in language comprehension relating to working memory training in healthy older adults. *International Journal of Geriatric Psychiatry*, 28(5), 539–546. doi:10.1002/gps.3859
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press.
- Cattell, R. (1971). The discovery of fluid and critallized general intelligence. Abilities: Their structure, growth and action: Boston: Houghton.
- 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), 193–9. 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.
- Colom, R., Flores-Mendoza, C., & Rebollo, I. (2003). Working memory and intelligence. *Personality & Individual Differences*, 34(1), 33.
- Colom, R., Shih, P. C., Flores-Mendoza, C., & Quiroga, M. Á. (2006). The real relationship between short-term memory and working memory. *Memory*, 14(7), 804-813.
- Colom, R., Abad, F. J., Quiroga, M. Á., Shih, P. C., & Flores-Mendoza, C. (2008). Working memory and intelligence are highly related constructs, but why? *Intelligence*, 36(6), 584–606.
- Colom, R., Haier, R. J., Head, K., Álvarez-Linera, J., Quiroga, M. Á., Shih, P. C., & Jung, R. E. (2009). Gray matter correlates of fluid, crystallized, and spatial intelligence: Testing the P-FIT model. *Intelligence*, 37(2), 124-135.

- Colom, R., Quiroga, M. Á., Shih, P. C., Martinez, K., Burgaleta, M., Martinez-Molina, A.,...Ramirez, I. (2010). Improvement in working memory is not related to increased intelligence scores. *Intelligence*, 38(5), 497–505. doi:10.1016/j.intell.2010.06.008
- Colom, R., Román, F. J., Abad, F. J., Shih, P. C., Privado, J., Froufe, M.,...Jaeggi, S. M. (2013). Adaptive n-back training does not improve fluid intelligence at the construct level: gains on individual tests suggest that training may enhance visuospatial processing. *Intelligence*, 41(5), 712–727.
- Conway, A. R. A., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences*, 7(12), 547.
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24(01), 87-114.
- Cowan, N., Saults, J. S., & Blume, C. L. (2014). Central and Peripheral Components of Working Memory Storage. *Journal of experimental psychology. General*.
- Davidson , J. E., & Kemp, I. A. (2011). Contemporary models of intelligence. In R. J. Sternberg & S. B. Kaufman (Eds.), *Cambridge handbook of intelligence* (pp. 58–84) . New York.
- Cambridge University Press Demetriou, A., Kyriakides, L., & Avraamidou, C. (2003). The missing link in the relations between intelligence and personality. *Journal of Research in Personality*, 37(6), 547–581.
- Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence. *Journal of experimental psychology. General*, 128(3), 309.
- Evans, J. S. B., & Stanovich, K. E. (2013). Dual-process theories of higher cognition advancing the debate. *Perspectives on Psychological Science*, 8(3), 223-241.
- Eysenck, H. J. (1950). *Dimensions of personality*, Vol. 5. Transaction Publishers.
- Farsides, T., & Woodfield, R. (2003). Individual differences and undergraduate academic success: The roles of personality, intelligence, and application. *Personality and Individual Differences*, 34(7), 1225–1243.
- Friedman, N., & Miyake, A. (2000). Differential roles for visuospatial and verbal working memory in situation model construction. *Journal of Experimental Psychology: General*, 129(1), 61-83.
- Freund, P. A., & Kasten, N. (2012). How smart do you think you are? A meta-analysis on the validity of self-estimates of cognitive ability. *Psychological bulletin*, 138(2), 296.
- Furnham, A. (2001). Self-estimates of intelligence: Culture and gender difference in self and other estimates of both general (g) and multiple intelligences. *Personality and Individual Differences*, 31(8), 1381-1405.
- Gates, N., & Valenzuela, M. (2010). Cognitive exercise and its role in cognitive function in older adults. *Current Psychiatry Reports*, 12(1), 20–27.
- Gobet, F., & Clarkson, G. (2004). Chunks in expert memory: Evidence for the magical number four... or is it two?. *Memory*, 12(6), 732-747.
- Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, 37(3), 424–438.
- Guilford, J., & Paul, J. (1967). The nature of human intelligence: McGraw-Hill New York.
- Haier, R. J., Colom, R., Schroeder, D. H., Condon, C. A., Tang, C., Eaves, E., & Head, K. (2009). Gray matter and intelligence factors: Is there a neuro-G?. *Intelligence*, 37(2), 136-144.
- Hossiep, R., Turck, D., & Hasella, M. (1999). *Bochumer Matrizentest (BOMAT) advanced*. Hogrefe.
- Jaeggi, S. M., Buschkuhl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences*, 105(19), 6829.

- Jaeggi, S. M., Buschkuhl, M., Perrig, W. J., & Meier, B. (2010). The concurrent validity of the N-back task as a working memory measure. *Memory*, 18(4), 394-412.
- Jaeggi, S. M., Buschkuhl, M., Jonides, J., & Shah, P. (2011). Short- and long-term benefits of cognitive training. *Proceedings of the National Academy of Sciences*, 108(25), 10081–6. doi:10.1073/pnas.1103228108
- Jaeggi, S. M., Buschkuhl, M., Shah, P., & Jonides, J. (2014). The role of individual differences in cognitive training and transfer. *Memory & Cognition*, 42(3), 464–480.
- Jaeggi, S. M., Studer-Luethi, B., Buschkuhl, M., Su, Y.-F., Jonides, J., & Perrig, W. J. (2010). The relationship between n-back performance and matrix reasoning — implications for training and transfer. *Intelligence*, 38(6), 625–635. doi:10.1016/j.intell.2010.09.001
- Jausovec, N., & Jausovec, K. (2012a). Sex differences in mental rotation and cortical activation patterns: Can training change them? *Intelligence*, 40(2), 151–162.
- Jausovec, N., & Jausovec, K. (2012b). Working memory training: improving intelligence-changing brain activity. *Brain and cognition*, 79(2), 96–106. doi:10.1016/j.bandc.2012.02.007
- Johnson, W., & Bouchard, T. J. (2005a). Constructive replication of the visual-perceptual-image rotation model in Thurstone's (1941) battery of 60 tests of mental ability. *Intelligence*, 33(4), 417–430.
- Johnson, W., & Bouchard, T. J. (2005b). The structure of human intelligence: it is verbal, perceptual, and image rotation (VPR), not fluid and crystallized. *Intelligence*, 33(4), 393–416.
- Johnson, W., te Nijenhuis, J., & Bouchard Jr, T. J. (2007). Replication of the hierarchical visual-perceptual-image rotation model in de Wolff and Buiten's (1963) battery of 46 tests of mental ability. *Intelligence*, 35(1), 69-81.
- Jung, R. E., & Haier, R. J. (2007). The Parieto-Frontal Integration Theory (P-FIT) of intelligence: converging neuroimaging evidence. *Behavioral and Brain Sciences*, 30(02), 135-154.
- Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W. & Engle, R. W. (2004). The generality of working memory capacity: a latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General*, 133(2), 189.
- Kane, M. J., Hambrick, D. Z., & Conway, A. R. A. (2005). Working memory capacity and fluid intelligence are strongly related constructs: comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, 131(1), 66–71.
- 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.
- Karbach, J., & Verhaeghen, P. (2014). Making Working Memory Work A Meta-Analysis of Executive-Control and Working Memory Training in Older Adults. *Psychological science*, 25(11), 2027-2037.
- Kaufman , S. B. ( 2009 ). Beyond general intelligence: dual-process theory of human intelligence (Unpublished Ph.D. dissertation). Yale University , New Haven, CT .
- Kaufman, S. B., DeYoung, C. G., Gray, J. R., Brown, J., & Mackintosh, N. (2009). Associative learning predicts intelligence above and beyond working memory and processing speed. *Intelligence*, 37(4), 374–382.
- Kaufman , S. B. (2011). Intelligence and the cognitive unconscious. In R. J. Sternberg & S. B. Kaufman (Eds.), *The Cambridge handbook of intelligence* (pp. 442–467) . New York: Cambridge University Press.
- Kaufman, S. B. (2013). *Ungifted: Redefining intelligence*. New York: Basic Books.

- Kelly, M. E., Loughrey, D., Lawlor, B. A., Robertson, I. H., Walsh, C., & Brennan, S. (2014). The impact of exercise on the cognitive functioning of healthy older adults: A systematic review and meta-analysis. *Ageing research reviews*, 16, 12-31.
- Klatzky, R. L., Giudice, N. A., Marston, J. R., Tietz, J., Golledge, R. G., & Loomis, J. M. (2008). An n-back task using vibrotactile stimulation with comparison to an auditory analogue. *Behavior research methods*, 40(1), 367-372.
- Kornilova, T. V., & Novikova, M. A. (2013). Self-assessed intelligence, psychometric intelligence, personality, and academic achievement: Two structural models. *Educational Achievement: Teaching Strategies, Psychological Factors and Economic Impact*. New York: Nova Science Publishers, 197-212.
- Kueider, A. M., Parisi, J. M., Gross, A. L., & Rebok, G. W. (2012). Computerized cognitive training with older adults: a systematic review. *PLoS ONE*, 7(7), e40588. doi:10.1371/journal.pone.0040588
- Kuhl, J. 2000a. A theory of self-development: affective fixation and the STAR model of personality disorders and related styles. In J. Heckhausen (Ed.), *Motivational psychology of human development: developing motivation and motivating development*. Oxford, England: Elsevier.
- Kuhl, J. (2000b). The volitional basis of personality systems interaction theory: applications in learning and treatment contexts. *International Journal of Educational Research*, 33(7), 665–703.
- Kuhl, J., & Kazén, M. (1997). *Persönlichkeits-Stil-und-Störungs-Inventar (PSSI)* [The Personality-Styles-and-Disorder-Inventory (PSDI): Manual]. Göttingen, Germany: Hogrefe.
- Kyllonen, C., P., Christal, & E., R. (1990). Reasoning ability is (little more than) working-memory capacity? (Vol. 14). Orlando, FL, ETATS-UNIS: Elsevier.
- Larson, G. E., & Saccuzzo, D. P. (1989). Cognitive correlates of general intelligence: Toward a process theory of g. *Intelligence*, 13(1), 5-31.
- Mackintosh, N. (2011). *IQ and human intelligence*. New York, USA: Oxford University Press.
- Mackintosh, N. J., & Bennett, E. S. (2003). The fractionation of working memory maps onto different components of intelligence. *Intelligence*, 31(6), 519-531.
- Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in psychology research: how often do they really occur? *Perspectives on Psychological Science*, 7(6), 537–542.
- Mangels, J. A., Butterfield, B., Lamb, J., Good, C., & Dweck, C. S. (2006). Why do beliefs about intelligence influence learning success? A social cognitive neuroscience model. *Social cognitive and affective neuroscience*, 1(2), 75-86.
- Marcek, V., & Urbanek, T. (2011). Philosophical and scientific concept of paradigm in psychology. *Ceskoslovenska psychologie*, 55(3), 222-233.
- Maslen, H., Faulmüller, N., & Savulescu, J. (2014). Pharmacological cognitive enhancement—how future neuroscientific research could advance ethical debate. *Frontiers in Systems Neuroscience*, 8, 107.
- Miller, G. A. "The magical number seven, plus or minus two: Some limits on our capacity for processing information. PSR 63, 81–97 (1956)." (2012).
- Mohamed, A. D. (2012). Modafinil has the potential for addiction. *AJOB Neuroscience*, 3(2), 36-38.
- Mohamed, A. D. (2014). The Effects of Modafinil on Convergent and Divergent Thinking of Creativity: A Randomized Controlled Trial. *The Journal of Creative Behavior*.
- Moody, D. E. (2009). Can intelligence be increased by training on a task of working memory? *Intelligence*, 37(4), 327–328.

- Morey, C. C., & Bieler, M. (2013). Visual short-term memory always requires general attention. *Psychonomic bulletin & review*, 20(1), 163-170.
- 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
- Nee, D. E. & Jonides, J. (2013). Neural evidence for a 3-state model of visual short-term memory. *NeuroImage*.
- Noack, H., Lövdén, M., & Schmiedek, F. (2014). On the validity and generality of transfer effects in cognitive training research. *Psychological research*, 1-17.
- Oberauer, K. (2002). Access to information in working memory: exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(3), 411.
- Oberauer, K., Schulze, R., Wilhelm, O. (2005). Working memory and intelligence—their correlation and their relation: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, 131(1), 61-65.
- Oberauer, K., Süs, H. M., Wilhelm, O., & Wittmann, W. W. (2008). Which working memory functions predict intelligence? *Intelligence*, 36(6), 641–652.
- Oelhafen, S., Nikolaidis, A., Padovani, T., Blaser, D., Koenig, T., & Perrig, W. J. (2013). Increased parietal activity after training of interference control. *Neuropsychologia*, 51(13), 2781–2790.
- 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), 775–778.
- Oztekin, I. & McElree, B. (2010). Relationship between measures of working memory capacity and the time course of short-term memory retrieval and interference resolution. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 36(2), 383–97. doi:10.1037/a0018029.
- Penner, I.-K., Vogt, A., Stöcklin, M., Gschwind, L., Opwis, K., & Calabrese, P. (2012). Computerised working memory training in healthy adults: a comparison of two different training schedules. *Neuropsychological Rehabilitation*, 22(5), 716–733. doi:10.1080/09602011.2012.686883
- Peters, M., Laeng, B., Latham, K., Jackson, M., Zaiyouna, R. & Richardson, C. (1995). A redrawn vandenberg and kuse mental rotations test-different versions and factors that affect performance. *Brain and cognition*, 28(1), 39–58. Elsevier.
- Pratkanis, A. R., Eskenazi, J., & Greenwald, A. G. (1994). What you expect is what you believe (but not necessarily what you get): A test of the effectiveness of subliminal self-help audiotapes. *Basic and Applied Social Psychology*, 15, 251–276. doi:10.1207/s15324834basp1503\_3
- Pugin, F., Metz, A. J., Stauffer, M., Wolf, M., Jenni, O. G., & Huber, R. (2014). Working memory training shows immediate and long-term effects on cognitive performance in children and adolescents. *F1000Research*, 3, 82.
- Rafi, A., & Samsudin, K. (2009). Practising mental rotation using interactive desktop mental rotation trainer (iDeMRT). *British Journal of Educational Technology*, 40(5), 889–900.
- Raven, J. C. (1990). *Advanced Progressive Matrices. Sets I, II*. Oxford: Oxford University Press.
- Redick, T. S., Shipstead, Z., Harrison, T. L., Hicks, K. L., Fried, D. E., Hambrick, D. Z.,...Engle, R. W. (2012). 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
- Redick, T. S., & Lindsey, D. R. (2013). Complex span and n-back measures of working memory: A meta-analysis. *Psychonomic bulletin & review*, 20(6), 1102-1113.

- Reijnders, J., van Heugten, C., & van Boxtel, M. (2012). Cognitive interventions in healthy older adults and people with mild cognitive impairment: a systematic review. *Ageing Research Reviews*, 12(1), 263–275.
- Salminen, T., Strobach, T., & Schubert, T. (2012). On the impacts of working memory training on executive functioning. *Frontiers in Human Neuroscience*, 6, 166. doi:10.3389/fnhum.2012.00166
- Salthouse, T. (2006). Mental exercise and mental aging. *Perspectives on Psychological Science*, 1(1), 68.
- 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, 27.
- Shah, P., Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing : An individual differences approach (Vol. 125). Washington, APA
- Shipstead, Z., Redick, T. S., & Engle, R. W. (2012). Is working memory training effective? *Psychological Bulletin*, 138(4), 628–54. doi:10.1037/a0027473
- Smith, M. E., & Farah, M. J. (2011). Are prescription stimulants “smart pills”? The epidemiology and cognitive neuroscience of prescription stimulant use by normal healthy individuals. *Psychological bulletin*, 137(5), 717.
- Soh, L., & Jacobs, K. E. (2013). The biasing effect of personality on self-estimates of cognitive abilities in males and females. *Personality and Individual Differences*, 55(2), 141-146.
- Soubelet, A., & Salthouse, T. A. (2011). Personality–cognition relations across adulthood. *Developmental Psychology*, 47(2), 303.
- Spearman, C. (1904). " General Intelligence," Objectively Determined and Measured. *The American journal of psychology*, 15(2), 201-292.
- Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied*, 74(11), 1.
- Spielberger, C. D., & Gorsuch, R. L. (1983). *State-trait anxiety inventory for adults: sampler set: manual, test, scoring key*. Menlo Park, CA: Mind Garden.
- Stankov, L., & Crawford, J. D. (1993). Ingredients of complexity in fluid intelligence. *Learning and Individual Differences*, 5(2), 73-111.
- Stanovich, K. E., & Toplak, M. E. (2012). Defining features versus incidental correlates of Type 1 and Type 2 processing. *Mind and Society*, 11 , 3–13.
- Stephenson, C. (2010). *Does training to increase working memory capacity improve fluid intelligence?* (Doctoral dissertation). UMI Dissertation Publishing (number: 3417011). Claremont Graduate University, Ann Arbor, Michigan.
- Stephenson, C. L., & Halpern, D. F. (2013). Improved matrix reasoning is limited to training on tasks with a visuospatial component. *Intelligence*, 41(5), 341–357.
- Sternberg, R. J. (2008). Increasing fluid intelligence is possible after all. *Proceedings of the National Academy of Sciences*, 105(19), 6791–6792.
- Stransky, D., Wilcox, L. M., & Dubrowski, A. (2010). Mental rotation: cross-task training and generalization. *Journal of Experimental Psychology: Applied*, 16(4), 349.
- Studer-Luethi, B., Jaeggi, S. M., Buschkuhl, M., & Perrig, W. J. (2012). Influence of neuroticism and conscientiousness on working memory training outcome. *Personality and Individual Differences*, 53(1), 44–49.
- Stumm, Sophie. "Intelligence, gender, and assessment method affect the accuracy of self-estimated intelligence." *British Journal of Psychology* 105.2 (2014): 243-253.

- Subramaniam, K., Luks, T. L., Fisher, M., Simpson, G. V., Nagarajan, S., & Vinogradov, S. (2012). Computerized cognitive training restores neural activity within the reality monitoring network in schizophrenia. *Neuron*, 73(4), 842–853.
- Süß, H. M., Oberauer, K., Wittmann, W. W., Wilhelm, O., & Schulze, R. (2002). Working-memory capacity explains reasoning ability—and a little bit more. *Intelligence*, 30(3), 261-288.
- Syzmanowicz, A., & Furnham, A. (2011). Gender differences in self-estimates of general, mathematical, spatial and verbal intelligence: Four meta analyses. *Learning and Individual Differences*, 21(5), 493-504.
- Tillman, C. M., Nyberg, L., & Bohlin, G. (2007). Working memory components and intelligence in children. *Intelligence*, 36(5), 394-402.
- Thompson, T. W., Waskom, M. L., Garel, K.-L. A., Cardenas-Iniguez, C., Reynolds, G. O., Winter, R.,...Gabrieli, J. D. (2013). Failure of working memory training to enhance cognition or intelligence. *PLoS one*, 8(5), e63614.
- Thurstone, L. (1938). Primary mental abilities. Wang, Z., Zhou, R., & Shah, P. (2014). Spaced cognitive training promotes training transfer. *Frontiers in Human Neuroscience*, 8, 217.
- Urban, K. R., & Gao, W. J. (2014). Performance enhancement at the cost of potential brain plasticity: neural ramifications of nootropic drugs in the healthy developing brain. *Frontiers in systems neuroscience*, 8.
- Wechsler, D. (2008). *Wechsler adult intelligence scale—Fourth Edition (WAIS-IV)*. San Antonio, TX: NCS Pearson.
- Wykes, T., Reeder, C., Landau, S., Everitt, B., Knapp, M., Patel, A., & Romeo, R. (2007). Cognitive remediation therapy in schizophrenia: Randomised controlled trial. *The British Journal of Psychiatry*, 190(5), 421–427.
- Yeager, D. S., & Dweck, C. S. (2012). Mindsets that promote resilience: When students believe that personal characteristics can be developed. *Educational Psychologist*, 47(4), 302-314.