Towards a Neuroscience of Computer Programming & Education

Andreas Ola Lidström
100224072

A thesis submitted in partial fulfilment of the requirements of the University of East Anglia for the degree of Doctor of Philosophy. Research undertaken in the School of Psychology, University of East Anglia.

30/09/2021

Abstract

Computer programming is fast becoming a required part of School curricula, but students find the topic challenging and university dropout rates are high. Observations suggest that hands-on keyboard typing improves learning, but quantitative evidence for this is lacking and the mechanisms are still unclear. Here we study neural and behavioral processes of programming in general, and Hands-on in particular. In project 1, we taught naïve teenagers programming in a classroom-like session, where one student in a pair typed code (Hands-on) while the other participated by discussion (Hands-off). They were scanned with fMRI 1-2 days later while evaluating written code, and their knowledge was tested again after a week. We find confidence and math grades to be important for learning, and easing of intrinsic inhibitions of parietal, temporal, and superior frontal activation to be a typical neural mechanism during programming, more so in stronger learners. Moreover, left inferior frontal cortex plays a central role; operculum integrates information from the dorsal and ventral streams and its intrinsic connectivity predicts confidence and long-term memory, while activity in Broca's area also reflects deductive reasoning. Hands-on led to greater confidence and memory retention. In project 2, we investigated the impact of feedback on motivation and reaction time in a rule-switching task. We find that feedback targeting personal traits increasingly impair performance and motivation over the experiment, and we find that activity in precentral gyrus and anterior insula decrease linearly over time during the personal feedback condition, implicating these areas in this effect. These findings promote hands-on learning and emphasize possibilities for feedback interventions on motivation. Future studies should investigate interventions for increasing Need for Cognition, the relationship between computer programming and second language learning (L2), and the role of explicit verbalization of knowledge for successful coding, given the language-like processing of code.

Table of Contents

List of Figures

List of Tables

Acknowledgements

Vetenskapsrådet & the taxpayers of Sweden – For funding the HOPE project.

Associate Professor Sara Bengtsson – My primary supervisor, for remembering me years later, giving me a chance, and for exemplary support.

Associate Professor Anna Eckerdal – My secondary supervisor, for instigating the HOPE project and for unending optimism.

Kristina von Hausswolff – For a great work designing and running the teaching sessions.

Professor Michael Thuné & Associate Professor Maria Weurlander – For interesting and relevant discussion during our journal clubs.

Professor William Penny – For invaluable feedback on our DCM approach.

Professor Ivanka Savic – For hiring me straight out of school and allowing me to first learn fMRI analysis.

Jasenko Dervisic & Amirhossein Manzouri – For keeping me sane while doing so.

The food delivery drivers of the pandemic – For keeping the service sector and me from collapsing.

Author's Declaration

I declare that the work contained in this thesis has not been submitted for any other award and that it is all my own work. I also confirm that this work fully acknowledges opinions, ideas, and contributions from the work of others.

The research in Chapter 2 has been previously presented in oral and poster formats:

Oral Presentations

Lidström, A. (2021). Towards a Neuroscience of Computer Programming & Education. **MINT - Centre for Subject-Didactive Research - Uppsala University, Sweden.**

Lidström, A. (2021). Towards a Neuroscience of Computer Programming & Education. **CAP – Cognition, Action & Perception Research Group - University of East Anglia, UK.**

Poster Presentations

Lidström, A., Bengtsson, SL. (2019). Adopting Relevant Concepts in Computer Programming. **OHBM 2019 - Organization for Human Brain Mapping – Rome, Italy.**

Lidström, A., Bengtsson, SL. (2019). Learning Concepts in Computer Programming. **MCC 2019 – 7th International Symposium on Motivational and Cognitive Control - Berlin, Germany.**

The Research in Chapter 3 is based on a study previously presented in another scientific publication:

Published Work

Appelgren, A., Bengtsson, SL. (2015) Feedback on Trait or Action Impacts on Caudate and Paracingulum Activity. **PLoS ONE 10(6): e0129714. doi:10.1371/journal.pone.0129714**

Any ethical clearance for the research presented in this thesis has been approved. Approval has been sought and granted by the regional Ethics committee EPN Uppsala, Sweden, Dnr 2017-178, and EPN Stockholm, Sweden Dnr 2014/10-31/2.

Name: Andreas Ola lidström

Signature: Andreas Lidström

Date: 30/9/2021

Word count: 38,331

Chapter 1
General Introduction

1.1. The impetus to study programming knowledge acquisition

Because we live in an increasingly computerized world where digital knowhow plays an ever more dominant role in society, many organizations are growing more cognizant of the importance of staying competitive on this front in the global community. In 2013 the European Union published the ‘International Computer and Information Literacy Study’ (ICILS) (Commission, 2014), where they cast doubt on the ‘digital natives’ idea that young people having grown up with computers would take more naturally to programming (Prensky, 2001). Instead, they emphasized the role of schooling in attaining the competencies outlined in the EU ‘Digital Competence Framework for Citizens’ (DIGICOMP) (Vuorikari, Punie, Carretero, & Van den Brande, 2016). This was followed by the 2017 report ‘Digital skills in the EU labour Market’ where it was identified that 40% of the EU workforce had “*little to no digital skill*” while predicting a future where nearly all jobs will require digital skills (Kiss, 2017). In accordance with this view, the Swedish government recently decided to add computer programming as a required part of the School curriculum (Regeringskansliet, 2017), following the lead of the United Kingdom, where computer education was introduced in the 2014 curriculum starting all the way down in stage 1 education (age 5-7) (Education, 2014), championing concepts such as ‘computational thinking’ (Wing, 2006). We have also seen the rise of hugely influential private charity organizations such as www.computingatschool.org.uk and www.raspberrypi.org. This means that it is more relevant than ever to understand how we should best train our future young programmers, especially given the trouble of high dropout rates that is widely thought to have been plaguing computer education for the longest time, which we will discuss in the next section. In a systematic review of introductory programming, the authors identify the need for quantitative studies on learning strategies, and investigations into the role of self-efficacy processes such as confidence, stress, and motivation (Luxton-Reilly et al., 2018). In addition, the field of education calls for interdisciplinary efforts between cognitive neuroscience and educational practices; 'Educational neuroscience' benefits both from the rigorous control of stimuli and tasks necessary in cognitive neuroscience, and from the ecological validity found in education research (Ansari, Coch, & De Smedt, 2011).

1.2. The problems in computer education

According to a report Prepared for the European Commission in 2013, the average student dropout rate at University level "*seems to be at around 19%*" based on Eurostat data of student enrolment between 1999 and 2010 (Hüsing et al., 2013). This state of affairs seems to be taken as a fact across the field of computer science education and a common claim is that there is something about the subject itself that sets it apart from other programs, even though other schools like engineering seems to be facing similar rates (Paura & Arhipova, 2016). This view has led to several attempts to explain why this seems to be the case, one example being the debate starting around

2010 about whether computer science differs from other subjects by illustrating that the students have both higher than usual rates of failing and high grades, creating what some argued is a "*characteristic*" bi-modal distribution of grades (Robins, 2010). The author argued against the idea that this is caused by an underlying split population of "programmers" and "non-programmers" entering the courses and subsequently perform according to unknown innate characteristics, as proposed by (Utting et al., 2013). Since "*decades of extensive research*" has failed in delivering these crucial interpersonal factors, they instead proposed a mechanism dubbed 'Learning Edge Momentum' that attempted to explain the bi-modal distribution of grades by proposing that inherent to the subject of computer science are nested concepts that build on each-other and that are self-reinforcing, resulting in a snowball-effect where the most affluent early on pulls further and further ahead. This argument places an emphasis on learning material and techniques, rather than on the learner themselves.

Since then, the whole issue of bi-modally distributed outcomes of introductory computer science courses has been largely abandoned, including by the original proponents of it (Fincher & Robins, 2019), but this debate did raise some interesting points for consideration that are worth discussing here. A 2013 paper brought attention to the way scientists and educators look at their data, claiming that it is possible to look at the same set of data on correlations between grades on different tests in early programming education and have it conform to your theory of choice, be it "geek genes", prior knowledge, "stumbling points" (defined as skills and concepts that can have a major impact on progress) or 'learning edge momentum', and that furthermore these concepts are not mutually exclusive (Ahadi & Lister, 2013). The authors liken this to the fable of the blind men and the elephant and suggest that there may in fact be a grander view with the potential to see the whole elephant rather than one of its parts. Unfortunately, they gave no insight on how to achieve this. The simple facts of human biases may well explain much of this whole debate. An empirical study of the final grade distributions for every undergraduate computer science class at the University of British Columbia between 1996 and 2013 (30,214 grades) found no statistical evidence for the grades being bi-modally distributed, claiming at most 5.8% of the cases examined fit the description of bimodality (Patitsas, Berlin, Craig, & Easterbrook, 2016). In the same paper they also showed in an experiment asking teachers to rate grade distributions as bimodal or not, that labeling ambiguous distributions as bimodal significantly correlated with being primed for bi-modality beforehand, believing in "geek-genes" and habitually looking at histograms of their class grades. Research of this type shows the importance of large quantitative analyses and the role of researcher bias in student evaluations.

The bimodality of grades aside, the relatively high dropout rate still remains to be explained. For example: A 2015 paper reported a dropout rate of 32.2% in Estonian first year 'Information and Communications Technology' (ICT) students, and they found significant differences in mathematics, academic achievement, expectations and satisfaction between the students that dropped out and those who stayed for the second year (Kori et al., 2015). This fits well with long established findings in the literature where the described predictors of achievement include mathematics and spatial reasoning tests (Choi-man, 1988), ‘general skills test battery’ and mathematics (Erdogan, Aydin, & Kabaca, 2008) and Paper-folding tasks, map-sketching and algorithmic thinking (Fincher et al., 2006). In addition to these performance measures there are a host of other environmental variables that could impact student performance such as student income impacted by parental aid or working during studies, parental education, educational planning, perceived quality of the education, satisfaction, stress, marital status, place of residence, demographics, social integration on and off campus, and the institutional characteristics of the school itself such as finances, faculty and other organizational structures (Kori et al., 2015).

Overall, a factor that is frequently associated with learning success in programming seems to be mathematically related skills, but factors that could be seen as responsive to both internal and environmental mechanisms is the degree of confidence (including positive self-assessment) and thinking about learning (metacognition) (Bennedsen & Caspersen, 2005; Byrne & Lyons, 2001; Holvikivi, 2010; White & Sivitanides, 2003). Next, we will discuss the role of internal states on learning improvement and, following that, other potential avenues more specific to programming language learning.

1.3. Mindset

One of the biggest recent phenomena in education that has been championed as being a “*national education priority*” (Rattan, Savani, Chugh, & Dweck, 2015) with “*profound effects*” on educational achievement is the so called ‘Growth Mindset’ theory proposed by the American psychologist Carol Dweck (Dweck, 2008a, 2008b). In this conception, people will fall somewhere in between the two extremes of either believing that their innate strengths and ability to learn is ‘fixed’, and thus unable to change, or they believe that they have the ability to improve with the right amount of effort put into the task (‘growth mindset’). Ultimately it boils down to a matter of attribution. When you encounter difficulties in your pursuit of a specific learning goal, you will either attribute the failure to achieve to your fixed lack of relevant cognitive ability, or to simply not having uncovered all the necessary methods and steps to arrive at the desired outcome, spurring further attempts according to this theory. This means that it is when encountering setbacks that the growth mindset theory should be the most applicable. A person with a growth mindset would see a setback as an area of opportunity to learn more and a new height to aspire to with new skills possible to attain, whereas a person with a fixed mindset would see an insurmountable cliff blocking of yet another part of progress in life. As for application in education, Dweck herself spells out proposed strategies for promoting a growth mindset in students in the classroom setting: "*praising students for the process they have engaged in—the effort they applied, the strategies they used, the choices they made, the persistence they displayed, and so on—yields more long-term benefits than telling them they are 'smart' when they succeed*", "*the teacher should portray challenges as fun and exciting, while portraying easy tasks as boring and less useful for the brain*" (Dweck, 2010). This coupled with explicit growth mindset promotion and the encouragement to introspect with regards to one's own goals and progress to date, with the help of journal writing for instance, to bolster the sense of achievement is encouraged. This type of goal-directed tracking of personal progress is something that is also promoted in other fields such as psychotherapy (Kolb & Boyatzis, 1970) and the broader field of self-help (Wright & Chung, 2001) and career coaching (Feldman, 2001).

Dweck was motivated to compile her theory based on her own observations going back to the 70:s. Clear ideas about where to focus investigation and interventions in the realm of education can be found in publications predating her codified work, for example in the paper (originally published in 1998) "Inside the Black Box" (Paul Black & Wiliam, 2010). The authors put forward ideas such as "*Feedback to any pupil should be about the particular qualities of his or her work, with advice on what he or she can do to improve and should avoid comparisons with other pupils*" and "*self-assessment by pupils, far from being a luxury, is in fact an essential component of formative assessment*". Here they already present a proto-view that mirrors the two mindsets proposed by Dweck but calling them the "Fixed-IQ" and "un-tapped potential" views. Unlike Dweck who leans heavily towards the potential for growth though, they take the position that the truth of which of these two views are more accurate in their words, "*clearly*" lies somewhere between these two extremes. They do however assert that "*the evidence is that ways of managing formative assessment which work with the assumptions of 'un-tapped potential' do help all pupils to learn and can give particular help to those who have previously fallen behind.*". This branch of inquiry which focuses on what is called 'formative assessment' has since evolved into what is now known as 'Assessment for Learning' (AfL). AfL is now a "*popular term at all levels of education*" and seen as "*an important way to improve student learning*" (McDowell, Sambell, & Davison, 2009). It is promoted by well-regarded institutions such as the 'Cambridge International Education teaching and Learning team'(https://www.cambridge-community.org.uk/professional-development/gswafl/index.html). However, Paul Black himself has been skeptical of where things have progressed to in regards to AfL, calling it "*a free brand name to attach to any practice*" (P Black, 2006). In addition, although the ideas of Dweck; that students with practice derive a significant benefit in learning outcomes partly because of the effect of this positive outlook on their own capacity for improvement are intriguing, the practical effects of mindset, and interventions to promote a 'growth' attitude, has recently been called into question. A 2018 meta-analysis described the impact of mindset as inconsistent, where most analyses yielded small or null effects. A relationship between mindset and achievement that the authors describe as "*very weak*" (Sisk, Burgoyne, Sun, Butler, & Macnamara, 2018). They even go so far as to argue that there is a strong argument to be made that resources might be better allocated away from the study and implementation of mindset interventions, focusing instead on other avenues of education research. Even if growth mindset might not be the strong predictor of outcome one could have hoped for, more cognitive neuroscientific studies on the matter could shed light on possible differences in processing strategy employed when encountering new information or problems to solve. Understanding the neural mechanisms may facilitate developing new tools to optimize education suited for different cognitive strategies.

The term ‘Mindset’ does not only refer to the theory proposed by Dweck however, but can be referred to more broadly as all the attitudinal and motivational states of mind that the student brings with them, including their level of stress, and how they relate to learning, captured by measurements such as Need For Cognition (NFC) (Petty, Brinol, Loersch, & McCaslin, 2009). Some people are more susceptible to stress than others (Vollrath, 2001), but it has been known for a long time that stress, and the hormones that regulate stress in mammals, have no simple relationship to learning and performance. The two types of stress hormone receptors in the human brain are not evenly distributed, and this system is set up to maximize adaptive behaviors in times of duress (De Kloet, Oitzl, & Joëls, 1999). What this means is that some level of stress can be beneficial for some types of tasks, but more deleterious to others (Schwabe & Wolf, 2013). NFC is a measurement going back to the paper “An Experimental Investigation of Need for Cognition” (Cohen, Stotland, & Wolfe, 1955) and has been showed to be a consistent and reliable instrument (Sadowski & Gulgoz, 1992). A number of different versions of this scale has been created over the years, even as late as 2018 (Lins de Holanda Coelho, HP Hanel, & J. Wolf, 2020). Examples of the questions in this 6-item instrument are “*I would prefer complex to simple problems*”, “*I really enjoy a task that involves coming up with new solutions to problems*” and “*I would prefer a task that is intellectual, difficult, and important to one that is somewhat important but does not require much thought*”. The nature of these questions is such that they ask whether you enjoy intellectual problems rather than if you could become better at them or even if you are any good at those types of problems to begin with (even though odds are that is the case if you enjoy the challenge). This would seem to sidestep the potential issue of biasing by trait neuroticism but might instead be biased by other traits such as openness to experience.

Another somewhat overlooked account, which finds a voice in the literature primarily up until the 1960s, is the notion that intrinsic motivation and confidence are stimulated by hands-on learning (Paris & Turner, 2012). Von Wright suggests, in his model of the logic of events, that the way a learner acts; how the individual approaches and deals with a task, will impact on the personal experience of success and failure (Von Wright, 1983). Studies find that students' confidence in lab-tasks is related to their reported experience of *"the opportunities you get, the will-power you have, what you actually do, and what you achieve"* including hands-on learning (Skogh, 2013; Swafford, Orr, & Hall, 2014). Cortese et al., 2016 find that confidence, when dissociated from performance, is reflected in altered neural activation of frontoparietal areas which, as the authors argue, implies a link between confidence and attention (Cortese, Amano, Koizumi, Kawato, & Lau, 2016). To dig deeper into the relevant neural aspects specific to these internal states, we first must examine the nature of programming itself, which is the focus of the next section. For further discussion of the applicability of mindset in general and more specifically to our data, please see the relevant sections of the discussion.

1.4. Computer programming and language structures

Different programming languages have been created to fulfill a multitude of requirements depending on the preferences of the creators and the purposes of the language. At the lowest end are the machine codes used to run instructions directly on various pieces of hardware like chips and processors, and above that, low-level programming ‘Assembly’ languages that are designed to be human-readable but are spartan and are designed for one specific computer architecture only. The most familiar programming languages like Java, Python, C# and Visual Basic are so-called ‘high-level’ languages that offer significant amounts of abstraction of the low-level details of actually running the computer, providing the user with a more goal-directed way of interacting with the machine and often incorporating more natural language-like elements.

Programming languages can be described as built upon three language-structures: Syntax, Semantics and Logical structures (Sebesta, 2004). The syntax is the formal grammar of the language. These are the rules that govern what variable names are valid, how brackets, parentheses, and operators like adding, subtracting and comparisons work, and in what order these elements are interpreted by the compiler that translates the code into instructions the computer can understand. The semantics is the meaning of each of the program statements that make up the code, dictating the outcome when the program is run. Thus, the semantics can be said to flow from the Syntax, based on the ordering of the elements in the code and their content, and this is ultimately what the programmer is concerned with.

Finally, the logical structures of the code are abstractions like ‘do-while’ and ‘for’-loops, and ‘if-then-else’ logic. The formal structure of these are of course determined by the syntax of the language, but the distinctive aspect of these logic structures is that they are what provides the program with the ability to ‘branch’ its operation based on certain conditions and change the direction of program execution based on previous conditions. Therefore, they form the program into more complex linguistic ‘shapes’ than simply top to bottom, line by line. Programming languages also require the user to engage in logical reasoning more generally since, within the sequences of instructions, they need to assign the correct variables and then procedurally change the values of these variables accordingly so as to reach the desired outcome at the end of execution (Folk, 1973; Louden, 1993).

There are many ideas about how computer programming is best learned, what concepts to start with and what to ignore for later (Curricula, 2013), and which language to even use in the first place (Mannila & de Raadt, 2006). One basic distinction to be made regardless of chosen language is whether to focus on teaching algorithms and think of programs as more or less linear sequences of problems to be solved, or to favor the use of newer paradigms like so-called ‘object-oriented’ programming (Nirosh, 2015). In the latter mode, programs are made up of several separate ‘objects’, where each can be thought of as its own sovereign program with its own variables and functions that, ideally, can only be accessed by other objects via dedicated channels by explicit permission. This formal separation is called ‘encapsulation’ and this paradigm of separate objects where multiple copies or ‘instances’ of the same object can exist at the same time is naturally favored for instance in game programming, where it makes intuitive sense to think of objects with different characteristics managing themselves more or less independently in the larger program that make up the game world (Louden, 1993).

One way forward in the pursuit of structuring effective education is to attempt to understand exactly how beginners adopt relevant concepts (Holvikivi, 2010; Teräs, 2007). According to this research students need not only know the syntax and the semantics of the language but may also need to adopt a "*programming way of thinking*" (Holvikivi, 2010). It should be clear from this brief introduction to the nature of programming languages that your intended ends will to a large extent dictate the means to get there, in the same way as starting by learning Latin if your intent is to learn English might be helpful in some regards, but not necessarily the most optimal for your intended use. However, since the fundamental nature of all programming languages is to some extent language-like, it seems only natural that one should try to leverage the vast amount of research on language processing to inform our teaching of programming languages. To this end we will now move into the realm of neuroscience and explore what brain researchers have to say, starting with language.

1.5. Language and the brain

If computer programs are structured in a similar way to other languages, then we should reasonably expect to see it processed in the brain by areas responsible for the various aspects of natural language learning and production. For a simplified picture of how language processing is proposed to work in the brain, we can conceptualize language information entering the brain either via the auditory cortex in the temporal lobe or via the visual cortex at the back of the brain if in written form. This information then travels forward through the brain following two rough pathways known as the dorsal and the ventral pathways, respectively, on its way towards the frontal lobe that is thought to contribute to our conscious experience of whatever we just encountered (Stuss, Picton, & Alexander, 2001). Two cortical structures particularly important for language processing are Wernicke's area in the temporal lobe, chiefly responsible for language comprehension (Mesulam, Thompson, Weintraub, & Rogalski, 2015), and Broca's area of the inferior frontal lobe, classically seen as responsible for language production (Blank, Scott, Murphy, Warburton, & Wise, 2002).

In his 2004 review Ullman looks into the literature of the neuroanatomical underpinnings of language processing and proposes a more detailed model of language wherein the mental lexicon and knowledge depend on the 'declarative memory system', based primarily on temporal lobe structures, and the mental grammar that handles complex rule-based combinations depend on frontal, basal ganglia, parietal and cerebellar structures, the 'procedural memory system' (Ullman, 2004). According to the Declarative/Procedural (DP) model of language, the Lexical/Declarative system is responsible for the learning and use of explicit knowledge with conscious access. This includes the largely arbitrary mental lexicon of word meanings (semantics), facts about objects and events, and categories and exceptions to rules. This system is supported by medial temporal lobe structures, primarily the hippocampus and surrounding regions (subiculum, entorhinal cortex, perirhinal cortex, and para-hippocampal gyrus), inferior, ventral, and superior temporal cortex but also relies on prefrontal areas, primarily Broca’s area in the ventrolateral prefrontal cortex (VL-PFC) (Brodmann areas 44, 45 and 47). This system is considered part of the ventral stream of visual processing.

The Grammatical/Procedural system on the other hand is responsible for implicit, rule-based syntax and grammar processing. This includes managing combinations, structures, sequences, and hierarchies. This system is thought to primarily be supported by the basal ganglia, especially the caudate nucleus, and Broca’s area, but it also includes the premotor cortex, the supplementary motor area (SMA) and pre-SMA, parietal cortex, especially the supramarginal gyrus (Brodmann area 40), superior temporal cortex and parts of the cerebellum (hemispheres, vermis, dentate nucleus) that is thought to be important for retrieval of information and attention. This system is considered part of the dorsal stream of visual processing. It is worth noting that superior temporal cortex and Broca’s area are implicated in both the Declarative and Procedural language systems. Thus, there could be functionally segregated areas within these regions, or they may share connections to the same neuronal populations. Both systems are highly interconnected, cooperatively responsible for language processing, and can even compensate for one another to an extent (Ullman, 2004).

Since Broca's area is implicated as important for both the Declarative and Procedural systems of language in the DP model, it is further hypothesized that Broca's area is part of at least two topographically organized basal ganglia thalamocortical circuits (Ullman, 2006); the anterior/ventral, and posterior/dorsal parts, corresponding roughly to Brodmann areas 45 and 44 respectively, where the anterior/ventral portion of Broca's area receives inputs from prefrontal and superior temporal areas and is primarily thought to be part of the (explicit) declarative system where its main function is the retrieval of lexical and semantic information stored in declarative memory and is therefore responsible for handling the understanding of meaning (semantics). Based primarily on inferences from animal models, Ullman argues that the posterior/dorsal portion of Broca's area receives inputs from parietal cortex and the dentate gyrus of the cerebellum. It is also connected to the SMA and pre-SMA (Ford, McGregor, Case, Crosson, & White, 2010), areas that have been implicated in the processing of sequential motor and cognitive information, rule-based grammar, linguistic and musical syntax, phonology, motor imagery, action control and temporality (Ullman, 2004). All these various types of information processes can be described as manipulation of sequential information and are thus part of the (implicit) procedural system in the DP model. This part of Broca is thought to be especially important for the understanding and use of grammar and complex structures (syntax) and the learning and processing of rules, sequences, and hierarchies. Given the large variation in types of information handled, Broca's area could be considered as a system for multimodal processing of memory, where the posterior/dorsal part of Broca's area could be seen as primarily responsible for what Ullman calls ‘procedural memory’, a system that is responsible for the acquisition and processing of sequences across multiple domains. The division within the Procedural system could thus be that the basal ganglia learn grammar and rules, and posterior Broca’s area uses grammar and rules.

If the processing of computer programming language is found to overlap with how natural language is processed, then perhaps we can settle whether programming should truly be thought of as a language (and hopefully stand to benefit from the field of language education) rather than something else but with language-like features, or perhaps if this distinction is ultimately meaningless. As discussed above, the other primary contender for explaining what programming is constituted of is mathematics, which is classically thought of as processed mainly in the parietal lobe where activity has been found for both arithmetic complexity and mathematical competence, and that also correlates with mathematical and numerical IQ (Grabner et al., 2007). It is also a long-standing debate as to whether the language system plays a role in processing deductive reasoning as argued by Mental logic theory (Rips, 1994, 2001), as opposed to probabilistic accounts (Oaksford & Chater, 2001, 2007) and Mental model theory (Johnson-Laird, 1980). We are therefore positioned to investigate if programming elicits activity in the temporal and frontal language areas, the parietal mathematics areas, or in both systems. The other major contention that we will deal with is regarding the physicality of programming. That is, the effects of physically interacting with the code through the medium of the computer and keyboard. To this end, we will now explore the field of ’Embodied Cognition’ research, and how it ties into education.

1.6. Embodied cognition – action, perception, and cognition

In the 2017 paper "The story so far: How embodied cognition advances our understanding of meaning-making", that research on language comprehension has the potential to give us "*deep insights*" into the foundational workings of knowledge acquisition and, further, that the sensorimotor systems have well-documented impacts on the processing of action-related words (Galetzka, 2017). The field of education already seems well disposed towards what is known as 'Embodied cognition', or sometimes 'Grounded cognition'. At its core, this line of research investigates cognition in a multitude of aspects that make up the human animal. It starts by postulating that our cognitive faculties are evolved to process the physical and social environments that we inhabit, and that they spring from the neural systems responsible for processing sensory information and goal-directed actions (Kiefer & Barsalou, 2013). Approaches to pedagogy based in this view aim to actively incorporate the physical body in the activity or multiple senses by including multimedia in some form (Leitan & Chaffey, 2014). We will discuss these approaches in more detail in the following sections.

The most relevant part of this field for our interests is the various motor theories, dealing with the possible ways that motor control (action), perception and cognition relate to each other. A 2016 paper proposes a system to classify the different models theorized to explain this relationship (Gentsch, Weber, Synofzik, Vosgerau, & Schütz-Bosbach, 2016), and they call this meta-theoretical framework 'Grounded Action Cognition'. They suggest sorting theories based on whether action, cognition and perception are ‘constitutional’ of one another; that they are needed in full or in part for the working of the others, or if they can function independently. It can also be the case that one is vital for a healthy acquisition of the other but that after this they can function independently. They argue that the only “*genuine grounding relation*” is when one of the faculties is ‘partially constituted’ by another, because if they are in a sense one and the same it is impossible to say which is grounded in which, and the concept of grounding becomes meaningless.

Three categories of models are described: ’Common coding’, ‘Internal’ and ‘Simulation’ theories. The common coding theories assume that there is some commonly coded representation of action that can be shared between motor and perceptual systems in order to plan actions. In this category, the ‘Ideomotor theory’ (Shin, Proctor, & Capaldi, 2010) holds that action and perception are separate but can be activated cooperatively, while the theory of ‘event coding’ (Hommel, Müsseler, Aschersleben, & Prinz, 2001) proposes a common code representing the structure of planned or perceived actions where this code is independent of both action and perceptual domains, serving instead as an indirect communication between the two that can be used to detect when action and perception overlap in their outcomes. Neither of these models are considered to be constitutive, meaning that perception, cognition and action are treated as separate but interacting processes in these frameworks.

The internal models propose that the motor system contains anticipatory models (Gentsch et al., 2016). This means that action, perception, and cognition are informed by their predictions. These anticipatory models are generated from motor commands but are not fully constituted by them because feedback forms an integral part. In this category, the ‘Motor control theory’ (Miall & Wolpert, 1996) holds that anticipatory models of what the sensory inputs will be after a planned physical action is calculated in the motor cortices, and later compared with the outcomes to optimize the movement. These mental models are not the same as the planned motor activity though but a separate activity, so this theory is thought of as partially constitutive. In ‘Predictive coding’ (Friston, 2005) models however, motor control and perceptual inferences are thought to be one and the same, but cognition based on exteroceptive modalities, like vision and hearing, are thought to be only partially constituted of the motor system. This is because it is postulated that unexpected sensory information is translated into a motor representation in order to potentially modify movement later. The ‘Emulation theory’ (Grush, 2004) on the other hand tries to explain the cognition of motor imagery by postulating that we emulate our bodies within motor centers, and that there are both dedicated and more general systems participating in the generation of this.

Finally, the simulation theories are simply based on the assumption that cognition is simulation-like. The ‘Mirror neuron theory’ (Gallese, Fadiga, Fogassi, & Rizzolatti, 1996) holds that the same population of neurons are responsible for processing both observed behavior in others and execution of those same sets of actions for oneself. This is in line with the ‘Motor imagery account' (Decety, Jeannerod, & Prablanc, 1989), where motor imagery is thought of as simulations of possible actions. In the theory of 'Perceptual symbol systems' (Barsalou, 1999) this is carried to its extreme, where cognition is thought to consist exclusively of activation patterns in our various sensory modalities. Cognition and perception share the same representations and is in itself a reactivation of sensory states, both in introspective cognition and simulation of the mental states of others. In all these three accounts, action and perception are merged in a mechanistic way, and they are thus fully constitutive.

Which one (or a combination) of these accounts are correct is presently unsettled, but the accumulation of evidence supporting complex interactions between perception, action and cognition is substantial (Gentsch et al., 2016). Next, we will outline what research into embodied cognition has to say specifically about language learning.

1.7. Motor actions and language learning

The idea that learning is an activity, and that the learning process should include practical elements has its roots far back in time. According to Comenius (Comenius, 1642) real knowledge and understanding are based on the individuals' own exploration - on their experience of the things themselves. "*The disciples shall soon learn to express what they know using both hand and mouth, therefore you (the teacher) must not leave anything until sufficiently absorbed by the disciples' eyes, ears, mind and memory*" (Larsson, 2007). Similar ideas crop up regularly in the educational literature as evident from scholarly pieces of Rousseau (Rousseau, 1762), Dewey; ‘My pedagogical creed’(Dewey, 1897), and Piaget, who points to the importance of students being active, and for them to seek learning themselves: "*In every field, action comes first, classification and conceptualization come later*." (Piaget & Bringuier, 1980). Empirical studies showing greater retention of knowledge when students actively manipulate tools in comparison to simply listening to the teacher supports this notion (Good, Feekes, & Shawd, 1993; Paris, Yambor, & Packard, 1998; Stohr-Hunt, 1996; Young, 2002), and when it comes to language there is evidence showing that gestures can in fact enhance memory formation in foreign language learning. Macedonia & von Kriegstein (2012) argue that classroom studies on second language learning show advantages of so-called ‘multimodal learning’, where the teaching is not constrained to reading and listening (Macedonia & von Kriegstein, 2012). They also describe four ways that gestures could impact language. First, that the physical enactment creates a motor trace in the memory-representation of the verbal information. Second, that doing things in a ‘wider’ way (involving more cognitive activities) can lead to better verbal memory. Third, that motor imagery (kinesthetic representations) is what leads to improved performance. And fourth, that the impact is primarily caused by increased perceptual- and attention-processes when moving or using objects. It should be noted that these are not mutually exclusive options.

Broadly, it could be the case that motor activity forms a part of the memory, strengthening it, or it could be that motor activity strengthens the memory during learning but does not form a constitutive part after.

A 2013 study used functional magnetic resonance imaging (fMRI) to investigate how meaningful and irrelevant gestures impacted learning of new made-up words, mimicking the process of learning a second language, in which the vast majority of the new words were nouns with distinct verb-usages such as ‘hammer’, ‘violin’, ‘spade’ and so on. They found no behavioral effects between the different conditions but performing meaningful gestures with each word during training produced larger activations in Broca’s area; inferior frontal cortex (BA47) and inferior temporal gyrus. Thus, they conclude that their findings challenge the existence of an ‘enactment effect’ of gestures, for learning single words at least, but that their imaging data shows that gestures lead to “*deeper semantic encoding of new words*”. This study only had people learn to associate new words with already memorized meanings however, so how gestures affect the learning of new concepts is left unanswered, but it also did not rule-out possibly beneficial effect of the motor action. Older studies focus not on learning words, but on remembering imperative commands like “clap your hands” or “put on the glove” (Bäckman, Nilsson, & Chalom, 1986; Nilsson, 2000). These studies are called ‘Subject Performed Task’ or ’SPT’ paradigms, where the aim is to examine the effect of performing a task rather than simply memorizing the command. Nilsson, 2000 found that recall in the SPT condition was generally superior, and that participants who performed the actions were less impaired by distractions that could divide attention. Bäckman et al., 1986 concluded that physical movement improves memory if it leads to a stronger integration of the verbs and nouns within the commands the participants were trying to remember.

As detailed above, there exist several proposals as to why sensorimotor processes could facilitate cognitive learning (Macedonia & von Kriegstein, 2012). To re-iterate; One proposal holds that hands-on learning forms ‘motor traces’ in the memory representation of the item, which strengthen the association formation and in turn facilitates retention (Engelkamp & Zimmer, 1985; Zimmer & Engelkamp, 1985), and this notion is based in Schema theory (Bartlett & Bartlett, 1995), or more specifically the multisensory learning theory (Shams & Seitz, 2008)**,** which are theories stipulating that we all make sense of new experiences by activating associative memory representations, or schemata, where prior multimodal experiences enable the reactivation of a range of unisensory components. Empirical data in support of this theory include that functional connectivity (strength of neural signaling) between various sensory cortices is strengthened after multimodal learning, when a simple unisensory stimulus is later being processed (Seitz & Dinse, 2007; Von Kriegstein & Giraud, 2006). Another example is research showing that activity in motor cortex is predictive of having learned words enriched by gestures or not when the knowledge is later retrieved (Mayer, Yildiz, Macedonia, & von Kriegstein, 2015).

A second proposal holds that hands-on learning benefits come about through enhanced perceptual and attentional processes that would take place when objects and actions interact (Bäckman et al., 1986). This idea relates to theories stipulating that attention is a limited resource, and that by directing awareness to relevant stimuli, we most efficiently process categorization and association of information (Broadbent & Ladefoged, 1959; Bundesen, 1990; Lavie, 1995; Treisman, 1964). Studies in support of this proposal are those that fail to find motor activation during tests of retention after hands-on learning, but instead find increased activation in task-related areas, such as the left inferior frontal- (Broca's area) and left temporal (Wernicke's area) regions after language learning with gestures (Krönke, Mueller, Friederici, & Obrig, 2013). It is worth noting however, that an alternate, closely related theory may account for these findings: The levels-of-processing theory (Craik & Lockhart, 1972). This theory stipulates that language areas will become more engaged following in-depth processing, since more complex learning involves deep analysis of the meaning of the study material (i.e. analysis of the semantics) that is rehearsed (Nyberg, 2002), which in turn leads to better recall.

Of particular relevance to the current work in this thesis are some recent studies that have tried to examine the embodied effects of actually typing on keyboards in typical lecture settings, where one study concluded that taking notes on a laptop seemed to lead to "*shallower processing*" compared to longhand writing (Mueller & Oppenheimer, 2014). The laptop students showed worse performance on both facts and conceptual questions, and the authors hypothesize that this could be because of the laptop-user's tendencies to transcribe the lecture verbatim instead of re-formulating, distilling, or summarizing the concepts themselves. Suggestively, this could be explained by the haptic feedback from the freehand writing leading to less energy/attention being spent on error-detection, thus freeing up mental resources for deeper reflections on the content. Another study also showed that free recall of word lists were significantly better when the lists were memorized by copying them with pen and paper compared to both laptop keyboard and iPad touchscreens (Mangen, Anda, Oxborough, & Brønnick, 2015). Participants were tested immediately after the memorization, first on free recall and, after that, recognition with lists with distraction words mixed in. Only the free recall differed significantly depending on whether a pen or a digital device had been used during encoding. The authors argue that this is experimental evidence for the importance of the sensorimotor attributes of writing and that writing is an "*embodied process*". Taken together, perhaps this could be one of the factors setting computer education apart from other subjects. Whatever the case, researchers in the fields of education and embodied cognition have been hard at work trying a seemingly endless number of approaches to improve learning outcomes. In the next section I will present a few of these efforts, including new types of media and technology.

1.8. Embodied cognition and programming

As we will see, educators are willing to try their hand at creating useful learning strategies based on the concepts from embodied cognition research, but since the basic science is not quite settled yet it lacks a strong, theoretically guided 'bottom up' approach, resulting in a lack of compelling evidence for effectiveness (P. Howard-Jones, 2014). Let me stress that this is by no means unusual when it comes to how science operates, at least in the early days of a field, but if we grant that embodied cognition approaches fall under the broader umbrella of education, educators could understandably be frustrated by the apparent lack of concrete, readily actionable outcomes on this front. Since we are in the digital age, it should also come as no surprise that researchers have been trying various multimedia solutions in conjunction with embodied paradigms of learning, to help students absorb new information. Some have also tried to provide approaches to help sort through the resulting mass of studies. In a review, the authors presented possible ways of measuring cognitive load of different proposed learning approaches, in an effort to discern the most optimal ones from the perspective of student experience, and they write that "*research in embodied cognition has inspired a number of studies on multimedia learning and instructional psychology*", and that while both subjective and physiological measurements of cognitive load do appear in recent studies on embodied cognition, judgements of learning might be better predictors of learning than cognitive load measurements (Skulmowski & Rey, 2017). An example of a multimedia study is the 2010 paper by Katai & Toth, where the authors try to integrate as many sensory modalities as possible using multimedia and 'the arts' (dance, music, rhythm and role-playing) in their programming education (Katai & Toth, 2010). This is an attempt to convey algorithmic thinking, and they do claim to show an improvement in sorting algorithm-knowledge when comparing their method with a conventional teaching group, but there is no way of discerning what part or parts of their approach is responsible and to what extent. It is also worth noting that this type of education is described as "*time-intensive*" by the authors. There has also been attempts to evaluate the effectiveness of hands-on education specifically in computer programming (Handur et al., 2016). The authors argue that this has been a "*less researched area*" and that "*adequate literature is not available on 'hands-on science' specifically for programming course[sic]*". They also concluded that their hands-on course led to improvements in performance, problem solving skills and logical thinking in their students.

As we have seen, there is both a demand and the potential for neuroscience to supply the basic understanding of the processes relevant for education in general, and programming in particular. In her 2010 paper, Jodi Tommerdahl lays out a hierarchical model of education science where neuroscience is the most basic research that then informs cognitive neuroscience, followed by psychological mechanisms, then educational theories, and finally, the classroom dynamics at the top (Tommerdahl, 2010). She wrote that it could take many years or even decades to establish what she called "the new science of evidence-based education" and predicted that the road between the lab and the classroom will be longer than what many researchers think. We have now dealt with learning, language, and programming from multiple angles but only in a tangentially related fashion. In the penultimate section, we will detail the existing work in the specific sub-field of neuroscience of computer programming learning.

1.9. Neuroimaging and computer programming

In neuroimaging more broadly, there are many studies dedicated to the formation of different types of memories like short term, long term, episodic, visual, motoric and lingual, but to my knowledge only seven studies on computer programming has been published as of the writing of this thesis (Duraes, Madeira, Castelhano, Duarte, & Branco, 2016; Floyd, Santander, & Weimer, 2017; Krueger et al., 2020; Lee et al., 2016; Peitek et al., 2018; Siegmund et al., 2014; Siegmund et al., 2017).

The first study to examine programming using neuroimaging methods was published in 2014, where the stated aim was to explore whether fMRI could be used to measure program comprehension (Siegmund et al., 2014). They found significant activity in five regions in the left hemisphere (Brodmann areas 6, 21, 40, 44 and 47) when contrasting code comprehension and a syntax error task. The authors broadly interpreted their findings as "*related to working memory, attention and language*". They have since performed two follow-ups to their original experiment. The intent of the first one was to examine the differences between 'bottom-up' comprehension and 'top-down' reading of code using 'semantic cues' (helpful variable and function names) and 'beacons' which are indicators of the purpose of the code (Siegmund et al., 2017). They were able to replicate the finding that Brodmann area 21 (temporal lobe), 40 (parietal lobe) and 44 (Broca's area, inferior frontal gyrus) in the left hemisphere were activated during code comprehension, however they were unable to replicate areas 6 (premotor cortex) and 47 (orbital inferior frontal cortex) possibly because of the low number of trials compared to the original experiment. They also showed bilateral de-activation in area 39 during comprehension with semantic cues as compared to bottom-up processing, but not compared to their rest condition. The second follow-up in 2018 was essentially a replication of the original study reaffirming the primary conclusions that areas linked to working memory, attention, and language processing are active during code reading (Peitek et al., 2018).

Another study into the underpinnings of programming aimed at examining the act of debugging code, a frequent task of professional programmers (Duraes et al., 2016). During fMRI, participants first used a joystick to mark lines of code suspected of containing errors, and subsequently decided if it was a bug or not by responding using a button. The goal was to examine the difference between bug suspicion and bug detection. The authors conclude, in accordance with Sigismund et al. 2014, that "*programming comprehension recruit reading and general language processing networks*", and that the medial frontal region is active during bug detection (confirmation) while insular cortex was processing suspicion more than confirmation.

Programming comprehension in self-reported novice and expert programmers has also been investigated using EEG (Lee et al., 2016). They focused their analysis on the higher Beta and Gamma frequencies. Using the same task developed by Siegmund et al. 2014, they found that Brodmann areas 6 (premotor cortex) and 44 (Broca's area, inferior frontal gyrus) were active during programming comprehension, and that the experts exhibited higher Beta and Gamma brainwave activation in the F3 (frontal lobe) and P8 (right parietal lobe) electrodes, which the authors interpret as linked to the "*superior programming comprehension*" of the expert group and that they "*therefore excels at digit encoding, short-term* *memory (solving simple programs in a short time), and SME (the ability to recall program functions after an extended period of time)"*. Though the authors expressed some hesitation about their findings, given the relatively sparse research in this field, the reported areas of activation are again in line with the results from Siegmund et al. 2014.

In the same vein, Floyd et al. 2017 published a paper where they tested whether a machine learning classifier could correctly separate code comprehension, code review and prose review from fMRI data in experts and more novice programmers (Floyd et al., 2017). The classifier could do so successfully approximately 80% of the time, and they argue that this implies that "*largely distinct neural representations*" are responsible for these tasks. The areas pushing the classifier towards a code task includes frontal cortex, occipital cortex, parietal cortex, and ventral temporal cortex. They further claim that based on the weights used by the classifier to distinguish code tasks from the prose task, that the areas responsible for code comprehension and review are highly correlated, meaning that the same areas are responsible for both evaluating the meaning of code in general and evaluating if a change is warranted or not. Their most interesting find, however, was that the classifier did a worse job distinguishing between code and prose in the experts, and that the accuracy was negatively correlated with computer science grade point averages. The authors interpreted this to mean that the more expert the programmer, the more language-like the code processing will be. Though intriguing, interpretation of these results is somewhat difficult because of the way the authors deigned their code tasks, where both code and written commentary on the code were present simultaneously on screen thus blurring the distinction between code and prose.

The latest paper to be published in this field is another study focused on distinguishing between code and prose and is the first study to investigate code being written while in the scanner (Krueger et al., 2020). The authors tested both short fill-in-the-blank type questions as well as longer free form responses, and they found that while prose writhing activated the classical left-hemisphere language areas in line with the previous studies, the long form code writing activated right-hemisphere areas associated with attention, memory, planning, and spatial ability. They thus concluded that writing code is a distinct neural process from code reading.

In addition to being supported by language areas, one would predict that hands-on computer programming requires typical action-related visuospatial information processing. One of the most influential theories within Cognitive Neuroscience; The dual pathway (Goodale & Milner, 1992) proposes that visuo-spatial actions make use of the ventral and the dorsal stream. The ventral stream transfers information from the visual cortex to Wernicke's and Broca's area about the meaning of objects and symbols, while the dorsal stream transfers information from the visual cortex to the parietal and superior frontal cortices about spatial mental models of relations, functions, and manipulation of objects (Milner & Goodale, 1995; Mishkin & Ungerleider, 1982). Although this theory is supported by many empirical studies, it is surprising how little is known of how information is integrated between the two streams (McIntosh & Schenk, 2009).

All these previous studies agree that solving programming problems does depend on language areas. It could however be more informative to focus on simple tasks with novice programmers to study early knowledge acquisition and processing rather than experts evaluating more complex problems. One example of such a benefit is, as Siegmund et al. 2014 suggests, that experts have a harder time ignoring meaning in code leading to less contrast between the syntax- and semantics-tasks. There is also less concern about the extent of practice or qualitative differences between the experiences of the participants with novices. Taking this into consideration, we now have all the pieces in place for designing our study of hands-on learning in novice, object-oriented pair-programming.

1.10. The current project and its rationale

Building a rigid field of study in this fashion, informed from the ground up instead of proposing theories ‘top down’ is arduous, but should have the potential to eventually produce results robust enough to warrant investing in implementing practice derived from them. Previous research in the field of programming education by Eckerdal and colleagues, point to learning benefits when students are engaged in hands-on learning, i.e. typing on the keyboard (Berglund & Eckerdal, 2015; Eckerdal et al., 2007; Höök & Eckerdal, 2015). It is often the case in the classroom that students work in pairs, where only one student at a time is the keyboard driver, so-called pair programming (Salleh, Mendes, & Grundy, 2011). However, the mechanisms behind any learning effects of physically interacting with the keyboard have not been investigated. Previous studies outlined above deal with the effects of motor activity on learning outcomes on word-meaning associations or actions that are symbolic of the very activity to be remembered. Keyboard typing involves potential motor impact when learning to associate more abstract knowledge in the absence of a clear ‘meaningful’ link between the motor actions themselves and the material to be learned, such as the case is when learning to program. This forms the basis for our project: To investigate the role of keyboard typing when learning language-like rules and facts of programming in a pair programming-setting. This thesis covers work that is one part of a collaborative interdisciplinary effort jointly established by my supervisor Sara Bengtsson and my co-supervisor Anna Eckerdal, under the project title ”Hands-On in Programming Education”, or ”HOPE”, where me and my supervisor will attempt to answer the neuroscientific questions, and my co-supervisor and her Ph.D. student Kristina von Hausswolff will cover the more traditional educational perspectives.

By training programming naïve teenagers in a classroom like setting and subsequently study their brain activation with fMRI in a systematic, hierarchical fashion, this research project attempts to address cognitive processes of programming language acquisition. In addition, we have investigated if there are any neural processes responsible for any increase in learning outcomes gained from physically typing on the computer keyboard during learning. More specifically, we have investigated the role of the actual motor involvement on programming knowledge, cognitive engagement, motivation, confidence, and long-term memory. I will present two studies in this thesis: In chapter 2, I will detail the main study, laying the groundwork for this thesis, encompassing several fMRI analysis methods in an attempt to answer the core questions in our part of the HOPE project. This will then be followed, in chapter 3, by a description of findings from the re-analysis of a previous fMRI dataset of a rule-switching task with the aim to address questions raised in the main study regarding the effect of feedback as a potential intervention on stress reduction and as a motivational booster.

Chapter 2

The neural correlates of hands-on computer programming learning

2.1. Hypotheses

This experiment was designed to answer a series of research questions, but the primary motivation behind its design was to investigate, in an interdisciplinary way, if the notion that working hands-on leads to better outcomes is true, and if so, how this advantage is instantiated in the brain in the case of learning to program. Students refer to programming as "*writing a computer program and to study what happens when the program is executed*" (Eckerdal, Thuné, & Berglund, 2005). Indeed, a programmer needs both theoretical knowledge about rules of syntax and branching logic, in addition to mastering the practical skills needed to implement and test new code. Whereas the social aspect of peer learning may be beneficial on its own, in the current study we address differences in underlying learning mechanisms between being hands-on (keyboard typing) and hands-off.

We test the following hypotheses:

(I) Learning computer programming hands-on will lead to better performance, better long-term memory for the task, and greater confidence for the task, than learning programming hands-off. We test if motor cortices, attentional processes, and/or language processes yield the additional support to the mental model of programming in those who learned the task hands-on.

(II) The various aspects required to successfully solve programming tasks require a hierarchical recruitment of multiple brain regions.

(III) Neural computations of programming code take place in dual streams originating in the visual cortex, and we investigate where information between the two streams is integrated.

(IV) The connectivity patterns between the active brain regions are relevant in addition to the activity levels of the individual brain regions when attempting to explain what constitutes the processing of programming knowledge and how other measured variables interact with it.

(V) Broca’s area is involved in deduction of logical reasoning.

(VI) Grades in mathematics and the participant’s confidence will contribute to predicting learning outcome.

2.2. Method

2.2.1. Experiment overview

The experiment consists of two parts: First, a learning session designed to teach a beginner the rudiments of Java programming where students work in pairs. We randomly ascribe one in the pair to type on the keyboard and the other student to sit next to the typist. We then test the acquired knowledge on day 1. This is followed by an fMRI session within two days, designed to mimic the same type of tasks encountered in the learning session while at the same time accommodating the constraints of the fMRI setting and the requirements of the planned statistical analyses. Importantly, during fMRI scanning, all participants respond with a single button press in each trial. Thus, regardless of learning condition, when brain images are acquired, the motor output is the same among the hands-on and hands-off group. Performance is again measured during fMRI scanning, as well as in a long-term memory test seven days after the learning session. In addition to performance outcome, we collect data on various intrinsic variables, such as stress and confidence (see below for a comprehensive description),and the participants' final grades from their primary education (year 9).

2.2.2. Participants

A total of 53 healthy volunteers with normal or corrected to normal vision and without prior programming experience were recruited from four

Swedish upper secondary schools around Stockholm in second- and third year natural science programs (age 17-18). 7 additional subjects (age 18-22) were recruited via an online platform for research volunteers (www.studentkaninen.se) in order to reach our desired minimum of 25 participants in each of our two conditions, since we wanted to ensure that our statistical power was higher than the n=17 participants that was used in the first ever study of programming using fMRI that found several significant activations (p<.05 FDR corrected) in language areas (Siegmund et al., 2014). 3 of the subjects later failed to show up for the fMRI experiment, a further 2 had to be excluded because of undisclosed dental braces and 1 due to equipment failure, making the final number of fMRI participants n=54 (age mean ± s.d. 17.7 ± 1.6 years; 31 females). The study was approved by the regional ethics committee EPN Uppsala, Sweden, Dnr 2017-178.

2.2.3. Hands-on vs. hands-off learning session and written test

For a more comprehensive description of the learning session and the written test, please see the following publications from our project collaborators: (Von Hausswolff & Eckerdal, 2018; Von Hausswolff, Eckerdal, & Thuné, 2020).

The experimental paradigm begins with a 4-hour long learning session, written by our collaborators in the field of computer science-education, the goal of which is for the participants to learn the basics of programming in Java using the Dr Java development environment (www.drjava.org). In this educational session, the students worked in pairs (one to four pairs per session). In a typical programming educational session, it is commonplace to work in pairs in a computer lab setting. Often, one of the students will be hands-on, (typing) while the other sits next to the driver discussing solutions hands-off, then switching places (Williams, Kessler, Cunningham, & Jeffries, 2000). We artificially imposed fixed hands-on or hands-off roles upon the participants, randomly assigned. The hands-on students were the only ones allowed to interact with the computer and the hands-off students had to merely follow along with the exercises and discuss the solutions. The final split between the two conditions, after the exclusions during the fMRI scanning, ended up as follows: Hands-On, n=25 (age mean ± s.d. 17.6 ± 1.5 years; 15 female), Hands-Off, n=29 (age mean ± s.d. 17.8 ± 1.7 years; 16 female)

The material in the learning session consisted of a projected pre-recorded lecture video that was following along a series of progressively more complex programming examples and tasks illustrating core programming concepts and examples of problem solving. This was done in the effort to make the lesson as similar as possible for all participants. The video was paused after each section to allow the students to work through the problems one by one and allow the teacher to interject with pointers and additional information or help if needed. A printed worksheet was also provided to go along with the lecture. The learning session was concluded with a written test tailored to evaluate the programming knowledge acquired, consisting of both multiple choice- and complete the program-type questions. We evaluated students' engagement with the study material by means of video recordings of each working pair. Parts of the videos were scored by three independent raters, resulting in a score for each participant indicating that they were >60% engaged, <40% engaged or about equal compared to their partner. Engagement here being a count of all comments, and gestures that were thought to indicated problem solving (Von Hausswolff et al., 2020).

2.2.4. Questionnaires

Participants rated their stress and motivation for the task itself, before and at the end of the learning session, on a 1-10 Likert scale. They also completed the 'Intrinsic Motivation Index' (IMI) (Deci & Ryan, 2013) and the 'Need for Cognition' (NFC) scale before the onset of the lesson day 1 (Cacioppo, Petty, & Feng Kao, 1984). The participants later filled out the 'Dweck Mindset' scale when they came back for the fMRI scan (Dweck, 2008b). Behavioral data were analyzed using Microsoft Excel v16.0, IBM SPSS Statistics v25 for regression analyses, and IBM SPSS AMOS v21 for structural equations modelling.

2.2.5. Long-term memory test

One week after the Learning session (7±1 days), the subjects were contacted either by phone or email and asked to take a written test of their retained programming knowledge. This test of Long-term Memory (LTM) had the exact same structure as the test at the conclusion of the lesson and was designed to be as similar as possible without being a direct copy of the same questions. Both the written test day 1 and the LTM test included 11 questions split into two parts: First, 8 multiple choice questions, and then 3 written questions where the students had to compose working programs from scratch.

2.2.6. fMRI paradigm

Within two days of the learning session the participants came back to undertake our fMRI programming tasks. Before entering the scanner, the participants were familiarized with the fMRI task by running 9 unique demo trials on a PC until they understood the different tasks. We presented the participants with a task paradigm that used the same style of programming questions included in the learning session and test but adapted for the constraints of fMRI and our experimental design. In total, we presented 120 trials of binary Yes/No-questions split evenly between four conditions (30 trials/condition) of increasing simulation complexity:

1. ‘Syntax’ - Relatively simple code where the participant evaluates whether the code could be compiled without errors, or not. There were three different possible errors: missing quotation marks, terminating semicolons, or variable declarations (“int”/”string”). This condition reflects a ‘shallow’ code-reading.
2. ‘Simulation’ - The exact same code as in the ‘Syntax’ task above but without errors and an added output line at the bottom. Participants evaluate whether the code generates the stated output, or not. Hence,

for successful performance, the participants are required to simulate the meaning of the code in this condition.

3. ‘Spatial Simulation’ – Code is presented together with a still image of an object (turtle) in a 2D grid. Here the image serves as the output. The participants evaluate whether the output image matches the code or not. The code uses function-calls like ‘move’ and ‘turn’ to shift the position of turtles on the screen. This condition requires the participants to simulate the meaning of the code, including the steps of the turtles, to decide if they have ended up in the correct positions.
4. ‘Iterative Spatial Simulation’ – The same as ‘Spatial simulation above but incorporating while-loops in the code. Hence, this condition puts more demand on keeping various code elements in mind (see figure 2.2).

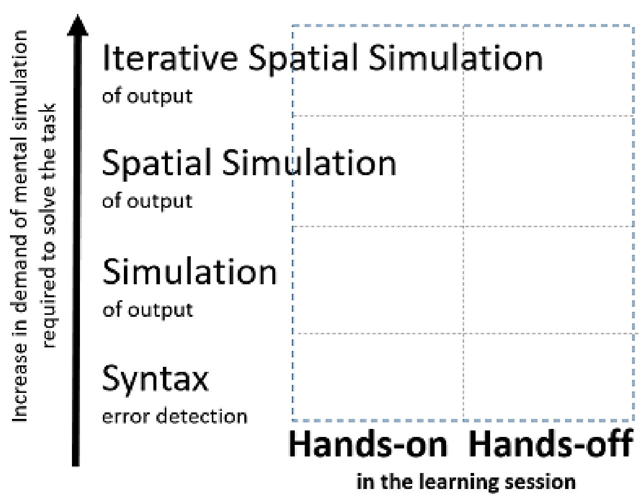


Fig. 2.1. Experimental design of the fMRI experiment. (2x4 factorial design)

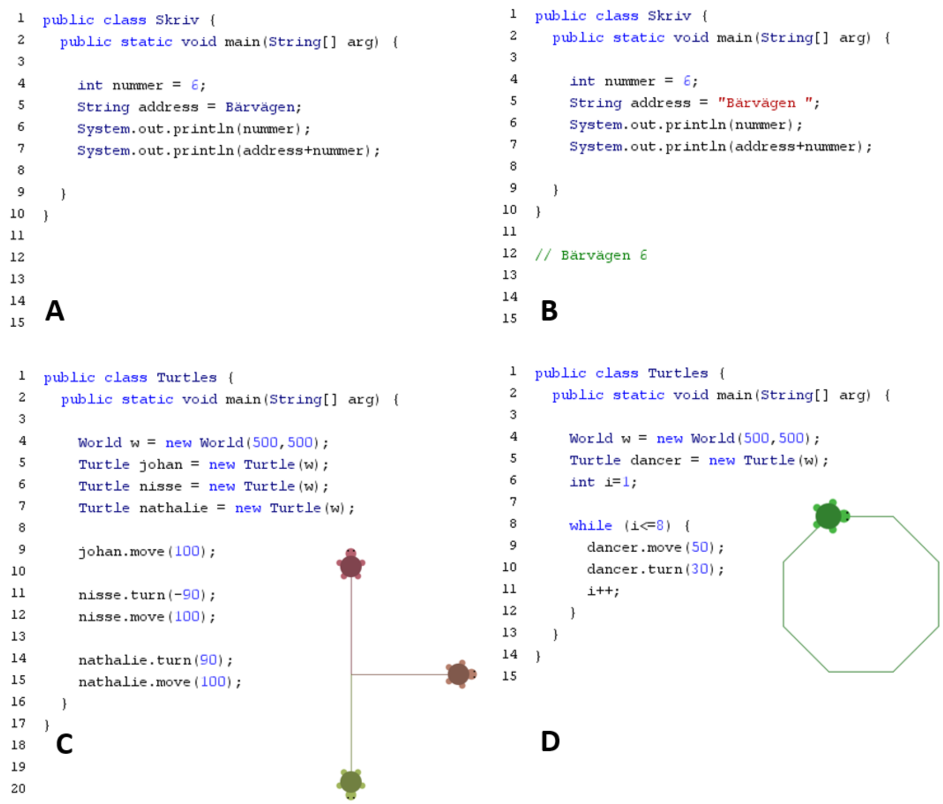


Fig. 2.2. Examples of the four fMRI tasks. A: Syntax. B: Simulation. C: Spatial simulation. D: Iterative spatial simulation.

The experiment was divided into 3 sessions of approximately 20 minutes each allowing the participant a minute or two of rest in the scanner between each session. The questions were presented in pseudorandom order (the same for all participants) with a balanced mix of all 4 question types in all 3 sessions and with equal amounts of ‘Yes’ and ‘No’ as the correct answer. Each trial begins with a 200ms written cue heralding the upcoming task, where the text “Syntax”, “Meaning”, or “Turtle” was displayed on screen. We chose to display the word “Meaning” to denote the simulation task since the verbal instruction to the participants was specifically to attend to the *meaning* of the code, unlike in the ‘syntax’ task where they were explicitly instructed to *ignore* the meaning and only look for the three possible errors, which were also explicitly listed for them. We also did not distinguish between the two spatial tasks in our instructions for the participants, meaning that the cue “Turtle” could be followed by either the ‘spatial’ or the ‘iterative spatial’ task. Our instruction was simply that the “Turtle” cue meant that they were about to be presented with code and an image output from the turtle-program they were familiar with from the programming lesson. The instructions are followed by a blank screen of a variable delay of 4-8 seconds in steps of 444 ms, producing 10 delay lengths. The purpose of the variable delay period is to enables us to study brain activity related to task preparation, or ‘task set’ (Bengtsson, Haynes, Sakai, Buckley, & Passingham, 2009; Sakai & Passingham, 2003). This activation should reflect task preparation and is not confounded by potentially unbalanced visual stimuli between conditions. The actual programming task is displayed for either 13 seconds for ‘simulation’ and ‘syntax’, or 20 seconds for the two spatial simulation tasks. After the task stimuli disappear from the screen, the participant has a 3 second window to answer either Yes or No by pressing left or right on a button box according to what is prompted on the screen, where what side represents Yes is also pseudorandom. This setup increases the likelihood that brain activation during task will reflect topic specific problem-solving, without being contaminated by response related motor preparation. After the answer prompt, a confidence scale of 1-4 is presented for 3 seconds after all 60 of the spatial simulation tasks (50% of total trials), asking the participants to indicate the confidence in the answer just entered. This is also done on the button box: 1=No confidence/guess, 2=Not really sure, 3=Pretty sure, and

4=No doubt. Each trial ends with an Inter-Trial Interval (ITI) of either 4 or 7.5 seconds to keep the trials out of phase with the repetition time of the scanning sequence (TR). The total length of a single trial, accounting for the variable delay, the two task lengths, the presence or absence of a confidence-prompt and the two ITI:s was between 24.5 and 42 seconds. During the variable delay and the ITI periods, the screen was black with a white fixation cross in the center of the screen.

2.2.7. fMRI data acquisition

Nuclear Magnetic Resonance Imaging was acquired on a 3 Tesla GE Discovery MR750 equipped with an 8-channel phased array receiving coil. Functional MRI was performed in three session of a gradient echo pulse sequence of 612 volumes, using 3 mm isotropic voxels, TE = 30 ms, TR = 2205 ms, FoV = 23 cm, 47 bottom-up interleaved axial slices, flip angle=70 deg, for a total of 1836 volumes (~70min.). 3D T1-weighted SPGR (Spoiled Gradient Echo pulse sequence) images was acquired with 1 mm isotropic voxels, TE = 3.06 ms, TR = 7.9 ms, TI = 450 ms, FoV = 24 cm, 176 axial slices, flip angle = 12 deg.

2.2.8. Image preprocessing

Conversion of DICOM files into NIFTI-format was done using the dcm2nii software (https://www.nitrc.org/projects/dcm2nii/). Preprocessing and all subsequent analysis was carried out using SPM12, version 7487 (Penny, Friston, Ashburner, Kiebel, & Nichols, 2011). The following procedure and settings were used:

1. Realign: Estimate

Quality: 0.9, Separation: 4mm, Smoothing: 5mm FWHM, Registering all 3 sessions to the first image in the session 1 time-series, Interpolation: 2nd Degree B-Spline.

2. Realign: Reslice

All fMRI images and the mean image, Interpolation: 4th Degree B-Spline, Masking enabled.

3. Coregister: Estimate

Reference image: T1-weighted image, Source image: resliced mean image, Other images: Resliced images, Objective function: ‘Normalised Mutual Information’, Separation [4 2] mm, Tolerances [0.02 0.02 0.02 0.001 0.001 0.001 0.01 0.01 0.01 0.001 0.001 0.001], Histogram Smoothing [7 7] mm.

4. Normalise: Estimate & Write

Images to align: T1-weighted image, Images to write: Resliced mean image & coregistered images (T1 & fMRI sessions), Bias regularisation: 0.0001, Bias FWHM cutoff: 60mm, Tissue probability map: *spm12\tpm\TPM.nii,* Affine Regularisation: ‘ICBM space template - European brains’, Warping regularisation: [0 0.001 0.5 0.05 0.2], Smoothness: 0, Sampling distance: 3mm, Bounding box: [-78 -112 -70 ; 78 76 85], Voxel sizes: [2 2 2] mm, Interpolation: 4th Degree B-Spline.

5. Smooth

Images to smooth: Normalised images, FWHM: [12 12 12] mm, No Implicit masking used.

2.2.9. GLM analysis

2.2.9.1. First level GLM model

The settings used for building the regressors for our GLM modelling was a Microtime resolution of 16 bins, and a Microtime onset of 8 bins. Each GLM model consisted of three sessions of 612 TRs of 2205ms each with the following conditions: ‘Cue_Syntax’, ‘Cue_Simulation’ and ‘Cue_Spatial’ (both spatial tasks) modelling the presentation of the cue. Six regressors modelling epochs of preparatory activation: after the cue and before task stimuli presentation categorized as either short or long: ‘Delay_4_Syntax’, ‘Delay_8_Syntax’, ‘Delay_4_Semantic’, ‘Delay_8_Semantic’, ‘Delay_4_Spatial’ and ‘Delay_8_Spatial’ and four regressors modelling the epochs of correctly answered events of Task: ‘Task_Syntax’, ‘Task_Semantic’, ‘Task_Spatial’, ‘Task_Iterative_spatial’. Button-presses, i.e., timings for input during the answer and confidence prompts were modelled as one regressor, and finally, the six Realignment Parameters (RPs) from the SPM realignment preprocessing step to account for head movements. The Button-presses and RPs were treated as regressors of no interest. As stated above, only correctly answered trials were modeled, and we combined the variable delay lengths into high (‘8’) and low (‘4’) for simplicity. The conditions labeled ‘Delay_4’ include delay periods of 4 to 5.776 seconds, and ‘Delay_8’ includes delays of 6.22 to 8 seconds. All this adds up to a first level model for each participant with 63 columns of regressors: 20 conditions, in 3 sessions, plus the three columns encoding the session means (see figure 2.3). The SPM settings for model specification and estimation were as follows: High-pass filter: 128 seconds (~7.8 mHz), to remove low frequency correlated noise. No Global normalization of image intensity. Masking threshold: 0.8, to exclude non-brain voxels from the statistics. Serial correlations accounted for by Auto-Regressive AR(1) modelling (removes further artefactual correlations from cardiac, respiratory, motion and hemodynamic sources). Classical (non-Bayesian) model estimation was used.

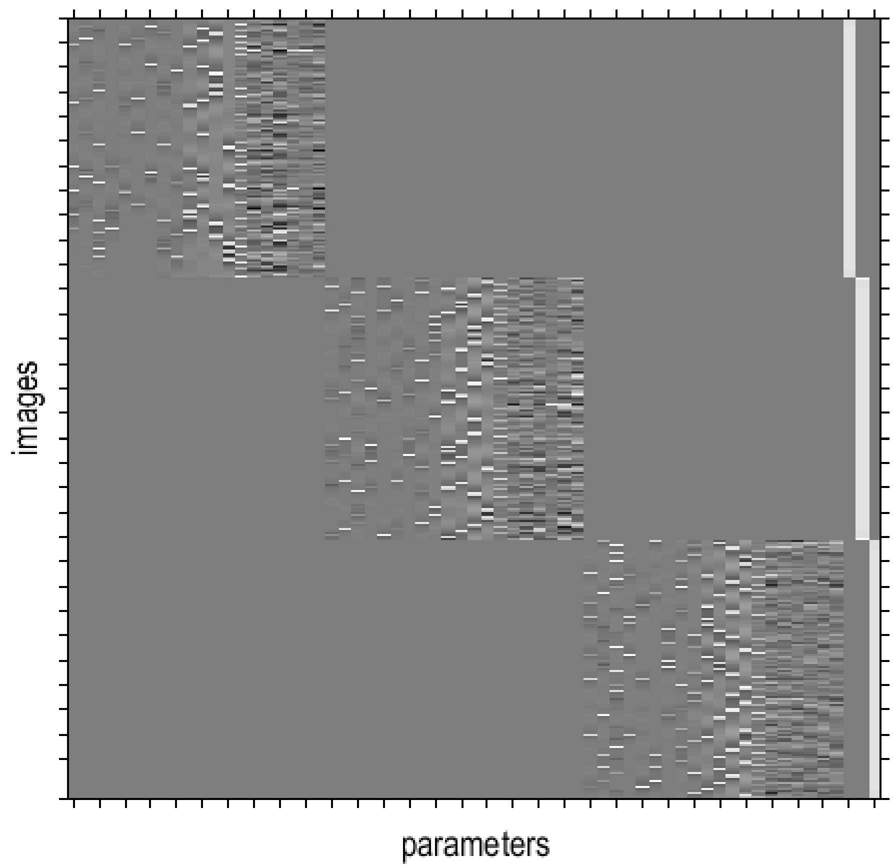


Fig 2.3. First level GLM design matrix.

2.2.9.2 Second level GLM models

T-contrast maps, either of mean activity for a single condition, or between conditions, were brought into a so-called ‘Second level’ analysis, meaning analysis of average group-level effects. Peak statistics for Family-Wise Error-corrected (FWE) activation clusters were identified for task effects ($p<0.05$). The t-maps were either corrected on a whole brain level or, when investigating particular hypotheses, corrected on a region of interest level (ROI). To get an insight into how programming knowledge is represented in the brain of a programming novice, we investigated the hierarchy of processing hypothesis in terms of simulation load: Simulation vs. Syntax, Spatial Simulation vs. Simulation, Iterative Simulation vs. Spatial Simulation. We did this both for the preparation phase, and the task phase.

2.2.9.3. Region of interest analysis (ROI)

In addition to the whole brain GLM analysis, hypothesis driven ROI analyses were conducted. The ROIs were applied to a contrast of all four conditions during task performance (Syntax, Simulation, Spatial Simulation, Iterative Spatial simulation) vs baseline: Hands-on vs. Hands-off. The initial threshold of this image was set to $p<0.05$ uncorrected, and we report findings significant at a $p<0.05$ family-wise error corrected level.

To test the multisensory theory that motor areas support the retention of programming knowledge in the Hands-on group as predicted by the studies of language learning discussed in chapter 1.7 to look for activations in motor cortices during task execution. Since no previous study has investigated non-overt motor activity on programming language learning, and since the participants in the hands-on condition interacted with the computer using both hands, we decided to use a bilateral ROI covering the whole motor cortex including SMA. We used the primary motor areas 4a and 4p from the anatomy toolbox (Eickhoff et al., 2007; Zilles & Amunts, 2010) in SPM right and left hemispheres combined (size of ROI: 2212 voxels), as well as the supplementary motor area (SMA) from AAL v3 (2020) toolbox, L and R (size of ROI: 3856 voxels). Given the size of our motor ROI we also tested the premotor coordinate (-24 –19 55) reported in (Mayer et al., 2015) to support language learning with both 4mm, and 8mm ROIs.

To test the hypothesis that attention towards an object increases in the hands-on condition, we created a ROI from the Neurosynth database (https://www.neurosynth.org/), a web-based platform for performing automated meta-analysis of published fMRI results. By using the search term ‘attention’ an image of 3131 voxels was created based on 1831 studies, spanning the fronto-parietal cortex.

To test the levels-of-processing theory we selected ROIs covering the areas reported in the fMRI studies of programming discussed in chapter 1.9 for the purpose of testing whether learning hands-on leads to deeper processing, manifested as stronger activations during task execution. We used left area 44 (838 voxels) and 45 (579 voxels) as defined in the anatomy toolbox in SPM to create a ROI for Broca’s area. The temporal language area in the left hemisphere was specified from the Neurosynth database (search term ‘language network’) as this area is not defined in the anatomy toolbox. The resulting image was based on 83 studies and contained 3687 voxels. One additional ROIs was also created: a parietal network ROI to test regions involved in mathematics, defined using the keywords ‘rotation’ (102 studies) and a ‘arithmetic’ (96 studies) comprising 4797 voxels.

2.2.9.4. GLM-behavioral correlations

On the second level, we correlated BOLD activity with confidence ratings, task accuracy and long-term memory scores by extracting the first eigenvalue from a 4mm radius sphere around the peaks significantly activated at the FWE-corrected level in the task contrasts.

2.2.9.5. Finite impulse response modelling (FIR)

It is well known that brain imaging contrasts may be confounded by differences in stimuli content between conditions. A way to minimize the effect of such confounds is to sample the hemodynamic response of a condition during the preparatory phase of the trial, after the cue but before the task stimuli is presented. During this phase, the screen is blank, and BOLD activation is thought to reflects the so called ‘task set’ activity. We investigated if we could confirm any differences between conditions in the task set, by conducting FIR analyses within the masks created by the statistical maps obtained from the task phase GLM. We implemented a FIR analysis to closely capture the hemodynamic response during the transition from the preparatory phase of the trial, after the cue has been shown, going into the task phase, where the code is presented, and task-execution takes place. A FIR analysis models the level of BOLD activity over time, split into any number of so-called ‘time-bins’ instead of modeling the whole BOLD response as predicted by events convolved with a hemodynamic response-function (Penny et al., 2011). Each time-bin becomes an estimated parameter encoding the amplitude of the response in each voxel, at that time from the onset of the modeled trial. By combining the parameter estimates for all the trials, we get the average response curve for that task, and we can then average this over all participants at the second level. We settled on ten time-bins of 2.2 seconds each, the smallest time increment we could sample, corresponding to the length of one TR, the time it takes the scanner to sample the whole brain volume once. Given the longest variable delay length of eight seconds, this would mean that the first four time-bins would cover the preparatory phase and the remaining six would cover the full 13 seconds of the short tasks and 66% of the long tasks. (see figure 2.4)

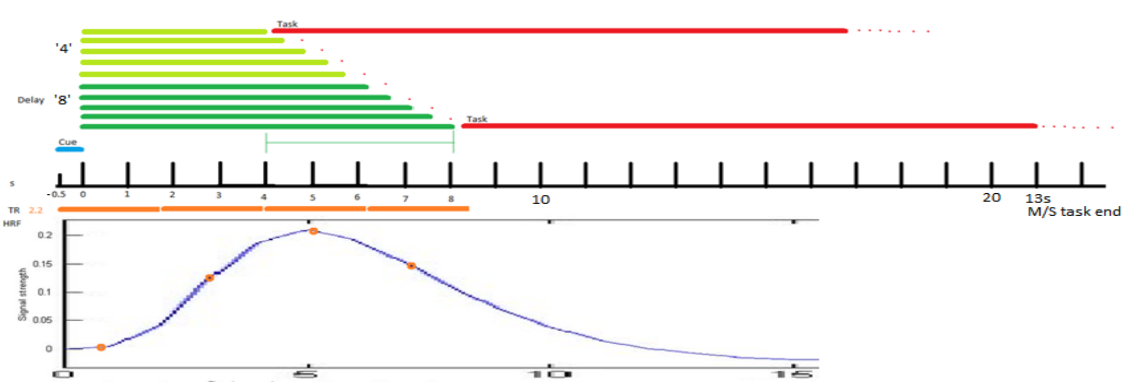


Fig. 2.4. Temporal map of one hypothetical trial where the onset of the Cue coincides with the beginning of a TR, and what parts of the idealized hemodynamic response curve would be sampled during the delay phase in

our FIR analysis. Top: Green bars show the lengths of the 10 different delay lengths and where the task subsequently is presented is shown by the red bars and cover 13 seconds (The length of the two short tasks). Middle: Timeline in seconds. The 4 first TRs of 2.2 seconds each are shown in orange. Bottom: The canonical HRF used in the SPM software package to convolve with the GLM models, scaled to match the timeline.

The ten varying delay lengths between the cue and the task presented a problem for this type of analysis as we need to pool the activity at each time-bin over all participants. The first two time-bins presents no problem as all delay lengths are in the delay phase during the first four seconds. When we get to the third time-bin however, the trials that had a variable delay of only four seconds are now presenting the task, so the activity measured at this point cannot be pooled with that of the trials that are still in delay. This means that for time-bin three, we only pool the eight highest delays, and similarly for time-bin four we only pool the four longest delays. Consequently, we lose accuracy in our measurement of BOLD activity as we approach the onset of the task because we unfortunately did not match the lengths of our trials with this type of modelling in mind when initially designing the fMRI experiment. The last time-bins (five to ten) correspond to the task as stated above but, as a consequence of the varying delay lengths, the activity in the earlier bins for trials with short delays were moved up, so that bin three from the two shortest delay trials are pooled into bin five (the first bin covering the start of the task phase), and so on. The decisions of whether to classify a bin for a trial with a certain delay into task or not (shift it up) when pooling the activity was based on where most of the time-bin fell in real time. If more than half the time-bin fell after trial onset, it was shifted up. Likewise, time-bins nine and ten fell outside of the task for short trials (13s) with shorter delays, so they were not pooled into any of the time-bins.

Pooled Bin:	**Delay**				**Task**					
delay (s)	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
4	1	2	-	-	3	4	5	6	7	8
4.4444	1	2	-	-	3	4	5	6	7	8
4.8888	1	2	3	-	4	5	6	7	8	-
5.3332	1	2	3	-	4	5	6	7	8	-
5.7776	1	2	3	-	4	5	6	7	8	-
6.222	1	2	3	-	4	5	6	7	8	9
6.6664	1	2	3	-	4	5	6	7	8	9
7.1108	1	2	3	4	5	6	7	8	9	10
7.5552	1	2	3	4	5	6	7	8	9	10
8	1	2	3	4	5	6	7	8	9	10

Table 2.1. Detailed breakdown of the shifted time-bins used for pooling the results of all the trials for each delay length in the FIR analysis. For each delay (Y axis) the 10 sampled time points after the cue (X axis) are sorted into Delay or Task, depending on the length of the delay. For example: for a trial with the 8 second delay, bins 1-4 are covering the delay phase, but for trials with the 4 second delay only bins 1 and 2 cover the delay, and the majority of the 3:rd time-bin is now sampling the beginning of the task presentation.

The first level design for the FIR analysis included only the onsets for correctly answered trials, where each set of ten columns (one for each bin in the FIR curve) model one trial from the presentation of the cue. In a second level group analysis, we used activity in primary visual cortex (averaged over all four tasks) as a diagnostic to ensure that the response curves we got from the shifted table of time-bins were reasonable for other areas as the visual cortex is supposed to show a strong, characteristic response to the onset of visual task-stimuli (Figure 2.13). We also compared this response with the canonical Hemodynamic Response-Function (HRF) used in SPM to convolve the GLM model (Figure 2.4).

2.2.10. Dynamic causal modelling (DCM)

2.2.10.1 Design overview

DCM was implemented to investigate the network dynamics between the regions identified in the standard GLM analysis. This not only allows for identification of relationships in the flow of neural activity between areas in the brain and how it changes with task switching, but also how the connections are influenced by other co-variables of interest. The DCM is based on a simplified GLM model comprising only three columns, plus the six motion parameters. The first column modeled the onsets of all spatial conditions. The second column modeled only the iterative trials, and the third column modeled the button-press events as a regressor of no interest. As each subject run was split into three sessions, the model for every subject was temporally concatenated as if part of one long session. The reason we chose to focus on the two spatial tasks was that combined, they were the hardest tasks for the participants to solve, while also requiring processing in all the brain areas relevant for the ‘Syntax’ and ‘simulation’ tasks as well, in addition to the ones specific for the spatial tasks to successfully solve the task. They can thus be considered to represent the "full" programming task, combining all the necessary elements learned during the teaching session.

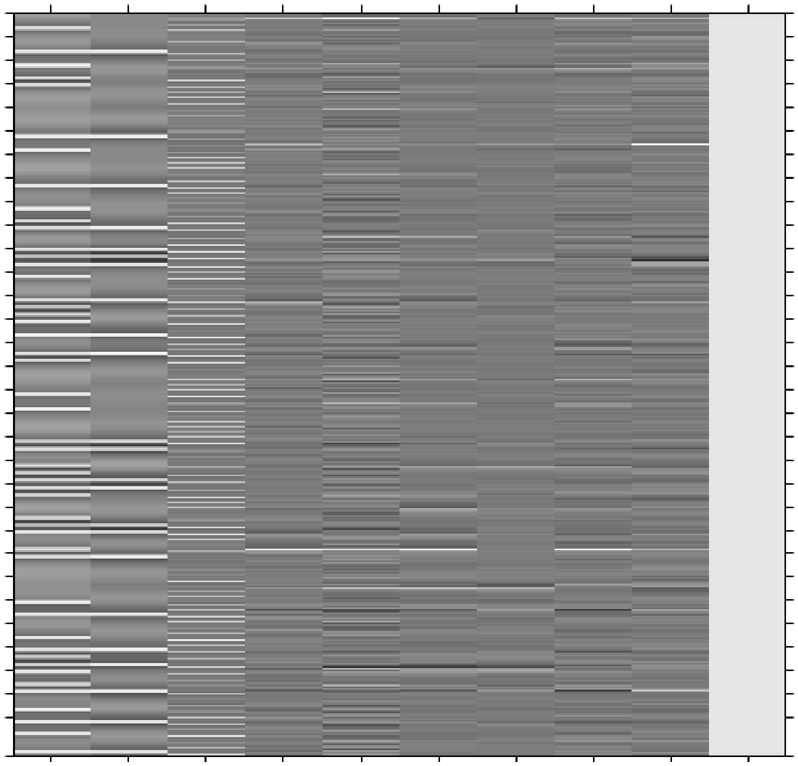


Fig. 2.5. Example simplified GLM design matrix showing the 2 conditions together in column 1, followed by the iterative spatial task alone in column 2, the button presses in column 3 and then the 6 motion realignment parameters followed by the constant term.

2.2.10.2. VOI specification

The data going into the DCM analysis from eight coordinates identified in our original GLM analysis was extracted from the simplified GLM model for each subject using the SPM Volume Of Interest (VOI) function. Each VOI was defined as a sphere of 8mm radius centered on the coordinate for the peak voxel in the thresholded group-level contrast map. A smaller sphere of 4mm radius was then automatically centered on the peak voxel of the simplified GLM within the larger sphere to allow for individual variation between subjects, and then masked to exclude voxels outside the brain using the mask automatically generated during preprocessing. The data in the VOI for each subject is thus the first eigenvariate of the timeseries derived from all voxels within the 4mm sphere (the time series corresponding to this eigenvariate is, in practice, very similar to taking the mean voxel time series in the sphere. It's routinely used instead of the mean, however, because oftentimes voxels in a sphere are not all reporting on the same computational or neurophysiological process – the eigenvariate effectively downweighs atypical voxels).

The eight VOIs were all nodes in the left hemisphere only, since one of our primary hypotheses regarded language processing and, in an effort, to reduce the size of the DCM, given that we wanted to include all eight nodes in a single model. The nodes selected were defined using the following coordinates:

Area	**VOI Coordinate**
Broca's area	-56 10 10
Medial Temporal Gyrus (MTG)	-52 -50 4
Middle Occipital Gyrus (MOG)	-32 -88 14
Superior Parietal Lobule (SPL)	-24 -70 44
Supramarginal Gyrus (SMG)	-56 -30 36
Superior Frontal Gyrus (SFG)	-22 2 58
Inferior Frontal Gyrus pars Opercularis (IFG-op)	-40 6 26
Inferior Frontal Gyrus pars Triangularis (IFG-tri.)	-42 32 16

Table 2.2. VOI coordinates for the DCM. (x, y, z, mm, MNI space)

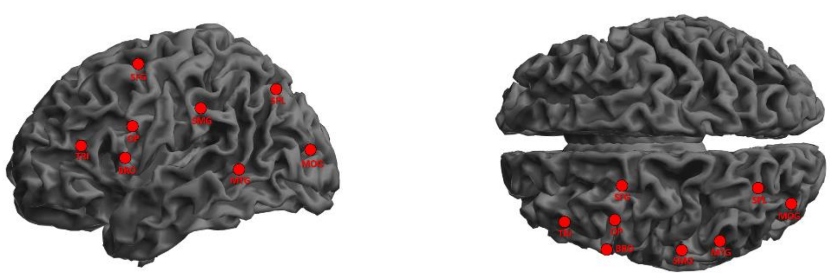


Fig. 2.6. Position of our 8 VOI nodes for the DCM in the left hemisphere. Coordinates are taken from the whole brain GLM analysis (p<0.05 FWE).

2.2.10.3. DCM analysis steps

The analysis steps for the DCM proceeded as follows: Twenty-eight different DCM:s (described in more detail in the next section) based on the same VOI data, were estimated for each subject, producing parameter weights encoding the strength of rate of change-influence exerted by the nodes on the other connected nodes, as defined in that particular model configuration (Zeidman, Jafarian, Corbin, et al., 2019). These individual DCMs were then combined into a single group-level analysis termed a 'GCM' (Group Causal Model) where the winning model is decided based on the relative (complexity-penalized) fit of the different models to the fMRI data in the VOIs in terms of log-evidence relative to the worst model. This is called 'Bayesian Model Selection' (BMS). The winning model was then analyzed at the group level using the 'Parametric Empirical Bayes'(PEB) approach in SPM. This provides an estimate of the group connectivity, and effects of covariates on those parameters, while still taking the variance of the individual parameter estimates from the first level DCMs into account (Zeidman, Jafarian, Seghier, et al., 2019). The complete PEB model was then further reduced after estimation by utilizing an automatic search over nested PEB models to switch off statistically unimportant connections (Penny et al., 2011). Eight covariates were added to our PEB model based on their relevance as determined by the ANOVA and Structural Equation Models of the behavioral data (detailed in the Results section). These were: Group (Hands-on/off), score on the long-term memory test, stress ratings from after the learning session, average change in reported confidence for correctly answered trials between the two conditions relevant for the DCM analysis (Confidence IS-S), average grades, and final grade in mathematics from high-school education (age 16), average scores on the tasks during fMRI scanning, and scores on the Need for Cognition scale.

The final step after identifying interesting parameters in the model is to perform ‘Leave-One-Out’ (LOO) cross-validation of the effects of the covariates included in the PEB model. The procedure here is to predict covariates from DCM parameters, using a leave-one-(subject)-out cross-validation method to avoid overfitting bias. This produces an unbiased estimated of how well the covariates can be predicted from DCM parameters (or combinations thereof). Four of our variables could be significantly predicted from our model: Group, Confidence loss, Average Grade, and LTM (at p=.058 only). After having identified these variables as significant, we estimated a new PEB model containing only these four, giving us a final model with our most robust parameter estimates. To further home in on our most solid findings we chose to only report parameters that have an estimated probability of being non-zero equal to or greater than 99%.

2.2.10.1 Rationale for DCM model construction

The gross underlying architecture of our DCM:s was based on the concept of Dorsal- and Ventral Stream processing, where the task-relevant information begins in the visual cortex and spreads forward towards the frontal cortices via either a dorsal pathway, concerned primarily with spatial relationships, or via a ventral pathway, concerned with object identification and processing of semantic information. We specified our driving input to the model (the strength of which is captured by “C-parameters”) as entering into all of the three most caudal regions (SPL, MOG and MTG) because we did not explicitly model primary visual cortex in an effort to reduce the number of nodes and thereby the complexity of our model. Our first hypothesis to be implemented into the structure of our DCMs was to test models where the dorsal and ventral streams connect via the most frontal nodes in our models or not, producing one set of closed loop models and one set of "open" models. The second consideration was the structure internal to the two pathways. We decided to group our nodes into a "fronto-parietal" network (SFG, SMG and SPL) and a "language" network (IFG-op., -IFG-tri. And Broca), where the connections formed triangles. These were then connected to each other via SFG to IFG in the front, and SPL to MOG to MTG in the back of the brain. Models with more interconnectivity between the two networks were also tested, including a "full" model (where all nodes connect to all other nodes for a total of 131 parameters to estimate) for a completely data-driven baseline for model evidence. All enabled connections in the models were bidirectional. The B-matrix (encoding which connections may change as a function of task) was evaluated by running versions of the basic models with and without B-parameters for the connections between the two networks and within them, and for the connections between MTG and the language network since the MTG node is one of the areas identified in previous fMRI experiments of programming as relevant to task performance. The results of the Bayesian Model Comparisons will be presented with the other results from the DCM in the next section, and for a complete account of the structures of the twenty-eight models please see appendix IV (Stephan et al., 2010).

2.3. Results

2.3.1. Preamble

We will first present statistics for all our variables by group (Hands-on/Hands-off). We then present relationships between the variables of behavioral data that we have collected: correlations, linear regression analyses and Structural equation modelling (SEM). Analyses of the fMRI data will then follow, beginning with standard GLM analyses, then the FIR analysis and finally, DCM modelling. We will present uncorrected t-tests for all our variables as a convenient way of representing the differences between the hands-on and hands-off groups on any one variable that may be of interest. We will also present our SEM to visualize our best model of how the variables relate causally to programming knowledge, but to ensure statistical robustness we will rely on our linear regression models and fMRI results for all our major conclusions. Our measured variables can broadly be divided into three categories for convenience:

1. Variables measuring our primary outcomes of the experiment: Result on the written test, performance in the fMRI scanner, the confidence in the fMRI performance, and Long-term memory tests.

2. Secondary variables derived from, or linked to, our primary variables: Stress and motivation after lesson, differences in stress and motivation beginning to end, score for question categories in

our three main tests, and the difference between the first test and the long-term memory test one week later.

3. Variables conceptualized as relatively static over our experimental timescale: NFC, Growth mindset, self-esteem, motivation and stress before the lesson, grades in math, languages, and final GPA.

Table 2.3 outlines the number of rows of code containing task-relevant information displayed in the four different fMRI conditions and table 2.4 lists t-tests for all our variables. Recall that the syntax- and simulation tasks contain identical code, so the difference in relevant lines here account for the addition of the output lines that the subject is tasked with comparing their simulated output to. The number of lines of text in the two spatial tasks were designed to be as similar as possible in length to preclude any observed differences in activity from reflecting reading effort as far as feasible. The number of lines between the two conditions was not significantly different as measured by a two-tailed t-test ($t(45)=-1.472$, $p=0.148$).

# code lines relevant to task solution		fMRI performance (%)	fMRI confidence (1-5)
Task	**mean ± SD**	**mean ± SD**	**mean ± SD**
Syntax	3.63±1.25	77.4 ±14.5	N/A
Simulation	4.97±1.52	80.1 ±11.7	N/A
Spatial simulation	8.33±1.56	81.7 ±16.1	3.1 ± 0.7
Iterative spatial simulation	9.20±2.82	67.2 ±12.7	2.6 ± 0.6

Table 2.3. Comparison of average performance and confidence in each of the four task conditions and the average number of lines of relevant code on screen for each of the tasks.

2.3.2. Primary outcome variables

There was a significant difference in confidence for the Hands-on (M= 3.05, SD= 0.46) and Hands-off (M= 2.76, SD= 0.63) conditions; t(52)= -1.894, p = 0.032 (one-tailed) (See table 2.4). There were no significant differences in any of the other primary outcome variables. These results suggest that our experimental learning condition does influence the confidence in participant’s answers in our fMRI test. Specifically, our results suggest that when learning in the Hands-on condition, participants were more confident and there is a weak trend towards better performance in the long-term memory test as well.

2.3.3. Secondary outcome variables

There was a significant difference in the following four variables: Confidence on the spatial fMRI test for the Hands-on ($M= 3.35$, $SD= 0.43$) and Hands-off ($M= 3.02$, $SD= 0.71$) conditions; $t(52)= -1.894$, $p = 0.032$ (one-tailed). Confidence on the iterative spatial fMRI test for the Hands-on ($M= 2.76$, $SD= 0.55$) and Hands-off ($M= 2.51$, $SD= 0.60$) conditions; $t(52)= -2.029$, $p = 0.024$ (one-tailed). Stress after learning for the Hands-on ($M= 2.28$, $SD= 2.19$) and Hands-off ($M= 3.36$, $SD= 2.09$) conditions; $t(51)= 1.830$, $p = 0.037$ (one-tailed). Simulation score on the LTM test for the Hands-on ($M= 8.73$, $SD= 1.32$) and Hands-off ($M= 7.98$, $SD= 1.75$) conditions; $t(52)= -1.746$, $p = 0.043$ (one-tailed) (See table 2.4). These results again suggest that our experimental learning condition does influence the confidence in participant’s answers. Specifically, our results suggest that when learning in the Hands-on condition, participants were more confident, less stressed and there is also a weak trend towards better performance in the long-term memory test here as well.

2.3.4. Covariate variables

No covariates reach significant differences between groups, but one of them show a weak trend: self-esteem for the Hands-on ($M= 22.20$, $SD= 5.60$) and Hands-off ($M= 19.59$, $SD= 5.69$) conditions; $t(52)= -1.696$, $p = 0.096$ (two-tailed) (See table 2.4). These results suggest that our learning outcomes are not influenced by the self-esteem of our participants.

	Hands-on		Hands-off				
Primary outcome Variables	**Mean**	**SD**	**mean**	**SD**	**t**	**p**	**dir.**
Confidence	3.05	0.46	2.76	0.63	-1.894	0.032*	1-t
LTM	37.7	7.45	34.09	8.57	-1.638	0.053~	1-t
Written test score	35.61	9.14	33.2	10.77	-0.878	0.192	1-t
fMRI score	93.88	14.92	90.76	13.51	-0.807	0.212	1-t
Secondary Variables							
Confidence on the spatial fMRI	3.35	0.43	3.02	0.71	-1.894	0.032*	1-t
Confidence on the iterative spatial fMRI	2.76	0.55	2.51	0.6	-2.029	0.024*	1-t
Stress after learning	2.28	2.19	3.36	2.09	1.83	0.037*	1-t
Simulation score on the LTM test	8.73	1.32	7.98	1.75	-1.746	0.043*	1-t
Confidence iterative minus spatial fMRI	-0.59	0.3	-0.51	0.4	-1.616	0.056~	1-t
Written answers score on the written test	19.3	3.3	17.16	5.9	-1.608	0.057~	1-t
Syntax score on the LTM test	6.33	0.94	5.87	1.31	-1.475	0.073	1-t
Iterative score on the LTM test	22.64	5.91	20.24	6.19	-1.449	0.076	1-t
Syntax score on the written test	6.74	1.59	6.11	2.27	-1.169	0.124	1-t
Iterative spatial score on the fMRI	20.8	4.05	19.76	3.61	-0.999	0.161	1-t
Syntax score on the fMRI test	23.92	4.68	22.83	4.15	-0.909	0.183	1-t
Iteration score on the written test	20.66	6.37	19.07	7.13	-0.858	0.197	1-t
LTM score minus written test score	-2.09	4.84	-0.89	5.49	0.847	0.200	1-t
Simulation score on the fMRI	24.44	3.95	23.62	3.26	-0.836	0.203	1-t
Stress before minus after learning	0.68	1.28	0.17	1.93	-1.119	0.268	2-t
Simulation score on the written test	8.21	1.97	8.03	2.02	-0.338	0.368	1-t
Multiple choice score on the written test	16.31	6.29	16.04	5.3	-0.169	0.433	1-t
Spatial score on the fMRI	24.72	4.7	24.55	5.01	-0.127	0.450	1-t
Motivation after learning	7.8	1.5	7.43	2.1	-0.733	0.467	2-t

Table 2.4…

Covariate Variables							
self-esteem	22.2	5.6	19.59	5.69	-1.696	0.096	2-t
Number of languages studied	3.36	0.64	3.17	0.38	-1.33	0.189	2-t
Art grade	18.2	1.7	17.5	2.75	-1.102	0.275	2-t
Sum of language grades	60.1	14.28	56.72	10.94	-0.982	0.33	2-t
Stress before learning	2.96	2.42	3.48	1.94	0.88	0.383	2-t
IMI Value/Usefulness	18.87	4.43	20.07	5.64	0.832	0.409	2-t
IMI Effort/Importance	9.04	4.93	7.89	6.62	-0.691	0.493	2-t
Motivation before learning	7.16	1.99	6.83	1.67	-0.667	0.508	2-t
IMI Pressure/Tension	-5.83	4.81	-5.25	5.42	0.397	0.693	2-t
Math grade	18.1	2.73	18.36	2.43	0.373	0.711	2-t
Average grades	18.07	1.5	17.99	1.8	-0.158	0.875	2-t
Growth-mindset	1.84	4.44	1.66	4.2	0.157	0.876	2-t
IMI Perceived Competence	22.61	5.54	22.32	7.32	-0.155	0.877	2-t
NFC	19.02	16.96	18.69	16.17	-0.073	0.942	2-t
IMI Interest/Enjoyment	22.48	5.49	22.36	8.31	-0.06	0.952	2-t
Average of language grades	17.83	2.02	17.8	1.95	-0.049	0.961	2-t

Table 2.4. t-tests, Hands-on vs. Hands-off. (=p<.5), (~=p<.6).*

2.3.5. Correlations

Here we present the most interesting correlations between our variables (table 2.5). (For a complete correlation table, see appendix III) Correlations of note: LTM correlates with performance scores, math grades and NFC, stress, and confidence (p=0.06). Confidence correlates additionally with motivation and stress (neg), NFC, math grades, and day 1 written test scores (p=0.06).

	LTM	Confidence	Written test	fMRI score	Motivation before	Motivation decrease	Stress after	Engagement	NFC	Entity score	Grade avg.	Grade math	Grade art	Grade Sum Lang.
LTM	1	0.22 (p=0.06)	0.83***	0.49***	0.20 (p=0.08)	-0.11	-0.42**	0.16	0.52***	0.1	0.28*	0.44**	0.13	0.18 (p=0.1)
Confidence		1	0.21 (p=0.06)	0.15	0.31*	0.17	-0.46***	-0.1	0.41**	-0.19 (p=0.09)	-0.02	0.23*	-0.179 (p=0.1)	0.094
Written test			1	0.48***	0.30*	-0.13	-0.33**	0.15	0.46***	0.15	0.27*	0.49***	0.08	0.26*
fMRI score				1	-0.1	-0.14	-0.28*	0.24	0.24*	-0.1	0.33**	0.43**	0.04	0.30*
Motivation before					1	0.47***	-0.32*	-0.02	0.49***	0.04	0.04	0.23*	0.09	0.19 (p=0.08)
Motivation decrease						1	0.08	-0.06	0.20 (p=0.08)	0.02	0.23*	0.15	0.31*	0.24*
Stress after							1	-0.1	-0.40**	-0.02	0.18	-0.14	0.23*	0.09
Engagement								1	0.18	-0.11	-0.08	0.11	-0.11	-0.09
NFC									1	-0.03	0.07	0.37**	-0.02	0.08
Entity score										1	0.03	-0.03	-0.01	-0.09
Grade avg.											1	0.64***	0.47***	0.58***
Grade math												1	0.14	0.47***
Grade art													1	0.32**
Grade Sum Lang.														1

Table 2.5. Partial correlations table of selected variables

2.3.6. Structural equation modeling (SEM)

Here we present an exploratory structural equation model, primarily to visualize the relationships between our most influential variables. This will form the basis for what variables to focus on in our analyses going forward. Maximum Likelihood (ML) estimation was used to compute the model. Number of distinct sample moments: 105, Number of distinct parameters to be estimated: 40, Degrees of freedom (105 – 40): 65, Chi-square = 113.864, Degrees of freedom = 65, Probability level = .000.

The model, totaling 25 variables; 14 observed and 11 unobserved (8 of which are error terms), allows us to assume that confidence plays a significant role in programming learning success. Confidence in our model is reflected in the experienced motivation coming into the class, the stress at the end of the class, and the confidence scores in the fMRI scanner. We note that Hands-on has a slight positive influence on confidence ($p=.08$). In addition, Need for Cognition scores reflect learning success, and we note that the math grade has a slight positive impact on programming knowledge with a p-value of .08. Programming knowledge in our model is reflected in accuracy on tasks in the fMRI scanner and the written test, LTM, and less drop in motivation after the class. For the estimated coefficients and approximated significance values see figure 2.7-8 and table 2.6.

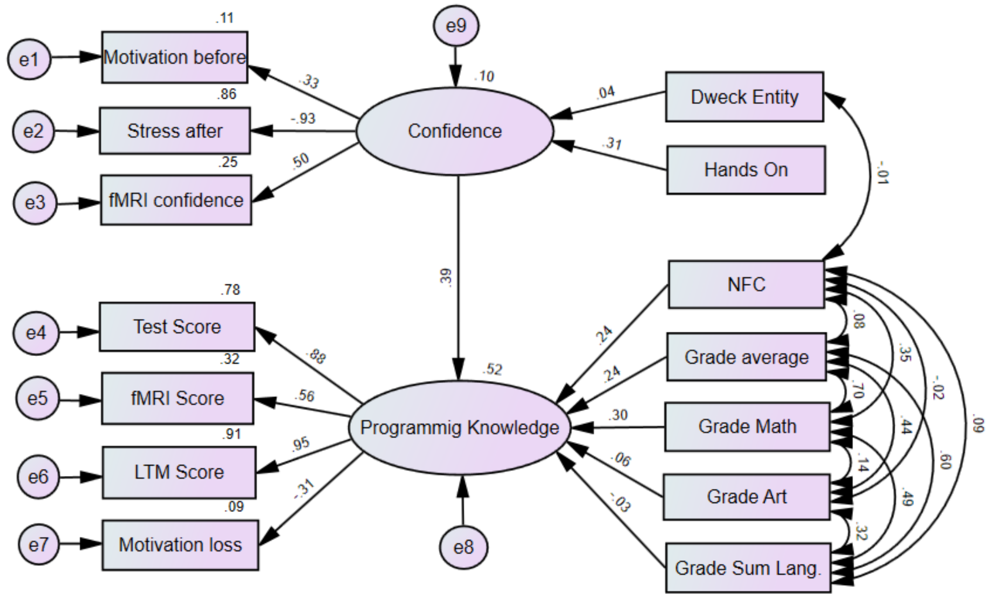


Fig. 2.7. SEM results (standardized). Straight arrows=standardized regression coefficients (change in SD / SD change of predictor). Curved arrows=correlations. Numbers above variables=R^2.

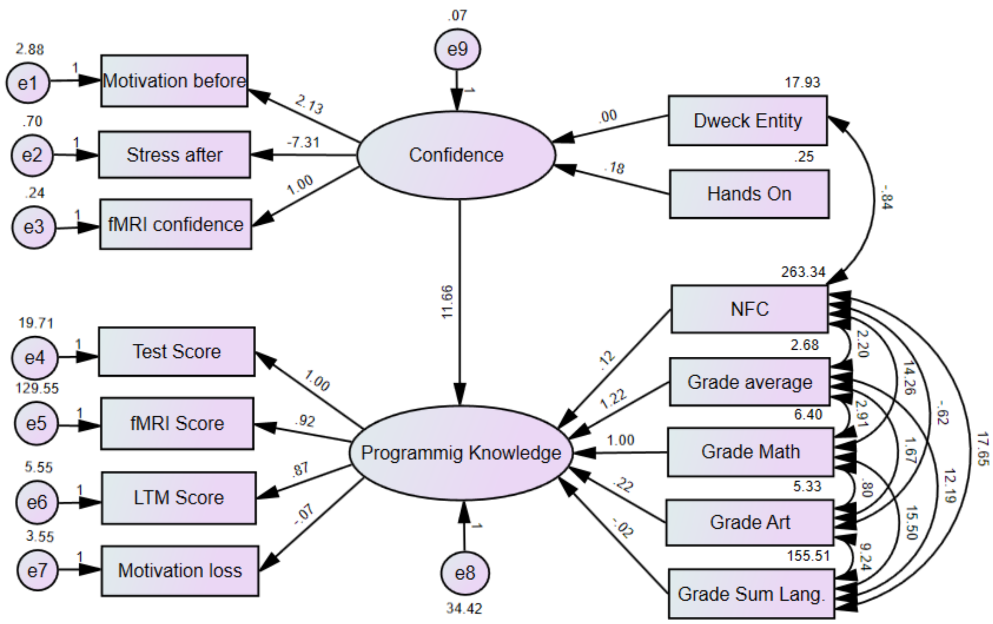


Fig. 2.8. SEM results (non-standardized). Straight arrows=unstandardized regression weights (change / single unit change of predictor). Curved arrows=covariance. Numbers above variables=variance.

			Estimate	S.E.	C.R.	P	Label
Confidence	<---	hands_on	.177	.101	1.752	.080	
Confidence	<---	Dweck_Entity	.003	.009	.277	.782	
Programmig_knowledge	<---	Confidence	11.659	4.559	2.558	.011	
Programmig_knowledge	<---	Grade_avg	1.222	.939	1.301	.193	
Programmig_knowledge	<---	NFC	.124	.060	2.057	.040	
Programmig_knowledge	<---	Grade_Sum_Lang	-.022	.089	-.250	.803	
Programmig_knowledge	<---	Grade_math	1.005	.567	1.773	.076	
Programmig_knowledge	<---	Grade_art	.218	.445	.490	.624	
Stress_a	<---	Confidence	-7.306	2.814	-2.596	.009	
fMRI_confidence	<---	Confidence	1.000				
Motivation_b	<---	Confidence	2.130	1.046	2.037	.042	
Test_sum	<---	Programmig_knowledge	1.000				
fMRI_score	<---	Programmig_knowledge	.917	.205	4.478	***	
LTM_sum	<---	Programmig_knowledge	.875	.095	9.257	***	
Motivation_b_a	<---	Programmig_knowledge	-.072	.032	-2.228	.026	

Table 2.6. SEM results table. Shows the estimated regression coefficients and the associated p-values for the model.

2.3.7 ANOVA, Hands-on vs. Hands-off

We test what the effect of learning condition (learning hands-on vs. hands-off) is on our 15 variables from our exploratory SEM using analysis of variance as an alternative to means testing. Our selected breakdown between the two conditions including the means and standard deviations of the two groups is presented in table 2.7. (For a table of group statistics, see Appendix II. For a full ANOVA table containing all our variables, see Appendix VII) Learning computer programming hands-on leads to greater confidence in one's task performance (p=0.032), lower stress after class (p=0.036), and better long-term memory scores (p=0.053). Note that all these effects were quite weak, so any conservative correction (like for instance Bonferroni) would discount these potentially interesting findings.

To ensure that our random group assignment did not bias our results going forward into our fMRI analyses, we tested engagement scores, NFC, Mindset, and grade averages. There were no differences between the two groups in engagement during the learning session ($p=0.27$), nor in scores on trait questionnaires (NFC: $p=0.47$, Mindset: $p=0.88$) or grades ($p=0.88$).

Variable	***Hands on n=25 (mean±SD)***	***Hands off n=29 (mean±SD)***	***One-way ANOVA***
Task confidence during fMRI (1-4)	3.05 ± 0.46	2.76 ± 0.63	*p=0.032 1-t
Written test score	89.0% ± 22.8%	83.0% ± 27.0%	p=0.19 1-t
fMRI task score	78.3% ± 12.4%	75.7% ± 11.3%	p=0.21 1-t
Long-term memory score	94.3% ± 18.5%	85.3% ± 21.5%	*p=0.053 1-t
Motivation before learning (1-10)	7.2 ± 2.0	6.8 ± 1.7	p=0.50 2-t
Stress before learning (1-10)	2.9 ± 2.4	3.4 ± 1.9	p=0.38 2-t
Stress after learning (1-10)	2.28 ± 2.189	3.35 ± 2.094	*p=0.036 1-t
Motivation loss (Before – After learning)	-0.64 ± 1.8	-0.34 ± 2.2	p=0.30 1-t
Engagement scores (-1,0,1)	-0.08 ± 0.57	0.10 ± 0.61	p=0.27 2-t
Need for cognition (±72)	19.0 ± 17.0	18.7 ± 16.2	p=0.47 2-t
Average grades (max 20)	18.1 ± 1.5	18.0 ± 1.8	p=0.88 2-t
Mathematics grade (max 20)	18.1 ± 2.7	18.4 ± 2.4	p=0.71 2-t
Grade sum languages	60.1 ± 14.2	56.7 ± 10.9	p=0.33 2-t
Grade Art (max 20)	18.2 ± 1.7	17.5 ± 2.8	p=0.28 2-t
Growth-mindset (±10)	1.8 ± 4.4	1.7 ± 4.2	p=0.88 2-t

Table 2.7. Selected one-way ANOVA results. Means and significance for each response variable as influenced by the learning condition factor (Hands On / Hands Off).

2.3.8. Regression analysis

Here we again selected our most interesting 15 variables, as determined by the SEM, and entered them into several regression models attempting to explain long-term memory performance, since this is our ultimate outcome variable. The variables can be broadly categorized as the main experimentally manipulated variable; hands-on, 'trait' variables; sex, NFC, mindset, motivation before, stress before, grade average, math grade, art grade and language grade, and variables shaped during the experiment; motivation change, stress change, fMRI confidence, fMRI confidence change and Engagement. To investigate which aspects that most contribute to successful LTM performance, we ran a stepwise backward regression analysis which resulted in 10 models (Adjusted R_2=.504, F(6, 47)=9.814, p<.000). The state variables best predicting LTM retention are confidence in doing the more complex Iterative Spatial task as compared to the Spatial task, stress after class, stress reduction after the class, and motivation reduction after the class. The 'trait' variables best predicting LTM are average grades and NFC. Variables that did not contribute to model fit, and thus were systematically excluded were motivation before class, average task confidence ratings, hands-on, sex, Dweck mindset score, Engagement, math-, language- and art grades. Test scores day 1, and performance accuracy during fMRI scanning were not included in the model as these are highly correlated with LTM scores. Recall that the LTM test was designed to be almost identical to the written test at the end of the teaching session but administered one week later, and that performance on the LTM test and the written test are highly correlated (R=.857, p<.000).

Model		Beta	t	Sig.	95% CI Lower	95% CI Upper
10						
	NFC	.410	3.683	.001	.083	.282
	Motivation reduction after lesson	-.248	-2.391	.021	-1.936	-.166
	Stress after lesson	-.381	-3.192	.003	-2.088	-.473
	Grade average	.315	2.962	.005	.483	2.532
	fMRI confidence reduction IS vs I	-.214	-2.095	.042	-8.711	-.174
	Stress reduction after lesson	-.245	-2.286	.027	-2.012	-.128

Table 2.8. Backwards regression analysis predicting LTM. An automated backward removal method was used (criterion: Probability of F-to-remove >= .100) resulting in 10 models. Adjusted R2=.504, F(6, 47)=9.814, p<.000. Variables excluded: Stress decrease, Grade math, fMRI Confidence, Motivation before lesson, Grade Sum of Languages, Female sex, Grade art, Hands On, Dweck Mindset, Engagement.

A similar but slightly different result is obtained when using a forward regression analysis. The best forward model out of 6 produced from the same selection of variables was also highly significant (Adjusted R^2=.504, $F(5, 48)=9.814$, $p<.000$) and is presented here for comparison, together with an additional alternative regression comprising only the variables that were not self-estimated by the participants.

Model		Beta	t	Sig.	95% CI Lower	95% CI Upper
6						
	NFC	.410	3.683	.001	.083	.282
	fMRI confidence reduction IS vs I	-.214	-2.095	.042	-8.711	-.174
	Stress after lesson	-.381	-3.192	.003	-2.088	-.473
	Grade average	.315	2.962	.005	.483	2.532
	Motivation reduction after lesson	-.248	-2.391	.021	-1.936	-.166
	Stress reduction after lesson	-.245	-2.286	.027	-2.012	-.128

Table 2.9. Forward Regression analysis predicting LTM. An automated forward removal method was used (criterion: Probability of F-to-enter <=.050). resulting in 5 models. Adjusted R2=.504, F(5, 48)=9.814, p<.000. Variables not entered: Hands On, Female sex, Dweck Entity score, Motivation before lesson, Grade Math, Grade art, Grade Sum of Languages, fMRI Confidence, Engagement.

Model		Beta	t	Sig.	95% CI Lower	95% CI Upper
	Hands On	.255	2.043	.023	.004	.507
	Female sex	-.202	-1.686	.049	-.444	.039
	Grade average	.054	.271	.394	-.348	.456
	Grade math	.549	2.775	.004	.150	.947
	Grade art	.073	.526	.301	-.207	.353
	Grade Sum of Languages	-.113	-.771	.222	-.409	.183
	fMRI Confidence	.013	.099	.461	-.251	.277
	fMRI confidence reduction IS vs I	-.041	-.299	.383	-.310	.230

Table 2.10. Linear regression analysis predicting LTM (no self-estimated variables). R=.647,R^2=.419 (.316 adjusted),SE=6.78295 F(8)=4.06, p=.001.

2.3.9. GLM & ROI analysis

When the students simulated the output of the code as compared to simply looking for syntax errors, left Broca's area and medial temporal gyrus [-52 -50 4] were significantly more active (figure 2.11). Note that the only difference in stimuli between the conditions was the added output line in Simulation trials. Notably, the temporal area was also more active during the task-set phase, suggesting that its activation is involved in the computation of task rules, rather than simply reflecting specific visual stimuli encountered, such as the number, or type of, words read (figure 2.10). As illustrated in table 2.11 and in the beta values of figure 2.9, the amplitude of the activation in Broca's area was not further increased in the two spatial conditions when compared to the simulation task, suggesting similar processing across all three conditions and casting doubt on the levels-of-processing-theory.

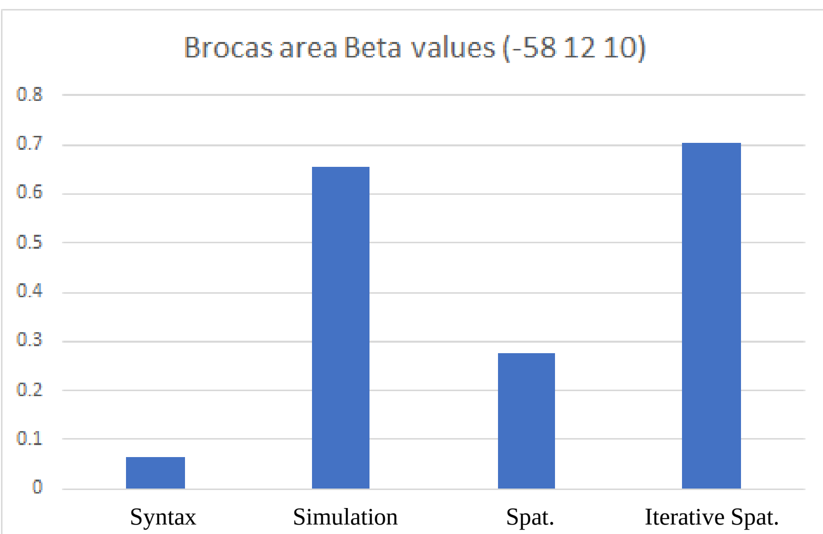


Fig. 2.9. Averaged beta-values for each task condition extracted from Broca's area [-58 12 10].

The right Cuneus and the occipital lobe were more active when the students simulated the output of the code as compared to simply looking for syntax errors. Whereas the right Cuneus is only active during task performance and not task set, the occipital lobe activation is also involved in simulating the upcoming task during the task-set phase, which again suggests that it is not simply involved in processing the visual stimuli input, but also relevant in more abstract computations (table 2.11, figure 2.10-12). For the participants to manage spatial objects, and logical structures such as the while-loops, additional activation in occipital, posterior parietal and superior frontal cortices are required. However, studying the task-set activity, we note that the frontal activations are missing, which suggests that these are less relevant for task rule setting, and more relevant for specific task execution processes such as keeping items in working memory and maintaining attention. It is more posterior activations that are seen during task-set; left superior and middle occipital gyrus, including calcarine cortex, cuneus, and precuneus, left and right supramarginal gyrus (including area 40), and left precentral gyrus (table 2.11, Figure 2.12) Figures 2.14-17 further illustrate the task-set activation, based on FIR modelling. To further corroborate task relevant activations, we note that BOLD amplitude of the left Lingual gyrus, right Cuneus, left and right Supramarginal gyrus, left precuneus, left Occipital pole, Left inferior frontal gyrus, left and right superior Parietal lobule, left middle frontal gyrus, left superior frontal gyrus, left supplementary motor cortex, right precentral gyrus, left and right middle Occipital gyrus, and the left inferior occipital gyrus correlated significantly with performance scores (fMRI task/LTM task). Significant correlations with confidence were also observed for approximately half of these areas. (see table 2.11)

Task Contrast	**Area**	**Peak**	**T**	**P (FWE)**	**Sig. Task-set activity**	**Cor. With Score / LTM**	**Cor. With conf.**
Simulation vs. Syntax	Left Broca's Area	-56 10 10	4.55	0.040	ns	ns	ns
	Left Middle Temporal Gyrus	-52 -50 4	5.40	0.003	svc *	ns	ns
	Left Lingual Gyrus	-2 -78 0	8.55	0.000	wb ***	SCORE *	ns
	Left Superior Temporal Gyrus	-50 -46 12	5.00	0.011	ns	ns	ns
	Left Precuneus	-4 -52 52	4.70	0.026	ns	ns	ns
	Right Cuneus	14 -98 12	10.82	0.000	wb ***	SCORE *	ns
	Right Occipital Fusiform Gyrus	28 -76 -14	8.49	0.000	svc **	ns	ns
Spatial Simulation vs. Simulation	Left Superior frontal Gyrus	-22 2 58	6.73	0.000	ns	ns	*

Table 2.11…

	Right Superior frontal Gyrus	26 12 54	7.47	0.000	ns	ns	ns
	Left Supramarginal Gyrus	-56 -30 36	6.92	0.000	ns	LTM **	ns
	Right Supramarginal Gyrus	60 -28 48	8.21	0.000	ns	LTM *	ns
	Left Precuneus	-10 -52 62	6.20	0.000	ns	LTM *	ns
	Left Lingual Gyrus	-14 -70 -6	17.47	0.000	wb ***	ns	ns
	Right Lingual Gyrus	10 -72 -4	10.97	0.000	svc ***	ns	ns
	Left Occipital Pole	-14 -96 22	10.03	0.000	wb ***	SCORE *	ns
	Left Middle Occipital Gyrus	-44 -76 18	5.14	0.007	svc ***	ns	ns
	Left Superior Parietal Lobule	-40 -40 54	4.76	0.023	ns	ns	ns
	Left Superior Parietal Lobule	-36 -44 62	4.64	0.032	ns	ns	ns

Table 2.11…

	Right cerebellum exterior	34 -42 -40	4.98	0.012	ns	ns	ns
	Right Inferior frontal Gyrus (p.op)	54 10 16	4.65	0.032	ns	ns	ns
Iterative Spatial Simulation vs. Spatial simulation	Left Inferior Frontal Gyrus (p.op)	-40 6 26	8.20	0.000	ns	ns	ns
	Left Inferior Frontal Gyrus (p.tri)	-42 32 16	7.70	0.000	ns	LTM *	*
	Left Superior Parietal Lobule	-24 -70 44	12.30	0.000	ns	SCORE *** & LTM **	*
	Right Superior Parietal Lobule	20 -66 52	10.70	0.000	ns	SCORE *** & LTM *	*
	Right Superior Parietal Lobule	28 -70 42	10.60	0.000	ns	SCORE *** & LTM **	*
	Left Middle Frontal Gyrus	-26 18 50	5.15	0.009	ns	SCORE *** & LTM *	ns
	Right Middle Frontal Gyrus	26 6 50	5.63	0.002	ns	SCORE *** & LTM **	*

Table 2.11…

	Right Middle Frontal Gyrus	52 36 26	6.96	0.000	ns	SCORE ***	ns
	Right Middle Frontal Gyrus	40 22 28	5.96	0.001	ns	ns	ns
	Left Superior frontal Gyrus	-22 6 52	4.89	0.019	ns	SCORE *** & LTM **	*
	Left Supplementary Motor Cortex	-4 22 48	6.40	0.000	ns	LTM *	*
	Right Precentral Gyrus	42 6 32	6.10	0.000	ns	SCORE ***	ns
	Left Middle Occipital Gyrus	-32 -88 14	9.20	0.000	svc ***	SCORE *** & LTM **	*
	Right Middle Occipital Gyrus	34 -80 10	9.99	0.000	ns	SCORE ***	ns
	Left Inferior Occipital Gyrus	-42 -68 -4	10.18	0.000	wb ***	SCORE * & LTM *	ns
	Left Cerebellum exterior	-8 -74 -32	6.86	0.000	ns	LTM *	**
	Left Cerebellum exterior	-24 -62 -32	5.04	0.012	ns	LTM **	**
	Right Cerebellum exterior	8 -76 -32	5.50	0.003	ns	ns	**
	Cerebellar Vermal Lobules VII-X	0 -52 -38	5.51	0.003	ns	LTM ***	*

Table 2.11. GLM results table, incorporating which activations also show significant Task-set activity and/or correlation with Score/LTM and

*Confidence. Ns=non-significant, wb=whole brain corrected statistics, svc=small volume corrected statistics. *=p<.05,**=p<.01,***=p<.001.*

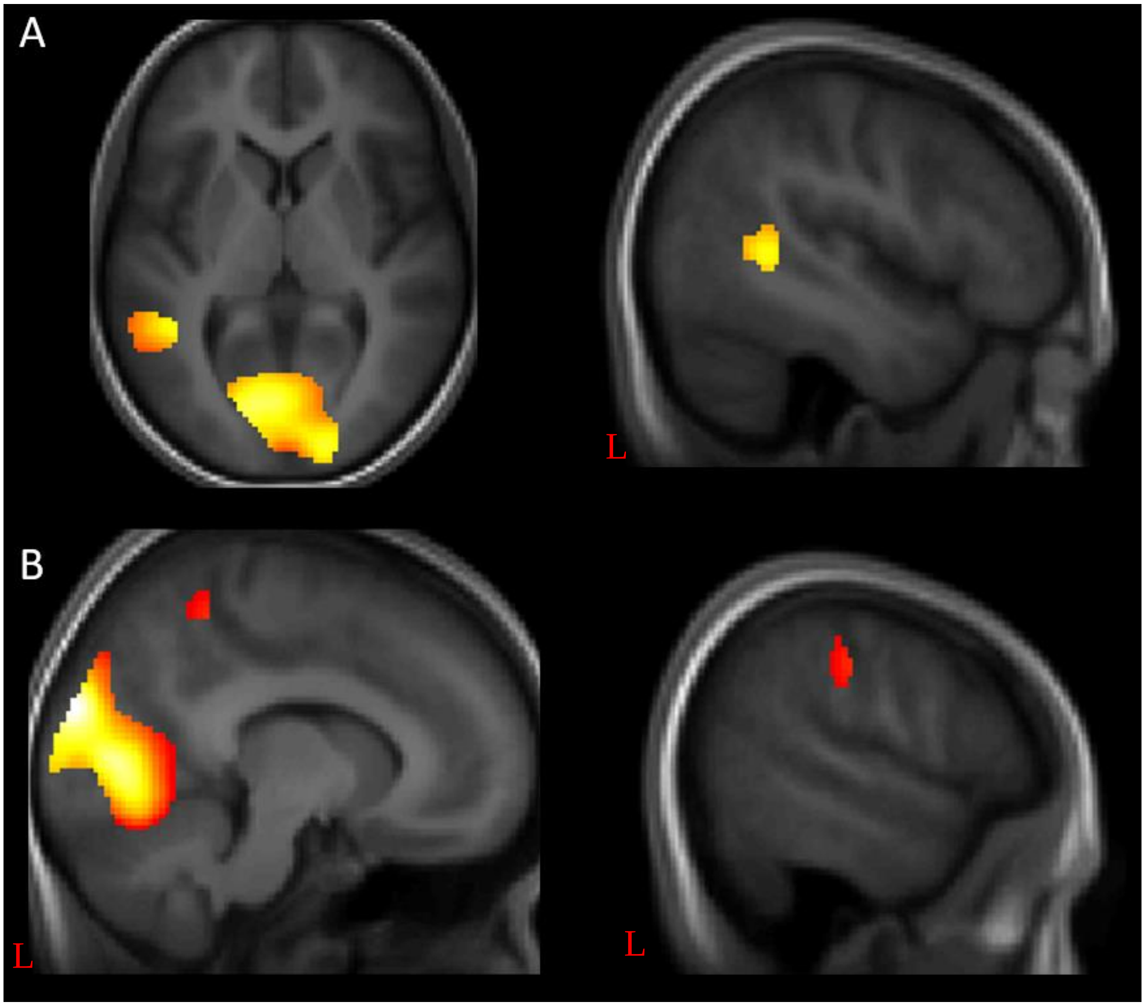


Fig. 2.10. Significant Task set activity during the delay phase (before task presentation). Display threshold: p=0.05 whole brain uncorrected, masked by the corresponding p=0.05 FWE-corrected Task activations. A: Semantic vs. Syntax contrast. B: Spatial vs. Semantic contrast.

In this first figure of task set activity we can clearly see patterns of activation matching the corresponding activity when subsequently solving the corresponding task (cf. 2.10.A vs. 2.11.A, and 2.10.B vs. 2.11.B). These areas in temporal-, parietal- and inferior frontal cortices also agree with the previous findings from the fMRI experiments investigating programming discussed in chapter 1.

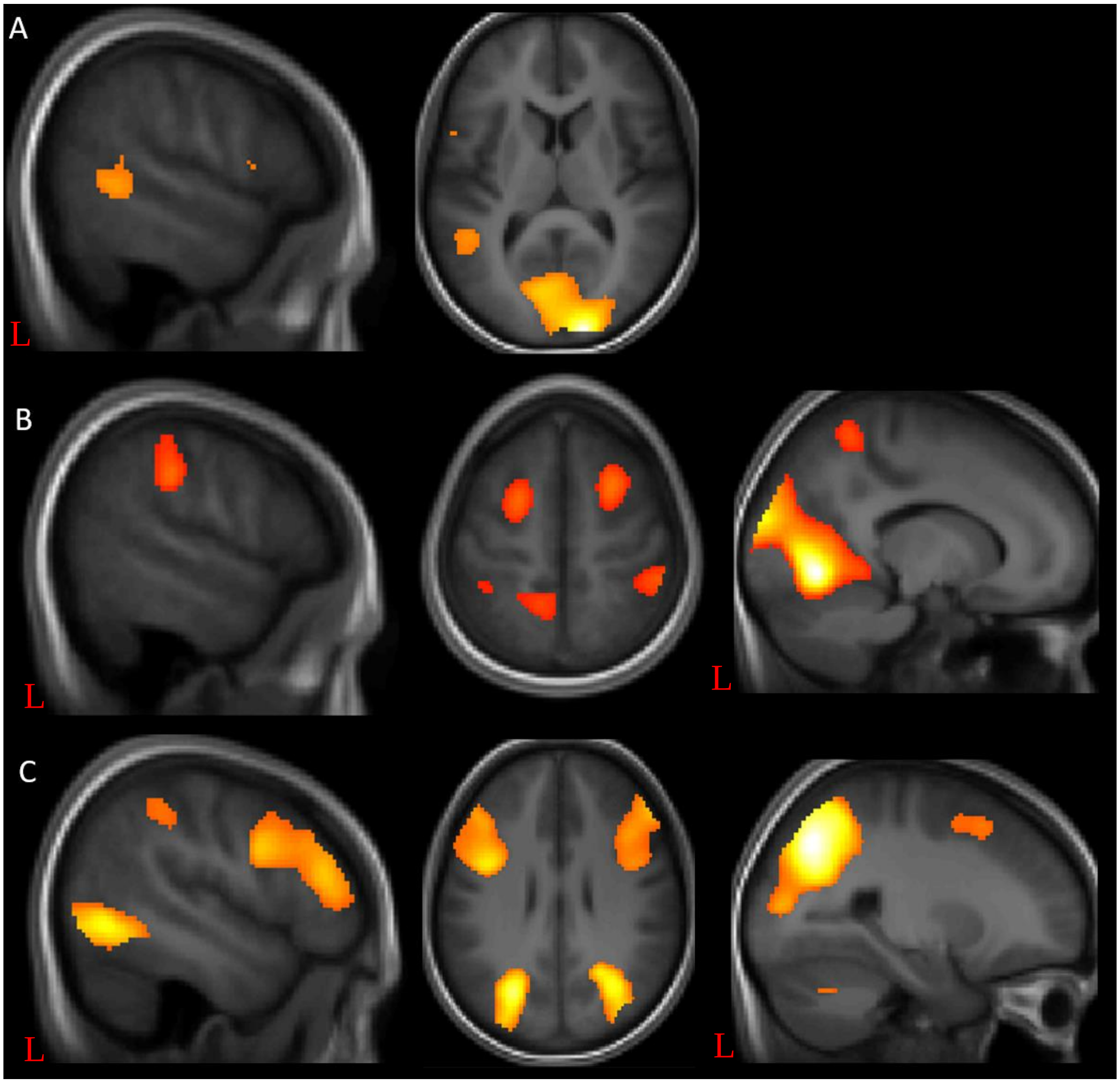


Fig. 2.11. Significant Task activity. Display threshold: p=0.05 FWE-corrected. A: Semantic vs. Syntax contrast. B: Spatial vs. Semantic contrast. C: Iterative Spatial vs. Spatial contrast.

To further test our hypothesis of hierarchical, task-dependent processing we superimposed the activation clusters from our GLM analyses in order to illustrate the spatial relationships between the various regions significantly activated during each of our task contrasts. Note that since these are contrasts between our conditions, it means that the new areas of activation are added to the previous, not that they take their place when solving the more complex tasks.

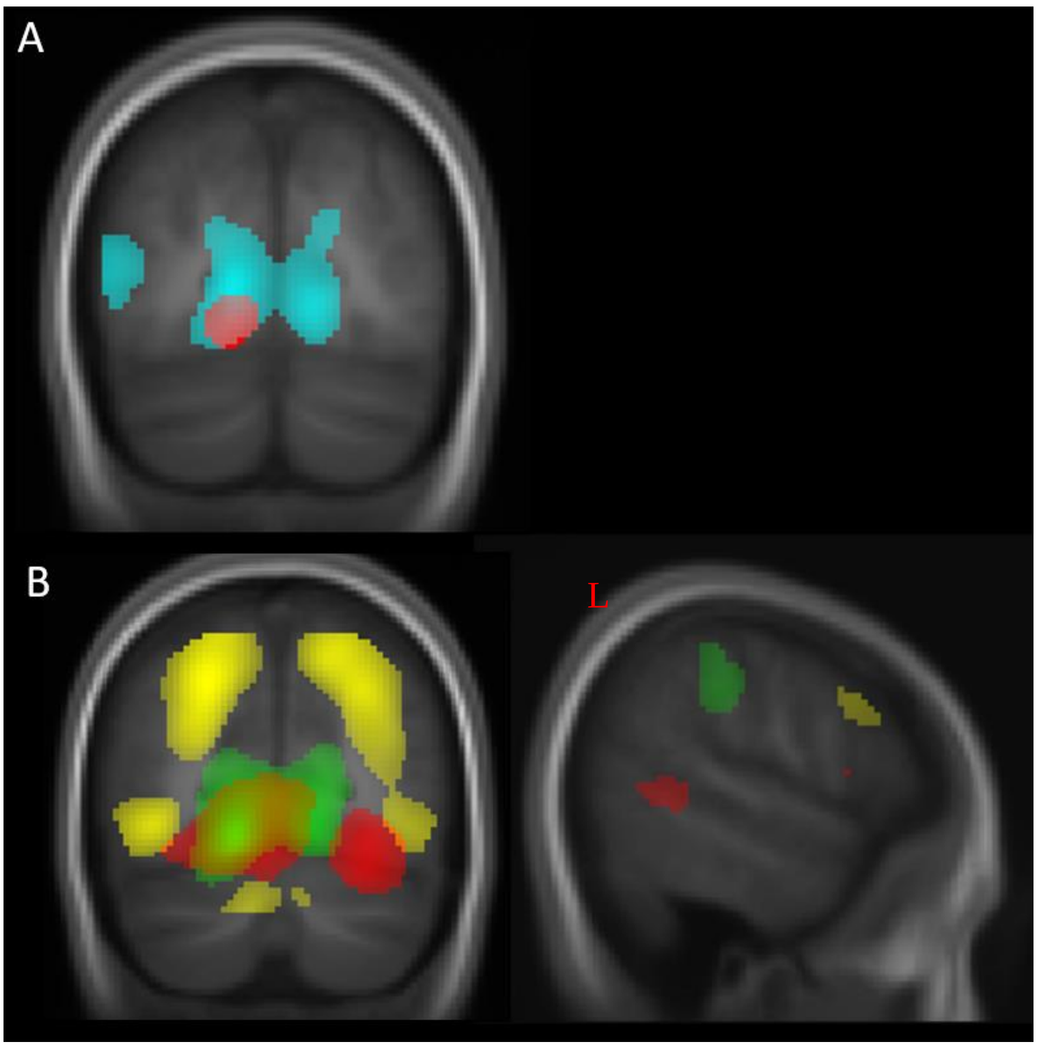


Fig. 2.12. Superposition of Significant Task set and Task. A: Task set activations. Red: Semantic vs. Syntax contrast. Cyan: Spatial vs. Semantic

contrast. Display threshold: p=0.05 whole brain uncorrected, masked by the corresponding p=0.05 FWE-corrected Task activations. B: Task activations. Red: Semantic vs. Syntax contrast. Green: Spatial vs. Semantic contrast. Yellow: Iterative Spatial vs. Spatial contrast. Display threshold: p=0.05 FWE-corrected.

All our contrasts looking for Hands-on vs. Hands-off differences in our initial GLM analysis reported no significant results that survived whole-brain statistics. Our ROI analysis that attempted to control for this also failed to produce any significant clusters that differed between the two learning conditions. Our ROI approach attempting to test the multisensory theory that motor areas support the retention of knowledge after Hands-on learning, gave no significant results in primary motor areas 4a and 4p ($p>0.05$ uncorrected). The premotor coordinate (-24 –19 55) from (Mayer et al., 2015) showed $t(52)=1.8$, $p=0.09$ (corrected 4mm sphere ROI), $t(52)=2.3$, $p=0.076$ (corrected sphere 8mm ROI). Testing the hypothesis that attention towards an object increases after Hands-on learning, gave no significant results in the attentional network ROI ($p>0.05$ uncorrected). Testing the levels-of-processing theory, showed in area 44: $t(52)=2.7$, $p=0.14$ corrected, area 45: $t(52)=2.5$, $p=0.14$ corrected, and the temporal language area: $p>0.05$ uncorrected (4 and 8mm spheres). We will return to the question regarding any effects of learning hands-on in the DCM section.

2.3.10. FIR analysis

As seen in the table of GLM results above, we matched significant task-set results from the preparatory phase with the significant whole-brain results from our task contrasts. For the areas where we observed significant task-set activation where we also observed that the level of activation for each subject significantly correlated with one of our primary outcome variables (score or LTM) we plotted the results of our FIR model to visualize the shape of the BOLD response across the trials. Recall that the FIR time-bins cover 2.2s each (1 TR), that bins 1-4 cover the Preparatory phase and bins 5-10 cover the Task. Note that the classification of time-bins into preparatory phase and task is based on real time, i.e., what is displayed on the screen at that moment, not based on the delayed BOLD response to the stimuli.

The first FIR curve tested was extracted from a peak coordinate in primary visual cortex as a control to make sure that the pooling of our time-bins produced reasonable data (Fig. 2.13.). The reason for choosing this area is that it is known to produce strong, reliable BOLD responses to visual stimuli and is the basis for the hemodynamic response model used in the SPM statistical package to model fMRI data.

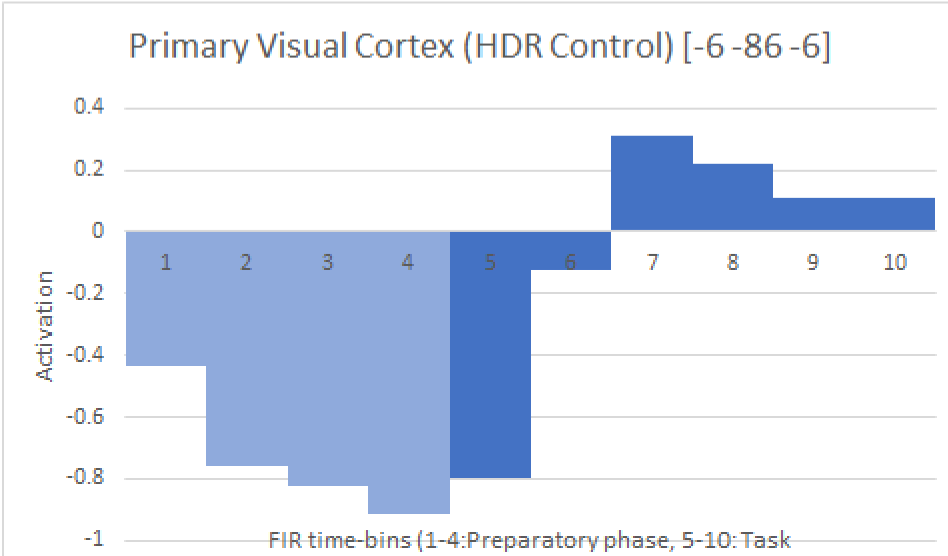


Fig. 2.13. FIR plot of primary visual cortex, used as a control to validate that the pooling of time-bins was done correctly. Shows the expected pattern of activation that is used as a basis for modelling the hemodynamic response in SPM.

We see a strong characteristic rise in activity after the presentation of the task stimuli at the start of time-bin 5, rising to a peak around 6.5 seconds later, as expected. We next extracted FIR curves for significant peak activations in our GLM analysis for all three of the task-contrasts, beginning with the simulation vs. syntax condition (Fig. 2.14.), followed by the spatial simulation vs. simulation only condition (Fig. 2.15.) and finally, the iterative spatial simulation vs. spatial simulation condition (Figs. 2.16-17.)

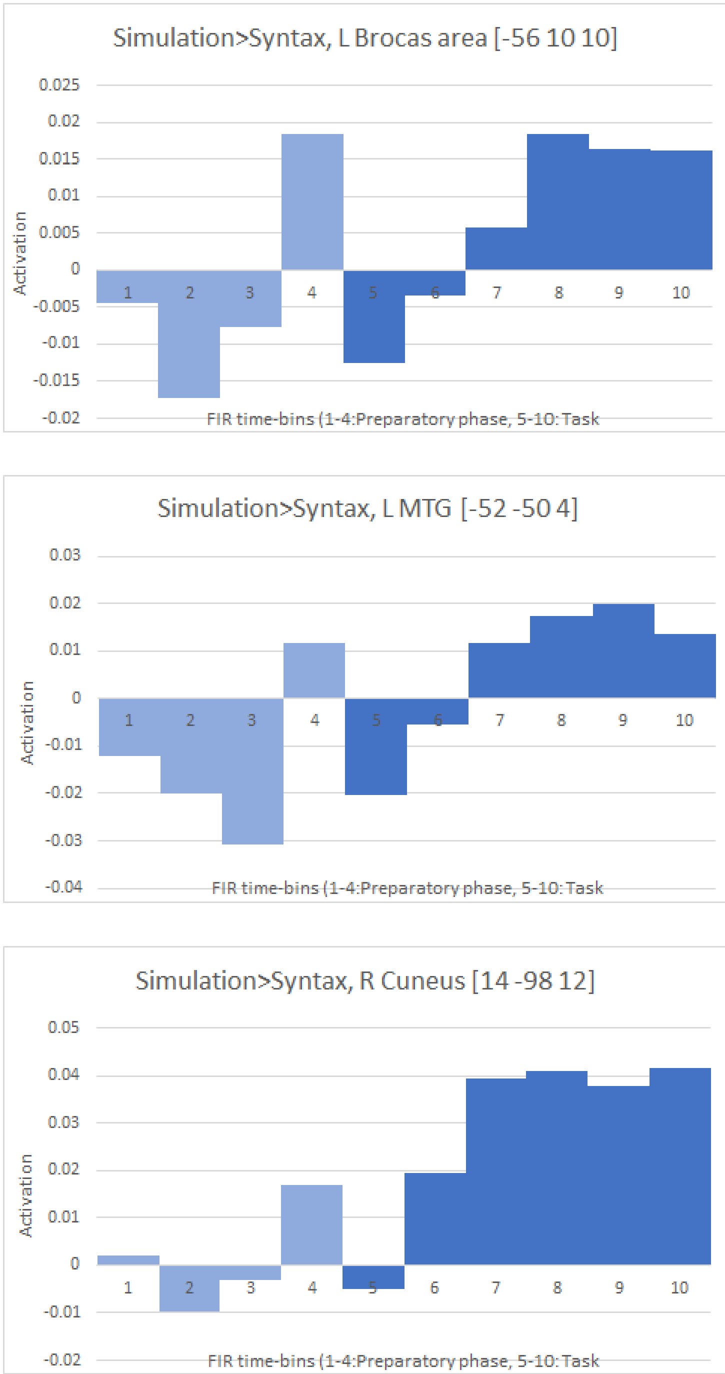
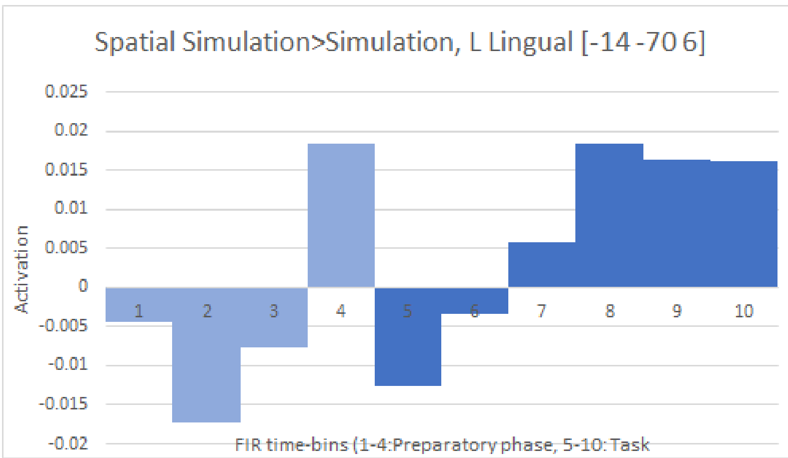


Fig. 2.14. FIR plots of the Semantic vs. Syntax contrast for Top: Left Broca's area [-56 10 10], Middle: Left Middle Temporal Gyrus [-52 -50 4], Bottom: Right Cuneus [14 -98 12].

Unlike in the visual cortex control, we can see an increase of activation peaking right before the presentation of the task stimuli, representing the task-set activity, followed by a drop before climbing anew while solving the actual task.



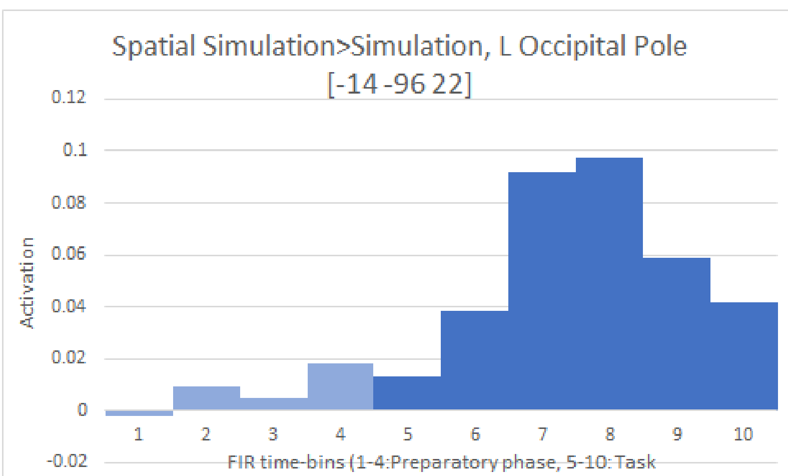


Fig. 2.15. FIR plots of the Spatial Simulation vs. Simulation contrast for Top: Left Lingual gyrus [-14 -70 6], Bottom: Left Occipital pole [-14 -96 22].

As can be seen in this figure of our spatial simulation vs simulation only contrast, some areas show stronger preparatory responses than others, even though they were all significantly activated. Our left occipital pole FIR model, for instance looks closer to our visual control than it does the activation in left lingual cortex, indicating that their processing is probably quite distinct.

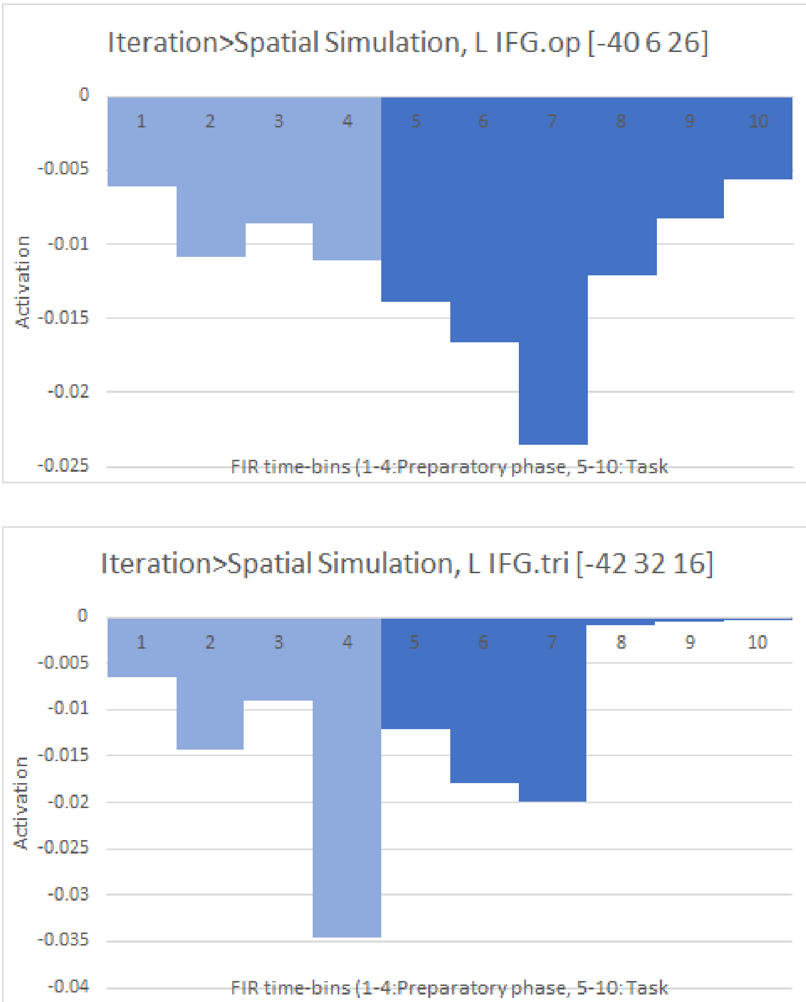
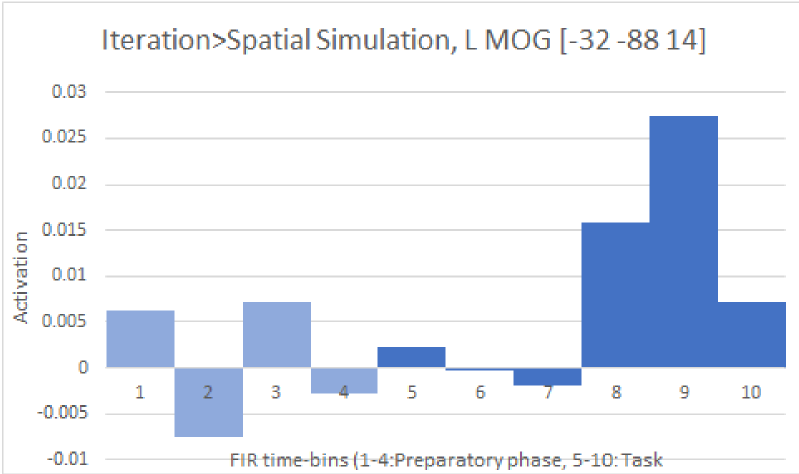


Fig. 2.16. FIR plots of the Iterative Spatial Simulation vs. Spatial Simulation contrast for Top: Left Inferior Frontal gyrus pars Opercularis [-40 6 26], Bottom: Left Inferior Frontal gyrus pars Triangularis [-42 32 16].



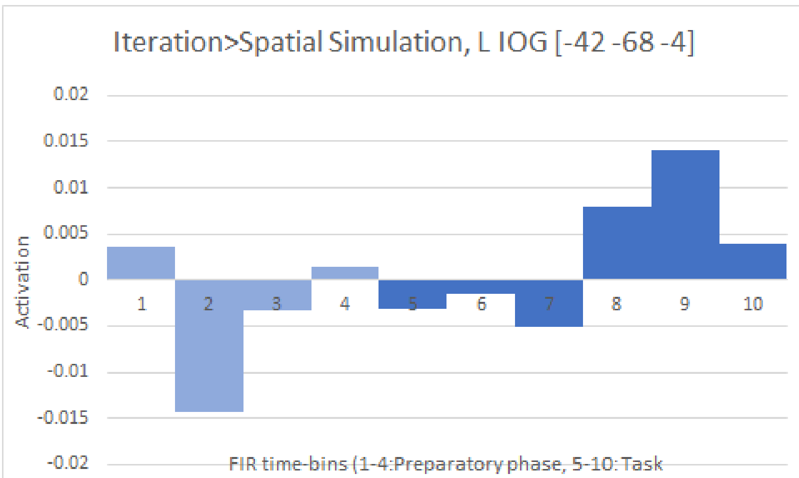


Fig. 2.17. FIR plots of the Iterative Spatial Simulation vs. Spatial Simulation contrast for Top: Left Middle Occipital gyrus [-32 -88 14], Bottom: Left Inferior Occipital gyrus [-42 -68 -4].

These last two figures cover four peaks in the iterative spatial condition. While the FIR plots for left MOG and left IOG look quite similar to the rest of our models, the shapes of the two left IFG responses look quite different. The starkest difference being that they are primarily inhibited during both the preparatory phase and during task processing.

2.3.11. DCM analysis

2.3.11.1 Bayesian model selection (BMS)

Here we present the results of our DCM analysis, beginning with model design, model comparisons and final model selection, followed by our estimated model parameters for our connections and covariates, and finally Leave-One-Out-validation of our covariates. As detailed in the methods section, our DCM is constructed out of the activation patterns in the significant regions identified in the GLM analysis and incorporates testing of our hypotheses both by means of comparing several differently structured DCMs with each-other, and by later incorporating the most promising variables identified in our regression models at a group-level DCM analysis called a PEB model.

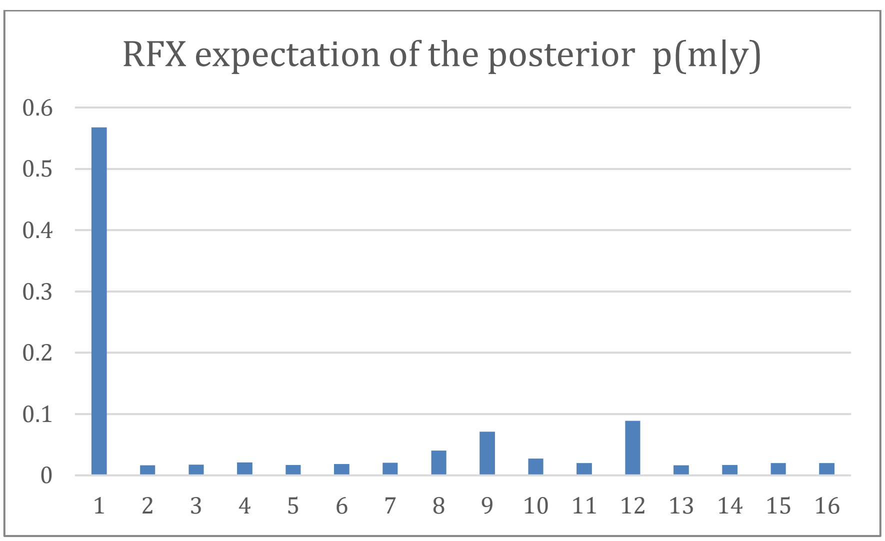


Fig. 2.18. Bayesian Model comparison: First 16 models (Basic structure). Probability of model given the data relative to the other models. Model 1 is the clear winner.

The first step is to run Bayesian Model Selection (BMS) on our list of proposed DCMs, where the BMS is based on the relative probability of each model given our input data. The winning model turned out to be the most complex one (The others can be said to be ‘nested’ models, where some connections in model 1 has been turned off). For the full description of all the models see appendix IV.

The percent of the variance explained by ‘model 1’ for each subject was only 4.53% on average with a standard deviation of 2.87 (min=0, max=9.37) which is low, as at least around 10% variance explained is recommended as a rule of thumb (Zeidman, Jafarian, Corbin, et al., 2019). ‘Model 1’ was updated to include 3 driving inputs (MOG, SPL and MTG), bidirectional extrinsic connections in the matrix of B parameters (connections affected by the task switching) and our input data into the model was optimized by adjusting the sampled voxels to the local maxima within an 8mm radius for each subject instead of sampling the exact coordinate for the fMRI activity at group level. This ‘updated model 1’ with individually optimized input data achieved a variance explained of 8% on average with a standard deviation of 4.13 (min=1.04, max=15.77)

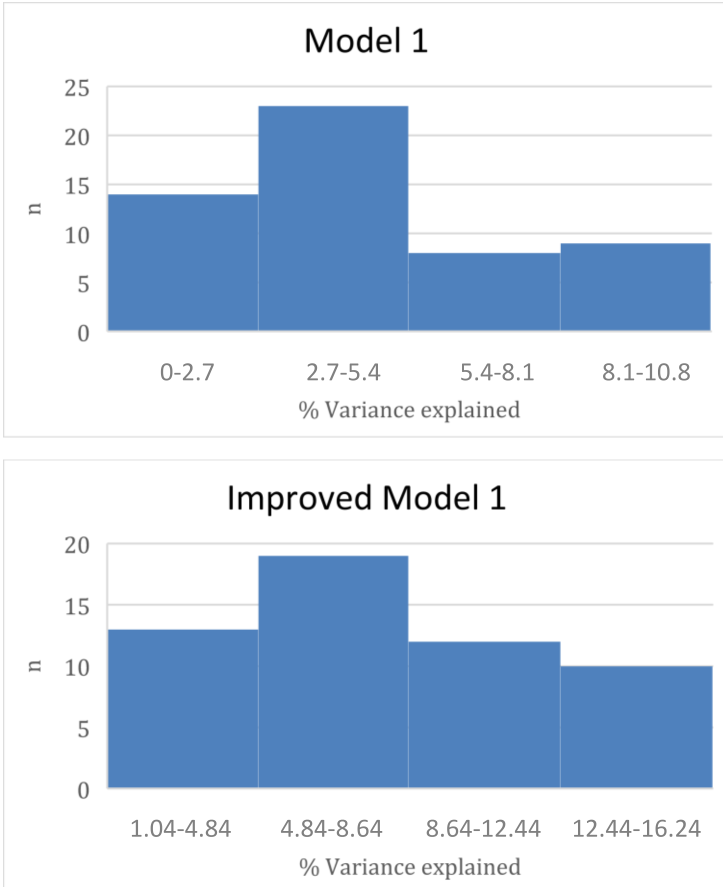


Fig. 2.19. Histograms showing the improvement in the distribution of % Variance explained after adjusting our winning DCM (model 1) and the input data. Top: initial "model 1". Bottom: "Improved model 1".

After having found the winner amongst our first generation of DCMs, as well as improved the model fit further by adopting the more flexible individual sampling of ROI data, we could design a second generation of DCMs to run BMS on, to select our final DCM that would then proceed to a group-level analysis. These final 12 models compared were all based on the 'updated model 1' and were designed to test hypotheses; whether three driving inputs was better than one, whether to model the intrinsic B parameters or not, if a ‘full’ model with all connections turned on (i.e., wholly data driven connectivity map) was better than our best ‘model 1’, and to test models with more direct communication between the language and attention networks. Figure 2.20. shows the result of this final round of BMS where, again, our ‘improved model 1’ is the clear winner. Figure 2.21. illustrates graphically the location of the nodes in our winning DCM and what connects are included in it.

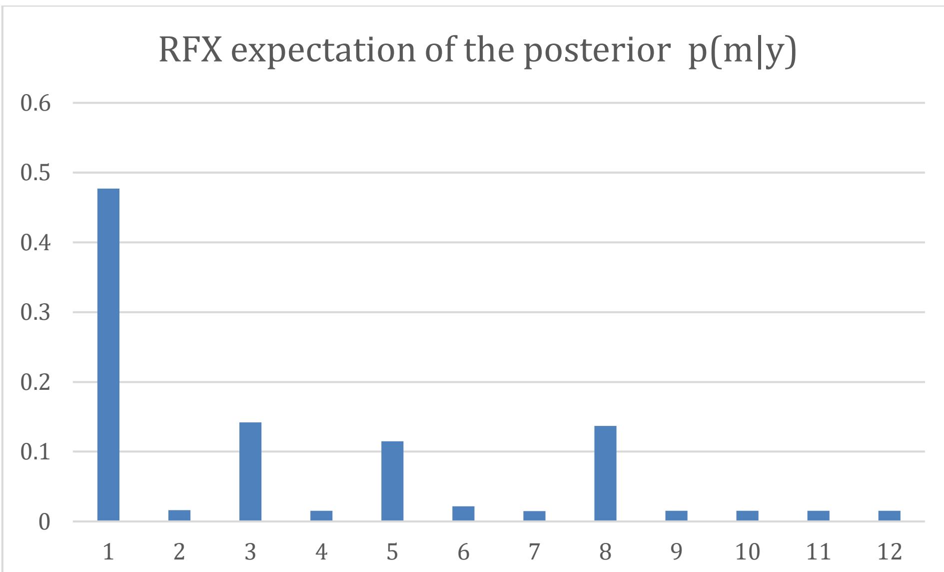


Fig. 2.20. *Bayesian Model comparison: Final 12 models. Probability of model given the data relative to the other models. The Improved Model 1 is the clear winner. Model descriptions: 1=Improved model 1. 2=one input only. 3=three inputs, no intrinsic B parameters. 4=one input only and intrinsic B parameters. 5=’Full’ model, all 128 parameters turned on with 3 inputs. 6=Model with all B parameters matching the A parameters turned on. 7=Top 42 highest probability connections from the ‘full’ model only (number of parameters matched to our model 1 to compare data driven model with same complexity). 8=Top 57 connections from the ‘full’ model*

(All connections significantly different from zero). 9=Model 1 but without the connection between SFG and IFG.op. 10=Model 1 but without connections to the SMG node. 11=Model 1 but with a bidirectional connection between MTG and SPL. 12=Model 1 but with a bidirectional connection between MTG and SFG.

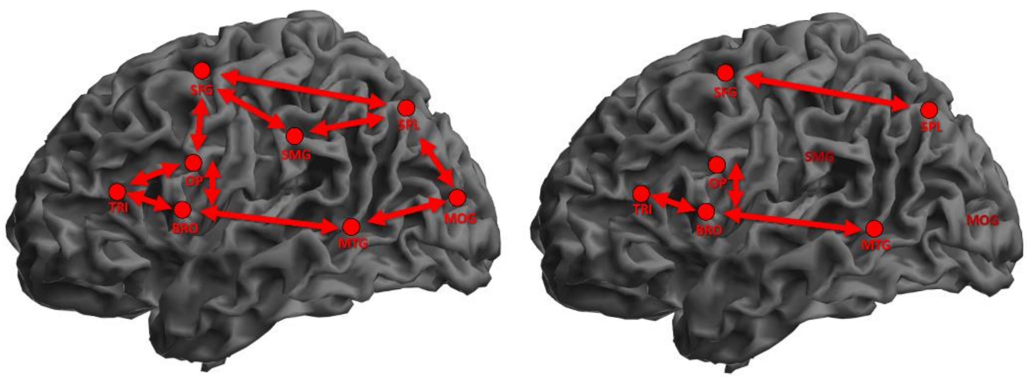


Fig. 2.21. Winning Model connections for the 'Improved Model 1'. Left: 'A parameters' – All connections active during task processing. Right: 'B parameters' – All connections that change connectivity when switching to the spatial tasks. Nodes in the model are all marked with their names. Red dots indicate that the parameter modeling self-inhibition for that node is present in the model. Arrows indicate a connection between two nodes (bi- or unidirectionality indicated by the arrowheads). There are three driving inputs into the model (MOG, SPL & MTG).

The Bayesian model comparison showed that our 'Improved model 1' outperformed the 'full' data-driven model, showing that it did not simply win by being more complex than its nested competitors. It also outperformed the models with some connections removed, demonstrating that we lose model evidence if we exclude the influence of the SMG node and that the communication between the dorsal network and the language network between the SFG and IFG.op node is supported by the model evidence. The models testing for more direct communication between the MTG and the two dorsal nodes SFG and SPL performed worse than our ‘model 1’ and so does not support the idea of more strong direct connections between the dorsal attention network and ventrolateral language areas in the temporal lobe.

2.3.11.2 Parametric empirical Bayes model (PEB)

With our winning DCM selected we proceed to model group-level effects, including the effects of covariates on our DCM using Parametric Empirical Bayes modelling (PEB). Like the BMS step described above, we did two generations of PEB modelling. The first generation was used to identify which covariate effects were strong enough to survive the Leave-One-Out significance testing (LOO), and a second PEB model was subsequently created with only those covariates. The results of our final PEB model includes only the four most significant covariables and this is presented here in figure 2.22, and table 2.12. (For the results of our initial PEB model and LOO validation see appendix V.) Positive parameter values for the covariates mean that the effect of the parameter in the DCM is positively correlated with the covariate. The contribution of a covariate to the DCM parameter is equal to the effect size times the covariate score. Interpretation of the parameters of the PEB model works as follows: for intrinsic (self-inhibitory) connections (e.g., Broca->Broca) a positive parameter means higher inhibition. For extrinsic connections (e.g., Broca->MTG) a positive parameter means excitatory connectivity. For the modulatory (B) parameters a positive parameter means increased connectivity when switching to the iterative spatial task.

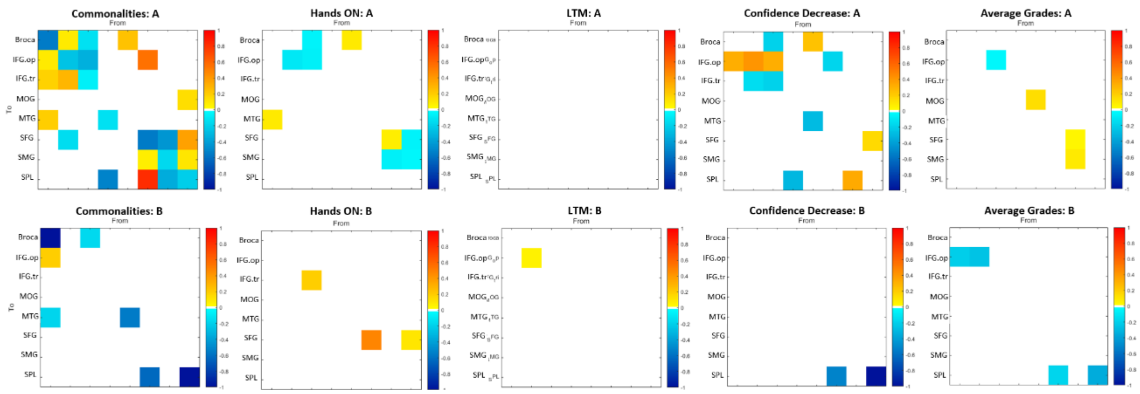


Fig. 2.22. Matrix of significant DCM PEB parameter estimates. Positive values indicate increased effective connectivity as measured in rate of change-influence (in Hz) on the connected node.

The results presented here in figure 2.22. is a visual representation of the strengths and directions of the PEB parameter estimates presented further on in table 2.12. The results are split up into ‘Commonalities’ meaning connections shared in common across all subjects and conditions, and then the covariates (hands-on learning, LTM score, Confidence decrease and average grades), indicating which connections are up or down-regulated in accordance with the covariate in question across subjects. Finally the results are also split between the ‘A-‘ and ‘B-matrices’ of the DCM, where ‘A’ is the common connections active while solving both the spatial tasks and ‘B’ encodes the connectivity changes when switching to the iterative task. The following images presented in figure 2.23. also illustrate the results of the

PEB graphically on the brain for easy comparison with the full model from 2.21., and for identifying global patterns of connection-directionality easier than with just the connectivity matrix in 2.22. Our final interpretations of these results and the upcoming LOO validation will ultimately be discussed in chapter 2.4.5.

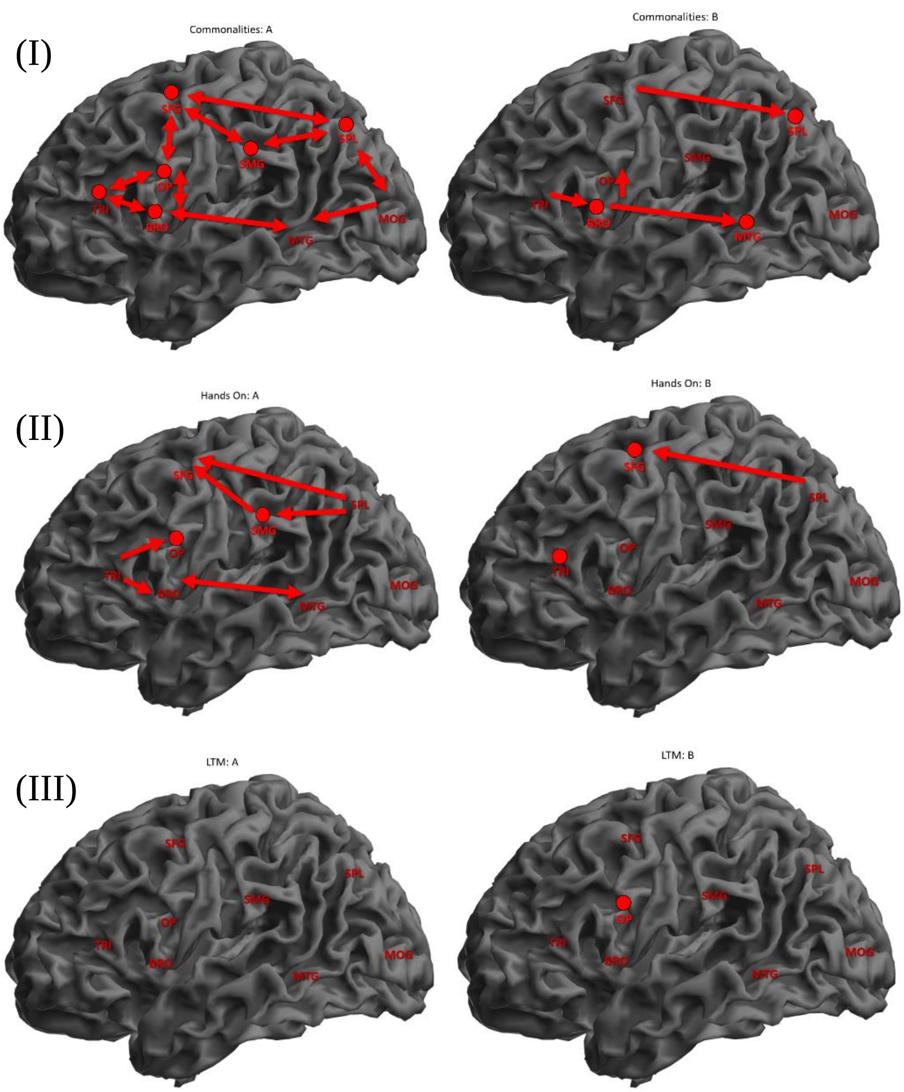


Fig. 2.23…

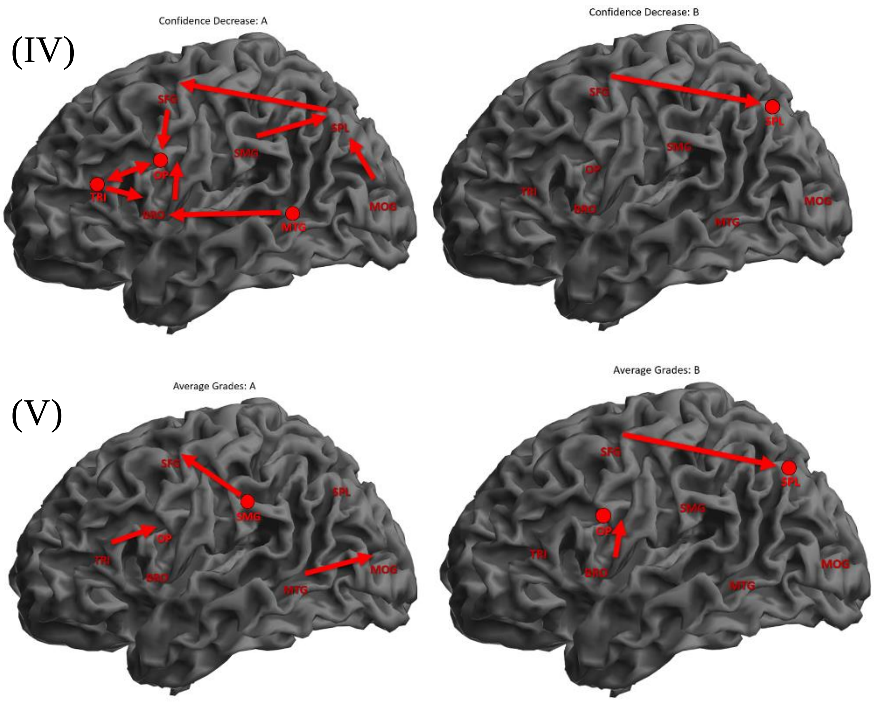


Fig. 2.23. DCM parameters in the reduced PEB model with >=.99 probability of being non-zero. Left: 'A parameters' – Shared between both tasks. Right: 'B parameters' – Connections affected when doing the Iterative Spatial Simulation task over and above the Spatial Simulation condition. Arrows: Indicate the directionality of the extrinsic connections. Dots: Signify intrinsic inhibitory connectivity. (I): Effective connectivity common to both spatial tasks. (II): Effective connectivity differences in subjects that learned Hands On. (III): Effective connectivity differences that correlate with LTM performance. (IV): Effective connectivity changes that correlate with confidence decrease when switching from the spatial to the iterative spatial task. (V): Effective connectivity changes that correlate with average grades.

	Connection	Dir.	Common.		Hands On		LTM		Confidence Decrease (IS-S)		Avg. Grades	
			A	B	A	B	A	B	A	B	A	B
Dorsal-Ventral link	SFG→ IFG.op	f	0.550						neg 0.184			
		b	neg 0.143									
Dorsal Pathway	MOG↓											
	SPL↓		neg 0.210	neg 0.985						neg 1.008		neg 0.346
	SMG↓		- 0.162		- 0.075						0.089	
	SFG↓		neg 0.559			0.480						
	MOG→ SPL	f	neg 0.519						neg 0.310			
		b	0.134									
	SPL→ SMG	f	0.099		neg 0.034							
		b	neg 0.344						0.343			
	SPL→ SFG	f	0.362		neg 0.060	0.095			0.148			
		b	0.825	neg 0.626						neg 0.521		neg 0.178
	SMG→ SFG	f	neg 0.439		0.063						0.044	
		b	0.065									

Table 2.12…

Ventral Pathway	**MTG↓**			neg 0.553					neg 0.285			
	Broca↓		neg 0.548	neg 1.160								
	IFG.op↓		neg 0.253		neg 0.098			0.051	0.418			neg 0.262
	IFG.tri↓		neg 0.072			0.200			neg 0.195			
	MOG→ MTG	f	neg 0.138									
		b									0.140	
	MTG→ Broca	f	0.237		0.094				0.247			
		b	0.198	neg 0.181	0.080							
	IFG.op→ Broca	f	0.084									
		b	0.088	0.197					0.326			neg 0.207
	Broca→ IFG.tri	f	0.175									
		b	neg 0.135	neg 0.159	neg 0.059				neg 0.191			
	IFG.op→ IFG.tri	f	0.292						neg 0.149			
		b	neg 0.334		neg 0.077				0.320		neg 0.039	

Table 2.12. Parameter estimates for all parameters with a p≥.99 of being non-zero (significantly contributing towards the total model evidence). Commonalities: the underlying connection strengths shared by both tasks, that is then modulated by the parameters for the covariates included in the model. A: Spatial task connection strengths. B: modification of connection strength when instead doing the iterative task. Covariates: Hands, LTM and Confidence Decrease are covariates in the group DCM (PEB model). Legend: '→'=Extrinsic connection, f=forward, b=backward. Forward direction is defined by the anatomical position of the nodes (y-coordinate)

and is indicated by the '→' arrow. '↓'=Intrinsic (Inhibitory) connection. 'neg'=negative parameter.

2.3.11.3. Leave one out validation (LOO)

The parameter estimates we presented thus far are all evaluated to be non-zero with a probability >.99 by means of a Bayesian approach, but as a final test of real-world validity the parameter estimates for our covariates can be tested further by performing Leave-One-Out-validation. LOO validation is an iterative statistical approach by which a new PEB model is calculated for each covariate separately, but where one subject is omitted in each iteration. The algorithm then compares the DCM connection strengths of the one subject to the group PEB model and estimates what the covariate in question should be based on the lone subjects' parameters. The accuracy of the estimated covariate values can then be correlated with the real values for all the subjects in turn, and this can be interpreted as a test of whether the covariates included in the DCM can be predicted from the measured connectivity strengths at significantly better than chance levels. As mentioned before, our final PEB model with four covariables came out of a first-generation LOO validation, and the LOO results of those four variables in our final model are presented here in figure 2.24.

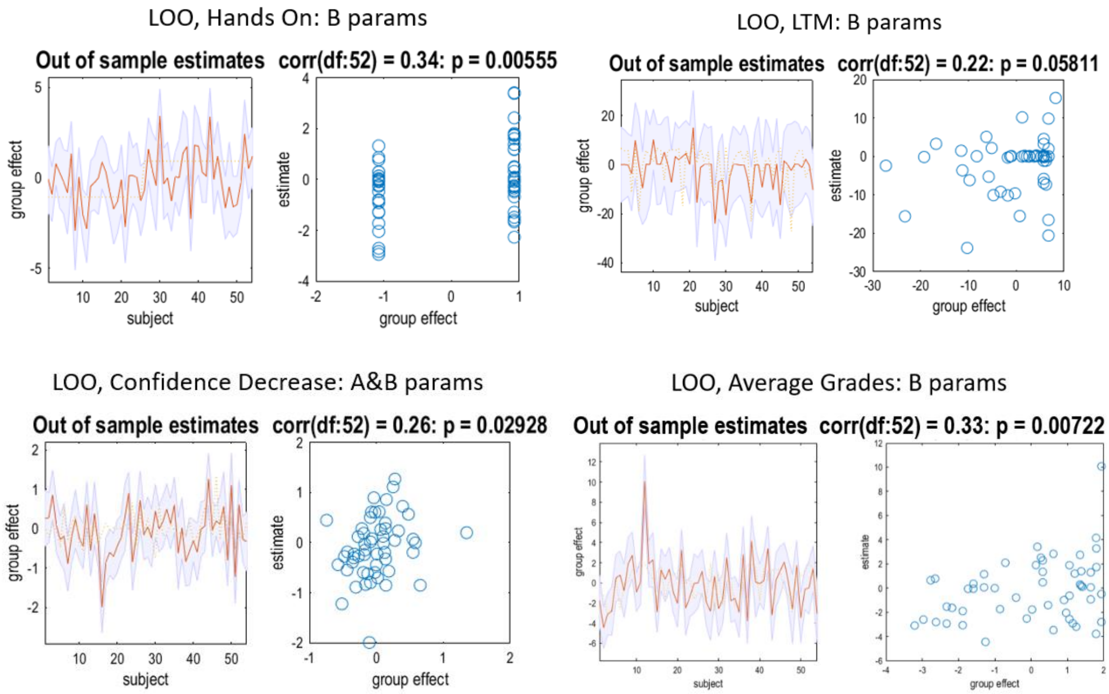


Figure 2.24. Results of significant LOO analyses and whether they were estimated from the B parameters or the A and B parameters. Correlation between the subject's actual variable values and the values estimated from their connectivity parameters in the DCM model. Left: Red line=Estimated values. Dotted orange line=actual values. Right: Scatter plot of the correlation.

The results for our final four covariables in the PEB model showed an average correlation of r=.288 (SD=.057) between the estimates and real values based on the DCM PEB parameters, indicating that the model has some validated predictive power (All showed significant p-values at <.05 except for LTM at p=.058). As stated before, we will discuss possible implications of this further in chapter 2.4.5.

2.4. Discussion

2.4.1. Hands-on vs. Hands-off and predicting success

Even with the uniformly high performance of our participants, we did find several significant behavioral and neural differences depending on whether they had learned programming hands-on or hands-off on the first day of training. The most interesting ones being that the students working hands-on rated their confidence higher in the fMRI task, and they reported less stress after the lesson, indicating that they felt more secure in their knowledge than the hands-off group. When it comes to actual measurements of performance however, we could not find strong support for a benefit, but we do see a trend where the hands-on group performed slightly better on both LTM score ($p=.054$) and written test score ($p=.057$). Looking at our regression analyses aimed at predicting LTM scores, we see that the group variable ‘hands on’ was excluded from the automatic regression analyses, only showing up as significant in our selection of non-self-estimated variables, indicating that hands on is not a strong predictor of LTM performance. It is worth noting however, that in these analyses, hands-on is second only to math grade, both in terms of beta value and in significance. The model with the most predictive power (the backwards regression model) singled out NFC, grade average, and stress before the lesson as the best a priori predictors, with motivation decrease across the lesson and the confidence decrease when moving to the hardest fMRI task as significant predictors. The fact that hands-on shows up as a significant predictor of LTM in our curated regression, which is also the one with the least explanatory power at an adjusted R^2=.316, can be interpreted as meaning that hands-on provides a good enough proxy in leu of better predictive variables, like NFC, stress and motivation, at least in the context of our relatively short timeframe of a 3-hour learning session. Looking at our SEM, we again see confidence and NFC jump out as significant contributors to our ‘programming knowledge’ latent variable. Grade average and math also show relatively big parameter estimates of contribution, but they do not reach significance. Like in the linear regressions, language grades, art grades and Growth Mindset contribute very little. To summarize these findings, we see that our regression models agree with our theoretical conception of the relationships between our variables in that the biggest contributors to programming knowledge formation seems to be confidence driven by working hands-on, NFC, grade average and math grade. This fits with the findings discussed in the introduction where mathematics was proposed as a good candidate for predicting good programming outcomes for students in multiple papers (Bennedsen & Caspersen, 2005; Byrne & Lyons, 2001; Choi-man, 1988; Erdogan et al., 2008; Fincher et al., 2006; Holvikivi, 2010; White & Sivitanides, 2003; Wilson & Shrock, 2001), even though one could predict this to be somewhat task dependent. It is worth noting that both mathematical ability and average grades are often used in research as rough approximations of IQ (Mitchell et al., 2020), and NFC has been shown to be modestly related to fluid intelligence in general (Fleischhauer et al., 2010). It is also important to stress that we found no significant difference in either engagement, self-esteem, or NFC between the groups, meaning that any significant group differences found was not unintentionally confounded by our random assignment of experimental condition.

2.4.2. Motor theories of learning

When it comes to evaluating the various proposals for enhanced learning due to using motoric actions during the learning, we cannot see a simple answer defeat all other alternatives, but rather a combination of influences. The only theory that we can confidently exclude based on our data is the schema theory-based ideas of motor traces being laid down in the motor cortex and that reactivation of these provide stronger associations and better performance (Bartlett & Bartlett, 1995; Shams & Seitz, 2008). We did not observe any significant activity in premotor cortex, primary motor cortex or the supplementary motor area during task preparation or execution in the participants who learned programming hands-on, not even with ROI analysis. The theories that predict enhanced perceptual or attentional processes are interesting to consider further since, although they came up short when it came to directly contrasting our two experimental groups, frontal and parietal areas were found to correlate with confidence, suggesting that perhaps over longer learning timespans, significant differences between groups would develop as confidence was found to be significantly higher in the hands-on group. We did observe several frontal regions activate during the more difficult tasks in our GLM that could perhaps be linked with enhanced attention, which correlated with fMRI score and/or LTM score, showing that they are relevant for successfully solving the task. The third theory that intrinsic motivation and confidence is the primary benefit of hands-on learning (Paris & Turner, 2012) is supported by our findings. Even though we did not see any significant differences between the groups in our GLM, the behavioral measures are in line with this theory and, as we will detail later, we did see significant effects of confidence and hands-on in our DCM model (detailed below), so there is support for relevant changes from our brain imaging data.

2.4.3. Hierarchical recruitment of brain regions

One of the most basic research questions for this project was how the programming knowledge required for successfully solving our tasks was instantiated in the brain. As shown by our GLM, we managed to replicate the preliminary findings from the other groups that pioneered the study of programming using fMRI (Duraes et al., 2016; Floyd et al., 2017; Krueger et al., 2020; Lee et al., 2016; Peitek et al., 2018; Siegmund et al., 2014; Siegmund et al., 2017), showing that at the core of programming are the classical language areas of the brain (Middle Temporal Gyrus and Broca’s area) together with parts of the occipital lobe. These are then supplemented by parietal, and yet more occipital areas when the complexity is increased to include visuo-spatial processing. This can thus be described as more of a task specific activation. Finally, when moving to the most complex task incorporating the iterative code structures, we see the recruitment of further parietal areas, but also frontal regions classically associated with working memory (D’Esposito, Postle, & Rypma, 2000). Since the more difficult tasks involve all the processes of the simpler tasks as well by design, we can see this clear hierarchical relationship of complexity emerge as illustrated in figure 2.12, where we recruit more regions in a roughly posterior to anterior direction as the tasks become more involved rather than see changes in activity within the same structures. It is also important to stress that this additional recruitment, especially going from the spatial to the iterative spatial task is not simply due to the increasing complexity in terms of lines of code containing task-relevant information, as the number of lines separating the two spatial tasks are not significantly different.

2.4.4. Task set – not simply language processing

The delay period in our experimental design that we analyzed using both standard GLM analysis and FIR modeling allows us to draw further inferences regarding the nature of the activity recorded in the areas found to activate during processing of the actual tasks themselves. The way this works is that if we compare the activations during the delay phase in figure 2.10 and the activations for the same contrasts during task processing in figure 2.11 (A & B), we see that there is an overlap in significant activity between the two phases. And when we plot the activity in the most significant activations in our FIR analysis, like left Broca's area and left MTG, as seen in figure 2.14, we can see that after the cue has been presented, activity in these areas increase before the onset of the task. The reason we can call this task-set activity (activity preparing for solving the upcoming task) is that it is triggered by the cue word that the participant has learned to associate with a particular task, and we then see the same areas used when solving that task show up in the delay phase. Perhaps more interesting is that since there is no text displayed on screen during the delay phase, we can confidently assert that the activity seen is not the result of reading lines of text, and since the classical language areas of Broca and Wernicke show this task-set activity, that can be seen as evidence that the activity in these areas when solving the task is not simply a reflection of reading comprehension, but that these areas are involved in processing the meaning and sequential logic of the tasks. As can be seen in figure 2.12A, although the task-set activity for the semantic and spatial tasks do overlap to some extent, they are by no means identical, showing that this activity is task-specific. That areas other than prefrontal or parietal cortexes may be involved in the processes of logic and deduction, especially Broca's area, is a relatively recent notion but research into these structures have shown evidence that Broca's area is involved in at least categorical arguments and perhaps sequential processing (Prado, 2018; Prado, Chadha, & Booth, 2011). Further evidence of this is the fact that there is a significant difference of activation in this area between the Semantic and the Syntax task, while we don't see more activity in this area in the two spatial tasks, as could be predicted if reading was the primary driver of the task related activations we observed, since there is more written text displayed in the two spatial conditions. We will also touch on this idea further in the next section dealing with our DCM results.

2.4.5. DCM & LOO validation

The first hypothesis that we could tackle using DCM was the network structure of the regions identified in our GLM analysis. The Bayesian model selection over our entire model space selected our 'improved model

1' as the best fit to our observed data, meaning that our best model followed the dual stream hypothesis of a ventral stream flowing forward into the classical language network, and a dorsal stream ultimately terminating in the frontal cortex. That this model won also indicates that the two streams do not significantly interconnect between the more posterior nodes in the network, as tested by less successful models connecting MTG to SFG or SPL, but that the two meet in the frontal lobe, in our winning model represented by the connection between IFG.op and SFG (figure 2.21). It is also worth noting that one of the other out-competed models was our 'full' completely data driven model, meaning that model 1 did not win by simply being more complex than its nested competitors by over-fitting.

Looking at the parameter estimates from our PEB model (figures 2.22-3) we can see that a slight majority of them are negative (38 neg. vs. 29 pos.), normally indicating that they inhibit their target node, but 16 of them are estimates of inhibitory 'intrinsic' connectivity within the eight nodes, meaning that there is an increase in activity due to these influences as they decrease the self-inhibition. Only 5 of the estimates for the intrinsic connectivity are positive (increased inhibition) meaning that we see almost twice as much excitatory connectivity compared to inhibitory during the processing of our programming tasks. We can also see that the connectivity pattern for confidence, and to a lesser extent hands-on are the closest match to the 'commonalities' pattern. This could be interpreted as evidence that high confidence and learning hands-on has a more 'global' effect. In the case of Hands-on learning, it means that the mode of learning causes connectivity changes in nearly all significant connections in our model. In the case of confidence, it means that the connectivity changes either come about as a consequence of the improved confidence from working hands-on, or that changes in all these connections are what produces the observed changes in confidence, as there is no way to determine causality from the DCM alone. It is also worth pointing out that both confidence and hands-on affects both the ventral and the dorsal pathways. Confidence is reflected in altered backward connections of the dorsal stream from superior frontal to parietal areas. Participants more confident in the iterative task tend towards having greater excitatory communication from parietal to frontal areas, and from SMG to SPL. Inhibitory connections are instead seen between some of the inferior frontal areas of the ventral stream. We were also able to predict whether participants had small or large changes in confidence when doing the iterative spatial task as compared to the spatial task with significant accuracy from the A and B parameter strengths in the LOO validation. For high LTM score we see only one change: Increased inhibition in IFG.op (Broca's area 44). Interestingly, this variable was the only one to affect only a single parameter but was also strong enough to get very close to significant prediction support from LOO validation at $p=.058$, showing a correlation of $R=.22$ between predicted and actual values of LTM performance. Unlike LTM, the other variables affect the connectivity in the frontoparietal network. One way to interpret this is that it does not look like it is the case that attentional processes instantiated in the frontoparietal network gives any advantage when it comes to long-term memory. Instead, whatever differentiates students with high LTM scores is found within the IFG language areas, also agreeing with the correlations we found between BOLD activity in left IFG and LTM score in our GLM analysis. Moreover, the DCM shows that greater LTM, confidence, and lower stress are all marked by a strengthened inhibition of intrinsic activation in the left inferior frontal cortex, Broca's area, or operculum during task performance. Looking closer at the B matrix for the commonalities, which indicates changes in connectivity when switching to the iterative spatial task, most notably, a change in Broca's area is observed. We see decreased self-inhibition and a reduced connectivity to MTG together with a weakened retrograde influence from SFG to SPL as well. This can be interpreted as an increase in the communication within the inferior frontal lobe, coupled to a reduction in retrograde information transfer, indicating that Broca's area activity is more involved in the processing of these iterative sequence tasks. This fits with the idea that IFG also processes sequential logic (Prado, 2018; Prado et al., 2011) but not producing a big enough change in BOLD response to show up in our GLM. For the average grade variable, we again see IFG.op (Broca's area 44), but connections within the dorsal network as well, meaning that grade average also has a somewhat more global effect.

Finally, when we return to the main objective of our experiment, namely differences when learning hands-on, we do see significant effects in our DCM model. Recall that there were no significant differences in BOLD amplitude in our whole brain exploratory analysis of hands-on vs. hands-off (GLM). However, from our PEB parameters we were able to predict whether a participant had learned in the Hands-on or the Hands-off condition with significant accuracy from the participants' B parameter strengths: $R=.34$, $p=.005$. Hands-on learning led to an increase in forward communication from parietal to frontal areas. In addition, it led to bidirectional inhibition between Broca and Wernicke's area, while there was an added excitatory communication from operculum to triangularis in the IFG. This means that the language processes in the ventral pathway are influenced both by Hands-on learning, and by confidence (as discussed above), which indicate that theses variables support depth of learning, an interpretation corroborated by the fact that we also see that regulation of activity in Broca's area predicting long-term memory retention of programming knowledge. One final reflection regarding our LOO validation results is that it is interesting to note that we were able to predict from out fMRI data variables separated in time, specifically the learning condition from the day before, and then projecting forward, the LTM score evaluated one week later.

2.4.6. Limitations

The most straightforward limitation in this study is that our sample size is too small for what is typically recommended for producing reliable results form a SEM. However, since we primarily used our SEM as a means to visualize our theoretical conception of the relationships between our variables, and since the results are backed up well by our significantly more robust traditional regression analyses, we feel justified in including it. A more fundamental issue is the inherent homogeneity of our sample population: all high performance, self-selected students from relatively affluent schools and regions. This might have been compensated for better if we had a better idea of how difficult to make the programming tasks.

They were designed from the start to mimic what you would typically find in introductory lessons of programming coursework and were even piloted in another school setting prior to our experiment, but evidently a slightly harder set of tasks could have produced a wider variance in our sample. This could possibly account for why we failed to find significant differences correlated to fMRI score in our DCM. Perhaps weaker students might also stand to benefit more from the benefits of working hands-on. Another major limitation is the constraints imposed by the fMRI setting itself. Not only does it force our hand in how we can design the tasks themselves, but the inherently uncomfortable process of being restrained in a prone position within the scanner means that we are automatically in a strict economy of time where we are forced to balance the number of conditions that we want to test versus how many repetitions we will need to secure enough statistical power to detect significant activations. Since we expected any differences to be small, we had to err on the side of caution, meaning that our experiment pushed the limit of what is tolerable for the test subjects. Our ROI analysis of hands-on vs. hands-off learning effects made use of rather large and broadly defined regions, especially our ROI covering the motor regions in both hemispheres which we attempted to partially counteract by including a smaller region centered on a reported finding from a previous study of language learning. The number of ROIs tested was also a potential issue arising from the exploratory nature of this experiment, but since we failed to find any significant results at our un-thresholded analysis further statistical corrections were irrelevant. Finally, we did not design our fMRI paradigm from the start with the intention to perform FIR analysis, meaning that our random length delay phases needlessly complicated the pooling of our subjects' data, necessitating further manual interventions that always has the possibility of introducing biases or errors that can be hard to identify.

2.4.7. Conclusions

To restate the answers to our research questions; **1:** We did not have enough statistical power to conclude that learning hands-on produced better outcomes in terms of BOLD amplitude. Our strongest finding is that confidence is significantly higher in the hands-on group during fMRI scanning, and that confidence in turn is predictive of better LTM score. **2:** Programming recruits a hierarchy of brain regions depending on the type and complexity of the task. **3:** Processing of computer code takes place in dual streams, and information interconnect in the frontal lobe. **4:** Language areas are involved in processing the logic (deductive output) of the tasks, not simply reading comprehension. **5:** Average grades and math are confirmed as pretty good predictors of future success in programming, but NFC is better. **6:** Relevant aspects of the programming knowledge are partly constituted by the strengths of interconnectivity between the relevant brain regions, not just the levels of activity within them.

In this study we have managed to replicate the previous findings showing that language areas are the most relevant activated regions when solving programming tasks (Duraes et al., 2016; Floyd et al., 2017; Krueger et al., 2020; Lee et al., 2016; Peitek et al., 2018; Siegmund et al., 2014; Siegmund et al., 2017), indicating that programming is indeed more akin to a language rather than pure math, even though math is a good predictor of success. This could point a way forward by possibly leveraging the knowledge from the various sub-fields of L2 language learning. We have also shown that complete programming novices, after only 3 hours of teaching, show significant differences in their functional connectivity between task relevant brain regions large enough to predict if they learned in the hands-on condition or not. A strength of the short timeframe and novice subjects is that we eliminate a lot of potential confounds that come with time and previous experiences, giving us good ground to stand on when attempting to capture the core requirements to learn computer programming. In the future we would like to see longitudinal studies of a similar kind to see if the small effects we have shown are further amplified over time. We would also like to see studies focused on the characteristics captured by the NFC scale, and possibly if there are effective interventions possible to promote them in students, especially in computer programming since it seems to be a good predictor of success and correlates strongly with confidence. From the results of this study, we can feel confident to recommend having programming students work hands-on, validating the prevailing consensus.

Chapter 3
Effects of feedback interventions on stress and motivation

3.1. Introduction

This chapter details the results from a re-analysis of the behavioral and fMRI data from the 2015 paper “Feedback on Trait or Action Impacts on Caudate and Paracingulum Activity” by Alva Appelgren and Sara Bengtsson. We will summarize the background regarding the experiment relevant to our new findings in the following sections, but for a complete description, see the original paper (Appelgren & Bengtsson, 2015).

In our programming experiment detailed in Chapter 2, we show that attitudinal states are important for successful learning and problem solving. We found significant positive correlations between performance on our long-term memory test (LTM), low stress, high confidence, and a high need for cognition (NFC). Whether NFC is possible to influence, either short term or lasting, remains to be explored. According to our theory of relationships laid out in our structural equations model (figure 2.9), motivation is influenced by confidence, which is itself a downstream consequence of innate dispositions, such as NFC, probably factoring in the subjective assessment of the student’s current level of aptitudes and interests regarding the task in question. If we assume that the current aptitude and interest of the students at the start of our programming experiment were at least close to equal, given that they were all programming naïve and volunteered out of personal interest, any variance in performance not due to more general cognitive abilities such as those measured by IQ must be due to continuing motivational factors. Recall that our sample was highly skewed with both average grade and math grade means around 18 ± 2 out of a maximum of 20 credits (See appendix I). With the analyses in this chapter, we investigate the dynamics of motivation across an experiment, and test how performance measures vary accordingly. Since the data set also incorporates two conditions of different feedback types, it has the potential to address feedback interventions on motivation and cognitive performance, and whether the two can be linked in a compelling way. We have previously shown, using the SEM in Chapter 2, that both stress and motivation are aspects of confidence. We found that motivation and stress were both correlated with performance in our programming study, we therefore hypothesize that if feedback can induce changes in motivation and/or stress we would expect to see that reflected in task performance here in the form of changes in reaction time, and that this effect would be compounded over time with more feedback received. We have also shown that activity in regions significantly activated during our programming task (SFG, IFG, SPL, MOG) were correlated with our confidence measures. These same regions, when tested in our DCM, also showed changes in interconnectivity correlated with robustness against confidence decrease when switching to our hardest programming task. Taken together, we therefore also hypothesize that any changes in motivation or stress induced by feedback should be reflected in brain activity in regions previously identified with confidence in our programming experiment.

3.1.1. Summary of results from the original paper

The original paper found that the participants displayed both higher accuracy and faster reaction times (RT) in the trials without feedback as compared to the trials with feedback. This effect seems to be driven by the bivalent rule condition, not by the much simpler univalent rule. It was found that trait feedback ("You are clever") could lead to increased motivation, but only in people with a growth mindset. On the other hand, it was found that task feedback ("Your choice was correct") led to both a higher motivation to continue with the task, and to less stress. Task feedback also led to better accuracy in the following non-feedback trials as compared to trait feedback. Many of the participants reported after the experiment that they found the trait feedback distracting or annoying and the article summarizes its finding with the following conclusion: "It turned out that trait feedback was less beneficial for motivation and performance improvement."

3.1.2. Hypotheses

(I) Feedback-induced changes in motivation and stress will impact reaction times.

(II) Changes in reaction times will progress over the course of the experiment.

(III) Brain regions previously implicated in confidence will show activations that correlate with reaction time changes.

3.2. Method

3.2.1. Experimental overview

This experiment was a typical rule-switching task (univalent vs bivalent rule), where each participant comes into the fMRI scanning lab for two visits with 18.5±6.45 days in between. At each visit, the procedure was the same, except that the type of feedback they received was changed for the second visit. The design is balanced so that half of the participants received the first type of feedback initially and the other half received the second type, so in the end all participants receive both types of feedback, meaning that this is a within-subject design. The two types of feedback were called 'Trait' feedback ("You are clever"/"wrong") and 'Task' feedback ("Your choice was correct"/"Wrong"), making this a type of priming experiment (Bengtsson & Penny, 2013).

Each visit begins with a self-rated motivation and stress measurements on 10-point Likert scales. This was then followed by 100 trials with feedback after each response. The participants were then asked to report their motivation and stress again. This was followed by a further 50 trials without any feedback. Finally, the participants were asked to rate their motivation and stress one final time.

There were two trial types in this experiment: 'Univalent' (30%) and 'Bivalent' (70%) rule trials. Each trial begins with a rule symbol, denoting what the correct response should be given the image subsequently presented in the response phase. For the univalent trials, the correct response for the symbol presented was always the same, but for the bivalent trials the response to any given image was dependent on which one of two possible bivalent rule symbols was presented before.

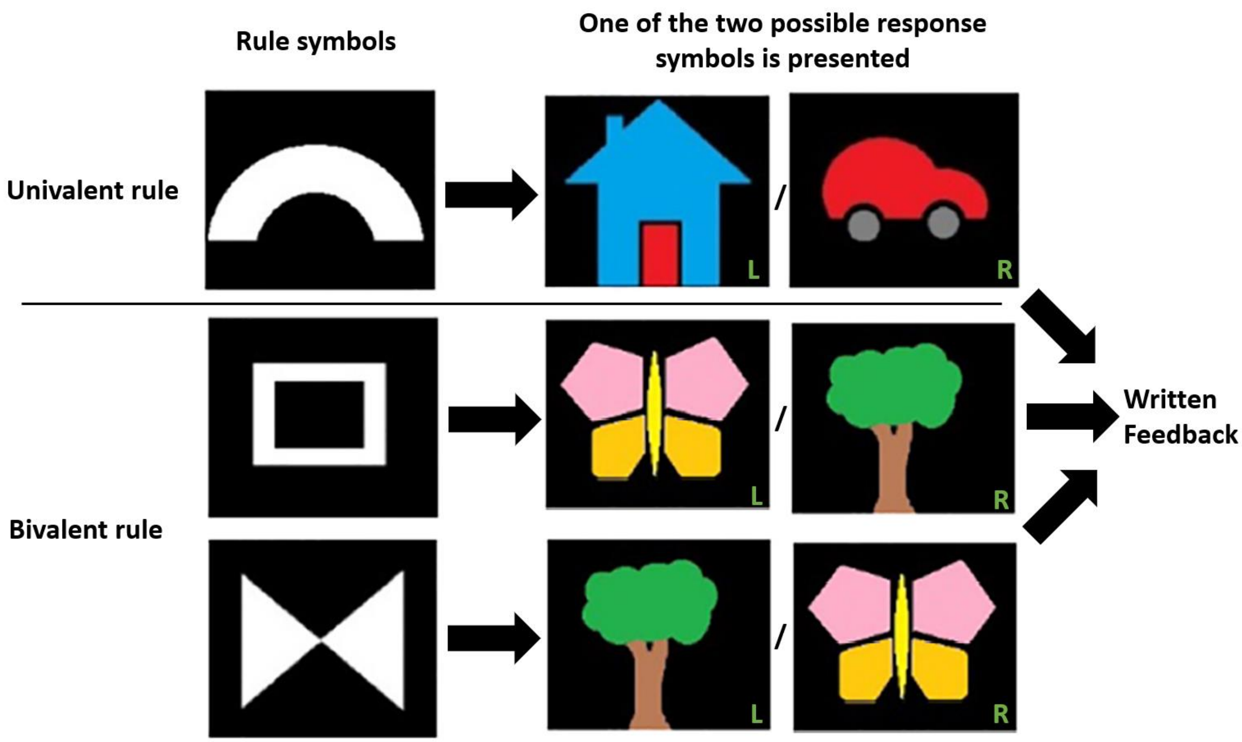


Fig. 3.1. Task description of the different rule- and response symbols. Correct responses indicated by the green letter. L=left button, R=right button.

3.2.2. Participants

20 healthy volunteers (age 24±5.6, 8 females, 13 native Swedish speakers and 7 native English speakers) took part in the study. All participants were right-handed and neurologically healthy. The participants completed a

practice session until they achieved greater than 60% task accuracy before beginning the experiment in proper, with a further 12 practice trials in the scanner to familiarize them with the button box and the visual display. The study was approved by the local ethics committee in Stockholm (EPN), Sweden, Dnr 2014/10-31/2.

3.2.3. Statistical analyses – behavioral data

Behavioral data was analyzed using Microsoft Excel v16.0, and IBM SPSS Statistics v25. To look deeper into motivational effect on performance, we investigated differences in RT depending on feedback type. Firstly, we investigate if the cognitive process of switching between two different rules was influenced by the external feedback. That is, if switching from the univalent rule to the bivalent rule or vice versa influenced the RT on the next trial. Next, we investigate the temporal dynamics of RT across the whole experimental session. To this end, we partitioned the session into 10 time-bins covering 10 trials each (figure 3.2) so we could compare the two feedback conditions against each other across time as opposed to in aggregate. We also calculated the effects of feedback and time using linear regression.

3.2.4. fMRI data acquisition

Nuclear Magnetic Resonance Imaging was acquired on a 3 Tesla GE Discovery MR750 equipped with an 8-channel phased array receiving coil. Functional MRI was performed in two session of echo-planar T2*-weighted imaging of 273 volumes, using 2 mm isotropic voxels, TE = 30 ms, TR = 2600 ms, FoV = 28.8 cm, 40 contiguous oblique slices, flip angle=90 deg, for a total of 546 volumes (~70min.). 3D T1-weighted SPGR (Spoiled Gradient Echo pulse sequence) images was acquired with 1 mm isotropic voxels, TE = 3.06 ms, TR = 7.9 ms, TI = 450 ms, FoV = 24 cm, 176 axial slices, flip angle = 12 deg.

3.2.5. Image preprocessing

The re-analysis made use of the already pre-processed images from the original study up to and including the first level GLM:s for each individual participant. The preprocessing and all subsequent analysis in the original paper was carried out using SPM12b (2014) (Penny et al., 2011).

3.2.6. GLM analysis with first order temporal modulation

Second level group GLM re-analysis was carried out using SPM12, version 7487 (Penny et al., 2011). This type of GLM analysis using what is called temporal modulation (t-mod) shows areas that show an activation pattern of ‘first order’ or linear relationship with time. ROI analysis was carried out using the Family-Wise Error-corrected (FWE) task activation clusters from the programming experiment (detailed in Chapter 2) to look for common processing patterns. Peak statistics on FWE-corrected activation clusters were investigated for areas showing first order temporal correlation ($p<.05$) in our ROI analysis, in addition to uncorrected whole-brain activations ($p<.001$) (see table 3.2).

3.3. Results

3.3.1. Reaction time

We found a highly significant difference in RT between switch and stay trials: Average RT (ms) for trait feedback: Switch=759.2 ± 209.8 Stay=693.4 ± 187.5 (t(21)=4.95, p=.0000674***). Average RT (ms) for task feedback: Switch=766.8 ± 197.5 Stay=691.2 ± 193.9 (t(21)=5.98, p=.0000075***). And if all feedback trials are combined (trait + task): Switch=803.6 ± 216.8 Stay=727.1 ± 194.4 (t(21)=7.57, p= .0000000011***). However, we did not find any significant differences in RT between switch and stay trials when comparing trials of task and trait feedback (p>0.3).

Trait feedback slows RT marginally, whereas task feedback leads to significantly faster reactions over the time of the experiment. Computing the delta RT (task-trait) therefore revealed the following: RT started out ~30ms slower during task- as compared to trait feedback, but over time this relationship flips to instead favor task feedback (figure 3.2). This correlation is significant (R= - .733, p = .008**). The same relationship holds for both switch- and stay trials (R=.8 p=.003** and R=.591 p=.036* respectively), but this is driven solely by the bivalent trials (R=.777 p=.004**) and not by the univalent trials (R=.099 p=.393). This phenomenon thus seems to be the strongest during higher cognitive load.

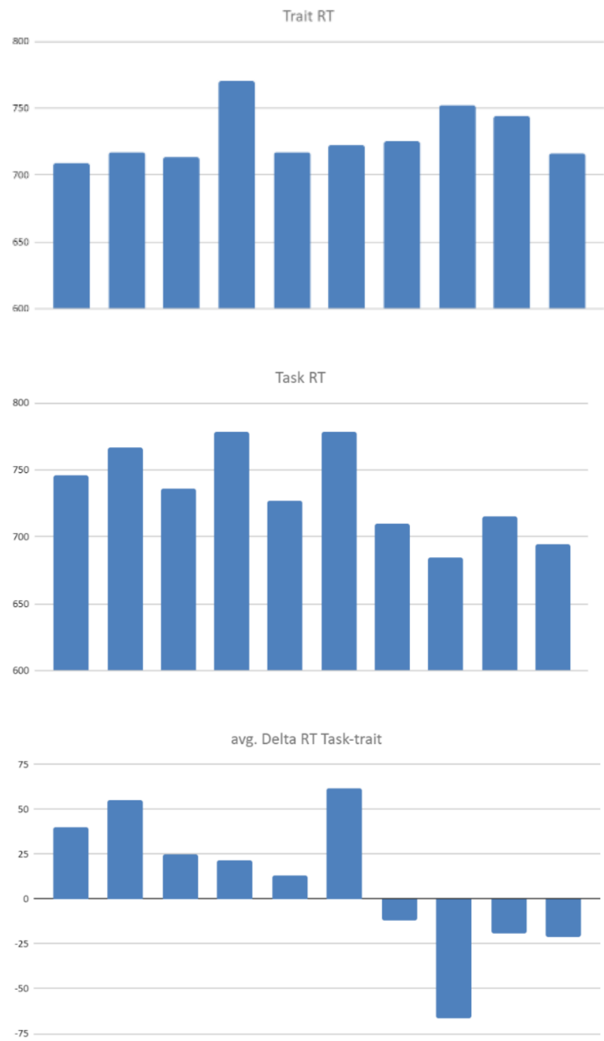


Fig. 3.2. Temporal dynamics of reaction time across 10 time-bins of 10 trials each. Left: Average RT during trait feedback over time. R=.292, p=.206. Right: Average RT during task feedback over time. R=-.676, p=.016. Bottom: Average delta RT (task-trait) over time. R=.732, p=.008**. (*=p<.05, **=p<.01)*

When comparing RT over time in the two feedback conditions Task and Trait, we find a significant negative time*feedback-interaction effect in a regression analysis predicting RT (Table 3.1).

Model		Beta	t	Sig.	95.0% CI Lower	95.0% CI Upper
1						
	Time	1.238	1.967	0.067	-0.887	23.625
	Feedback	1.078	2.507	0.023*	8.788	104.978
	Interactrion	-1.901	-2.583	0.020*	-17.196	-1.694

*Table 3.1. Linear regression analysis. Effect of Time, Feedback and Time*Feedback-interaction on reaction time. (*=p<.05)*

The results from looking at the time-bins shows that reaction time decreases over time in the task feedback condition, as would be expected in a repetitive learning paradigm; RT during task feedback significantly correlates with time: R=-.676, p=.016*. This learning effect is absent in the trait feedback condition; RT during trait feedback did not significantly correlate with time: R=.292, p=.206 (figure 3.2).

3.3.2. GLM analysis using first order temporal modulation

We used t-mod to look for brain regions that show activations that display a negative linear relationship with time (areas where activation strength decreases over the experimental session) to link the results of our time-bin analysis of RT to the task related activity in the brains of the participants. The activity in these areas can thus be implicated in the change in RT over the session. We investigated both positive and negative t-mod (increased and degreased activity with time, respectively), and in both feedback conditions (Trait, and Task). No contrast was significant at the whole-brain FWE-corrected statistical threshold. The only contrast showing significant activations in our ROI analysis was the negative t-mod for the bivalent Trait feedback condition. The GLM results are presented below in table 3.2 and figure 3.3-4. We observed significant activations in left precentral gyrus when using ROI:s of significant task activations from our programming experiment.

Area	Peak	T	p (FWE)
FWE-corrected Task activations form programming experiment ROI			
L precentral gyrus**	-60 8 24	4.88	0.000
L precentral gyrus**	-60 10 28	4.85	0.000
p<0.001 uncorrected peaks			
L precentral gyrus	-60 8 24	4.88	0.070
L precentral gyrus	-40 -16 62	4.37	0.304
R middle frontal gyrus	50 34 18	3.6	0.951
L superior frontal gyrus	-20 -8 74	3.31	0.998
L inferior frontal gyrus.op	-44 16 26	3.7	0.907
L inferior frontal gyrus.op	-58 20 20	3.38	0.994
L inferior frontal gyrus.op	46 20 26	4.19	0.457
R inferior frontal gyrus.op	60 20 24	3.42	0.991
R inferior frontal gyrus.op	32 30 0	4.15	0.497
L anterior Insula	-34 20 0	4.36	0.307
R anterior Insula	32 22 -6	4	0.648
L caudate	-12 14 4	3.62	0.944
R caudate	10 16 6	3.95	0.703
L Putamen	-16 10 -6	3.43	0.989
L Thalamus	-8 -12 2	3.62	0.942
L supramarginal gyrus	-50 -28 44	4.44	0.251
R supramarginal gyrus	38 -32 38	3.37	0.995
R precentral gyrus	54 14 34	4.02	0.629
L postcentral gyrus	-42 -36 60	3.25	0.999
L superior parietal lobule	-32 -50 66	3.27	0.999
R occipital fusiform gyrus	34 -64 -14	3.84	0.801

Table 3.2. Significant FWE-corrected (p<.05) negative t-mod activations after small-volume-correction, using ROI:s from our programming experiment. Contrast: Bivalent trials in the Trait feedback condition (70

trials). Whole-brain uncorrected results at p<0.001 are also presented. (=p<.05, **=p<.01)*

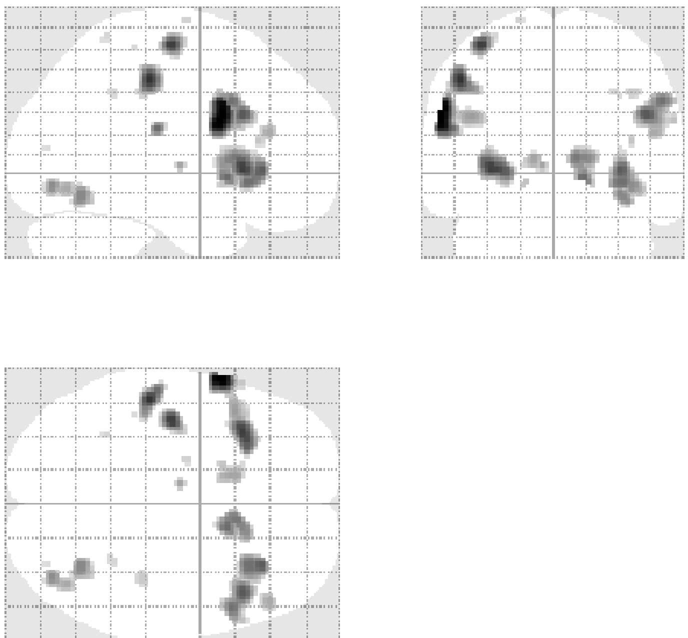


Fig. 3.3. Glass brain, t-mod GLM, uncorrected (p<.001). Average activity during bivalent trials with personal feedback only (70 trials).

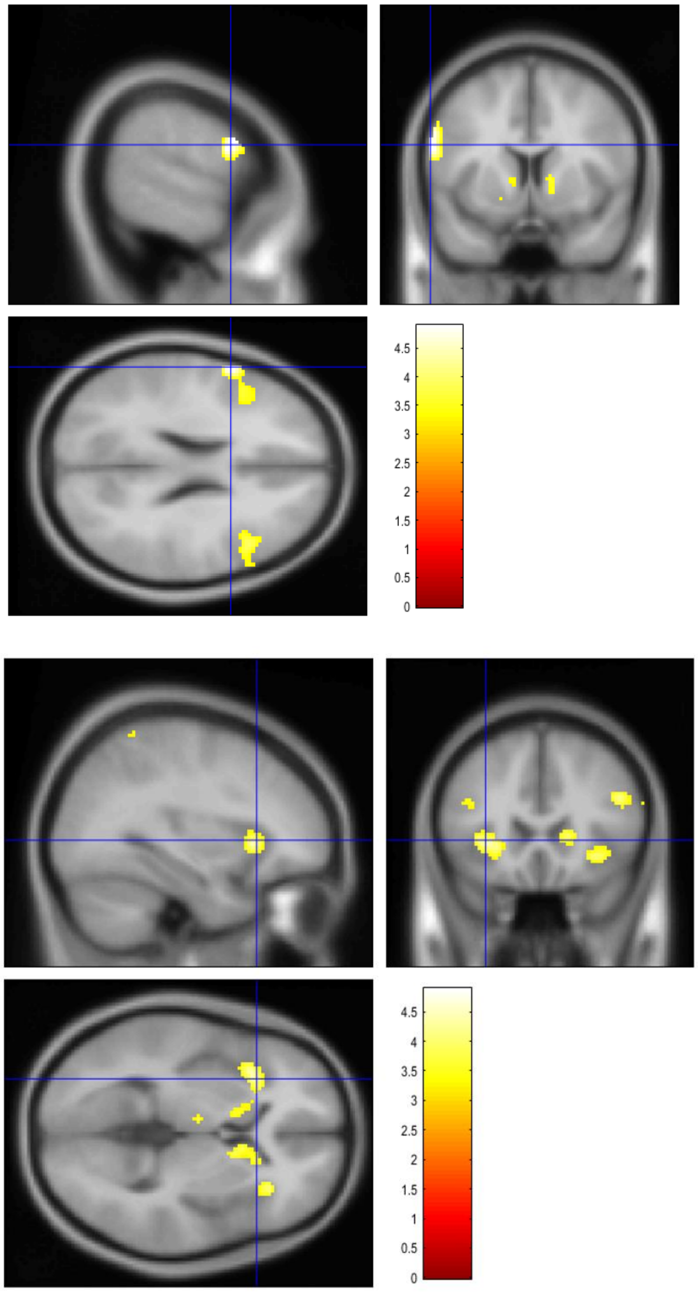


Fig. 3.4. Selected t-mod GLM results, Average activity during bivalent trials with personal feedback only (70 trials). Top: Significant FWE-corrected ROI activation (p<.05), Left precentral gyrus [-60 8 24]. Bottom: Additional non-significant peak of interest (t=4.36), Left anterior Insula [-34 20 0].

3.4. Discussion

3.4.1. Motivational interventions

The t-mod GLM showed two areas where activation changes over time, suggesting that they might be responsible for the shift in RT:s we observed over the course of the experiment. The strongest activity was seen in the left precentral gyrus, a peak coordinate falling within the Brodmann area 44 mask of the SPM Anatomy toolbox, denoting part of Broca's area. This is in line with the findings from our experiment in Chapter 2, where we saw significant changes in connectivity within Broca's area in participants that recorded a smaller drop in confidence when switching to the hardest programming task (table 2.12). The precentral gyrus has been shown to be involved in auditory feedback processing (Christoffels, Formisano, & Schiller, 2007) as well as successful inhibition of behavior in a stop-signal experiment investigating the effect of motivation (in the form of a monetary incentive) on how positive incentives may actually impair behavioral performance (Padmala & Pessoa, 2010). Taken together these findings paint a picture of the precentral gyrus as a region not only evaluating auditory feedback, but also involved in balancing behavior in accordance with external motivational manipulations in the form of incentives. A second interesting area of similar statistical threshold, but not included in the activations ROI from our programming experiment, was the left anterior insular cortex, a region that is associated with a wide array of cognitive processes including interoception, emotions, language processing, perception, salience, and consciousness, and aberrations of the insular cortex has also been linked to anxiety and mood disorders (Gasquoine, 2014).

In our conception from our programming study, stress and motivation are both reflected in confidence, one of the measures that significantly predicted success in our SEM and ANOVA (tables 2.6-7). As stated in the introduction, the previously published findings from this experiment shows that feedback interventions have the potential to influence both the stress and motivation of the participants, suggesting that they experienced decreased confidence, possibly reflected in Broca's area activity change over time. They also expressed annoyance with the trait-based feedback, possibly reflected in the activity in insular cortex. That the type of feedback is the important factor and specifically, that task-based feedback is preferred, is a point in agreement with the Growth Mindset theory as we have discussed in Chapters 1 and 2.

It seems clear that motivational aspects can have both direct and indirect effects on performance and that the wrong type of interventions can potentially deteriorate a student's motivation to engage with tasks. Sun & Rueda, 2012 investigated if different aspects of engagement (behavioral, emotional and cognitive), as measured by a scale developed in 2004 (Fredricks, Blumenfeld, & Paris, 2004; Moore & Lippman, 2006), are influenced by factors such as situational interest, computer self-efficacy, or self-regulation. They concluded that situational interest seemed to be key to engagement (Sun & Rueda, 2012). These three motivational and learning factors have previously been linked to student engagement levels in multiple studies (Bates & Khasawneh, 2007; Dembo, Junge, & Lynch, 2006; Kanuka, 2005). Engagement is defined as the "*quality or effort students make to perform well and achieve desired outcomes*" (Hu & Kuh, 2002; J. C. Richardson & Newby, 2006; J. T. Richardson, Long, & Foster, 2004). In their paper, Sun & Rueda found both that "*interest and self-regulation were significantly correlated with all types of engagement*" and that "*interest was only a significant predictor of emotional engagement; self-regulation was a significant predictor of all types of engagement*". They suggest that it is important therefore to facilitate emotional engagement by increasing student interest, but it is not clear how best to achieve this. It might even be argued that that interest follows from emotional engagement and not vice versa. It is important to note that this study was in the context of online learning where the students had to take responsibility for their own education, and so the fact that self-regulation significantly predicted all three types of engagement should perhaps come as no surprise. It might be that students with a greater ability to self-regulate either compensate for any lack of interest or perhaps somehow 'manufacture' genuine interest in their effort to reach a higher desired goal, like a good passing grade.

3.4.2. Limitations

The primary limitation of this study is the lack of statistical power because of the small sample size. Additional measurements of variables like confidence and NFC would also have enabled us to draw closer links between our two experiments.

3.4.3. Conclusions

In conclusion: The type of intervention matters. The wrong type of feedback is not only perceived as distracting but manifests as an observable decline in motivation, resulting in a performance drop, in this case an increased reaction time towards the end of the experimental session. We note brain activity in Broca’s area that decreases over time, a region linked with confidence in our study of programming. Since this effect was observed only when the participants received personal trait feedback, it could be the case that direct task-related feedback is one of the contributing factors mediating the benefit of hands-on interaction with programming observed in chapter 2, or perhaps more likely; That reaction times normally improve over the course of the experiment due to repetition, but the introduction of trait-based feedback significantly retards this improvement. This would also fit with the previous published paper’s finding that reaction times were faster in the no feedback-condition. The fact that we observed activity changes in Broca’s area and, to a lesser extent, insular cortex, as a function of feedback condition suggest that they are involved in the processing of motivational aspects, like confidence and stress. Since motivation can be significantly affected by interventions in the form of written feedback, we would like to see future experiments with more longitudinal measures of stress, motivation, and confidence, perhaps separated into multiple motivational aspects for greater clarity like those discussed above. The purpose of this is ultimately to test additional types of interventions on confidence.

Chapter 4
General Discussion

4.1. Summary of Findings

We have shown that programming knowledge is primarily instantiated in the classical language processing areas of the brain, with additional parietal and prefrontal areas recruited during spatial and iterative tasks. Learning computer programming while typing on the keyboard (hands-on), results in better performance on a long-term memory test, and produces higher levels of confidence in the participants. This benefit appears to be mediated by increased attentional and/or motivational processes rather than involvement of the motor cortices. Motivation can in turn be adversely affected by external feedback based on personality traits. We also identified the NFC construct as a strong predictor of future performance in addition to math grade, and grade average, but not Growth Mindset.

4.2. Theoretical Implications

Since we failed to find any direct links between motor cortex activity and learning condition in our programming tasks, we must conclude that any benefits observed in our study, must stem from some other source. And, since our findings suggest that the benefits are mediated by attention and motivation, ultimately manifesting in greater confidence, this gives us at least these three variables to use as measures to gauge the effectiveness of any other proposed embodied cognition-based intervention. In other words, if our findings are correct and transferable, then we would expect to see a similar pattern of enhanced attention, motivation and confidence in other similar interventions that claim to provide a positive effect on learning outcomes. Also, the fact that we were able to distinguish the students that had learned in the hands-on condition based on the connectivity patterns in their brains, after such a short training session, suggests that this type of connectivity analysis may be useful to evaluate students of varying levels of expertise and backgrounds. The potential advantage would be the ability to detect small but significant changes in brain connectivity after an intervention, perhaps regardless of previous experience level, because these small changes can then be compared between the groups instead of the baseline connectivity. In the next section, we will discuss our findings as they relate to the two personality traits included in this study (Growth Mindset and NFC).

4.3. Mindset, Antifragility & Need for Cognition

Perhaps the most prominent theory of attitudinal effect on learning is the so-called 'Growth Mindset' theory proposed by the American psychologist Carol Dweck, as discussed in Chapter 1. The central thesis as laid out in the book "Mindset: The new psychology of success" has been cited more than 12000 times according to Google Scholar. It is highly improbable that any given educator has not been exposed to the idea of Growth Mindset in one form or another. In Sweden for example, Mindset is promoted by the national agency for education (https://www.skolverket.se/skolutveckling/forskning-ochutvarderingar/artiklar-om-forskning/berom-for-talang-till-skada-for-motivationen), the biggest labor unions for teachers, 'Lärarförbundet' (https://www.lararforbundet.se/bloggar/forstelararbloggen/elever-kan-naa-sin-fulla-potential-med-hjalp-av-dynamiskt-mindset-paa-alla-nivaaer) and 'Lärarnas Riksförbund' (https://skolvarlden.se/artiklar/tank-ratt-och-lyckas-i-skolan), together with forces from the private education sector with companies offering seminars and other professional development in the realm of Mindset such as 'Lärar Fortbildning AB' (https://www.lararfortbildning.se/grundskola/uppdrag/inspirera/mitt-mindset) and 'Medalgon Utbildning AB' (https://www.mynewsdesk.com/se/brainsmart/pressreleases/stort-intresse-foer-growth-mindset-livslaangt-laerande-1229682).

The proliferation of corporate entities engaged in the promotion of any trend in education should perhaps provoke reflection in the mind of any skeptical consumer of research literature. A cursory search online yields many sometimes interconnected private companies and think-tanks promoting initiatives based on Growth Mindset in any language you care to read, like the recently rebranded 2015 organization 'Mindset Scholars Network' , since March 2021 now known as the 'Student experience research network' (SERN). They are a partner to the 2010 organization 'Project for Education Research that Scales' (PERTS) whose original purpose was to leverage the field of mindset studies in education; "*Inspired and excited by the potential of these low-cost, low-time interventions, we wondered: Why aren't all schools doing these?*" (https://www.perts.net/press_kit). Many of these organizations also seems to have a clear activist side to them. One example of this is the spotlight video series from SERN described by them as "*designed to elevate stories of scholars and intermediary organizations using research to promote equity and inclusion*.", where they for example praise the decision by the University of California Board of Regents to suspend their SAT/ACT requirement for applicants until 2024 (https://www.universityofcalifornia.edu/press-room/university-california-board-regents-approves-changes-standardized-testing-requirement). That many initiatives and companies in this realm are directly or indirectly tied to activism in one form or another is perhaps unsurprising given that the probable aspirations of the members is to improve the conditions of people in education, but it has to be the case that any successful theory of intervention should stand on its own merits, and be based in empirical evidence of actual effectiveness. Stepping away from quantitative performance evaluations in the form of SAT scores for example, as in the case of University of California, seems like a bad idea if we want to accurately evaluate and/or promote learning. Ideas that center the teacher and their methods as the primary determinant of learning outcomes, instead of the students and the study material, also run the risk of putting an unmanageable load on the shoulders of educators. There is even recent evidence from a meta-analysis of teacher evaluations that suggest that teacher quality does not seem to play a significant role in producing good learning outcomes, at least in university courses (Uttl, White, & Gonzalez, 2017). They found that the correlation between student's evaluations of their teacher (SET) and learning explain at best 10% of the variance in learning outcome, and that previous meta-analyses showed inflated effects due to small sample sizes and publication bias. SET scores are also further biased by the Dunning-Krueger effect, that is that one's ability to evaluate performance and measure learning is dependent on students' intelligence, ability and prior relevant knowledge (Dunning, 2011). They conclude that when re-analyzing the datasets, after adjusting for sample sizes, the estimated SET/learning correlations in nearly all the reported studies dropped to near zero.

Another recent meta-analysis by Sisk et al. focusing specifically of the effect of interventions on Growth Mindset also landed in strong critical conclusions. They found the correlation of growth mindset with achievement to be only r = .1 and the effect of the intervention to promote a growth mindset on achievement had an effect size of Cohen's d = .08 (Sisk et al., 2018).

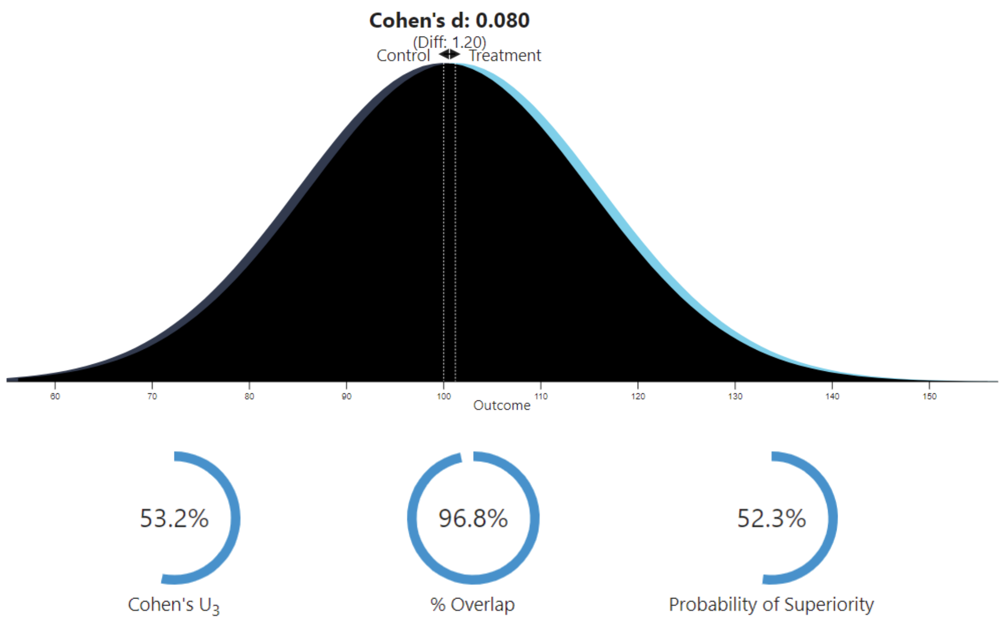


Fig. 4.1. Overlapping normal distributions showing the difference in means between treatment and control group produced by an effect size of d=.08. "Cohen's U_3*": % of treatment group above the mean of control group. "%*

Overlap": area covered by both groups. Probability of Superiority: Odds of a randomly selected subject from treatment group scoring higher than a randomly selected subject from the control group. "Interpreting Cohen's d Effect Size: An Interactive Visualization", (https://rpsychologist.com/cohend/), by Kristoffer Magnusson.

The authors summarize their findings by opposing the claims of previous researchers claiming such interventions can lead to "*large gains in student achievement*" and "*striking effects on educational achievement*" (Yeager & Walton, 2011). They conclude that the interventions were "*non-significant for adolescents, typical students, and students facing situational challenges (transitioning to a new school, experiencing stereotype threat)*" They did however state that there might be weak support for claims that these types of interventions could be of benefit to the students with high risk of academic failure and/or economically disadvantaged backgrounds. They advise caution however, citing a low number of effect sizes contributing to this finding, the small insignificant difference between high- and non-risk students, and the smaller sample sizes of low SES students as compared to the other groups. Interestingly, Carol Dweck herself has recently responded to criticism and even levied her own concerns with the directions her field has taken in an interview with Jon Severs for tes.com in April of 2020: (https://www.tes.com/news/growth-mindset-where-did-it-go-wrong). Regarding failures to replicate her findings she responded that "*Growth mindset is even more complex than we imagined*," adding that the major issue with her theory in her own opinion is the problem of how to correctly implement it into an effective practice. Dweck even adds "*We, as* *originators of the concept, believe it could work under certain circumstances in the classroom, with the correct implementation, and we want to make that journey. We have not completed that journey, but we are on it and we have decided to try and work it out with the educators*.”, indicating that much of the work towards effective implementation strategies still lies before us. In our own study, we found that the self-rated growth mindset of our sample of students did not correlate with any other variable in our experiment, and we were unable to find any significant relationship between growth mindset and our outcome variables measuring performance on our learning task. We must however be mindful of the strong homogeneity of our sample, consisting of self-enrolled, motivated students from a school with high grade requirements for acceptance, as discussed in chapter 2. Taken together, it seems reasonable to believe that if we are to find any meaningful application of the theory of mindset set forth by Dweck, it would be among low performance students (for whatever reason), since if they are already performing well (by any metric) there is, by definition, a smaller window of possible improvement to be gained in the first place. It is never a waste of time to thoroughly investigate promising new tools and theories. If they are found to have small effects, potentially only in certain populations, that at least points the way towards more targeted and potentially more refined niche applications with possibly greater effects.

Another intellectual trend is the notion of ‘Antifragility’ Put forward in the 2012 book “Antifragile: Things That Gain From Disorder” (Taleb, 2012), and later incorporated into a multitude of applications like physics (Naji, Ghodrat, Komaie-Moghaddam, & Podgornik, 2014), molecular biology (Danchin, Binder, & Noria, 2011), risk analysis (Aven, 2015), urban- and transport planning (Levin, Brodfuehrer, & Kroshl, 2014), engineering (Verhulsta, 2014) and computer science (Jones, 2014), but for our purposes the 2018 book "The Coddling of the American Mind: How Good Intentions and Bad Ideas Are Setting Up a Generation for Failure" (Haidt & Lukianoff, 2018) is the most relevant to the field of education. To summarize, the concept of antifragility is that the opposite of something being fragile is not that it does not break when stressed, but that it ends up stronger than before it encountered the obstacle in the first place. Two examples cited of this are bones healing thicker after a fracture and the immune system encountering a new pathogen and learning to fight it off. Heidt and Lukianoff argue that children are particularly antifragile in more ways than one, and that the right kinds of stressors are vital for producing well-functioning, psychologically stable adults. This extends to being able to handle setbacks and how to not let failure define you, echoing the ideas of introspection with regards to personal performance in the growth mindset theory. The ideas put forward in "The Coddling of the American Mind" is essentially an application of strategies from the field of Cognitive Behavioral Therapy, where the goal is to train individuals in how to conceptualize and contextualize situations and emotional reactions in such a way as to not be a hindrance in achieving whatever goals you have set out for yourself beforehand, including learning objectives in an educational setting (Beck & Beck, 1995; Schacter, Gilbert, Wegner, & Hood, 2011). Once again it seems probable that high-functioning students already apply something approaching these methods already, either consciously or by disposition, so once again one could expect any potential gains to be had primarily in low-performing populations.

Since the questions in a Dweck mindset instrument are straight to the point, asking things like if you believe that intelligence is something that you can improve and whether talented people need to have practiced a lot, it stands to reason that absent any empirical knowledge in this area the average person would have to base their answers on their own experiences and ideas that would be influenced by their intrinsic personality traits. A person high in trait neuroticism would be expected to give more pessimistic answers in general than a person low in neuroticism, rating their own ability to improve lower than it probably is (Costa Jr & McCrae, 2008). A different, but equally interesting, measure of attitudes towards learning is the Need For Cognition scale (NFC). NFC has been found to correlate positively with fluid intelligence (Fleischhauer et al., 2010), and although NFC and trait openness were found to be strongly related (Fleischhauer et al., 2010; Pacini & Epstein, 1999; Sadowski & Cogburn, 1997), they are not entirely overlapping constructs, as their correlation patterns to other personality traits differ; NFC was found to be more strongly correlated with emotional stability (low trait neuroticism) than openness was, and novelty and experience seeking was less correlated with NFC than with openness (Fleischhauer et al., 2010). NFC can thus be said to measure something distinct from the big five personality traits. To our knowledge there does not exist any research attempting any type of intervention to modify people’s attitudes as measured by the NFC scale. Why this has never been attempted when there is such an abundance of research into other seemingly closely related aspects of mindset and personality is unclear. Perhaps it is simply based on what gets conceptualized as a potentially flexible mindset versus a more fixed personality trait early on in any field of study and simply never reflected on again. We found NFC to significantly contribute to programming learning in our SEM (table 2.6). NFC also correlated with LTM, confidence, written test, fMRI test and lower stress (table 2.5), and it was our best predictor of LTM in our regression modeling (table 2.8).

4.3. Wider Implications & Future Directions

The very much still nascent field of educational neuroscience presents a wide array of promising avenues to pursue, as well as a host of questions and limitations to address for it to progress into a useful and broadly accepted avenue of inquiry. The application of functional magnetic resonance imaging to the study of real-time activity in the human brain only started in the 1990s with the first studies of activation in the visual cortex (Kwong et al., 1992). The field of brain imaging, and related research by cognitive neuroscientists and psychologists has therefore only been around for some 30 years, as opposed to the philosophical traditions of say the field of education that stretches back hundreds if not thousands of years. Modern neuroscience is arguably still in its infancy, with the fundamental operation of pretty much all neural processes still largely a mystery, and educational neuroscience barely even born yet. To begin to answer the question of what brain imaging can bring to a field such as education let us first turn to a point of philosophy. If we accept that any and all functioning of our mind is the emergent properties of our brain, (and we would argue that all available evidence indicates that this is indeed the case) and if we believe in the predictability of physical laws of cause and effect, then it follows that if we had the ability to understand how the brain does what it does, that is a full mechanistic explanation for how any given population of neurons give rise to the cognitive processes (and subjective experiences of what it is like to experience their input if they are conscious), we would be able to know what the limitations and possibilities of learning is, on average, or for any given individual, and consequently what would give us the most workable interventions. Our goal should be to strive towards this complete theory of mind. I would argue that any sufficiently accurate model of how our brain interacts with the world could never lead to adverse outcomes, only equal to less informed but nonetheless effective practices or better, and I see no credible case to exclude imaging or any other future methodology from contributing to education.

Since the 2000:s, a great many papers have delved into what the future prospects for education might be, what questions they hope to see addressed and what problems they foresee (Ansari & Coch, 2006; Ansari et al., 2011; Geake & Cooper, 2003; Goswami, 2006; Howard-Jones, 2010; P. Howard-Jones, 2014; P. A. Howard-Jones, 2014; Immordino-Yang & Damasio, 2007; Sousa, 2010; Tommerdahl, 2010; Varma, McCandliss, & Schwartz, 2008). The consensus that emerges is that though there have been growing calls for collaborations between cognitive neuroscience and education and positive expectations overall, the expected scope and immediacy of findings that can directly be implemented into practice is unrealistic and short-sighted. Instead, authors like Ansari argue that what is needed first and foremost is the implementation of an 'infrastructure' of collaborations, training, funding, and research program creation of an interdisciplinary character. This merging of the two fields is what they call *"the emerging field of 'Mind, Brain and Education'"*.

When it comes to the inherent methodological limitations of the current state of brain imaging the biggest to date is arguably the tradeoff between spatial resolution and temporal resolution. There are several ways to acquire fMRI images of the brain, but all of them must make sacrifices based on what the researcher chooses to focus on. If we know of a restricted area of the brain to image, we can leverage the smaller field of view to instead increase the spatial resolution (up to the limits set by the magnetic field strength of the scanner) meaning smaller voxel sizes. Thus, one direction of improvement is moving to scanners with higher field strengths (Ladd et al., 2018). The drawback of this being the size and cost of the machines going up as a result. As the number of voxels sampled goes up the repetition time (TR) between each full volume sampled goes up, meaning we lose temporal resolution. The fact that these constraints make themselves more felt as the size of the captured volume increases means that once whole-brain exploratory experiments have been sufficiently evaluated, researchers can (depending on their hypotheses) mitigate them to some extent by focusing in on only the previously reported areas.

Another concern with imaging, especially when it comes to fMRI, is ecological validity in the laboratory setting. The recent advances in virtual reality (VR) technology have been argued to be a potentially powerful tool to both enhance the ecological validity of experiments and enable entirely new types of psychological experiments that would be physically and/or ethically impossible without this technology (Reggente et al., 2018). Other technologies like functional near-infrared spectroscopy (FNIRS) and portable versions of magnetoencephalography (MEG) are also constantly developed to further combat the physical constraints imposed by traditional MEG and MRI machines. If we accept the claim that doing science in a naturalistic setting (most commonly in a classroom in the case of education), has the advantage of possibly revealing up until then previously unthought of questions, possibilities and perspectives that can lead to the formation of new scientific hypotheses or amendments to the ideas that the researchers went into the classroom to investigate initially, we then have a few different angles to consider: First, it is the norm rather than the exception in any science that more questions are generally generated than answers whatever methodology you apply. One should also consider how excited we ought to be at the prospect of discovering unexpected big fundamental challenges to our established knowledgebase, given that education is such an old and well-studied field. In other words: if any groundbreaking issue were to be discovered principally because of the nature of classroom studies, it could be viewed as an indictment of the shortcomings of previously established theory. Another aspect to consider when advocating for ecological validity is the possibility that any significant finding could be particular to that setting or combination of setting and student population, essentially a form of sampling- or selection bias, so in order to account for any such criticism the prudent researcher would have to perform the same interventions or information gathering from a large, varied sampling, thus approaching a situation not radically different from a controlled experiment anyway. Thus, the major distinction between ecological validity and an experimental setting can be viewed as a matter of how many or few variables we want to control for. From an experimental point of view this is essentially indistinguishable from allowing uncontrolled variance into our other variables, thus the tradeoff is in effect reduced confidence in the estimates of the effect sizes of whatever else we are attempting to measure in the students. If we were to discover some previously unknown variable that seems potentially relevant to our area of inquiry, we would still have to design a controlled experiment to isolate that effect later if we did not have enough information in our current study to accurately gauge the relative importance of the new find. A solid counterargument is that the complex relationships that emerge out of the social interaction of different people, including those between the teacher and the student is not easily captured in a controlled laboratory setting. The individual expectations and social pressures alone are clearly wildly different. One solution to this issue might be that if we accept that controlled experiments will eventually have to be performed anyway if we want to provide compelling statistical evidence for our results, then we should strive for efficiency and transferability of results at the same time, by putting effort into designing the type of matching, interdisciplinary experiments that we have attempted to do in this project, allowing us to test the outcomes both in a classroom setting and under laboratory conditions at the same time. The idea being that if we have successfully designed both experiments to be as similar as possible and to maximally account for all relevant variables known to us at the time, then any disparities between the two setting would automatically give us a hint that something might have been lost in translation when moving away from the social setting, but if we find that the results do line up, we can be fairly confident in concluding that the psychosocial factors linked to the setting does not contribute any statistically significant information to our studied phenomena. If we accept this then theoretically, given enough information, we should in the future be able to device a purely experimental design accounting for enough variables to, effectively, rescue the ecological validity that we lost when moving into the lab. On this view, ecological validity is a function of predictive validity, given enough theoretical knowledge. In practice, this will probably not be feasible across the board and, depending on what we want to study, there are surely more effective ways to accomplish some things in a naturalistic setting. When it comes to brain imaging, it should also be mentioned that the association patterns laid down in the brain after learning something has been shown to potentially activate in whatever circumstance a relevant stimulus appears later, meaning that there is evidence indicating that at least some of the insights gained from brain imaging are generalizable across settings. Perhaps the most interesting example is the study that found the so-called “*Halle berry neurons*”; A group of neurons around the hippocampus that responded both to images of Halle Berry, but also to her name in text on a screen (Quiroga, Reddy, Kreiman, Koch, & Fried, 2005). It should be noted however that these neurons might be more general and would also have responded to other faces or names sharing some similarity to Halle Berry.

While fMRI studies of the brain have some 30 years behind them and are still considered a relatively new domain of inquiry, an even younger domain with the potential to further complicate the fields of cognitive neuroscience and psychology is genetics, or more specifically, large-scale quantitative analyses of the genetic impact of the combined small effects of thousands of individual points of variation in our genome called single nucleotide polymorphisms or ‘SNP:s’ (Brookes, 1999) that have gotten steadily more reliable as access to bigger and better datasets of genetics coupled to outcome variables continue to grow due to the reduction in cost of both whole genome sequencing over time (Sboner, Mu, Greenbaum, Auerbach, & Gerstein, 2011), and the recent proliferation of consumer-targeted DNA testing companies like 23andme and Ancestry.com (Even if the rising complexity and therefore computational demands of this work offsets the overall cost-savings on sequencing somewhat) (Allyse, Robinson, Ferber, & Sharp, 2018). A recent indication of what we might expect to see more of in the future from this field is the paper “Dissecting polygenic signals from genome-wide association studies on human behaviour” (Abdellaoui & Verweij, 2021). Genome-wide association studies (GWAS) is a way to leverage large data sets of genomics data to statistically model the combined effects of all SNP:s in any one subject on variables of interest, deriving so-called polygenic scores for each variable. Essentially, calculating a probabilistic risk for expressing any given outcome that has some amount of explainable genetic etiology (Dudbridge, 2013). In the paper by Abdellaoui and Verweij they review the field of calculating polygenic scores and compare the calculated genetic contributions to a multitude of material, psychological and social outcomes with the gold standard of estimating heritability: twin studies (Knopik, Neiderhiser, DeFries, & Plomin, 2017). They find that as the statistics gets more refined, and more high-quality data is collated, the heritability estimates for each trait approaches the values derived from twin-studies even though they struggle to reach 50% at present. The heritability estimates from twin-studies for most of the reported traits are in the 30-50% range with a few notable outliers, namely psychological disorders like ADHD, Autism, Bipolar disorder, and schizophrenia around 75% heritability and adult IQ approaching 100%. It would be surprising if the GWAS and twin-studies numbers would not continue to converge, and as more and more traits are investigated, and more and more genetic data is gathered we inch closer and closer to enable the long-standing goal of 'personalized medicine' (Hamburg & Collins, 2010), it seems reasonable to think that it is also possible to achieve the equivalent 'personalized education' as a part of your own 'personal psychology'.

We have already discussed the type of studies we would like to see as a follow up to our experiments in chapters 2 and 3, but I should also incorporate what has been discussed above in this chapter and summarize my concluding visions of the future. I would like to see longitudinal studies of computer programming learning to see if the connectivity-changes we observed after learning hands-on are robust or only a short-term phenomenon. This would also allow us to test if constructs like NFC or Dweck's mindset have effects over long timespans or not. If possible, multimodal imaging and future advances in the technology could also allow for more specific results, including microstructure analyses of cortical layers and similar internal layout of other brain regions in accordance with the advances in the study of learning brain plasticity as discussed in "Plasticity in Gray and White" (Zatorre, Fields, & Johansen-Berg, 2012). Since we also found that language processing is central to the task of programming, I would like to see further research into what lessons can be learned from the second language learning (L2) field, even though we did not detect any significant benefit of previous language aptitude in the form of language grades in our experiment. There is already some support for this hypothesis: A recent paper titled "Relating Natural Language Aptitude to Individual Differences in Learning Programming Languages" found that language aptitude (measured by the Modern Language Aptitude Test (MLAT)) was "*a robust predictor of all of the Python learning outcomes. Specifically, learning rate*" (Prat, Madhyastha, Mottarella, & Kuo, 2020). Their findings indicate that although programming accuracy is not very explainable by the MLAT (explaining less than 10% of the variance), learning rate is (at around 40% variance explained). They found that general cognition (Fluid reasoning, working memory updating, working memory span, and inhibitory control) accounted for the biggest chunk of programming accuracy, explaining a little under 60%, which might explain why we did not see any effect of language aptitude, as we were evaluating our participant's program accuracy, and they all had to display similar enough learning rates to successfully learn during our limited lesson time. The MLAT measurement used in their study has been showed to correlate with IQ (Cox, Lynch, Mendes, & Zhai, 2019). They summarize previous work by saying "*In sum, most evidence has suggested a fairly strong positive link between nonverbal IQ and aptitude (Biedroń & Szczepaniak,* *2012; Granena & Long, 2013; Sternberg, Ehrman, & Grigorenko, 2000; Xiang et al., 2012)*". In their experiment, which is an investigation into Spanish–English bilinguals in the United States, they find positive relationships between language proficiency and nonverbal IQ measures, specifically with the sub-components of aptitude for grammatical inferencing and sound–symbol association measures. They also say "*Moreover, the same aptitude components also varied with nonverbal IQ, demonstrating that nonverbal abilities may also play a role in aptitude in uninstructed bilinguals*". Another study also found an aptitude/IQ correlation of r= .5, perhaps driven by combined verbal/nonverbal IQ measures in the studies he summarized (Li, 2016). These findings indicate that the distinction between language aptitude and general cognition in the paper by Prat et al. is perhaps not as clear cut as presented, which means that future research into the links between L2 learning and programming should also ideally take care to thoroughly map out the role of general cognitive performance.

4.4. Conclusions

To conclude this thesis, let me first start with a point regarding our findings concerning the importance of confidence as a mediator of future programming success: In our experimental setup we are unable to distinguish whether confidence is the proverbial chicken or the egg. What we mean by this is that either working hands-on leads to higher confidence through some process, and that confidence produces the connectivity changes we detected in the brain, ultimately resulting in increased performance, or working hands-on lead to the changes we observed, and the strength of that learned ability is then evaluated introspectively as a high confidence by the subjects. From a statistical point of view, the relationship is the same, but we would like to see an attempt to clarify the causal chain in the future. I would also like to see more research into individual differences between students, and how they interact with the course material or study techniques. Interventions for traits other than mindset, like NFC should also be investigated and language aptitude should be studied further, particularly when it comes to making sure that students don't fall behind in the coursework. Finally, our take home messages for educators: We recommend hands-on learning in computer programming and encourage future research on this front. Building confidence in the students is important, and we found links between confidence and attention and motivational aspects that warrant future investigation. NFC is a promising personality trait for future studies and, since programming is processed like a language, look to L2 learning research for possible inspiration for pedagogy.

References

Abdellaoui, A., & Verweij, K. J. (2021). Dissecting polygenic signals from genome-wide association studies on human behaviour. *Nature Human Behaviour*, 1-9.

Ahadi, A., & Lister, R. (2013, 2013). *Geek genes, prior knowledge, stumbling points and learning edge momentum: parts of the one elephant?*

Allyse, M. A., Robinson, D. H., Ferber, M. J., & Sharp, R. R. (2018). *Direct-to-consumer testing 2.0: emerging models of direct-to-consumer genetic testing.* Paper presented at the Mayo clinic proceedings.

Ansari, D., & Coch, D. (2006). Bridges over troubled waters: Education and cognitive neuroscience. *Trends in cognitive sciences, 10*(4), 146-151.

Ansari, D., Coch, D., & De Smedt, B. (2011). Connecting Education and Cognitive Neuroscience: Where will the journey take us? *Educational Philosophy and Theory, 43*(1), 37-42. doi:https://doi.org/10.1111/j.1469-5812.2010.00705.x

Appelgren, A., & Bengtsson, S. L. (2015). Feedback on Trait or Action Impacts on Caudate and Paracingulum Activity. *PloS one, 10*(6), e0129714.

Aven, T. (2015). The concept of antifragility and its implications for the practice of risk analysis. *Risk analysis, 35*(3), 476-483.

Bäckman, L., Nilsson, L.-G., & Chalom, D. (1986). New evidence on the nature of the encoding of action events. *Memory & Cognition, 14*(4), 339-346.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and brain sciences, 22*(4), 577-660.

Bartlett, F. C., & Bartlett, F. C. (1995). *Remembering: A study in experimental and social psychology*: Cambridge University Press.

Bates, R., & Khasawneh, S. (2007). Self-efficacy and college students' perceptions and use of online learning systems. *Computers in Human Behavior, 23*(1), 175-191.

Beck, J. S., & Beck, A. T. (1995). *Cognitive therapy: Basics and beyond*: Guilford press New York.

Bengtsson, S. L., Haynes, J.-D., Sakai, K., Buckley, M. J., & Passingham, R. E. (2009). The representation of abstract task rules in the human prefrontal cortex. *Cerebral Cortex, 19*(8), 1929-1936.

Bengtsson, S. L., & Penny, W. D. (2013). Self-associations influence task-performance through Bayesian inference. *Frontiers in Human Neuroscience, 7*, 490.

Bennedsen, J., & Caspersen, M. E. (2005, 2005). *An investigation of potential success factors for an introductory model-driven programming course*.

Berglund, A., & Eckerdal, A. (2015, 2015). *Learning Practice and Theory in Programming Education: Students' Lived Experience*.

Biedroń, A., & Szczepaniak, A. (2012). Working memory and short-term memory abilities in accomplished multilinguals. *The Modern Language Journal, 96*(2), 290-306.

Black, P. (2006). Assessment for learning: where is it now? Where is it going. *Improving student learning through assessment*, 9-20.

Black, P., & Wiliam, D. (2010). Inside the black box: Raising standards through classroom assessment. *Phi delta kappan, 92*(1), 81-90.

Blank, S. C., Scott, S. K., Murphy, K., Warburton, E., & Wise, R. J. (2002). Speech production: Wernicke, Broca and beyond. *Brain, 125*(8), 1829-1838.

Broadbent, D. E., & Ladefoged, P. (1959). Auditory perception of temporal order. *The Journal of the Acoustical Society of America, 31*(11), 1539-1539.

Brookes, A. J. (1999). The essence of SNPs. *Gene, 234*(2), 177-186.

Bundesen, C. (1990). A theory of visual attention. *Psychological review, 97*(4), 523.

Byrne, P., & Lyons, G. (2001, 2001). *The effect of student attributes on success in programming*.

Cacioppo, J. T., Petty, R. E., & Feng Kao, C. (1984). The efficient assessment of need for cognition. *Journal of personality assessment, 48*(3), 306-307.

Choi-man, C. (1988). Correlates of problem-solving in programming. *Chinese University of Hong Kong Educational Journal, 16*(2), 185-190.

Christoffels, I. K., Formisano, E., & Schiller, N. O. (2007). Neural correlates of verbal feedback processing: an fMRI study employing overt speech. *Human brain mapping, 28*(9), 868-879.

Cohen, A. R., Stotland, E., & Wolfe, D. M. (1955). An experimental investigation of need for cognition. *The Journal of Abnormal and Social Psychology, 51*(2), 291.

Comenius, J.-A. (1642). *Opera didactica omnia*: Laurentius de Geer.

Commission, E. (2014). The International Computer and Information Literacy Study (ICILS), Main findings and implications for education policies in Europe: European Commision.

Cortese, A., Amano, K., Koizumi, A., Kawato, M., & Lau, H. (2016). Multivoxel neurofeedback selectively modulates confidence without changing perceptual performance. *Nature communications, 7*(1), 1-18.

Costa Jr, P. T., & McCrae, R. R. (2008). *The Revised Neo Personality Inventory (neo-pi-r)*: Sage Publications, Inc.

Cox, J. G., Lynch, J. M., Mendes, N., & Zhai, C. (2019). On bilingual aptitude for learning new languages: The roles of linguistic and nonlinguistic individual differences. *Language Learning, 69*(2), 478-514.

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of verbal learning and verbal behavior, 11*(6), 671-684.

Curricula, J. T. F. o. C. (2013). *Computer Science Curricula 2013*: ACM/Association for Computing Machinery.

D'Esposito, M., Postle, B. R., & Rypma, B. (2000). Prefrontal cortical contributions to working memory: evidence from event-related fMRI studies. *Executive control and the frontal lobe: Current issues,* 3-11.

Danchin, A., Binder, P. M., & Noria, S. (2011). Antifragility and tinkering in biology (and in business) flexibility provides an efficient epigenetic way to manage risk. *Genes, 2*(4), 998-1016.

De Kloet, E. R., Oitzl, M. S., & Joëls, M. (1999). Stress and cognition: are corticosteroids good or bad guys? *Trends in neurosciences, 22*(10), 422-426.

Decety, J., Jeannerod, M., & Prablanc, C. (1989). The timing of mentally represented actions. *Behavioural brain research, 34*(1-2), 35-42.

Deci, E. L., & Ryan, R. M. (2013). *Intrinsic motivation and self-determination in human behavior*: Springer Science & Business Media.

Dembo, M. H., Junge, L., & Lynch, R. (2006). Becoming a self-regulated learner: Implications for web-based education. *Web-based learning: Theory, research, and practice*, 185-202.

Dewey, J. (1897). *My pedagogic creed*: EL Kellogg & Company.
Dudbridge, F. (2013). Power and predictive accuracy of polygenic risk scores. *PLoS genetics*, *9*(3), e1003348.
Dunning, D. (2011). The Dunning–Kruger effect: On being ignorant of one's own ignorance. *Advances in experimental social psychology* (Vol. 44, pp. 247-296): Elsevier.
Duraes, J., Madeira, H., Castelhano, J., Duarte, C., & Branco, M. C. (2016, 2016). *WAP:* Understanding the brain at software debugging.
Dweck, C. (2008a). Brainology: Transforming students' motivation to learn. *Independent school, 67*(2), 110-119.
Dweck, C. (2008b). *Mindset: The new psychology of success*: Random House Digital, Inc.
Dweck, C. (2010). Even geniuses work hard. *Educational Leadership, 68*(1), 16-20.
Eckerdal, A., McCartney, R., Moström, J. E., Sanders, K., Thomas, L., & Zander, C. (2007, 2007). From Limen to Lumen: computing students in liminal spaces.
Eckerdal, A., Thuné, M., & Berglund, A. (2005, 2005). What does it take to learn'programming thinking'?
Education, D. f. (2014). National curriculum in England: computing programmes of study. *GOV.UK*.
Eickhoff, S. B., Paus, T., Caspers, S., Grosbras, M.-H., Evans, A. C., Zilles, K., & Amunts, K. (2007). Assignment of functional activations to probabilistic cytoarchitectonic areas revisited. *Neuroimage, 36*(3), 511-521.
Engelkamp, J., & Zimmer, H. D. (1985). Motor programs and their relation to semantic memory. *German Journal of psychology*.
Erdogan, Y., Aydin, E., & Kabaca, T. (2008). Exploring the Psychological Predictors of Programming Achievement. *Journal of Instructional Psychology, 35*(3).
Feldman, D. C. (2001). Career coaching: What HR professionals and managers need to know. *Human Resource Planning, 24*(2).
Fincher, S., & Robins, A. (2019). *The Cambridge handbook of computing education research*: Cambridge University Press.
Fincher, S., Robins, A., Baker, B., Box, I., Cutts, Q., de Raadt, M., . . . Lister, R. (2006, 2006). Predictors of success in a first programming course.
Fleischhauer, M., Enge, S., Brocke, B., Ullrich, J., Strobel, A., & Strobel, A. (2010). Same or different? Clarifying the relationship of need for cognition to personality and intelligence. *Personality and Social Psychology Bulletin, 36*(1), 82-96.
Floyd, B., Santander, T., & Weimer, W. (2017, 2017). *Decoding the representation of code in the brain: An fMRI study of code review and expertise*.
Folk, M. J. (1973). Influences of Developmental Level on a Child's Ability to Learn Concepts of Computer Programming. *Dissertation Abstracts International, 34(3)*(1125a).
Ford, A., McGregor, K. M., Case, K., Crosson, B., & White, K. D. (2010). STRUCTURAL CONNECTIVITY OF BROCA'S AREA AND MEDIAL FRONTAL CORTEX. *Neuroimage, 52*(4), 1230-1237. doi:10.1016/j.neuroimage.2010.05.018

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of educational research, 74*(1), 59-109.

Friston, K. (2005). A theory of cortical responses. *Philosophical transactions of the Royal Society B: Biological sciences, 360*(1456), 815-836.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain, 119*(2), 593-609.

Gasquoine, P. G. (2014). Contributions of the insula to cognition and emotion. *Neuropsychology review, 24*(2), 77-87.

Geake, J., & Cooper, P. (2003). Cognitive Neuroscience: implications for education? *Westminster Studies in Education, 26*(1), 7-20.

Gentsch, A., Weber, A., Synofzik, M., Vosgerau, G., & Schütz-Bosbach, S. (2016). Towards a common framework of grounded action cognition: Relating motor control, perception and cognition. *Cognition, 146*, 81-89.

Good, L. A., Feekes, J., & Shawd, B. (1993). Let your fingers do the talking: Hands-on language learning through signing. *Childhood Education, 70*(2), 81-83.

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in neurosciences, 15*(1), 20-25.

Goswami, U. (2006). Neuroscience and education: from research to practice? *Nature reviews neuroscience, 7*(5), 406-413.

Grabner, R. H., Ansari, D., Reishofer, G., Stern, E., Ebner, F., & Neuper, C. (2007). Individual differences in mathematical competence predict parietal brain activation during mental calculation. *Neuroimage, 38*(2), 346-356.

Granena, G., & Long, M. (2013). *Sensitive periods, language aptitude, and ultimate L2 attainment* (Vol. 35): John Benjamins Publishing.

Grush, R. (2004). The emulation theory of representation: Motor control, imagery, and perception. *Behavioral and brain sciences, 27*(3), 377-396.

Haidt, J., & Lukianoff, G. (2018). *The coddling of the American mind: How good intentions and bad ideas are setting up a generation for failure*: Penguin UK.

Hamburg, M. A., & Collins, F. S. (2010). The path to personalized medicine. *New England Journal of Medicine, 363*(4), 301-304.

Handur, V., Kalwad, P. D., Patil, M. S., Garagad, V. G., Yeligar, N., Pattar, P., . . . Joshi, G. (2016). *Integrating Class and Laboratory with Hands-on Programming: Its Benefits and Challenges.* Paper presented at the 2016 IEEE 4th International Conference on MOOCs, Innovation and Technology in Education (MITE).

Holvikivi, J. (2010). Conditions for successful learning of programming skills *Key competencies in the knowledge society* (pp. 155-164): Springer.

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding (TEC): A framework for perception and action planning. *Behavioral and brain sciences, 24*(5), 849-878.

Höök, L. J., & Eckerdal, A. (2015, 2015). On the bimodality in an introductory programming course: An analysis of student performance factors.

Howard-Jones, P. (2010). *Introducing neuroeducational research: Neuroscience, education and the brain from contexts to practice*: Taylor & Francis.

Howard-Jones, P. (2014). Neuroscience and education: A review of educational interventions and approaches informed by neuroscience. *Education Endowment Foundation, Millbank, UK*.

Howard-Jones, P. A. (2014). Neuroscience and education: myths and messages. *Nature reviews neuroscience, 15*(12), 817-824.

Hu, S., & Kuh, G. D. (2002). Being (dis) engaged in educationally purposeful activities: The influences of student and institutional characteristics. *Research in higher education, 43*(5), 555-575.

Hüsing, T., B. Korte, W., Fonstad, N., Lanvin, B., Cattaneo, G., Kolding, M., . . . van Welsum, D. (2013). e-Leadership: e-Skills for Competitiveness and Innovation Vision, Roadmap and Foresight Scenarios, Final Report: European Commision.

Immordino-Yang, M. H., & Damasio, A. (2007). We feel, therefore we learn: The relevance of affective and social neuroscience to education. *Mind, brain, and education, 1*(1), 3-10.

Johnson-Laird, P. N. (1980). Mental models in cognitive science. *Cognitive science, 4*(1), 71-115.

Jones, K. H. (2014). Engineering antifragile systems: A change in design philosophy. *Procedia computer science, 32*, 870-875.

Kanuka, H. (2005). An exploration into facilitating higher levels of learning in a text-based internet learning environment using diverse instructional strategies. *Journal of Computer-Mediated Communication, 10*(3), JCMC1032.

Katai, Z., & Toth, L. (2010). Technologically and artistically enhanced multi-sensory computer-programming education. *Teaching and Teacher Education, 26*(2), 244-251. doi:10.1016/j.tate.2009.04.012

Kiefer, M., & Barsalou, L. W. (2013). 15 grounding the human conceptual system in perception, action, and internal states. *Action science: Foundations of an emerging discipline*, 381-401.

Kiss, M. (2017). Digital skills in the EU labour market: European Parliament Research Service (EPRS).

Knopik, V. S., Neiderhiser, J. M., DeFries, J. C., & Plomin, R. (2017). *Behavioral genetics*: Worth Publishers, Macmillan Learning.

Kolb, D. A., & Boyatzis, R. E. (1970). Goal-setting and self-directed behavior change. *Human Relations, 23*(5), 439-457.

Kori, K., Pedaste, M., Tõnisson, E., Palts, T., Altin, H., Rantsus, R., . . . Rüütmann, T. (2015, 2015). *First-year dropout in ICT studies*.

Krönke, K.-M., Mueller, K., Friederici, A. D., & Obrig, H. (2013). Learning by doing? The effect of gestures on implicit retrieval of newly acquired words. *Cortex, 49*(9), 2553-2568.

Krueger, R., Huang, Y., Liu, X., Santander, T., Weimer, W., & Leach, K. (2020). *Neurological divide: an fMRI study of prose and code writing.* Paper presented at the 2020 IEEE/ACM 42nd International Conference on Software Engineering (ICSE).

Kwong, K. K., Belliveau, J. W., Chesler, D. A., Goldberg, I. E., Weisskoff, R. M., Poncelet, B. P., . . . Turner, R. (1992). Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proceedings of the National Academy of Sciences, 89*(12), 5675-5679.

Ladd, M. E., Bachert, P., Meyerspeer, M., Moser, E., Nagel, A. M., Norris, D. G., . . . Zaiss, M. (2018). Pros and cons of ultra-high-field MRI/MRS for human application. *Progress in Nuclear Magnetic Resonance Spectroscopy, 109*, 1-50. doi:https://doi.org/10.1016/j.pnmrs.2018.06.001

Larsson, L. (2007). Recension av Johan Amos Comenius: Orbis Sensualium Pictus. Den synliga världen. Första upplagan på svenska och latin 1682 i faksimil med nyöversättning, inledning, tre efterord och bibliografi. Red. Lars Lindström. *LexicoNordica*, *14*, 209–223.

Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human perception and performance, 21*(3), 451.

Lee, S., Matteson, A., Hooshyar, D., Kim, S., Jung, J., Nam, G., & Lim, H. (2016, 2016). *Comparing Programming Language Comprehension between Novice and Expert Programmers Using EEG Analysis*.

Leitan, N. D., & Chaffey, L. (2014). Embodied cognition and its applications: A brief review. *Sensoria: A Journal of Mind, Brain & Culture, 10*(1), 3-10.

Levin, J. S., Brodfuehrer, S. P., & Kroshl, W. M. (2014). *Detecting antifragile decisions and models lessons from a conceptual analysis model of service life extension of aging vehicles.* Paper presented at the 2014 IEEE International Systems Conference Proceedings.

Li, S. (2016). The construct validity of language aptitude: A meta-analysis. *Studies in second language acquisition, 38*(4), 801-842.

Lins de Holanda Coelho, G., HP Hanel, P., & J. Wolf, L. (2020). The very efficient assessment of need for cognition: Developing a six-item version. *Assessment, 27*(8), 1870-1885.

Louden, K. C. (1993). Programming Languages Principles and Practice. PWS: Kent Publishing Company.

Luxton-Reilly, A., Albluwi, I., Becker, B. A., Giannakos, M., Kumar, A. N., Ott, L., . . . Szabo, C. (2018). *Introductory programming: a systematic literature review*. Paper presented at the Proceedings Companion of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education.

Macedonia, M., & von Kriegstein, K. (2012). Gestures enhance foreign language learning. *Biolinguistics, 6*(3-4), 393-416.

Mangen, A., Anda, L. G., Oxborough, G. H., & Brønnick, K. (2015). Handwriting versus Keyboard Writing: Effect on Word Recall. *Journal of Writing Research, 7*(2).

Mannila, L., & de Raadt, M. (2006). *An objective comparison of languages for teaching introductory programming.* Paper presented at the Proceedings of the 6th Baltic Sea conference on Computing education research: Koli Calling 2006.

Mayer, K. M., Yildiz, I. B., Macedonia, M., & von Kriegstein, K. (2015). Visual and motor cortices differentially support the translation of foreign language words. *Current Biology*, *25*(4), 530-535.

McDowell, L., Sambell, K., & Davison, G. (2009). Assessment for learning: a brief history and review of terminology.

McIntosh, R. D., & Schenk, T. (2009). Two visual streams for perception and action: current trends.

Mesulam, M.-M., Thompson, C. K., Weintraub, S., & Rogalski, E. J. (2015). The Wernicke conundrum and the anatomy of language comprehension in primary progressive aphasia. *Brain, 138*(8), 2423-2437.

Miall, R. C., & Wolpert, D. M. (1996). Forward models for physiological motor control. *Neural networks, 9*(8), 1265-1279.

Milner, A., & Goodale, M. (1995). The Visual Brain in Action. *Oxford Psychology Series, 27*.

Mishkin, M., & Ungerleider, L. G. (1982). Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. *Behavioural brain research, 6*(1), 57-77.

Mitchell, B. L., Cuéllar-Partida, G., Grasby, K. L., Campos, A. I., Strike, L. T., Hwang, L.-D., . . . Martin, N. G. (2020). Educational attainment polygenic scores are associated with cortical total surface area and regions important for language and memory. *Neuroimage, 212*, 116691.

Moore, K. A., & Lippman, L. H. (2006). *What do children need to flourish?: Conceptualizing and measuring indicators of positive development* (Vol. 3): Springer Science & Business Media.

Mueller, P. A., & Oppenheimer, D. M. (2014). The pen is mightier than the keyboard: Advantages of longhand over laptop note taking. *Psychological science, 25*(6), 1159-1168.

Naji, A., Ghodrat, M., Komaie-Moghaddam, H., & Podgornik, R. (2014). Asymmetric Coulomb fluids at randomly charged dielectric interfaces: Anti-fragility, overcharging and charge inversion. *The Journal of chemical physics, 141*(17), 174704.

Nilsson, L.-G. (2000). Remembering actions and words. *The Oxford handbook of memory*, 137-148.

Nirosh. (2015). Introduction to Object Oriented Programming Concepts (OOP) and More. Retrieved from https://www.codeproject.com/Articles/22769/Introduction-to-Object-Oriented-Programming-Concep

Nyberg, L. (2002). Levels of processing: A view from functional brain imaging. *Memory, 10*(5-6), 345-348.

Oaksford, M., & Chater, N. (2001). The probabilistic approach to human reasoning. *Trends in cognitive sciences, 5*(8), 349-357.

Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*: Oxford University Press.

Pacini, R., & Epstein, S. (1999). The relation of rational and experiential information processing styles to personality, basic beliefs, and the ratio-bias phenomenon. *Journal of personality and social psychology, 76*(6), 972.

Padmala, S., & Pessoa, L. (2010). Interactions between cognition and motivation during response inhibition. *Neuropsychologia, 48*(2), 558-565.

Paris, S. G., & Turner, J. C. (2012). Situated motivation *Student motivation, cognition, and learning* (pp. 229-254): Routledge.

Paris, S. G., Yambor, K. M., & Packard, B. W.-L. (1998). Hands-on biology: A museum-school-university partnership for enhancing students' interest and learning in science. *The Elementary School Journal, 98*(3), 267-288.

Patitsas, E., Berlin, J., Craig, M., & Easterbrook, S. (2016, 2016). *Evidence that computer science grades are not bimodal*.
Paura, L., & Arhipova, I. (2016). Student dropout rate in engineering education study program. *Eng. Rural Dev, 2016*, 641-646.
Peitek, N., Siegmund, J., Apel, S., Kästner, C., Parnin, C., Bethmann, A., . . . Brechmann, A. (2018). A look into programmers' heads. *IEEE Transactions on Software Engineering, 46*(4), 442-462.
Penny, W. D., Friston, K. J., Ashburner, J. T., Kiebel, S. J., & Nichols, T. E. (2011). *Statistical parametric mapping: the analysis of functional brain images*: Elsevier.
Petty, R. E., Brinol, P., Loersch, C., & McCaslin, M. J. (2009). The need for cognition *Handbook of individual differences in social behavior.* (pp. 318-329). New York, NY, US: The Guilford Press.
Piaget, J., & Bringuier, J. C. (1980). *Conversations with Jean Piaget*: University of Chicago Press.
Prado, J. (2018). The relationship between deductive reasoning and the syntax of language in Broca's area: A review of the neuroimaging literature. *LAnnee psychologique, 118*(3), 289-315.
Prado, J., Chadha, A., & Booth, J. R. (2011). The brain network for deductive reasoning: a quantitative meta-analysis of 28 neuroimaging studies. *Journal of cognitive neuroscience, 23*(11), 3483-3497.
Prat, C. S., Madhyastha, T. M., Mottarella, M. J., & Kuo, C.-H. (2020). Relating natural language aptitude to individual differences in learning programming languages. *Scientific reports, 10*(1), 1-10.
Prensky, M. (2001). Digital natives, digital immigrants part 1. *On the horizon, 9*(5), 1-6.
Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C., & Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature, 435*(7045), 1102-1107.
Rattan, A., Savani, K., Chugh, D., & Dweck, C. S. (2015). Leveraging Mindsets to Promote Academic Achievement: Policy Recommendations. *Perspectives on Psychological Science, 10*(6), 721-726. doi:10.1177/1745691615599383
Regeringskansliet. (2017). Stärkt digital kompetens i skolans styrdokument, Promemoria 2017-03-09: Regeringskansliet.
Reggente, N., Essoe, J. K.-Y., Aghajan, Z. M., Tavakoli, A. V., McGuire, J. F., Suthana, N. A., & Rissman, J. (2018). Enhancing the ecological validity of fMRI memory research using virtual reality. *Frontiers in neuroscience, 12*, 408.
Richardson, J. C., & Newby, T. (2006). The role of students' cognitive engagement in online learning. *American Journal of Distance Education, 20*(1), 23-37.
Richardson, J. T., Long, G. L., & Foster, S. B. (2004). Academic engagement in students with a hearing loss in distance education. *Journal of Deaf Studies and Deaf Education, 9*(1), 68-85.
Rips, L. J. (1994). *The psychology of proof: Deductive reasoning in human thinking*: Mit Press.
Rips, L. J. (2001). Two kinds of reasoning. *Psychological science, 12*(2), 129-134.
Robins, A. (2010). Learning edge momentum: A new account of outcomes in CS1. *Computer Science Education, 20*(1), 37-71.

Rousseau, J.-J. (1762). The social contract (1762): Londres.
Sadowski, C. J., & Cogburn, H. E. (1997). Need for cognition in the big-five factor structure. *The Journal of Psychology, 131*(3), 307-312.
Sadowski, C. J., & Gulgoz, S. (1992). Internal consistency and test-retest reliability of the Need for Cognition Scale. *Perceptual and Motor Skills, 74*(2), 610-610.
Sakai, K., & Passingham, R. E. (2003). Prefrontal interactions reflect future task operations. *Nature neuroscience, 6*(1), 75-81.
Salleh, N., Mendes, E., & Grundy, J. (2011). Empirical studies of pair programming for CS/SE teaching in higher education: A systematic literature review. *IEEE Transactions on Software Engineering, 37*(4), 509-525.
Sboner, A., Mu, X. J., Greenbaum, D., Auerbach, R. K., & Gerstein, M. B. (2011). The real cost of sequencing: higher than you think! *Genome biology, 12*(8), 1-10.
Schacter, D., Gilbert, D., Wegner, D., & Hood, B. M. (2011). *Psychology: European Edition*: Macmillan International Higher Education.
Schwabe, L., & Wolf, O. T. (2013). Stress and multiple memory systems: from ‘thinking’to ‘doing’. *Trends in cognitive sciences, 17*(2), 60-68.
Sebesta, R. W. (2004). *Concepts of programming languages*: Pearson Education India.
Seitz, A. R., & Dinse, H. R. (2007). A common framework for perceptual learning. *Current opinion in neurobiology, 17*(2), 148-153.
Shams, L., & Seitz, A. R. (2008). Benefits of multisensory learning. *Trends in cognitive sciences, 12*(11), 411-417.
Shin, Y. K., Proctor, R. W., & Capaldi, E. J. (2010). A review of contemporary ideomotor theory. *Psychological bulletin, 136*(6), 943.
Siegmund, J., Kästner, C., Apel, S., Parnin, C., Bethmann, A., Leich, T., . . . Brechmann, A. (2014, 2014). *Understanding understanding source code with functional magnetic resonance imaging*.
Siegmund, J., Peitek, N., Parnin, C., Apel, S., Hofmeister, J., Kästner, C., . . . Brechmann, A. (2017, 2017). *Measuring neural efficiency of program comprehension*.
Sisk, V. F., Burgoyne, A. P., Sun, J., Butler, J. L., & Macnamara, B. N. (2018). To what extent and under which circumstances are growth mind-sets important to academic achievement? Two meta-analyses. *Psychological science*, 0956797617739704.
Skogh, I.-B. (2013). 5. STUDENTS’ENCOUNTER WITH TECHNOLOGY EDUCATION. *Technology Teachers as Researchers: Philosophical and Empirical Technology Education Studies in the Swedish TUFF Research School*, 79.
Skulmowski, A., & Rey, G. D. (2017). Measuring Cognitive Load in Embodied Learning Settings. *Frontiers in Psychology, 8*. doi:10.3389/fpsyg.2017.01191
Sousa, D. A. (2010). *Mind, brain, & education: Neuroscience implications for the classroom*: Solution Tree Press.
Stephan, K. E., Penny, W. D., Moran, R. J., den Ouden, H. E., Daunizeau, J., & Friston, K. J. (2010). Ten simple rules for dynamic causal modeling. *Neuroimage, 49*(4), 3099-3109.

Sternberg, R., Ehrman, M., & Grigorenko, E. (2000). A Theory-Based Approach to the Measurement of Foreign Language Learning Ability: The Canal-F Theory and Test. *The Modern Language Journal, 84*(3).

Stohr-Hunt, P. M. (1996). An analysis of frequency of hands-on experience and science achievement. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching, 33*(1), 101-109.

Stuss, D. T., Picton, T. W., & Alexander, M. P. (2001). Consciousness, self-awareness, and the frontal lobes.

Sun, J. C. Y., & Rueda, R. (2012). Situational interest, computer self-efficacy and self-regulation: Their impact on student engagement in distance education. *British journal of educational technology, 43*(2), 191-204.

Swafford, C., Orr, M., & Hall, D. (2014). *Building confidence through hands-on activities.* Paper presented at the Proceedings of the 2014 ASEE Gulf-Southwest Conference.

Taleb, N. N. (2012). *Antifragile: Things that gain from disorder* (Vol. 3): Random House Incorporated.

Teräs, M. (2007). Intercultural learning and hybridity in the culture laboratory. *University of Helsinki, Department of Education*.

Tommerdahl, J. (2010). A model for bridging the gap between neuroscience and education. *Oxford Review of Education, 36*(1), 97-109.

Treisman, A. M. (1964). Selective attention in man. *British medical bulletin, 20*(1), 12-16.

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition, 92*(1-2), 231-270.

Ullman, M. T. (2006). Is Broca's area part of a basal ganglia thalamocortical circuit? *Cortex, 42*(4), 480-485.

Utting, I., Tew, A. E., McCracken, M., Thomas, L., Bouvier, D., Frye, R., . . . Sorva, J. (2013). *A fresh look at novice programmers' performance and their teachers' expectations.* Paper presented at the Proceedings of the ITiCSE working group reports conference on Innovation and technology in computer science education-working group reports.

Uttl, B., White, C. A., & Gonzalez, D. W. (2017). Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. *Studies in Educational Evaluation, 54*, 22-42. doi:https://doi.org/10.1016/j.stueduc.2016.08.007

Varma, S., McCandliss, B. D., & Schwartz, D. L. (2008). Scientific and pragmatic challenges for bridging education and neuroscience. *Educational researcher, 37*(3), 140-152.

Verhulsta, E. (2014). Applying systems and safety engineering principles for antifragility. *Procedia computer science, 32*, 842-849.

Vollrath, M. (2001). Personality and stress. *Scandinavian Journal of Psychology, 42*(4), 335-347.

Von Hausswolff, K., & Eckerdal, A. (2018). *Measuring Programming Knowledge in a Research Context.* Paper presented at the 2018 IEEE Frontiers in Education Conference (FIE).

Von Hausswolff, K., Eckerdal, A., & Thuné, M. (2020). *Learning to program hands-on: a controlled study.* Paper presented at the Koli Calling'20: Proceedings of the 20th Koli Calling International Conference on Computing Education Research.

Von Kriegstein, K., & Giraud, A.-L. (2006). Implicit multisensory associations influence voice recognition. *PLoS biology, 4*(10), e326.

Von Wright, G. H. (1983). Norms, truth and logic.

Vuorikari, R., Punie, Y., Carretero, S., & Van den Brande, L. (2016). DigComp 2.0: The Digital Competence Framework for Citizens, Update Phase 1: The Conceptual Reference Model: Joint Research Centre, European Commision.

White, G., & Sivitanides, M. (2003). An empirical investigation of the relationship between success in mathematics and visual programming courses. *Journal of Information Systems Education, 14*(4), 409.

Wilke, M. (2012). An alternative approach towards assessing and accounting for individual motion in fMRI timeseries. *Neuroimage, 59*(3), 2062-2072. doi:10.1016/j.neuroimage.2011.10.043

Wilke, M. (2014). Isolated assessment of translation or rotation severely underestimates the effects of subject motion in fMRI data. *PloS one, 9*(10), e106498.

Williams, L., Kessler, R. R., Cunningham, W., & Jeffries, R. (2000). Strengthening the case for pair programming. *IEEE software, 17*(4), 19-25.

Wilson, B. C., & Shrock, S. (2001). Contributing to success in an introductory computer science course: a study of twelve factors. *Acm sigcse bulletin, 33*(1), 184-188.

Wing, J. M. (2006). Computational thinking. *Communications of the ACM, 49*(3), 33-35.

Wright, J., & Chung, M. C. (2001). Mastery or mystery? Therapeutic writing: A review of the literature. *British Journal of Guidance & Counselling, 29*(3), 277-291.

Xiang, H., Dediu, D., Roberts, L., Oort, E. v., Norris, D. G., & Hagoort, P. (2012). The structural connectivity underpinning language aptitude, working memory, and IQ in the perisylvian language network. *Language Learning, 62*, 110-130.

Yeager, D. S., & Walton, G. M. (2011). Social-psychological interventions in education: They're not magic. *Review of educational research, 81*(2), 267-301.

Young, M. R. (2002). Experiential learning= hands-on+ minds-on. *Marketing Education Review*, *12*(1), 43-51.

Zatorre, R. J., Fields, R. D., & Johansen-Berg, H. (2012). Plasticity in gray and white: neuroimaging changes in brain structure during learning. *Nature neuroscience, 15*(4), 528-536. doi:10.1038/nn.3045

Zeidman, P., Jafarian, A., Corbin, N., Seghier, M. L., Razi, A., Price, C. J., & Friston, K. J. (2019). A guide to group effective connectivity analysis, part 1: first level analysis with DCM for fMRI. *Neuroimage, 200*, 174-190.

Zeidman, P., Jafarian, A., Seghier, M. L., Litvak, V., Cagnan, H., Price, C. J., & Friston, K. J. (2019). A guide to group effective connectivity analysis, part 2: Second level analysis with PEB. *Neuroimage, 200*, 12-25.

Zilles, K., & Amunts, K. (2010). Centenary of Brodmann's map—conception and fate. *Nature reviews neuroscience, 11*(2), 139-145.

Zimmer, H. D., & Engelkamp, J. (1985). An attempt to distinguish between kinematic and motor memory components. *Acta psychologica, 58*(1), 81-106.

Appendix

Appendix I - Descriptive statistics table

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation	Skewness		Kurtosis	
						Statistic	Std. Error	Statistic	Std. Error
diff_Test_LTM	54	-15.75	10.5	-1.4444	5.18335	-0.622	0.325	0.663	0.639
Entity_score	54	-9	10	-1.74	4.274	0.538	0.325	-0.144	0.639
fMRI_confidence	54	1	4	2.9	0.569	-0.812	0.325	1.133	0.639
fMRI_confidence_diff_t2_t	59	-1.29	0.8	-0.5238	0.35262	0.816	0.311	2.365	0.613
fMRI_confidence_turtle	59	1.03	4	3.1154	0.66073	-1.095	0.311	0.944	0.613
fMRI_confidence_turtle2	59	1.03	4	2.5916	0.61048	-0.385	0.311	-0.039	0.613
fMRI_meaning_score	54	12	29	24	3.582	-1.821	0.325	3.826	0.639
fMRI_score	54	43	109	92.2	14.131	-1.646	0.325	2.968	0.639
fMRI_syntax_score	54	13	30	23.33	4.396	-0.71	0.325	-0.09	0.639
fMRI_turtle_score	54	10	30	24.63	4.822	-1.649	0.325	2.026	0.639
fMRI_turtle2_score	54	7	28	20.24	3.821	-0.854	0.325	1.553	0.639
Grade_art	54	10	20	17.8241	2.3314	-0.992	0.325	0.947	0.639
Grade_avg	54	13.333	20	18.0281	1.652894	-0.738	0.325	-0.317	0.639
Grade_Avg_Lang	54	12.5	20	17.814	1.9652	-0.829	0.325	-0.026	0.639
Grade_math	54	10	20	18.2407	2.55273	-1.629	0.325	2.486	0.639
Grade_Sum_Lang	54	37.5	95	58.287	12.58766	1.097	0.325	1.271	0.639
IMI_EffortImportance	54	-10	19	8.5556	5.83957	-0.737	0.325	1.549	0.639
IMI_InterestEnjoyment	54	-5	32	22.3889	6.91298	-1.306	0.325	3.347	0.639
IMI_PercievedCompetence	54	-2	33	22.2963	6.40351	-1.11	0.325	3.212	0.639
IMI_PressreTension	54	-12	11	-5.3333	5.1725	1.084	0.325	0.865	0.639
IMI_ValueUsefulness	54	9	28	19.4259	5.10125	-0.314	0.325	-0.442	0.639
LTM_loop	54	4	26.5	21.3519	6.12729	-1.402	0.325	1.37	0.639
LTM_sequence	54	2	9.5	8.3287	1.59807	-2.081	0.325	4.809	0.639
LTM_sum	54	8.5	44	35.7616	8.20112	-1.604	0.325	2.243	0.639
LTM_syntax	54	2.5	8	6.081	1.16432	-1.088	0.325	1.709	0.639
Motivation_a	56	2	10	7.6071	1.78558	-0.728	0.319	0.776	0.628
Motivation_b	57	3	10	6.9649	1.81231	-0.244	0.316	-0.861	0.623
Motivation_b_a	57	-4	7	-0.5088	1.99215	0.73	0.316	2.383	0.623
N_Lang	54	3	5	3.2593	0.52071	1.927	0.325	3.021	0.639
NFC	54	-33	48	18.8426	16.38333	-0.821	0.325	0.839	0.639
RB_selfEsteem	54	1	28	20.8	5.747	-1.377	0.325	2.396	0.639
Stress_a	56	1	10	2.8036	2.14408	1.24	0.319	1.18	0.628
Stress_b	57	1	10	3.2895	2.16495	1.073	0.316	0.816	0.623
Stress_b_a	54	-4	5	0.4074	1.66562	-0.023	0.325	1.942	0.639
Test_sum	54	5	44	34.3171	10.02916	-1.131	0.325	0.561	0.639

Appendix II - Group statistics table

Group Statistics					
Variable		N	Mean	Std. Deviatio	Std. Error
NFC	HON	25	19.02	16.96	3.39
	HOFF	29	18.69	16.17	3.00
Entity_score	HON	25	-1.84	4.44	0.89
	HOFF	29	-1.66	4.20	0.78
RB_selfEsteem	HON	25	22.20	5.60	1.12
	HOFF	29	19.59	5.69	1.06
Stress_b	HON	25	2.96	2.42	0.48
	HOFF	29	3.48	1.94	0.36
Stress_a	HON	25	2.28	2.19	0.44
	HOFF	28	3.36	2.09	0.40
Stress_b_a	HON	25	0.68	1.28	0.26
	HOFF	29	0.17	1.93	0.36
Motivation_b	HON	25	7.16	1.99	0.40
	HOFF	29	6.83	1.67	0.31
Motivation_a	HON	25	7.80	1.50	0.30
	HOFF	28	7.43	2.10	0.40
Motivation_b_a	HON	25	-0.64	1.78	0.36
	HOFF	29	-0.34	2.21	0.41
IMI_InterestEnjoyment	HON	23	22.48	5.49	1.15
	HOFF	28	22.36	8.31	1.57
IMI_PercievedCompetence	HON	23	22.61	5.54	1.16
	HOFF	28	22.32	7.32	1.38
IMI_EffortImportance	HON	23	9.04	4.93	1.03
	HOFF	28	7.89	6.62	1.25
IMI_PressreTension	HON	23	-5.83	4.81	1.00
	HOFF	28	-5.25	5.42	1.02
IMI_ValueUsefulness	HON	23	18.87	4.43	0.92
	HOFF	28	20.07	5.64	1.07
Grade_avg	HON	25	18.07	1.50	0.30
	HOFF	29	17.99	1.80	0.33
Grade_math	HON	25	18.10	2.73	0.55
	HOFF	29	18.36	2.43	0.45

Grade_art	HON	25	18.20	1.70	0.34
	HOFF	29	17.50	2.75	0.51
Grade_Sum_Lang	HON	25	60.10	14.28	2.86
	HOFF	29	56.72	10.94	2.03
Grade_Avg_Lang	HON	25	17.83	2.02	0.40
	HOFF	29	17.80	1.95	0.36
Test_sum	HON	25	35.61	9.14	1.83
	HOFF	29	33.20	10.77	2.00
fMRI_score	HON	25	93.88	14.92	2.98
	HOFF	29	90.76	13.51	2.51
fMRI_meaning_score	HON	25	24.44	3.95	0.79
	HOFF	29	23.62	3.26	0.61
fMRI_syntax_score	HON	25	23.92	4.68	0.94
	HOFF	29	22.83	4.15	0.77
fMRI_turtle_score	HON	25	24.72	4.70	0.94
	HOFF	29	24.55	5.01	0.93
fMRI_turtle2_score	HON	25	20.80	4.05	0.81
	HOFF	29	19.76	3.61	0.67
fMRI_confidence	HON	25	3.05	0.46	0.09
	HOFF	29	2.76	0.63	0.12
fMRI_confidence_turtle	HON	25	3.35	0.43	0.09
	HOFF	29	3.02	0.71	0.13
fMRI_confidence_turtle2	HON	25	2.76	0.55	0.11
	HOFF	29	2.51	0.60	0.11
fMRI_confidence_diff_t2_t	HON	25	-0.59	0.30	0.06
	HOFF	29	-0.51	0.40	0.07
LTM_sum	HON	25	37.70	7.45	1.49
	HOFF	29	34.09	8.57	1.59
diff_Test_LTM	HON	25	-2.09	4.84	0.97
	HOFF	29	-0.89	5.49	1.02
N_Lang	HON	25	3.36	0.64	0.13
	HOFF	29	3.17	0.38	0.07

Appendix III - Full correlation table

Correlations

		Hands_on	Female	NFC	Entity_score	RB_selfEsteem	Stress_b	Stress_a	Stress_b_a	Motivation_b	Motivation_a	Motivation_b_a	IMI_InterestEnjoyment
Hands_on	R	1	0.049	0.010	-0.022	0.229	-0.121	-0.248	0.153	0.092	0.102	-0.074	0.009
	Sig. (2-tailed)		0.727	0.942	0.876	0.096	0.383	0.073	0.268	0.508	0.467	0.595	0.952
	N	54	54	54	54	54	54	53	54	54	53	54	51
Female	R	0.049	1	-.352**	0.097	-.314*	.323*	0.185	0.190	0.054	-0.044	0.149	0.012
	Sig. (2-tailed)	0.727		0.009	0.486	0.021	0.017	0.185	0.169	0.701	0.753	0.281	0.936
	N	54	54	54	54	54	54	53	54	54	53	54	51
NFC	R	0.010	-.352**	1	-0.028	.391**	-.344*	-.399**	0.067	.491**	.306*	0.147	.294*
	Sig. (2-tailed)	0.942	0.009		0.842	0.003	0.011	0.003	0.630	0.000	0.026	0.288	0.036
	N	54	54	54	54	54	54	53	54	54	53	54	51
Entity_score	R	-0.022	0.097	-0.028	1	0.064	-0.039	-0.018	-0.034	0.039	0.020	-0.020	0.010
	Sig. (2-tailed)	0.876	0.486	0.842		0.644	0.777	0.901	0.809	0.777	0.885	0.884	0.945
	N	54	54	54	54	54	54	53	54	54	53	54	51
RB_selfEsteem	R	0.229	-.314*	.391**	0.064	1	-.415**	-.499**	0.025	0.168	-0.049	-0.017	-0.047
	Sig. (2-tailed)	0.096	0.021	0.003	0.644		0.002	0.000	0.860	0.226	0.729	0.904	0.745
	N	54	54	54	54	54	54	53	54	54	53	54	51
Stress_b	R	-0.121	.323*	-.344*	-0.039	-.415**	1	.666**	.385**	-.295*	-.283*	-0.058	-0.226
	Sig. (2-tailed)	0.383	0.017	0.011	0.777	0.002		0.000	0.004	0.026	0.034	0.666	0.100
	N	54	54	54	54	54	57	56	54	57	56	57	54
Stress_a	R	-0.248	0.185	-.399**	-0.018	-.499**	.666**	1	-.385**	-.285*	-.391**	0.103	-.389**
	Sig. (2-tailed)	0.073	0.185	0.003	0.901	0.000	0.000		0.004	0.033	0.003	0.452	0.004
	N	53	53	53	53	53	56	56	53	56	56	56	53
Stress_b_a	R	0.153	0.190	0.067	-0.034	0.025	.385**	-.385**	1	0.052	0.148	-0.104	0.209
	Sig. (2-tailed)	0.268	0.169	0.630	0.809	0.860	0.004	0.004		0.707	0.290	0.455	0.142
	N	54	54	54	54	54	54	53	54	54	53	54	51
Motivation_b	R	0.092	0.054	.491**	0.039	0.168	-.295*	-.285*	0.052	1	.541**	.430**	.386**
	Sig. (2-tailed)	0.508	0.701	0.000	0.777	0.226	0.026	0.033	0.707		0.000	0.001	0.004
	N	54	54	54	54	54	57	56	54	57	56	57	54
Motivation_a	R	0.102	-0.044	.306*	0.020	-0.049	-.283*	-.391**	0.148	.541**	1	-.460**	.774**
	Sig. (2-tailed)	0.467	0.753	0.026	0.885	0.729	0.034	0.003	0.290	0.000		0.000	0.000
	N	53	53	53	53	53	56	56	53	56	56	56	53
Motivation_b_a	R	-0.074	0.149	0.147	-0.020	-0.017	-0.058	0.103	-0.104	.430**	-.460**	1	-.328*
	Sig. (2-tailed)	0.595	0.281	0.288	0.884	0.904	0.666	0.452	0.455	0.001	0.000		0.016
	N	54	54	54	54	54	57	56	54	57	56	57	54
IMI_InterestEnjoyment	R	0.009	0.012	.294*	0.010	-0.047	-0.226	-.389**	0.209	.386**	.774**	-.328*	1
	Sig. (2-tailed)	0.952	0.936	0.036	0.945	0.745	0.100	0.004	0.142	0.004	0.000	0.016	
	N	51	51	51	51	51	54	53	51	54	53	54	54
IMI_PercievedCompetence	R	0.022	-0.086	.408**	0.003	.352*	-.294*	-.385**	0.058	0.194	.444**	-.387**	.604**
	Sig. (2-tailed)	0.877	0.547	0.003	0.981	0.011	0.031	0.004	0.685	0.161	0.001	0.004	0.000
	N	51	51	51	51	51	54	53	51	54	53	54	54
IMI_EffortImportance	R	0.098	-0.013	0.167	0.117	0.153	0.050	-0.145	0.211	0.136	.289*	-.342*	.410**
	Sig. (2-tailed)	0.493	0.927	0.242	0.415	0.283	0.719	0.301	0.136	0.329	0.036	0.011	0.002
	N	51	51	51	51	51	54	53	51	54	53	54	54
IMI_PressreTension	R	-0.057	-0.072	-0.066	-0.059	-.529**	.282*	.347*	-0.083	-0.117	0.094	0.044	0.022
	Sig. (2-tailed)	0.693	0.615	0.644	0.681	0.000	0.039	0.011	0.563	0.400	0.505	0.752	0.874
	N	51	51	51	51	51	54	53	51	54	53	54	54
IMI_ValueUsefulness	R	-0.118	-.276*	0.249	0.026	-0.075	-.302*	-0.146	-0.137	.406**	.470**	-0.022	.568**
	Sig. (2-tailed)	0.409	0.050	0.078	0.855	0.602	0.027	0.296	0.338	0.002	0.000	0.875	0.000
	N	51	51	51	51	51	54	53	51	54	53	54	54
Grade_avg	R	0.022	0.043	0.082	0.052	.272*	0.152	0.177	-0.041	0.035	-0.182	-0.020	-0.266
	Sig. (2-tailed)	0.875	0.756	0.555	0.710	0.046	0.273	0.205	0.770	0.801	0.192	0.886	0.059
	N	54	54	54	54	54	54	53	54	54	53	54	51
Grade_math	R	-0.052	-0.007	.347*	0.004	.326*	-0.092	-0.144	0.016	0.206	0.086	-0.113	-0.016
	Sig. (2-tailed)	0.711	0.961	0.010	0.979	0.016	0.507	0.303	0.906	0.134	0.542	0.415	0.914
	N	54	54	54	54	54	54	53	54	54	53	54	51
Grade_art	R	0.151	0.242	-0.017	-0.009	-0.079	0.161	0.225	-0.083	0.091	-0.203	0.256	-0.236
	Sig. (2-tailed)	0.275	0.077	0.905	0.951	0.568	0.244	0.106	0.550	0.515	0.145	0.062	0.096
	N	54	54	54	54	54	54	53	54	54	53	54	51
N_Lang	R	0.181	0.070	0.080	-0.107	0.163	0.011	0.002	0.006	0.205	0.011	0.140	-0.050
	Sig. (2-tailed)	0.189	0.615	0.565	0.441	0.239	0.940	0.989	0.963	0.138	0.938	0.313	0.726
	N	54	54	54	54	54	54	53	54	54	53	54	51
Grade_Sum_Lang	R	0.135	0.107	0.088	-0.074	0.203	0.065	0.094	-0.047	0.190	-0.035	0.114	-0.097
	Sig. (2-tailed)	0.330	0.442	0.526	0.595	0.141	0.638	0.504	0.735	0.168	0.801	0.411	0.499
	N	54	54	54	54	54	54	53	54	54	53	54	51
Grade_Avg_Lang	R	0.007	0.099	0.059	0.022	0.179	0.109	0.185	-0.109	0.085	-0.094	0.026	-0.132
	Sig. (2-tailed)	0.961	0.477	0.673	0.876	0.195	0.434	0.184	0.432	0.543	0.503	0.851	0.356
	N	54	54	54	54	54	54	53	54	54	53	54	51
Test_syntax	R	0.160	-0.079	.308*	0.144	0.237	-0.180	-0.215	0.001	0.179	.408**	-.368**	.401**
	Sig. (2-tailed)	0.248	0.569	0.024	0.298	0.084	0.193	0.123	0.994	0.195	0.002	0.006	0.004
	N	54	54	54	54	54	54	53	54	54	53	54	51

Correlations

		IMI_PercievedCompetence	IMI_EffortImportance	IMI_PressreTension	IMI_ValueUsefulness	Grade_avg	Grade_math	Grade_art	N_Lang	Grade_Sum_Lang	Grade_Avg_Lang	Test_syntax	Test_sequence
Hands_on	R	0.022	0.098	-0.057	-0.118	0.022	-0.052	0.151	0.181	0.135	0.007	0.160	0.047
	Sig. (2-tailed)	0.877	0.493	0.693	0.409	0.875	0.711	0.275	0.189	0.330	0.961	0.248	0.737
	N	51	51	51	51	54	54	54	54	54	54	54	54
Female	R	-0.086	-0.013	-0.072	-.276*	0.043	-0.007	0.242	0.070	0.107	0.099	-0.079	-0.152
	Sig. (2-tailed)	0.547	0.927	0.615	0.050	0.756	0.961	0.077	0.615	0.442	0.477	0.569	0.274
	N	51	51	51	51	54	54	54	54	54	54	54	54
NFC	R	.408**	0.167	-0.066	0.249	0.082	.347*	-0.017	0.080	0.088	0.059	.308*	.409**
	Sig. (2-tailed)	0.003	0.242	0.644	0.078	0.555	0.010	0.905	0.565	0.526	0.673	0.024	0.002
	N	51	51	51	51	54	54	54	54	54	54	54	54
Entity_score	R	0.003	0.117	-0.059	0.026	0.052	0.004	-0.009	-0.107	-0.074	0.022	0.144	0.156
	Sig. (2-tailed)	0.981	0.415	0.681	0.855	0.710	0.979	0.951	0.441	0.595	0.876	0.298	0.259
	N	51	51	51	51	54	54	54	54	54	54	54	54
RB_selfEsteem	R	.352*	0.153	-.529**	-0.075	.272*	.326*	-0.079	0.163	0.203	0.179	0.237	.397**
	Sig. (2-tailed)	0.011	0.283	0.000	0.602	0.046	0.016	0.568	0.239	0.141	0.195	0.084	0.003
	N	51	51	51	51	54	54	54	54	54	54	54	54
Stress_b	R	-.294*	0.050	.282*	-.302*	0.152	-0.092	0.161	0.011	0.065	0.109	-0.180	-.401**
	Sig. (2-tailed)	0.031	0.719	0.039	0.027	0.273	0.507	0.244	0.940	0.638	0.434	0.193	0.003
	N	54	54	54	54	54	54	54	54	54	54	54	54
Stress_a	R	-.385**	-0.145	.347*	-0.146	0.177	-0.144	0.225	0.002	0.094	0.185	-0.215	-.324*
	Sig. (2-tailed)	0.004	0.301	0.011	0.296	0.205	0.303	0.106	0.989	0.504	0.184	0.123	0.018
	N	53	53	53	53	53	53	53	53	53	53	53	53
Stress_b_a	R	0.058	0.211	-0.083	-0.137	-0.041	0.016	-0.083	0.006	-0.047	-0.109	0.001	-0.163
	Sig. (2-tailed)	0.685	0.136	0.563	0.338	0.770	0.906	0.550	0.963	0.735	0.432	0.994	0.240
	N	51	51	51	51	54	54	54	54	54	54	54	54
Motivation_b	R	0.194	0.136	-0.117	.406**	0.035	0.206	0.091	0.205	0.190	0.085	0.179	0.216
	Sig. (2-tailed)	0.161	0.329	0.400	0.002	0.801	0.134	0.515	0.138	0.168	0.543	0.195	0.118
	N	54	54	54	54	54	54	54	54	54	54	54	54
Motivation_a	R	.444**	.289*	0.094	.470**	-0.182	0.086	-0.203	0.011	-0.035	-0.094	.408**	.341*
	Sig. (2-tailed)	0.001	0.036	0.505	0.000	0.192	0.542	0.145	0.938	0.801	0.503	0.002	0.012
	N	53	53	53	53	53	53	53	53	53	53	53	53
Motivation_b_a	R	-.387**	-.342*	0.044	-0.022	-0.020	-0.113	0.256	0.140	0.114	0.026	-.368**	-.291*
	Sig. (2-tailed)	0.004	0.011	0.752	0.875	0.886	0.415	0.062	0.313	0.411	0.851	0.006	0.033
	N	54	54	54	54	54	54	54	54	54	54	54	54
IMI_InterestEnjoyment	R	.604**	.410**	0.022	.568**	-0.266	-0.016	-0.236	-0.050	-0.097	-0.132	.401**	.352*
	Sig. (2-tailed)	0.000	0.002	0.874	0.000	0.059	0.914	0.096	0.726	0.499	0.356	0.004	0.011
	N	54	54	54	54	51	51	51	51	51	51	51	51
IMI_PercievedCompetence	R	1	.527**	-.356**	0.221	0.043	.321*	-0.203	0.189	0.178	0.058	.555**	.657**
	Sig. (2-tailed)		0.000	0.008	0.108	0.764	0.022	0.154	0.185	0.210	0.686	0.000	0.000
	N	54	54	54	54	51	51	51	51	51	51	51	51
IMI_EffortImportance	R	.527**	1	-0.092	0.252	0.031	0.200	-0.155	-0.022	-0.024	-0.001	.282*	.316*
	Sig. (2-tailed)	0.000		0.509	0.066	0.832	0.160	0.278	0.880	0.868	0.995	0.045	0.024
	N	54	54	54	54	51	51	51	51	51	51	51	51
IMI_PressreTension	R	-.356**	-0.092	1	0.184	-0.259	-.413**	0.172	-0.197	-.303*	-.321*	-0.254	-0.235
	Sig. (2-tailed)	0.008	0.509		0.184	0.066	0.003	0.227	0.165	0.031	0.022	0.072	0.097
	N	54	54	54	54	51	51	51	51	51	51	51	51
IMI_ValueUsefulness	R	0.221	0.252	0.184	1	-0.136	0.006	-0.081	-.298*	-0.268	-0.073	0.107	0.251
	Sig. (2-tailed)	0.108	0.066	0.184		0.340	0.967	0.570	0.034	0.057	0.612	0.455	0.075
	N	54	54	54	54	51	51	51	51	51	51	51	51
Grade_avg	R	0.043	0.031	-0.259	-0.136	1	.702**	.442**	0.234	.597**	.857**	.347*	.319*
	Sig. (2-tailed)	0.764	0.832	0.066	0.340		0.000	0.001	0.089	0.000	0.000	0.010	0.019
	N	51	51	51	51	54	54	54	54	54	54	54	54
Grade_math	R	.321*	0.200	-.413**	0.006	.702**	1	0.137	0.208	.492**	.682**	.518**	.536**
	Sig. (2-tailed)	0.022	0.160	0.003	0.967	0.000		0.322	0.132	0.000	0.000	0.000	0.000
	N	51	51	51	51	54	54	54	54	54	54	54	54
Grade_art	R	-0.203	-0.155	0.172	-0.081	.442**	0.137	1	0.240	.321*	0.263	0.021	0.127
	Sig. (2-tailed)	0.154	0.278	0.227	0.570	0.001	0.322		0.080	0.018	0.054	0.880	0.359
	N	51	51	51	51	54	54	54	54	54	54	54	54
N_Lang	R	0.189	-0.022	-0.197	-.298*	0.234	0.208	0.240	1	.882**	0.225	0.244	0.077
	Sig. (2-tailed)	0.185	0.880	0.165	0.034	0.089	0.132	0.080		0.000	0.101	0.075	0.581
	N	51	51	51	51	54	54	54	54	54	54	54	54
Grade_Sum_Lang	R	0.178	-0.024	-.303*	-0.268	.597**	.492**	.321*	.882**	1	.656**	.411**	0.199
	Sig. (2-tailed)	0.210	0.868	0.031	0.057	0.000	0.000	0.018	0.000		0.000	0.002	0.148
	N	51	51	51	51	54	54	54	54	54	54	54	54
Grade_Avg_Lang	R	0.058	-0.001	-.321*	-0.073	.857**	.682**	0.263	0.225	.656**	1	.464**	.270*
	Sig. (2-tailed)	0.686	0.995	0.022	0.612	0.000	0.000	0.054	0.101	0.000		0.000	0.049
	N	51	51	51	51	54	54	54	54	54	54	54	54
Test_synt	R	.555**	.282*	-0.254	0.107	.347*	.518**	0.021	0.244	.411**	.464**	1	.638**
	Sig. (2-tailed)	0.000	0.045	0.072	0.455	0.010	0.000	0.880	0.075	0.002	0.000		0.000
	N	51	51	51	51	54	54	54	54	54	54	54	54

Correlations

		Test_loops	Test_not_write	Test_write	Test_sum	fMRI_score	fMRI_meaning_score	fMRI_syntax_score	fMRI_turtle_score	fMRI_turtle2_score	fMRI_yes_score	fMRI_no_score	fMRI_confidence
Hands_on	R	0.118	0.023	0.218	0.121	0.111	0.115	0.125	0.018	0.137	0.118	0.080	0.254
	Sig. (2-tailed)	0.395	0.866	0.114	0.384	0.424	0.407	0.367	0.900	0.323	0.397	0.564	0.064
	N	54	54	54	54	54	54	54	54	54	54	54	54
Female	R	-0.190	-0.120	-0.213	-0.174	-0.137	-0.137	-0.046	-0.192	-0.084	-0.135	-0.108	-0.192
	Sig. (2-tailed)	0.170	0.387	0.122	0.209	0.322	0.323	0.742	0.164	0.547	0.330	0.436	0.165
	N	54	54	54	54	54	54	54	54	54	54	54	54
NFC	R	.447**	.458**	.369**	.444**	0.242	0.158	0.101	.322*	0.223	0.240	0.188	.413**
	Sig. (2-tailed)	0.001	0.000	0.006	0.001	0.078	0.254	0.466	0.018	0.105	0.080	0.173	0.002
	N	54	54	54	54	54	54	54	54	54	54	54	54
Entity_score	R	0.163	0.207	0.104	0.170	-0.079	-0.126	-0.164	-0.072	0.105	-0.087	-0.055	-0.182
	Sig. (2-tailed)	0.239	0.133	0.454	0.220	0.569	0.365	0.235	0.604	0.451	0.533	0.694	0.188
	N	54	54	54	54	54	54	54	54	54	54	54	54
RB_selfEsteem	R	.397**	.327*	.420**	.394**	.309*	0.197	.273*	0.213	.374**	.320*	0.229	.362**
	Sig. (2-tailed)	0.003	0.016	0.002	0.003	0.023	0.153	0.046	0.122	0.005	0.018	0.096	0.007
	N	54	54	54	54	54	54	54	54	54	54	54	54
Stress_b	R	-.448**	-.431**	-.348**	-.418**	-0.165	-0.029	-0.194	-0.130	-0.196	-0.060	-0.222	-.354**
	Sig. (2-tailed)	0.001	0.001	0.010	0.002	0.233	0.834	0.159	0.349	0.156	0.666	0.106	0.009
	N	54	54	54	54	54	54	54	54	54	54	54	54
Stress_a	R	-.323*	-.308*	-.308*	-.328*	-.276*	-0.105	-.287*	-0.211	-.321*	-0.132	-.339*	-.459**
	Sig. (2-tailed)	0.018	0.025	0.025	0.017	0.045	0.456	0.038	0.129	0.019	0.347	0.013	0.001
	N	53	53	53	53	53	53	53	53	53	53	53	53
Stress_b_a	R	-0.209	-0.188	-0.134	-0.173	0.114	0.082	0.110	0.071	0.130	0.049	0.147	0.129
	Sig. (2-tailed)	0.129	0.174	0.334	0.210	0.411	0.554	0.429	0.611	0.350	0.726	0.289	0.351
	N	54	54	54	54	54	54	54	54	54	54	54	54
Motivation_b	R	.289*	.306*	0.201	.274*	-0.098	-0.232	-0.068	-0.059	0.006	-0.110	-0.066	.304*
	Sig. (2-tailed)	0.034	0.025	0.145	0.045	0.479	0.092	0.627	0.672	0.965	0.429	0.637	0.025
	N	54	54	54	54	54	54	54	54	54	54	54	54
Motivation_a	R	.398**	.386**	.410**	.422**	0.028	-0.104	0.008	0.078	0.096	0.005	0.041	0.149
	Sig. (2-tailed)	0.003	0.004	0.002	0.002	0.843	0.461	0.955	0.579	0.496	0.969	0.772	0.287
	N	53	53	53	53	53	53	53	53	53	53	53	53
Motivation_b_a	R	-0.246	-0.183	-.389**	-.296*	-0.246	-0.218	-0.123	-.278*	-0.213	-.317*	-0.126	0.105
	Sig. (2-tailed)	0.073	0.184	0.004	0.030	0.073	0.114	0.377	0.042	0.121	0.019	0.363	0.449
	N	54	54	54	54	54	54	54	54	54	54	54	54
IMI_InterestEnjoyment	R	.349*	.346*	.380**	.385**	-0.012	-0.136	0.005	0.085	-0.030	-0.036	0.013	0.220
	Sig. (2-tailed)	0.012	0.013	0.006	0.005	0.936	0.340	0.970	0.552	0.835	0.803	0.930	0.121
	N	51	51	51	51	51	51	51	51	51	51	51	51
IMI_PercievedCompetence	R	.620**	.572**	.674**	.660**	.290*	0.100	0.196	.404**	0.242	0.237	0.271	0.231
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.039	0.487	0.169	0.003	0.088	0.094	0.054	0.104
	N	51	51	51	51	51	51	51	51	51	51	51	51
IMI_EffortImportance	R	0.231	0.155	.376**	0.275	0.003	0.011	-0.039	0.113	-0.098	0.062	-0.050	0.046
	Sig. (2-tailed)	0.103	0.278	0.007	0.051	0.983	0.941	0.787	0.429	0.495	0.666	0.728	0.748
	N	51	51	51	51	51	51	51	51	51	51	51	51
IMI_PressureTension	R	-0.239	-0.243	-0.243	-0.259	-.355*	-0.254	-.297*	-.317*	-.331*	-.349*	-.279*	-.336*
	Sig. (2-tailed)	0.091	0.086	0.086	0.067	0.011	0.072	0.035	0.024	0.018	0.012	0.048	0.016
	N	51	51	51	51	51	51	51	51	51	51	51	51
IMI_ValueUsefulness	R	.284*	0.219	.278*	0.263	-0.142	-0.148	-0.099	-0.071	-0.182	-0.107	-0.140	0.106
	Sig. (2-tailed)	0.043	0.122	0.048	0.062	0.321	0.302	0.488	0.619	0.202	0.453	0.327	0.461
	N	51	51	51	51	51	51	51	51	51	51	51	51
Grade_avg	R	.362**	.371**	.334*	.377**	.397**	.308*	.374**	.308*	.362**	.444**	0.266	0.013
	Sig. (2-tailed)	0.007	0.006	0.014	0.005	0.003	0.023	0.005	0.024	0.007	0.001	0.052	0.926
	N	54	54	54	54	54	54	54	54	54	54	54	54
Grade_math	R	.540**	.585**	.486**	.574**	.486**	.335*	.289*	.517**	.499**	.550**	.319*	0.233
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000	0.013	0.034	0.000	0.000	0.000	0.019	0.090
	N	54	54	54	54	54	54	54	54	54	54	54	54
Grade_art	R	0.077	0.143	0.000	0.082	0.045	0.136	-0.020	0.044	0.007	0.009	0.067	-0.177
	Sig. (2-tailed)	0.579	0.304	0.998	0.558	0.745	0.328	0.886	0.750	0.960	0.947	0.629	0.199
	N	54	54	54	54	54	54	54	54	54	54	54	54
N_Lang	R	0.105	0.147	0.104	0.135	0.170	0.111	0.168	0.129	0.167	0.100	0.194	0.125
	Sig. (2-tailed)	0.448	0.290	0.456	0.331	0.220	0.423	0.226	0.352	0.227	0.472	0.160	0.369
	N	54	54	54	54	54	54	54	54	54	54	54	54
Grade_Sum_Lang	R	0.263	.305*	0.252	.299*	.326*	0.212	.334*	0.254	.301*	.311*	0.265	0.104
	Sig. (2-tailed)	0.055	0.025	0.066	0.028	0.016	0.123	0.014	0.063	0.027	0.022	0.053	0.453
	N	54	54	54	54	54	54	54	54	54	54	54	54
Grade_Avg_Lang	R	.370**	.390**	.350**	.395**	.386**	0.245	.415**	.298*	.344*	.472**	0.221	0.034
	Sig. (2-tailed)	0.006	0.004	0.010	0.003	0.004	0.074	0.002	0.029	0.011	0.000	0.108	0.810
	N	54	54	54	54	54	54	54	54	54	54	54	54
Test_synt	R	.661**	.714**	.736**	.771**	.479**	.319*	.333*	.532**	.417**	.514**	.339*	0.159
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000	0.019	0.014	0.000	0.002	0.000	0.012	0.251
	N	54	54	54	54	54	54	54	54	54	54	54	54

Correlations											
		fMRI_confidence_turtle	fMRI_confidence_turtle2	fMRI_confidence_diff_t2_t	LTM_syntax	LTM_sequence	LTM_loop	LTM_not_write	LTM_write	LTM_sum	diff_Test_LTM
Hands_on	R	.271*	0.219	-0.105	0.200	0.235	0.197	0.234	0.172	0.222	-0.117
	Sig. (2-tailed)	0.048	0.112	0.450	0.146	0.087	0.153	0.089	0.213	0.107	0.401
	N	54	54	54	54	54	54	54	54	54	54
Female	R	-0.185	-0.176	0.028	0.121	-0.141	-0.240	-0.104	-0.231	-0.189	-0.036
	Sig. (2-tailed)	0.180	0.203	0.842	0.382	0.311	0.080	0.455	0.093	0.170	0.795
	N	54	54	54	54	54	54	54	54	54	54
NFC	R	.386**	.403**	0.001	0.165	.451**	.491**	.367**	.490**	.478**	0.103
	Sig. (2-tailed)	0.004	0.003	0.997	0.233	0.001	0.000	0.006	0.000	0.000	0.461
	N	54	54	54	54	54	54	54	54	54	54
Entity_score	R	-0.231	-0.103	0.228	-0.119	0.144	0.150	0.076	0.142	0.123	0.133
	Sig. (2-tailed)	0.092	0.460	0.097	0.393	0.299	0.279	0.584	0.304	0.375	0.337
	N	54	54	54	54	54	54	54	54	54	54
RB_selfEsteem	R	.333*	.355**	0.012	.310*	.405**	.360**	.398**	.318*	.392**	0.142
	Sig. (2-tailed)	0.014	0.008	0.929	0.023	0.002	0.008	0.003	0.019	0.003	0.306
	N	54	54	54	54	54	54	54	54	54	54
Stress_b	R	-.329*	-.369**	-0.016	-0.249	-.292*	-.367**	-.475**	-0.208	-.366**	-0.229
	Sig. (2-tailed)	0.012	0.005	0.903	0.070	0.032	0.006	0.000	0.132	0.006	0.096
	N	57	57	57	54	54	54	54	54	54	54
Stress_a	R	-.337*	-.359**	0.014	-.325*	-.345*	-.403**	-.527**	-0.232	-.423**	0.004
	Sig. (2-tailed)	0.011	0.007	0.917	0.018	0.011	0.003	0.000	0.094	0.002	0.980
	N	56	56	56	53	53	53	53	53	53	53
Stress_b_a	R	0.156	0.096	-0.110	0.029	-0.044	-0.023	0.017	-0.051	-0.021	-.302*
	Sig. (2-tailed)	0.260	0.490	0.429	0.836	0.751	0.871	0.903	0.714	0.878	0.027
	N	54	54	54	54	54	54	54	54	54	54
Motivation_b	R	.293*	.284*	-0.062	0.071	0.165	0.179	0.224	0.103	0.176	0.252
	Sig. (2-tailed)	0.027	0.032	0.648	0.610	0.234	0.196	0.103	0.461	0.204	0.067
	N	57	57	57	54	54	54	54	54	54	54
Motivation_a	R	0.150	0.103	-0.106	0.036	.338*	.308*	0.253	.293*	.305*	.330*
	Sig. (2-tailed)	0.269	0.451	0.439	0.800	0.013	0.025	0.068	0.033	0.026	0.016
	N	56	56	56	53	53	53	53	53	53	53
Motivation_b_a	R	0.085	0.165	0.124	-0.185	-.389**	-.297*	-0.165	-.405**	-.324*	-0.061
	Sig. (2-tailed)	0.532	0.220	0.357	0.181	0.004	0.029	0.232	0.002	0.017	0.660
	N	57	57	57	54	54	54	54	54	54	54
IMI_InterestEnjoyment	R	0.229	0.154	-0.170	0.260	.301*	0.247	0.239	0.268	.281*	.299*
	Sig. (2-tailed)	0.096	0.267	0.218	0.065	0.032	0.080	0.092	0.058	0.046	0.033
	N	54	54	54	51	51	51	51	51	51	51
IMI_PercievedCompetence	R	.291*	0.174	-0.253	.501**	.686**	.613**	.593**	.606**	.663**	0.221
	Sig. (2-tailed)	0.033	0.208	0.065	0.000	0.000	0.000	0.000	0.000	0.000	0.118
	N	54	54	54	51	51	51	51	51	51	51
IMI_EffortImportance	R	0.056	-0.052	-0.196	.339*	.420**	.281*	0.190	.411**	.341*	-0.011
	Sig. (2-tailed)	0.685	0.708	0.155	0.015	0.002	0.045	0.183	0.003	0.014	0.938
	N	54	54	54	51	51	51	51	51	51	51
IMI_PressreTension	R	-.364**	-.326*	0.132	-.367**	-0.260	-0.197	-.326*	-0.140	-0.250	-0.102
	Sig. (2-tailed)	0.007	0.016	0.342	0.008	0.065	0.166	0.020	0.326	0.077	0.474
	N	54	54	54	51	51	51	51	51	51	51
IMI_ValueUsefulness	R	0.196	0.081	-0.231	-0.005	0.144	0.149	0.050	0.192	0.139	.289*
	Sig. (2-tailed)	0.156	0.558	0.093	0.972	0.314	0.296	0.729	0.178	0.331	0.040
	N	54	54	54	51	51	51	51	51	51	51
Grade_avg	R	0.101	-0.068	-.285*	0.197	.374**	.417**	0.252	.479**	.413**	0.076
	Sig. (2-tailed)	0.466	0.627	0.037	0.153	0.005	0.002	0.066	0.000	0.002	0.584
	N	54	54	54	54	54	54	54	54	54	54
Grade_math	R	.365**	0.079	-.496**	.299*	.491**	.555**	.423**	.568**	.552**	0.236
	Sig. (2-tailed)	0.007	0.569	0.000	0.028	0.000	0.000	0.001	0.000	0.000	0.086
	N	54	54	54	54	54	54	54	54	54	54
Grade_art	R	-0.136	-0.194	-0.086	0.010	0.116	0.137	0.007	0.208	0.127	-0.043
	Sig. (2-tailed)	0.328	0.161	0.538	0.945	0.402	0.323	0.960	0.131	0.362	0.760
	N	54	54	54	54	54	54	54	54	54	54
N_Lang	R	0.173	0.083	-0.160	0.120	0.168	0.098	0.116	0.108	0.123	0.066
	Sig. (2-tailed)	0.211	0.550	0.247	0.386	0.225	0.481	0.405	0.438	0.376	0.634
	N	54	54	54	54	54	54	54	54	54	54
Grade_Sum_Lang	R	0.190	0.025	-.285*	0.165	.280*	0.214	0.172	0.254	0.238	0.201
	Sig. (2-tailed)	0.168	0.857	0.037	0.233	0.040	0.120	0.214	0.064	0.083	0.144
	N	54	54	54	54	54	54	54	54	54	54
Grade_Avg_Lang	R	0.129	-0.061	-.322*	0.149	.301*	.271*	0.162	.337*	.282*	.318*
	Sig. (2-tailed)	0.352	0.660	0.018	0.283	0.027	0.048	0.243	0.013	0.039	0.019
	N	54	54	54	54	54	54	54	54	54	54
Test_synt	R	.275*	0.038	-.410**	.430**	.589**	.552**	.474**	.584**	.588**	.561**
	Sig. (2-tailed)	0.044	0.787	0.002	0.001	0.000	0.000	0.000	0.000	0.000	0.000
	N	54	54	54	54	54	54	54	54	54	54

Correlations

		Hands_on	Female	NFC	Entity_score	RB_selfEsteem	Stress_b	Stress_a	Stress_b_a	Motivation_b	Motivation_a	Motivation_b_a	IMI_InterestEnjoyment
Test_sequence	R	0.047	-0.152	.409**	0.156	.397**	-.401**	-.324*	-0.163	0.216	.341*	-.291*	.352*
	Sig. (2-tailed)	0.737	0.274	0.002	0.259	0.003	0.003	0.018	0.240	0.118	0.012	0.033	0.011
	N	54	54	54	54	54	54	53	54	54	53	54	51
Test_loops	R	0.118	-0.190	.447**	0.163	.397**	-.448**	-.323*	-0.209	.289*	.398**	-0.246	.349*
	Sig. (2-tailed)	0.395	0.170	0.001	0.239	0.003	0.001	0.018	0.129	0.034	0.003	0.073	0.012
	N	54	54	54	54	54	54	53	54	54	53	54	51
Test_not_write	R	0.023	-0.120	.458**	0.207	.327*	-.431**	-.308*	-0.188	.306*	.386**	-0.183	.346*
	Sig. (2-tailed)	0.866	0.387	0.000	0.133	0.016	0.001	0.025	0.174	0.025	0.004	0.184	0.013
	N	54	54	54	54	54	54	53	54	54	53	54	51
Test_write	R	0.218	-0.213	.369**	0.104	.420**	-.348**	-.308*	-0.134	0.201	.410**	-.389**	.380**
	Sig. (2-tailed)	0.114	0.122	0.006	0.454	0.002	0.010	0.025	0.334	0.145	0.002	0.004	0.006
	N	54	54	54	54	54	54	53	54	54	53	54	51
Test_sum	R	0.121	-0.174	.444**	0.170	.394**	-.418**	-.328*	-0.173	.274*	.422**	-.296*	.385**
	Sig. (2-tailed)	0.384	0.209	0.001	0.220	0.003	0.002	0.017	0.210	0.045	0.002	0.030	0.005
	N	54	54	54	54	54	54	53	54	54	53	54	51
fMRI_score	R	0.111	-0.137	0.242	-0.079	.309*	-0.165	-.276*	0.114	-0.098	0.028	-0.246	-0.012
	Sig. (2-tailed)	0.424	0.322	0.078	0.569	0.023	0.233	0.045	0.411	0.479	0.843	0.073	0.936
	N	54	54	54	54	54	54	53	54	54	53	54	51
fMRI_meaning_score	R	0.115	-0.137	0.158	-0.126	0.197	-0.029	-0.105	0.082	-0.232	-0.104	-0.218	-0.136
	Sig. (2-tailed)	0.407	0.323	0.254	0.365	0.153	0.834	0.456	0.554	0.092	0.461	0.114	0.340
	N	54	54	54	54	54	54	53	54	54	53	54	51
fMRI_syntax_score	R	0.125	-0.046	0.101	-0.164	.273*	-0.194	-.287*	0.110	-0.068	0.008	-0.123	0.005
	Sig. (2-tailed)	0.367	0.742	0.466	0.235	0.046	0.159	0.038	0.429	0.627	0.955	0.377	0.970
	N	54	54	54	54	54	54	53	54	54	53	54	51
fMRI_turtle_score	R	0.018	-0.192	.322*	-0.072	0.213	-0.130	-0.211	0.071	-0.059	0.078	-.278*	0.085
	Sig. (2-tailed)	0.900	0.164	0.018	0.604	0.122	0.349	0.129	0.611	0.672	0.579	0.042	0.552
	N	54	54	54	54	54	54	53	54	54	53	54	51
fMRI_turtle2_score	R	0.137	-0.084	0.223	0.105	.374**	-0.196	-.321*	0.130	0.006	0.096	-0.213	-0.030
	Sig. (2-tailed)	0.323	0.547	0.105	0.451	0.005	0.156	0.019	0.350	0.965	0.496	0.121	0.835
	N	54	54	54	54	54	54	53	54	54	53	54	51
fMRI_yes_score	R	0.118	-0.135	0.240	-0.087	.320*	-0.060	-0.132	0.049	-0.110	0.005	-.317*	-0.036
	Sig. (2-tailed)	0.397	0.330	0.080	0.533	0.018	0.666	0.347	0.726	0.429	0.969	0.019	0.803
	N	54	54	54	54	54	54	53	54	54	53	54	51
fMRI_no_score	R	0.080	-0.108	0.188	-0.055	0.229	-0.222	-.339*	0.147	-0.066	0.041	-0.126	0.013
	Sig. (2-tailed)	0.564	0.436	0.173	0.694	0.096	0.106	0.013	0.289	0.637	0.772	0.363	0.930
	N	54	54	54	54	54	54	53	54	54	53	54	51
fMRI_confidence	R	0.254	-0.192	.413**	-0.182	.362**	-.354**	-.459**	0.129	.304*	0.149	0.105	0.220
	Sig. (2-tailed)	0.064	0.165	0.002	0.188	0.007	0.009	0.001	0.351	0.025	0.287	0.449	0.121
	N	54	54	54	54	54	54	53	54	54	53	54	51
fMRI_confidence_turtle	R	.271*	-0.185	.386**	-0.231	.333*	-.329*	-.337*	0.156	.293*	0.150	0.085	0.229
	Sig. (2-tailed)	0.048	0.180	0.004	0.092	0.014	0.012	0.011	0.260	0.027	0.269	0.532	0.096
	N	54	54	54	54	54	57	56	54	57	56	57	54
fMRI_confidence_turtle2	R	0.219	-0.176	.403**	-0.103	.355**	-.369**	-.359**	0.096	.284*	0.103	0.165	0.154
	Sig. (2-tailed)	0.112	0.203	0.003	0.460	0.008	0.005	0.007	0.490	0.032	0.451	0.220	0.267
	N	54	54	54	54	54	57	56	54	57	56	57	54
fMRI_confidence_diff_t2_t	R	-0.105	0.028	0.001	0.228	0.012	-0.016	0.014	-0.110	-0.062	-0.106	0.124	-0.170
	Sig. (2-tailed)	0.450	0.842	0.997	0.097	0.929	0.903	0.917	0.429	0.648	0.439	0.357	0.218
	N	54	54	54	54	54	57	56	54	57	56	57	54
LTM_syntax	R	0.200	0.121	0.165	-0.119	.310*	-0.249	-.325*	0.029	0.071	0.036	-0.185	0.260
	Sig. (2-tailed)	0.146	0.382	0.233	0.393	0.023	0.070	0.018	0.836	0.610	0.800	0.181	0.065
	N	54	54	54	54	54	54	53	54	54	53	54	51
LTM_sequence	R	0.235	-0.141	.451**	0.144	.405**	-.292*	-.345*	-0.044	0.165	.338*	-.389**	.301*
	Sig. (2-tailed)	0.087	0.311	0.001	0.299	0.002	0.032	0.011	0.751	0.234	0.013	0.004	0.032
	N	54	54	54	54	54	54	53	54	54	53	54	51
LTM_loop	R	0.197	-0.240	.491**	0.150	.360**	-.367**	-.403**	-0.023	0.179	.308*	-.297*	0.247
	Sig. (2-tailed)	0.153	0.080	0.000	0.279	0.008	0.006	0.003	0.871	0.196	0.025	0.029	0.080
	N	54	54	54	54	54	54	53	54	54	53	54	51
LTM_not_write	R	0.234	-0.104	.367**	0.076	.398**	-.475**	-.527**	0.017	0.224	0.253	-0.165	0.239
	Sig. (2-tailed)	0.089	0.455	0.006	0.584	0.003	0.000	0.000	0.903	0.103	0.068	0.232	0.092
	N	54	54	54	54	54	54	53	54	54	53	54	51
LTM_write	R	0.172	-0.231	.490**	0.142	.318*	-0.208	-0.232	-0.051	0.103	.293*	-.405**	0.268
	Sig. (2-tailed)	0.213	0.093	0.000	0.304	0.019	0.132	0.094	0.714	0.461	0.033	0.002	0.058
	N	54	54	54	54	54	54	53	54	54	53	54	51
LTM_sum	R	0.222	-0.189	.478**	0.123	.392**	-.366**	-.423**	-0.021	0.176	.305*	-.324*	.281*
	Sig. (2-tailed)	0.107	0.170	0.000	0.375	0.003	0.006	0.002	0.878	0.204	0.026	0.017	0.046
	N	54	54	54	54	54	54	53	54	54	53	54	51
diff_Test_LTM	R	-0.117	-0.036	0.103	0.133	0.142	-0.229	0.004	-.302*	0.252	.330*	-0.061	.299*
	Sig. (2-tailed)	0.401	0.795	0.461	0.337	0.306	0.096	0.980	0.027	0.067	0.016	0.660	0.033
	N	54	54	54	54	54	54	53	54	54	53	54	51

Correlations

		IMI_PercievedCompetence	IMI_EffortImportance	IMI_PressreTension	IMI_ValueUsefulness	Grade_avg	Grade_math	Grade_art	N_Lang	Grade_Sum_Lang	Grade_Avg_Lang	Test_syntax	Test_sequence
Test_sequence	R	.657**	.316*	-0.235	0.251	.319*	.536**	0.127	0.077	0.199	.270*	.638**	1
	Sig. (2-tailed)	0.000	0.024	0.097	0.075	0.019	0.000	0.359	0.581	0.148	0.049	0.000	
	N	51	51	51	51	54	54	54	54	54	54	54	54
Test_loops	R	.620**	0.231	-0.239	.284*	.362**	.540**	0.077	0.105	0.263	.370**	.661**	.893**
	Sig. (2-tailed)	0.000	0.103	0.091	0.043	0.007	0.000	0.579	0.448	0.055	0.006	0.000	0.000
	N	51	51	51	51	54	54	54	54	54	54	54	54
Test_not_write	R	.572**	0.155	-0.243	0.219	.371**	.585**	0.143	0.147	.305*	.390**	.714**	.849**
	Sig. (2-tailed)	0.000	0.278	0.086	0.122	0.006	0.000	0.304	0.290	0.025	0.004	0.000	0.000
	N	51	51	51	51	54	54	54	54	54	54	54	54
Test_write	R	.674**	.376**	-0.243	.278*	.334*	.486**	0.000	0.104	0.252	.350**	.736**	.897**
	Sig. (2-tailed)	0.000	0.007	0.086	0.048	0.014	0.000	0.998	0.456	0.066	0.010	0.000	0.000
	N	51	51	51	51	54	54	54	54	54	54	54	54
Test_sum	R	.660**	0.275	-0.259	0.263	.377**	.574**	0.082	0.135	.299*	.395**	.771**	.927**
	Sig. (2-tailed)	0.000	0.051	0.067	0.062	0.005	0.000	0.558	0.331	0.028	0.003	0.000	0.000
	N	51	51	51	51	54	54	54	54	54	54	54	54
fMRI_score	R	.290*	0.003	-.355*	-0.142	.397**	.486**	0.045	0.170	.326*	.386**	.479**	.486**
	Sig. (2-tailed)	0.039	0.983	0.011	0.321	0.003	0.000	0.745	0.220	0.016	0.004	0.000	0.000
	N	51	51	51	51	54	54	54	54	54	54	54	54
fMRI_meaning_score	R	0.100	0.011	-0.254	-0.148	.308*	.335*	0.136	0.111	0.212	0.245	.319*	.304*
	Sig. (2-tailed)	0.487	0.941	0.072	0.302	0.023	0.013	0.328	0.423	0.123	0.074	0.019	0.026
	N	51	51	51	51	54	54	54	54	54	54	54	54
fMRI_syntax_score	R	0.196	-0.039	-.297*	-0.099	.374**	.289*	-0.020	0.168	.334*	.415**	.333*	.384**
	Sig. (2-tailed)	0.169	0.787	0.035	0.488	0.005	0.034	0.886	0.226	0.014	0.002	0.014	0.004
	N	51	51	51	51	54	54	54	54	54	54	54	54
fMRI_turtle_score	R	.404**	0.113	-.317*	-0.071	.308*	.517**	0.044	0.129	0.254	.298*	.532**	.545**
	Sig. (2-tailed)	0.003	0.429	0.024	0.619	0.024	0.000	0.750	0.352	0.063	0.029	0.000	0.000
	N	51	51	51	51	54	54	54	54	54	54	54	54
fMRI_turtle2_score	R	0.242	-0.098	-.331*	-0.182	.362**	.499**	0.007	0.167	.301*	.344*	.417**	.383**
	Sig. (2-tailed)	0.088	0.495	0.018	0.202	0.007	0.000	0.960	0.227	0.027	0.011	0.002	0.004
	N	51	51	51	51	54	54	54	54	54	54	54	54
fMRI_yes_score	R	0.237	0.062	-.349*	-0.107	.444**	.550**	0.009	0.100	.311*	.472**	.514**	.454**
	Sig. (2-tailed)	0.094	0.666	0.012	0.453	0.001	0.000	0.947	0.472	0.022	0.000	0.000	0.001
	N	51	51	51	51	54	54	54	54	54	54	54	54
fMRI_no_score	R	0.271	-0.050	-.279*	-0.140	0.266	.319*	0.067	0.194	0.265	0.221	.339*	.405**
	Sig. (2-tailed)	0.054	0.728	0.048	0.327	0.052	0.019	0.629	0.160	0.053	0.108	0.012	0.002
	N	51	51	51	51	54	54	54	54	54	54	54	54
fMRI_confidence	R	0.231	0.046	-.336*	0.106	0.013	0.233	-0.177	0.125	0.104	0.034	0.159	0.095
	Sig. (2-tailed)	0.104	0.748	0.016	0.461	0.926	0.090	0.199	0.369	0.453	0.810	0.251	0.493
	N	51	51	51	51	54	54	54	54	54	54	54	54
fMRI_confidence_turtle	R	.291*	0.056	-.364**	0.196	0.101	.365**	-0.136	0.173	0.190	0.129	.275*	0.189
	Sig. (2-tailed)	0.033	0.685	0.007	0.156	0.466	0.007	0.328	0.211	0.168	0.352	0.044	0.172
	N	54	54	54	54	54	54	54	54	54	54	54	54
fMRI_confidence_turtle2	R	0.174	-0.052	-.326*	0.081	-0.068	0.079	-0.194	0.083	0.025	-0.061	0.038	0.001
	Sig. (2-tailed)	0.208	0.708	0.016	0.558	0.627	0.569	0.161	0.550	0.857	0.660	0.787	0.995
	N	54	54	54	54	54	54	54	54	54	54	54	54
fMRI_confidence_diff_t2_t	R	-0.253	-0.196	0.132	-0.231	-.285*	-.496**	-0.086	-0.160	-.285*	-.322*	-.410**	-.322*
	Sig. (2-tailed)	0.065	0.155	0.342	0.093	0.037	0.000	0.538	0.247	0.037	0.018	0.002	0.017
	N	54	54	54	54	54	54	54	54	54	54	54	54
LTM_syntax	R	.501**	.339*	-.367**	-0.005	0.197	.299*	0.010	0.120	0.165	0.149	.430**	.481**
	Sig. (2-tailed)	0.000	0.015	0.008	0.972	0.153	0.028	0.945	0.386	0.233	0.283	0.001	0.000
	N	51	51	51	51	54	54	54	54	54	54	54	54
LTM_sequence	R	.686**	.420**	-0.260	0.144	.374**	.491**	0.116	0.168	.280*	.301*	.589**	.806**
	Sig. (2-tailed)	0.000	0.002	0.065	0.314	0.005	0.000	0.402	0.225	0.040	0.027	0.000	0.000
	N	51	51	51	51	54	54	54	54	54	54	54	54
LTM_loop	R	.613**	.281*	-0.197	0.149	.417**	.555**	0.137	0.098	0.214	.271*	.552**	.833**
	Sig. (2-tailed)	0.000	0.045	0.166	0.296	0.002	0.000	0.323	0.481	0.120	0.048	0.000	0.000
	N	51	51	51	51	54	54	54	54	54	54	54	54
LTM_not_write	R	.593**	0.190	-.326*	0.050	0.252	.423**	0.007	0.116	0.172	0.162	.474**	.739**
	Sig. (2-tailed)	0.000	0.183	0.020	0.729	0.066	0.001	0.960	0.405	0.214	0.243	0.000	0.000
	N	51	51	51	51	54	54	54	54	54	54	54	54
LTM_write	R	.606**	.411**	-0.140	0.192	.479**	.568**	0.208	0.108	0.254	.337*	.584**	.794**
	Sig. (2-tailed)	0.000	0.003	0.326	0.178	0.000	0.000	0.131	0.438	0.064	0.013	0.000	0.000
	N	51	51	51	51	54	54	54	54	54	54	54	54
LTM_sum	R	.663**	.341*	-0.250	0.139	.413**	.552**	0.127	0.123	0.238	.282*	.588**	.848**
	Sig. (2-tailed)	0.000	0.014	0.077	0.331	0.002	0.000	0.362	0.376	0.083	0.039	0.000	0.000
	N	51	51	51	51	54	54	54	54	54	54	54	54
diff_Test_LTM	R	0.221	-0.011	-0.102	.289*	0.076	0.236	-0.043	0.066	0.201	.318*	.561**	.452**
	Sig. (2-tailed)	0.118	0.938	0.474	0.040	0.584	0.086	0.760	0.634	0.144	0.019	0.000	0.001
	N	51	51	51	51	54	54	54	54	54	54	54	54

Correlations

		Test_loops	Test_not_write	Test_write	Test_sum	fMRI_score	fMRI_meaning_score	fMRI_syntax_score	fMRI_turtle_score	fMRI_turtle2_score	fMRI_yes_score	fMRI_no_score	fMRI_confidence
Test_sequence	R	.893**	.849**	.897**	.927**	.486**	.304*	.384**	.545**	.383**	.454**	.405**	0.095
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000	0.026	0.004	0.000	0.004	0.001	0.002	0.493
	N	54	54	54	54	54	54	54	54	54	54	54	54
Test_loops	R	1	.946**	.898**	.983**	.493**	0.252	.388**	.529**	.472**	.511**	.365**	0.258
	Sig. (2-tailed)		0.000	0.000	0.000	0.000	0.066	0.004	0.000	0.000	0.000	0.007	0.060
	N	54	54	54	54	54	54	54	54	54	54	54	54
Test_not_write	R	.946**	1	.765**	.948**	.485**	0.262	.324*	.530**	.506**	.511**	.352**	0.225
	Sig. (2-tailed)	0.000		0.000	0.000	0.000	0.056	0.017	0.000	0.000	0.000	0.009	0.102
	N	54	54	54	54	54	54	54	54	54	54	54	54
Test_write	R	.898**	.765**	1	.930**	.501**	.293*	.443**	.542**	.382**	.497**	.390**	0.195
	Sig. (2-tailed)	0.000	0.000		0.000	0.000	0.032	0.001	0.000	0.004	0.000	0.004	0.158
	N	54	54	54	54	54	54	54	54	54	54	54	54
Test_sum	R	.983**	.948**	.930**	1	.524**	.294*	.404**	.570**	.477**	.537**	.394**	0.224
	Sig. (2-tailed)	0.000	0.000	0.000		0.000	0.031	0.002	0.000	0.000	0.000	0.003	0.103
	N	54	54	54	54	54	54	54	54	54	54	54	54
fMRI_score	R	.493**	.485**	.501**	.524**	1	.858**	.802**	.915**	.817**	.866**	.894**	0.166
	Sig. (2-tailed)	0.000	0.000	0.000	0.000		0.000	0.000	0.000	0.000	0.000	0.000	0.232
	N	54	54	54	54	54	54	54	54	54	54	54	54
fMRI_meaning_score	R	0.252	0.262	.293*	.294*	.858**	1	.647**	.755**	.538**	.755**	.756**	0.024
	Sig. (2-tailed)	0.066	0.056	0.032	0.031	0.000		0.000	0.000	0.000	0.000	0.000	0.861
	N	54	54	54	54	54	54	54	54	54	54	54	54
fMRI_syntax_score	R	.388**	.324*	.443**	.404**	.802**	.647**	1	.571**	.487**	.642**	.764**	0.122
	Sig. (2-tailed)	0.004	0.017	0.001	0.002	0.000	0.000		0.000	0.000	0.000	0.000	0.380
	N	54	54	54	54	54	54	54	54	54	54	54	54
fMRI_turtle_score	R	.529**	.530**	.542**	.570**	.915**	.755**	.571**	1	.757**	.807**	.806**	0.224
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000	0.000	0.000		0.000	0.000	0.000	0.104
	N	54	54	54	54	54	54	54	54	54	54	54	54
fMRI_turtle2_score	R	.472**	.506**	.382**	.477**	.817**	.538**	.487**	.757**	1	.739**	.702**	0.167
	Sig. (2-tailed)	0.000	0.000	0.004	0.000	0.000	0.000	0.000	0.000		0.000	0.000	0.228
	N	54	54	54	54	54	54	54	54	54	54	54	54
fMRI_yes_score	R	.511**	.511**	.497**	.537**	.866**	.755**	.642**	.807**	.739**	1	.550**	0.168
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000		0.000	0.224
	N	54	54	54	54	54	54	54	54	54	54	54	54
fMRI_no_score	R	.365**	.352**	.390**	.394**	.894**	.756**	.764**	.806**	.702**	.550**	1	0.126
	Sig. (2-tailed)	0.007	0.009	0.004	0.003	0.000	0.000	0.000	0.000	0.000	0.000		0.365
	N	54	54	54	54	54	54	54	54	54	54	54	54
fMRI_confidence	R	0.258	0.225	0.195	0.224	0.166	0.024	0.122	0.224	0.167	0.168	0.126	1
	Sig. (2-tailed)	0.060	0.102	0.158	0.103	0.232	0.861	0.380	0.104	0.228	0.224	0.365	
	N	54	54	54	54	54	54	54	54	54	54	54	54
fMRI_confidence_turtle	R	.335*	.305*	.291*	.318*	.273*	0.143	0.170	.349**	0.242	.296*	0.191	.954**
	Sig. (2-tailed)	0.013	0.025	0.033	0.019	0.045	0.304	0.220	0.010	0.078	0.030	0.166	0.000
	N	54	54	54	54	54	54	54	54	54	54	54	54
fMRI_confidence_turtle2	R	0.163	0.129	0.089	0.118	0.048	-0.093	0.062	0.087	0.086	0.028	0.056	.954**
	Sig. (2-tailed)	0.240	0.352	0.523	0.397	0.728	0.505	0.658	0.534	0.536	0.840	0.689	0.000
	N	54	54	54	54	54	54	54	54	54	54	54	54
fMRI_confidence_diff_t2_t	R	-.307*	-.312*	-.353**	-.352**	-.389**	-.397**	-0.190	-.456**	-.274*	-.462**	-0.236	-0.068
	Sig. (2-tailed)	0.024	0.022	0.009	0.009	0.004	0.003	0.169	0.001	0.045	0.000	0.085	0.626
	N	54	54	54	54	54	54	54	54	54	54	54	54
LTM_syntax	R	.450**	.389**	.531**	.484**	.343*	0.215	.342*	.349**	0.232	.380**	0.232	.282*
	Sig. (2-tailed)	0.001	0.004	0.000	0.000	0.011	0.118	0.011	0.010	0.091	0.005	0.092	0.039
	N	54	54	54	54	54	54	54	54	54	54	54	54
LTM_sequence	R	.770**	.675**	.833**	.796**	.467**	.300*	.355**	.507**	.398**	.504**	.328*	0.106
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000	0.028	0.008	0.000	0.003	0.000	0.015	0.445
	N	54	54	54	54	54	54	54	54	54	54	54	54
LTM_loop	R	.849**	.786**	.808**	.847**	.538**	.328*	.373**	.586**	.513**	.504**	.447**	0.217
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000	0.016	0.005	0.000	0.000	0.000	0.001	0.114
	N	54	54	54	54	54	54	54	54	54	54	54	54
LTM_not_write	R	.769**	.722**	.704**	.759**	.448**	0.208	.391**	.446**	.450**	.402**	.389**	.301*
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.001	0.132	0.003	0.001	0.001	0.003	0.004	0.027
	N	54	54	54	54	54	54	54	54	54	54	54	54
LTM_write	R	.768**	.684**	.813**	.792**	.527**	.385**	.332*	.605**	.445**	.546**	.391**	0.116
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000	0.004	0.014	0.000	0.001	0.000	0.003	0.403
	N	54	54	54	54	54	54	54	54	54	54	54	54
LTM_sum	R	.848**	.774**	.841**	.857**	.542**	.334*	.397**	.586**	.494**	.529**	.431**	0.223
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000	0.014	0.003	0.000	0.000	0.000	0.001	0.105
	N	54	54	54	54	54	54	54	54	54	54	54	54
diff_Test_LTM	R	.560**	.609**	.468**	.579**	0.157	0.040	0.154	0.176	0.142	0.202	0.080	0.081
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.258	0.773	0.265	0.203	0.306	0.142	0.564	0.560
	N	54	54	54	54	54	54	54	54	54	54	54	54

Correlations											
		fMRI_confidence_turtle	fMRI_confidence_turtle2	fMRI_confidence_diff_t2_t	LTM_syntax	LTM_sequence	LTM_loop	LTM_not_write	LTM_write	LTM_sum	diff_Test_LTM
Test_sequence	R	0.189	0.001	-.322*	.481**	.806**	.833**	.739**	.794**	.848**	.452**
	Sig. (2-tailed)	0.172	0.995	0.017	0.000	0.000	0.000	0.000	0.000	0.000	0.001
	N	54	54	54	54	54	54	54	54	54	54
Test_loops	R	.335*	0.163	-.307*	.450**	.770**	.849**	.769**	.768**	.848**	.560**
	Sig. (2-tailed)	0.013	0.240	0.024	0.001	0.000	0.000	0.000	0.000	0.000	0.000
	N	54	54	54	54	54	54	54	54	54	54
Test_not_write	R	.305*	0.129	-.312*	.389**	.675**	.786**	.722**	.684**	.774**	.609**
	Sig. (2-tailed)	0.025	0.352	0.022	0.004	0.000	0.000	0.000	0.000	0.000	0.000
	N	54	54	54	54	54	54	54	54	54	54
Test_write	R	.291*	0.089	-.353**	.531**	.833**	.808**	.704**	.813**	.841**	.468**
	Sig. (2-tailed)	0.033	0.523	0.009	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	N	54	54	54	54	54	54	54	54	54	54
Test_sum	R	.318*	0.118	-.352**	.484**	.796**	.847**	.759**	.792**	.857**	.579**
	Sig. (2-tailed)	0.019	0.397	0.009	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	N	54	54	54	54	54	54	54	54	54	54
fMRI_score	R	.273*	0.048	-.389**	.343*	.467**	.538**	.448**	.527**	.542**	0.157
	Sig. (2-tailed)	0.045	0.728	0.004	0.011	0.000	0.000	0.001	0.000	0.000	0.258
	N	54	54	54	54	54	54	54	54	54	54
fMRI_meaning_score	R	0.143	-0.093	-.397**	0.215	.300*	.328*	0.208	.385**	.334*	0.040
	Sig. (2-tailed)	0.304	0.505	0.003	0.118	0.028	0.016	0.132	0.004	0.014	0.773
	N	54	54	54	54	54	54	54	54	54	54
fMRI_syntax_score	R	0.170	0.062	-0.190	.342*	.355**	.373**	.391**	.332*	.397**	0.154
	Sig. (2-tailed)	0.220	0.658	0.169	0.011	0.008	0.005	0.003	0.014	0.003	0.265
	N	54	54	54	54	54	54	54	54	54	54
fMRI_turtle_score	R	.349**	0.087	-.456**	.349**	.507**	.586**	.446**	.605**	.586**	0.176
	Sig. (2-tailed)	0.010	0.534	0.001	0.010	0.000	0.000	0.001	0.000	0.000	0.203
	N	54	54	54	54	54	54	54	54	54	54
fMRI_turtle2_score	R	0.242	0.086	-.274*	0.232	.398**	.513**	.450**	.445**	.494**	0.142
	Sig. (2-tailed)	0.078	0.536	0.045	0.091	0.003	0.000	0.001	0.001	0.000	0.306
	N	54	54	54	54	54	54	54	54	54	54
fMRI_yes_score	R	.296*	0.028	-.462**	.380**	.504**	.504**	.402**	.546**	.529**	0.202
	Sig. (2-tailed)	0.030	0.840	0.000	0.005	0.000	0.000	0.003	0.000	0.000	0.142
	N	54	54	54	54	54	54	54	54	54	54
fMRI_no_score	R	0.191	0.056	-0.236	0.232	.328*	.447**	.389**	.391**	.431**	0.080
	Sig. (2-tailed)	0.166	0.689	0.085	0.092	0.015	0.001	0.004	0.003	0.001	0.564
	N	54	54	54	54	54	54	54	54	54	54
fMRI_confidence	R	.954**	.954**	-0.068	.282*	0.106	0.217	.301*	0.116	0.223	0.081
	Sig. (2-tailed)	0.000	0.000	0.626	0.039	0.445	0.114	0.027	0.403	0.105	0.560
	N	54	54	54	54	54	54	54	54	54	54
fMRI_confidence_turtle	R	1	.849**	-.404**	.344*	0.191	.302*	.362**	0.215	.312*	0.121
	Sig. (2-tailed)		0.000	0.002	0.011	0.166	0.026	0.007	0.119	0.022	0.381
	N	59	59	59	54	54	54	54	54	54	54
fMRI_confidence_turtle2	R	.849**	1	0.140	0.193	0.021	0.121	0.219	0.016	0.122	0.035
	Sig. (2-tailed)	0.000		0.289	0.161	0.881	0.385	0.112	0.907	0.381	0.802
	N	59	59	59	54	54	54	54	54	54	54
fMRI_confidence_diff_t2_t	R	-.404**	0.140	1	-.273*	-.294*	-.321*	-0.261	-.341*	-.335*	-0.151
	Sig. (2-tailed)	0.002	0.289		0.046	0.031	0.018	0.057	0.012	0.013	0.276
	N	59	59	59	54	54	54	54	54	54	54
LTM_syntax	R	.344*	0.193	-.273*	1	.539**	.486**	.616**	.498**	.610**	-0.028
	Sig. (2-tailed)	0.011	0.161	0.046		0.000	0.000	0.000	0.000	0.000	0.839
	N	54	54	54	54	54	54	54	54	54	54
LTM_sequence	R	0.191	0.021	-.294*	.539**	1	.861**	.765**	.885**	.915**	0.093
	Sig. (2-tailed)	0.166	0.881	0.031	0.000		0.000	0.000	0.000	0.000	0.503
	N	54	54	54	54	54	54	54	54	54	54
LTM_loop	R	.302*	0.121	-.321*	.486**	.861**	1	.876**	.905**	.984**	0.083
	Sig. (2-tailed)	0.026	0.385	0.018	0.000	0.000		0.000	0.000	0.000	0.552
	N	54	54	54	54	54	54	54	54	54	54
LTM_not_write	R	.362**	0.219	-0.261	.616**	.765**	.876**	1	.640**	.891**	0.059
	Sig. (2-tailed)	0.007	0.112	0.057	0.000	0.000	0.000		0.000	0.000	0.672
	N	54	54	54	54	54	54	54	54	54	54
LTM_write	R	0.215	0.016	-.341*	.498**	.885**	.905**	.640**	1	.919**	0.077
	Sig. (2-tailed)	0.119	0.907	0.012	0.000	0.000	0.000	0.000		0.000	0.578
	N	54	54	54	54	54	54	54	54	54	54
LTM_sum	R	.312*	0.122	-.335*	.610**	.915**	.984**	.891**	.919**	1	0.076
	Sig. (2-tailed)	0.022	0.381	0.013	0.000	0.000	0.000	0.000	0.000		0.585
	N	54	54	54	54	54	54	54	54	54	54
diff_Test_LTM	R	0.121	0.035	-0.151	-0.028	0.093	0.083	0.059	0.077	0.076	1
	Sig. (2-tailed)	0.381	0.802	0.276	0.839	0.503	0.552	0.672	0.578	0.585	
	N	54	54	54	54	54	54	54	54	54	54

Appendix IV - DCM models

Detailed description of the A, B & C matrixes of all DCM models entered into Bayesian model selection.

First round of BMS (16 models):

model **1**

	Broca	IFG_op	IFG_tri	MOG	MTG	SFG	SMG	SPL
Broca	1	1	1	0	1	0	0	0
IFG_op	1	1	1	0	0	1	0	0
IFG_tri	1	1	1	0	0	0	0	0
MOG	0	0	0	1	1	0	0	1
MTG	1	0	0	1	1	0	0	0
SFG	0	1	0	0	0	1	1	1
SMG	0	0	0	0	0	1	1	1
SPL	0	0	0	1	0	1	1	1

	Broca	IFG_op	IFG_tri	MOG	MTG	SFG	SMG	SPL
Broca	0	0	0	0	1	0	0	0
IFG_op	1	0	0	0	0	0	0	0
IFG_tri	1	0	0	0	0	0	0	0
MOG	0	0	0	0	0	0	0	0
MTG	0	0	0	0	0	0	0	0
SFG	0	0	0	0	0	0	0	1
SMG	0	0	0	0	0	0	0	0
SPL	0	0	0	0	0	0	0	0

	Visual	Loop
Broca	0	0
IFG_op	0	0
IFG_tri	0	0
MOG	1	0
MTG	0	0
SFG	0	0
SMG	0	0
SPL	0	0

model **2**

	Broca	IFG_op	IFG_tri	MOG	MTG	SFG	SMG	SPL
Broca	1	0	1	0	1	0	0	0
IFG_op	0	1	1	0	0	1	0	0

IFG_tri	1	1	1	0	0	0	0	0
MOG	0	0	0	1	1	0	0	1
MTG	1	0	0	1	1	0	0	0
SFG	0	1	0	0	0	1	1	1
SMG	0	0	0	0	0	1	1	1
SPL	0	0	0	1	0	1	1	1

	Broca	**IFG_op**	**IFG_tri**	**MOG**	**MTG**	**SFG**	**SMG**	**SPL**
Broca	0	0	0	0	0	0	0	0
IFG_op	0	0	0	0	0	0	0	0
IFG_tri	0	0	0	0	0	0	0	0
MOG	0	0	0	0	0	0	0	0
MTG	0	0	0	0	0	0	0	0
SFG	0	0	0	0	0	0	0	0
SMG	0	0	0	0	0	0	0	0
SPL	0	0	0	0	0	0	0	0

	Visual	**Loop**
Broca	0	0
IFG_op	0	0
IFG_tri	0	0
MOG	1	0
MTG	0	0
SFG	0	0
SMG	0	0
SPL	0	0

model **3**

	Broca	**IFG_op**	**IFG_tri**	**MOG**	**MTG**	**SFG**	**SMG**	**SPL**
Broca	1	0	1	0	1	0	0	0
IFG_op	0	1	1	0	0	1	0	0
IFG_tri	1	1	1	0	0	0	0	0
MOG	0	0	0	1	1	0	0	1
MTG	1	0	0	1	1	0	0	0
SFG	0	1	0	0	0	1	1	1
SMG	0	0	0	0	0	1	1	1
SPL	0	0	0	1	0	1	1	1

	Broca	**IFG_op**	**IFG_tri**	**MOG**	**MTG**	**SFG**	**SMG**	**SPL**
Broca	0	0	0	0	0	0	0	0
IFG_op	0	0	0	0	0	0	0	0
IFG_tri	0	0	0	0	0	0	0	0
MOG	0	0	0	0	0	0	0	0
MTG	0	0	0	0	0	0	0	0
SFG	0	0	0	0	0	0	0	1

SMG	0	0	0	0	0	0	0	0
SPL	0	0	0	0	0	0	0	0

	Visual	**Loop**
Broca	0	0
IFG_op	0	0
IFG_tri	0	0
MOG	1	0
MTG	0	0
SFG	0	0
SMG	0	0
SPL	0	0

model **4**

	Broca	**IFG_op**	**IFG_tri**	**MOG**	**MTG**	**SFG**	**SMG**	**SPL**
Broca	1	0	1	0	1	0	0	0
IFG_op	0	1	1	0	0	1	0	0
IFG_tri	1	1	1	0	0	0	0	0
MOG	0	0	0	1	1	0	0	1
MTG	1	0	0	1	1	0	0	0
SFG	0	1	0	0	0	1	1	1
SMG	0	0	0	0	0	1	1	1
SPL	0	0	0	1	0	1	1	1

	Broca	**IFG_op**	**IFG_tri**	**MOG**	**MTG**	**SFG**	**SMG**	**SPL**
Broca	0	0	0	0	1	0	0	0
IFG_op	0	0	0	0	0	0	0	0
IFG_tri	0	0	0	0	0	0	0	0
MOG	0	0	0	0	0	0	0	0
MTG	0	0	0	0	0	0	0	0
SFG	0	0	0	0	0	0	0	0
SMG	0	0	0	0	0	0	0	0
SPL	0	0	0	0	0	0	0	0

	Visual	**Loop**
Broca	0	0
IFG_op	0	0
IFG_tri	0	0
MOG	1	0
MTG	0	0
SFG	0	0
SMG	0	0
SPL	0	0

model **5**

	Broca	**IFG_op**	**IFG_tri**	**MOG**	**MTG**	**SFG**	**SMG**	**SPL**
Broca	1	0	1	0	1	0	0	0

IFG_op	0	1	1	0	0	1	0	0
IFG_tri	1	1	1	0	0	0	0	0
MOG	0	0	0	1	1	0	0	1
MTG	1	0	0	1	1	0	0	0
SFG	0	1	0	0	0	1	1	1
SMG	0	0	0	0	0	1	1	1
SPL	0	0	0	1	0	1	1	1

	Broca	**IFG_op**	**IFG_tri**	**MOG**	**MTG**	**SFG**	**SMG**	**SPL**
Broca	0	0	0	0	0	0	0	0
IFG_op	0	0	0	0	0	0	0	0
IFG_tri	1	0	0	0	0	0	0	0
MOG	0	0	0	0	0	0	0	0
MTG	0	0	0	0	0	0	0	0
SFG	0	0	0	0	0	0	0	0
SMG	0	0	0	0	0	0	0	0
SPL	0	0	0	0	0	0	0	0

	Visual	**Loop**
Broca	0	0
IFG_op	0	0
IFG_tri	0	0
MOG	1	0
MTG	0	0
SFG	0	0
SMG	0	0
SPL	0	0
model	**6**	

	Broca	**IFG_op**	**IFG_tri**	**MOG**	**MTG**	**SFG**	**SMG**	**SPL**
Broca	1	0	1	0	1	0	0	0
IFG_op	0	1	1	0	0	1	0	0
IFG_tri	1	1	1	0	0	0	0	0
MOG	0	0	0	1	1	0	0	1
MTG	1	0	0	1	1	0	0	0
SFG	0	1	0	0	0	1	1	1
SMG	0	0	0	0	0	1	1	1
SPL	0	0	0	1	0	1	1	1

	Broca	**IFG_op**	**IFG_tri**	**MOG**	**MTG**	**SFG**	**SMG**	**SPL**
Broca	0	0	0	0	1	0	0	0
IFG_op	0	0	0	0	0	0	0	0
IFG_tri	1	0	0	0	0	0	0	0
MOG	0	0	0	0	0	0	0	0
MTG	0	0	0	0	0	0	0	0

SFG	0	0	0	0	0	0	0	1
SMG	0	0	0	0	0	0	0	0
SPL	0	0	0	0	0	0	0	0

	Visual	**Loop**
Broca	0	0
IFG_op	0	0
IFG_tri	0	0
MOG	1	0
MTG	0	0
SFG	0	0
SMG	0	0
SPL	0	0
model	**7**	

	Broca	**IFG_op**	**IFG_tri**	**MOG**	**MTG**	**SFG**	**SMG**	**SPL**
Broca	1	1	1	0	1	0	0	0
IFG_op	1	1	1	0	0	1	0	0
IFG_tri	1	1	1	0	0	0	0	0
MOG	0	0	0	1	1	0	0	1
MTG	1	0	0	1	1	0	0	0
SFG	0	1	0	0	0	1	1	1
SMG	0	0	0	0	0	1	1	1
SPL	0	0	0	1	0	1	1	1

	Broca	**IFG_op**	**IFG_tri**	**MOG**	**MTG**	**SFG**	**SMG**	**SPL**
Broca	0	0	0	0	0	0	0	0
IFG_op	0	0	0	0	0	0	0	0
IFG_tri	0	0	0	0	0	0	0	0
MOG	0	0	0	0	0	0	0	0
MTG	0	0	0	0	0	0	0	0
SFG	0	0	0	0	0	0	0	0
SMG	0	0	0	0	0	0	0	0
SPL	0	0	0	0	0	0	0	0

	Visual	**Loop**
Broca	0	0
IFG_op	0	0
IFG_tri	0	0
MOG	1	0
MTG	0	0
SFG	0	0
SMG	0	0
SPL	0	0
model	**8**	

	Broca	IFG_op	IFG_tri	MOG	MTG	SFG	SMG	SPL
Broca	1	1	1	0	1	0	0	0
IFG_op	1	1	1	0	0	1	0	0
IFG_tri	1	1	1	0	0	0	0	0
MOG	0	0	0	1	1	0	0	1
MTG	1	0	0	1	1	0	0	0
SFG	0	1	0	0	0	1	1	1
SMG	0	0	0	0	0	1	1	1
SPL	0	0	0	1	0	1	1	1

	Broca	IFG_op	IFG_tri	MOG	MTG	SFG	SMG	SPL
Broca	0	0	0	0	0	0	0	0
IFG_op	0	0	0	0	0	0	0	0
IFG_tri	0	0	0	0	0	0	0	0
MOG	0	0	0	0	0	0	0	0
MTG	0	0	0	0	0	0	0	0
SFG	0	0	0	0	0	0	0	1
SMG	0	0	0	0	0	0	0	0
SPL	0	0	0	0	0	0	0	0

	Visual	Loop
Broca	0	0
IFG_op	0	0
IFG_tri	0	0
MOG	1	0
MTG	0	0
SFG	0	0
SMG	0	0
SPL	0	0
model	**9**	

	Broca	IFG_op	IFG_tri	MOG	MTG	SFG	SMG	SPL
Broca	1	1	1	0	1	0	0	0
IFG_op	1	1	1	0	0	1	0	0
IFG_tri	1	1	1	0	0	0	0	0
MOG	0	0	0	1	1	0	0	1
MTG	1	0	0	1	1	0	0	0
SFG	0	1	0	0	0	1	1	1
SMG	0	0	0	0	0	1	1	1
SPL	0	0	0	1	0	1	1	1

	Broca	IFG_op	IFG_tri	MOG	MTG	SFG	SMG	SPL
Broca	0	0	0	0	1	0	0	0
IFG_op	0	0	0	0	0	0	0	0
IFG_tri	0	0	0	0	0	0	0	0

MOG	0	0	0	0	0	0	0	0
MTG	0	0	0	0	0	0	0	0
SFG	0	0	0	0	0	0	0	0
SMG	0	0	0	0	0	0	0	0
SPL	0	0	0	0	0	0	0	0

	Visual	**Loop**
Broca	0	0
IFG_op	0	0
IFG_tri	0	0
MOG	1	0
MTG	0	0
SFG	0	0
SMG	0	0
SPL	0	0

model **10**

	Broca	**IFG_op**	**IFG_tri**	**MOG**	**MTG**	**SFG**	**SMG**	**SPL**
Broca	1	1	1	0	1	0	0	0
IFG_op	1	1	1	0	0	1	0	0
IFG_tri	1	1	1	0	0	0	0	0
MOG	0	0	0	1	1	0	0	1
MTG	1	0	0	1	1	0	0	0
SFG	0	1	0	0	0	1	1	1
SMG	0	0	0	0	0	1	1	1
SPL	0	0	0	1	0	1	1	1

	Broca	**IFG_op**	**IFG_tri**	**MOG**	**MTG**	**SFG**	**SMG**	**SPL**
Broca	0	0	0	0	0	0	0	0
IFG_op	0	0	0	0	0	0	0	0
IFG_tri	1	0	0	0	0	0	0	0
MOG	0	0	0	0	0	0	0	0
MTG	0	0	0	0	0	0	0	0
SFG	0	0	0	0	0	0	0	0
SMG	0	0	0	0	0	0	0	0
SPL	0	0	0	0	0	0	0	0

	Visual	**Loop**
Broca	0	0
IFG_op	0	0
IFG_tri	0	0
MOG	1	0
MTG	0	0
SFG	0	0
SMG	0	0

SPL	0	0
model	**11**	

	Broca	**IFG_op**	**IFG_tri**	**MOG**	**MTG**	**SFG**	**SMG**	**SPL**
Broca	1	1	1	0	1	0	0	0
IFG_op	1	1	1	0	0	1	0	0
IFG_tri	1	1	1	0	0	0	0	0
MOG	0	0	0	1	1	0	0	1
MTG	1	0	0	1	1	0	0	0
SFG	0	1	0	0	0	1	1	1
SMG	0	0	0	0	0	1	1	1
SPL	0	0	0	1	0	1	1	1

	Broca	**IFG_op**	**IFG_tri**	**MOG**	**MTG**	**SFG**	**SMG**	**SPL**
Broca	0	0	0	0	0	0	0	0
IFG_op	1	0	0	0	0	0	0	0
IFG_tri	0	0	0	0	0	0	0	0
MOG	0	0	0	0	0	0	0	0
MTG	0	0	0	0	0	0	0	0
SFG	0	0	0	0	0	0	0	0
SMG	0	0	0	0	0	0	0	0
SPL	0	0	0	0	0	0	0	0

	Visual	**Loop**
Broca	0	0
IFG_op	0	0
IFG_tri	0	0
MOG	1	0
MTG	0	0
SFG	0	0
SMG	0	0
SPL	0	0
model	**12**	

	Broca	**IFG_op**	**IFG_tri**	**MOG**	**MTG**	**SFG**	**SMG**	**SPL**
Broca	1	1	1	0	1	0	0	0
IFG_op	1	1	1	0	0	1	0	0
IFG_tri	1	1	1	0	0	0	0	0
MOG	0	0	0	1	1	0	0	1
MTG	1	0	0	1	1	0	0	0
SFG	0	1	0	0	0	1	1	1
SMG	0	0	0	0	0	1	1	1
SPL	0	0	0	1	0	1	1	1

	Broca	**IFG_op**	**IFG_tri**	**MOG**	**MTG**	**SFG**	**SMG**	**SPL**
Broca	0	0	0	0	0	0	0	0

IFG_op	1	0	0	0	0	0	0	0
IFG_tri	1	0	0	0	0	0	0	0
MOG	0	0	0	0	0	0	0	0
MTG	0	0	0	0	0	0	0	0
SFG	0	0	0	0	0	0	0	0
SMG	0	0	0	0	0	0	0	0
SPL	0	0	0	0	0	0	0	0

	Visual	**Loop**
Broca	0	0
IFG_op	0	0
IFG_tri	0	0
MOG	1	0
MTG	0	0
SFG	0	0
SMG	0	0
SPL	0	0
model	**13**	

	Broca	**IFG_op**	**IFG_tri**	**MOG**	**MTG**	**SFG**	**SMG**	**SPL**
Broca	1	0	0	0	1	0	0	0
IFG_op	0	1	1	0	0	1	0	0
IFG_tri	0	1	1	0	0	0	0	0
MOG	0	0	0	1	1	0	0	1
MTG	1	0	0	1	1	0	0	0
SFG	0	1	0	0	0	1	1	1
SMG	0	0	0	0	0	1	1	1
SPL	0	0	0	1	0	1	1	1

	Broca	**IFG_op**	**IFG_tri**	**MOG**	**MTG**	**SFG**	**SMG**	**SPL**
Broca	0	0	0	0	0	0	0	0
IFG_op	0	0	0	0	0	0	0	0
IFG_tri	0	0	0	0	0	0	0	0
MOG	0	0	0	0	0	0	0	0
MTG	0	0	0	0	0	0	0	0
SFG	0	0	0	0	0	0	0	0
SMG	0	0	0	0	0	0	0	0
SPL	0	0	0	0	0	0	0	0

	Visual	**Loop**
Broca	0	0
IFG_op	0	0
IFG_tri	0	0
MOG	1	0
MTG	0	0

SFG	0	0
SMG	0	0
SPL	0	0

model **14**

	Broca	**IFG_op**	**IFG_tri**	**MOG**	**MTG**	**SFG**	**SMG**	**SPL**
Broca	1	0	0	0	1	0	0	0
IFG_op	0	1	1	0	0	1	0	0
IFG_tri	0	1	1	0	0	0	0	0
MOG	0	0	0	1	1	0	0	1
MTG	1	0	0	1	1	0	0	0
SFG	0	1	0	0	0	1	1	1
SMG	0	0	0	0	0	1	1	1
SPL	0	0	0	1	0	1	1	1

	Broca	**IFG_op**	**IFG_tri**	**MOG**	**MTG**	**SFG**	**SMG**	**SPL**
Broca	0	0	0	0	0	0	0	0
IFG_op	0	0	0	0	0	0	0	0
IFG_tri	0	0	0	0	0	0	0	0
MOG	0	0	0	0	0	0	0	0
MTG	0	0	0	0	0	0	0	0
SFG	0	0	0	0	0	0	0	1
SMG	0	0	0	0	0	0	0	0
SPL	0	0	0	0	0	0	0	0

	Visual	**Loop**
Broca	0	0
IFG_op	0	0
IFG_tri	0	0
MOG	1	0
MTG	0	0
SFG	0	0
SMG	0	0
SPL	0	0

model **15**

	Broca	**IFG_op**	**IFG_tri**	**MOG**	**MTG**	**SFG**	**SMG**	**SPL**
Broca	1	0	0	0	1	0	0	0
IFG_op	0	1	1	0	0	1	0	0
IFG_tri	0	1	1	0	0	0	0	0
MOG	0	0	0	1	1	0	0	1
MTG	1	0	0	1	1	0	0	0
SFG	0	1	0	0	0	1	1	1
SMG	0	0	0	0	0	1	1	1
SPL	0	0	0	1	0	1	1	1

	Broca	**IFG_op**	**IFG_tri**	**MOG**	**MTG**	**SFG**	**SMG**	**SPL**

Broca	0	0	0	0	1	0	0	0
IFG_op	0	0	0	0	0	0	0	0
IFG_tri	0	0	0	0	0	0	0	0
MOG	0	0	0	0	0	0	0	0
MTG	0	0	0	0	0	0	0	0
SFG	0	0	0	0	0	0	0	0
SMG	0	0	0	0	0	0	0	0
SPL	0	0	0	0	0	0	0	0

	Visual	**Loop**
Broca	0	0
IFG_op	0	0
IFG_tri	0	0
MOG	1	0
MTG	0	0
SFG	0	0
SMG	0	0
SPL	0	0

model **16**

	Broca	**IFG_op**	**IFG_tri**	**MOG**	**MTG**	**SFG**	**SMG**	**SPL**
Broca	1	0	0	0	1	0	0	0
IFG_op	0	1	1	0	0	1	0	0
IFG_tri	0	1	1	0	0	0	0	0
MOG	0	0	0	1	1	0	0	1
MTG	1	0	0	1	1	0	0	0
SFG	0	1	0	0	0	1	1	1
SMG	0	0	0	0	0	1	1	1
SPL	0	0	0	1	0	1	1	1

	Broca	**IFG_op**	**IFG_tri**	**MOG**	**MTG**	**SFG**	**SMG**	**SPL**
Broca	0	0	0	0	1	0	0	0
IFG_op	0	0	0	0	0	0	0	0
IFG_tri	0	0	0	0	0	0	0	0
MOG	0	0	0	0	0	0	0	0
MTG	0	0	0	0	0	0	0	0
SFG	0	0	0	0	0	0	0	1
SMG	0	0	0	0	0	0	0	0
SPL	0	0	0	0	0	0	0	0

	Visual	**Loop**
Broca	0	0
IFG_op	0	0
IFG_tri	0	0
MOG	1	0

MTG	0	0
SFG	0	0
SMG	0	0
SPL	0	0

Second round of BMS (12 models):

model **1** Best model: 3 inputs, bidirectional B and diagonal B

Best Model [DCM_turtle_X_bidirB_centeredU_3C_diagB(full.B.C.null)_m0001]

	Broca	IFG_op	IFG_tri	MOG	MTG	SFG	SMG	SPL
Broca	1	1	1	0	1	0	0	0
IFG_op	1	1	1	0	0	1	0	0
IFG_tri	1	1	1	0	0	0	0	0
MOG	0	0	0	1	1	0	0	1
MTG	1	0	0	1	1	0	0	0
SFG	0	1	0	0	0	1	1	1
SMG	0	0	0	0	0	1	1	1
SPL	0	0	0	1	0	1	1	1

	Broca	IFG_op	IFG_tri	MOG	MTG	SFG	SMG	SPL
Broca	1	1	1	0	1	0	0	0
IFG_op	1	1	0	0	0	0	0	0
IFG_tri	1	0	1	0	0	0	0	0
MOG	0	0	0	0	0	0	0	0
MTG	1	0	0	0	1	0	0	0
SFG	0	0	0	0	0	1	0	1
SMG	0	0	0	0	0	0	0	0
SPL	0	0	0	0	0	1	0	1

	Visual	Loop
Broca	0	0
IFG_op	0	0
IFG_tri	0	0
MOG	1	0
MTG	1	0
SFG	0	0
SMG	0	0
SPL	1	0

model **2** 1 input, no diagonals

[DCM_turtle_X_bidirB_centeredU_3C_diagB(full.B.C.null)_m0002]

	Broca	IFG_op	IFG_tri	MOG	MTG	SFG	SMG	SPL
Broca	1	1	1	0	1	0	0	0

IFG_op	1	1	1	0	0	1	0	0
IFG_tri	1	1	1	0	0	0	0	0
MOG	0	0	0	1	1	0	0	1
MTG	1	0	0	1	1	0	0	0
SFG	0	1	0	0	0	1	1	1
SMG	0	0	0	0	0	1	1	1
SPL	0	0	0	1	0	1	1	1

	Broca	**IFG_op**	**IFG_tri**	**MOG**	**MTG**	**SFG**	**SMG**	**SPL**
Broca	1	1	1	0	1	0	0	0
IFG_op	1	1	0	0	0	0	0	0
IFG_tri	1	0	1	0	0	0	0	0
MOG	0	0	0	0	0	0	0	0
MTG	1	0	0	0	1	0	0	0
SFG	0	0	0	0	0	1	0	1
SMG	0	0	0	0	0	0	0	0
SPL	0	0	0	0	0	1	0	1

	Visual	**Loop**
Broca	0	0
IFG_op	0	0
IFG_tri	0	0
MOG	1	0
MTG	0	0
SFG	0	0
SMG	0	0
SPL	0	0
model	**3**	3 inputs, no diagonals [DCM_turtle_X_bidirB_centeredU_3C_diagB(full.B.C.null)_m0003]

	Broca	**IFG_op**	**IFG_tri**	**MOG**	**MTG**	**SFG**	**SMG**	**SPL**
Broca	1	1	1	0	1	0	0	0
IFG_op	1	1	1	0	0	1	0	0
IFG_tri	1	1	1	0	0	0	0	0
MOG	0	0	0	1	1	0	0	1
MTG	1	0	0	1	1	0	0	0
SFG	0	1	0	0	0	1	1	1
SMG	0	0	0	0	0	1	1	1
SPL	0	0	0	1	0	1	1	1

	Broca	**IFG_op**	**IFG_tri**	**MOG**	**MTG**	**SFG**	**SMG**	**SPL**
Broca	0	1	1	0	1	0	0	0
IFG_op	1	0	0	0	0	0	0	0
IFG_tri	1	0	0	0	0	0	0	0
MOG	0	0	0	0	0	0	0	0

MTG	1	0	0	0	0	0	0	0
SFG	0	0	0	0	0	0	0	1
SMG	0	0	0	0	0	0	0	0
SPL	0	0	0	0	0	1	0	0

	Visual	**Loop**
Broca	0	0
IFG_op	0	0
IFG_tri	0	0
MOG	1	0
MTG	1	0
SFG	0	0
SMG	0	0
SPL	1	0

model **4** 1 input and diagonals [DCM_turtle_bidirB_centeredU] [DCM_turtle_X_bidirB_centeredU_3C_diagB(full.B.C.null)_m0004]

	Broca	**IFG_op**	**IFG_tri**	**MOG**	**MTG**	**SFG**	**SMG**	**SPL**
Broca	1	1	1	0	1	0	0	0
IFG_op	1	1	1	0	0	1	0	0
IFG_tri	1	1	1	0	0	0	0	0
MOG	0	0	0	1	1	0	0	1
MTG	1	0	0	1	1	0	0	0
SFG	0	1	0	0	0	1	1	1
SMG	0	0	0	0	0	1	1	1
SPL	0	0	0	1	0	1	1	1

	Broca	**IFG_op**	**IFG_tri**	**MOG**	**MTG**	**SFG**	**SMG**	**SPL**
Broca	0	1	1	0	1	0	0	0
IFG_op	1	0	0	0	0	0	0	0
IFG_tri	1	0	0	0	0	0	0	0
MOG	0	0	0	0	0	0	0	0
MTG	1	0	0	0	0	0	0	0
SFG	0	0	0	0	0	0	0	1
SMG	0	0	0	0	0	0	0	0
SPL	0	0	0	0	0	1	0	0

	Visual	**Loop**
Broca	0	0
IFG_op	0	0
IFG_tri	0	0
MOG	1	0
MTG	0	0
SFG	0	0
SMG	0	0

SPL	0	0

model **5** The Monster

	Broca	**IFG_op**	**IFG_tri**	**MOG**	**MTG**	**SFG**	**SMG**	**SPL**
Broca	1	1	1	1	1	1	1	1
IFG_op	1	1	1	1	1	1	1	1
IFG_tri	1	1	1	1	1	1	1	1
MOG	1	1	1	1	1	1	1	1
MTG	1	1	1	1	1	1	1	1
SFG	1	1	1	1	1	1	1	1
SMG	1	1	1	1	1	1	1	1
SPL	1	1	1	1	1	1	1	1

	Broca	**IFG_op**	**IFG_tri**	**MOG**	**MTG**	**SFG**	**SMG**	**SPL**
Broca	1	1	1	1	1	1	1	1
IFG_op	1	1	1	1	1	1	1	1
IFG_tri	1	1	1	1	1	1	1	1
MOG	1	1	1	1	1	1	1	1
MTG	1	1	1	1	1	1	1	1
SFG	1	1	1	1	1	1	1	1
SMG	1	1	1	1	1	1	1	1
SPL	1	1	1	1	1	1	1	1

	Visual	**Loop**
Broca	0	0
IFG_op	0	0
IFG_tri	0	0
MOG	1	0
MTG	1	0
SFG	0	0
SMG	0	0
SPL	1	0

model 6 DCM_Model1_full_B

	Broca	**IFG_op**	**IFG_tri**	**MOG**	**MTG**	**SFG**	**SMG**	**SPL**
Broca	1	1	1	0	1	0	0	0
IFG_op	1	1	1	0	0	1	0	0
IFG_tri	1	1	1	0	0	0	0	0
MOG	0	0	0	1	1	0	0	1
MTG	1	0	0	1	1	0	0	0
SFG	0	1	0	0	0	1	1	1
SMG	0	0	0	0	0	1	1	1
SPL	0	0	0	1	0	1	1	1

	Broca	**IFG_op**	**IFG_tri**	**MOG**	**MTG**	**SFG**	**SMG**	**SPL**
Broca	1	1	1	0	1	0	0	0

IFG_op	1	1	1	0	0	1	0	0
IFG_tri	1	1	1	0	0	0	0	0
MOG	0	0	0	1	1	0	0	1
MTG	1	0	0	1	1	0	0	0
SFG	0	1	0	0	0	1	1	1
SMG	0	0	0	0	0	1	1	1
SPL	0	0	0	1	0	1	1	1

	Visual	**Loop**
Broca	0	0
IFG_op	0	0
IFG_tri	0	0
MOG	1	0
MTG	1	0
SFG	0	0
SMG	0	0
SPL	1	0

model **7** Monster top 42

				FROM				
	Broca	**IFG_op**	**IFG_tri**	**MOG**	**MTG**	**SFG**	**SMG**	**SPL**
Broca	1	0	0	1	1	0	0	0
IFG_op	0	1	1	1	1	0	0	1
IFG_tri	0	0	1	1	0	0	0	0
MOG	0	0	1	1	1	0	1	0
MTG	0	0	1	1	1	0	0	0
SFG	0	1	1	1	1	1	0	0
SMG	0	0	1	1	1	0	1	0
SPL	0	0	1	1	1	1	1	1

	Broca	**IFG_op**	**IFG_tri**	**MOG**	**MTG**	**SFG**	**SMG**	**SPL**
Broca	0	0	0	0	0	0	0	0
IFG_op	0	0	0	1	0	0	0	0
IFG_tri	0	0	0	0	0	0	0	0
MOG	0	1	0	1	1	1	0	1
MTG	0	0	0	0	1	0	0	0
SFG	0	0	0	1	0	0	0	0
SMG	0	0	0	0	0	0	0	0
SPL	0	0	0	0	1	0	0	1

	Visual	**Loop**
Broca	0	0
IFG_op	0	0
IFG_tri	0	0
MOG	1	0

MTG	1	0
SFG	0	0
SMG	0	0
SPL	1	0

model **8** Monster top 57 (all prob>0)

				FROM				
	Broca	**IFG_op**	**IFG_tri**	**MOG**	**MTG**	**SFG**	**SMG**	**SPL**
Broca	1	0	1	1	1	1	1	0
IFG_op	1	1	1	1	1	0	1	1
IFG_tri	1	0	1	1	1	0	0	0
MOG	0	1	1	1	1	0	1	1
MTG	1	0	1	1	1	0	0	0
SFG	1	1	1	1	1	1	0	1
SMG	0	0	1	1	1	1	1	1
SPL	0	1	1	1	1	1	1	1

	Broca	**IFG_op**	**IFG_tri**	**MOG**	**MTG**	**SFG**	**SMG**	**SPL**
Broca	0	0	0	0	0	0	0	0
IFG_op	0	0	0	1	0	0	0	0
IFG_tri	0	0	0	0	0	0	0	0
MOG	0	1	0	1	1	1	0	1
MTG	0	0	0	0	1	0	0	0
SFG	0	0	0	1	0	0	0	0
SMG	0	0	0	0	0	0	0	0
SPL	0	0	0	0	1	0	0	1

	Visual	**Loop**
Broca	0	0
IFG_op	0	0
IFG_tri	0	0
MOG	1	0
MTG	1	0
SFG	0	0
SMG	0	0
SPL	1	0

model 9 **1_cut** no conn between SFG and op

	Broca	**IFG_op**	**IFG_tri**	**MOG**	**MTG**	**SFG**	**SMG**	**SPL**
Broca	1	1	1	0	1	0	0	0
IFG_op	1	1	1	0	0	0	0	0
IFG_tri	1	1	1	0	0	0	0	0
MOG	0	0	0	1	1	0	0	1
MTG	1	0	0	1	1	0	0	0
SFG	0	0	0	0	0	1	1	1

SMG	0	0	0	0	0	1	1	1
SPL	0	0	0	1	0	1	1	1

	Broca	**IFG_op**	**IFG_tri**	**MOG**	**MTG**	**SFG**	**SMG**	**SPL**
Broca	1	1	1	0	1	0	0	0
IFG_op	1	1	0	0	0	0	0	0
IFG_tri	1	0	1	0	0	0	0	0
MOG	0	0	0	0	0	0	0	0
MTG	1	0	0	0	1	0	0	0
SFG	0	0	0	0	0	1	0	1
SMG	0	0	0	0	0	0	0	0
SPL	0	0	0	0	0	1	0	1

	Visual	**Loop**
Broca	0	0
IFG_op	0	0
IFG_tri	0	0
MOG	1	0
MTG	1	0
SFG	0	0
SMG	0	0
SPL	1	0
model	**10**	1_noSMG

				FROM				
	Broca	**IFG_op**	**IFG_tri**	**MOG**	**MTG**	**SFG**	**SMG**	**SPL**
Broca	1	1	1	0	1	0	0	0
IFG_op	1	1	1	0	0	1	0	0
IFG_tri	1	1	1	0	0	0	0	0
MOG	0	0	0	1	1	0	0	1
MTG	1	0	0	1	1	0	0	0
SFG	0	1	0	0	0	1	0	1
SMG	0	0	0	0	0	0	0	0
SPL	0	0	0	1	0	1	0	1

	Broca	**IFG_op**	**IFG_tri**	**MOG**	**MTG**	**SFG**	**SMG**	**SPL**
Broca	1	1	1	0	1	0	0	0
IFG_op	1	1	0	0	0	0	0	0
IFG_tri	1	0	1	0	0	0	0	0
MOG	0	0	0	0	0	0	0	0
MTG	1	0	0	0	1	0	0	0
SFG	0	0	0	0	0	1	0	1
SMG	0	0	0	0	0	0	0	0
SPL	0	0	0	0	0	1	0	1

	Visual	**Loop**

Broca	0	0
IFG_op	0	0
IFG_tri	0	0
MOG	1	0
MTG	1	0
SFG	0	0
SMG	0	0
SPL	1	0
model	**11**	1_MTGtoSPL

				FROM				
	Broca	**IFG_op**	**IFG_tri**	**MOG**	**MTG**	**SFG**	**SMG**	**SPL**
Broca	1	1	1	0	1	0	0	0
IFG_op	1	1	1	0	0	1	0	0
IFG_tri	1	1	1	0	0	0	0	0
MOG	0	0	0	1	1	0	0	1
MTG	1	0	0	1	1	0	0	1
SFG	0	1	0	0	0	1	1	1
SMG	0	0	0	0	0	1	1	1
SPL	0	0	0	1	1	1	1	1

	Broca	**IFG_op**	**IFG_tri**	**MOG**	**MTG**	**SFG**	**SMG**	**SPL**
Broca	1	1	1	0	1	0	0	0
IFG_op	1	1	0	0	0	0	0	0
IFG_tri	1	0	1	0	0	0	0	0
MOG	0	0	0	0	0	0	0	0
MTG	1	0	0	0	1	0	0	0
SFG	0	0	0	0	0	1	0	1
SMG	0	0	0	0	0	0	0	0
SPL	0	0	0	0	0	1	0	1

	Visual	**Loop**
Broca	0	0
IFG_op	0	0
IFG_tri	0	0
MOG	1	0
MTG	1	0
SFG	0	0
SMG	0	0
SPL	1	0
model	**12**	1_MTGtoSFG

				FROM				
	Broca	**IFG_op**	**IFG_tri**	**MOG**	**MTG**	**SFG**	**SMG**	**SPL**
Broca	1	1	1	0	1	0	0	0
IFG_op	1	1	1	0	0	1	0	0

IFG_tri	1	1	1	0	0	0	0	0
MOG	0	0	0	1	1	0	0	1
MTG	1	0	0	1	1	1	0	0
SFG	0	1	0	0	1	1	1	1
SMG	0	0	0	0	0	1	1	1
SPL	0	0	0	1	0	1	1	1

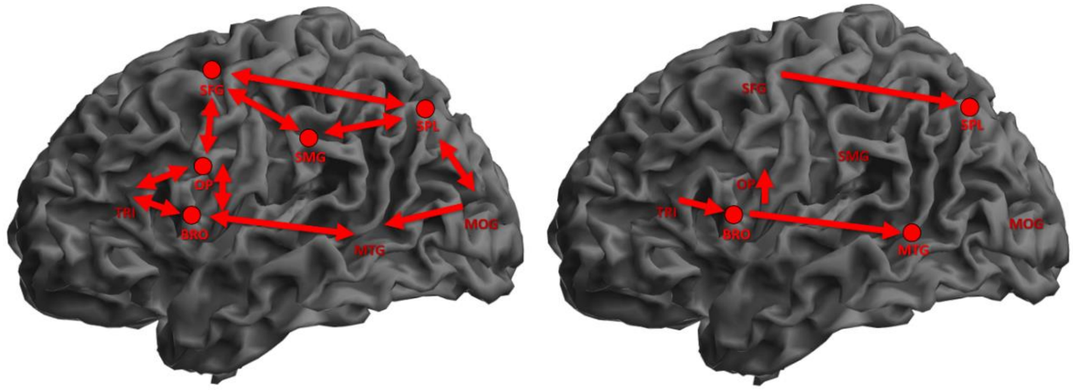
	Broca	**IFG_op**	**IFG_tri**	**MOG**	**MTG**	**SFG**	**SMG**	**SPL**
Broca	1	1	1	0	1	0	0	0
IFG_op	1	1	0	0	0	0	0	0
IFG_tri	1	0	1	0	0	0	0	0
MOG	0	0	0	0	0	0	0	0
MTG	1	0	0	0	1	0	0	0
SFG	0	0	0	0	0	1	0	1
SMG	0	0	0	0	0	0	0	0
SPL	0	0	0	0	0	1	0	1

	Visual	**Loop**
Broca	0	0
IFG_op	0	0
IFG_tri	0	0
MOG	1	0
MTG	1	0
SFG	0	0
SMG	0	0
SPL	1	0

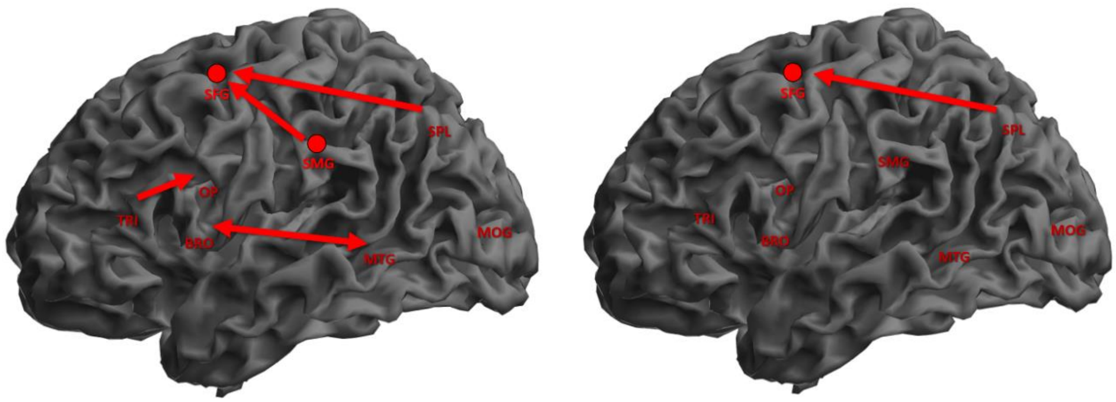
Appendix V - First generation PEB model with corresponding LOO validation.

Connection	commonalities		Hands		LTM		Stress After		Conf IS-S		Math grade		Grade avg	
	S	I	S	I	S	I	S	I	S	I	S	I	S	I
Broca -> Broca	-0.551	-1.14						0.157						
Broca -> IFG_op	0.96	0.167							0.319					-0.2
Broca -> IFG_tri	0.156													
Broca -> MTG	0.23	-0.156	-0.84		0.13									
IFG_op -> Broca	0.89						-0.27							
IFG_op -> IFG_op	-0.256					0.34			0.435				-0.08	-0.16
IFG_op -> IFG_tri	0.33	-		-	-0.7	-		-	-0.167	-		-		-
IFG_op -> SFG	-0.151	-		-		-	0.34	-	-0.29	-		-		-
IFG_tri -> Broca	-0.146	-0.179							-0.157					
IFG_tri -> IFG_op	-0.334	-	0.69	-		-		-	0.366	-		-	-0.06	-
IFG_tri -> IFG_tri					-0.19				-0.283					
MOG -> MOG		-		-		-	-0.79	-		-		-		-
MOG -> MTG	-0.148	-		-		-		-		-	0.38	-		-
MOG -> SPL	-0.527	-		-		-	0.92	-	-0.271	-		-		-
MTG -> Broca	0.234		-0.97											
MTG -> MOG		-		-	-0.12	-	0.4	-		-	0.91	-	0.12	-
MTG -> MTG		-0.489							-0.341					
SFG -> IFG_op	0.553	-		-	-0.1	-		-		-		-		-
SFG -> SFG	-0.573		0.114	-0.454										
SFG -> SMG	0.58	-		-	-0.9	-		-		-	0.24	-		-
SFG -> SPL	0.833	-0.66					-0.76			-0.651	0.64	-0.11		-0.16
SMG -> SFG	-0.457	-	-0.71	-		-		-		-	0.37	-	0.06	-
SMG -> SMG	-0.165	-	0.71	-		-	0.44	-		-	0.59	-	0.08	-
SMG -> SPL	-0.358	-		-		-		-	0.356	-		-		-
SPL -> MOG	0.139	-					-0.31							-
SPL -> SFG	0.371		0.82	-0.95			-0.41		0.149					
SPL -> SMG	0.99	-		-		-		-		-		-		-
SPL -> SPL	-0.211	-1.15								-1.187		-0.14		-0.28

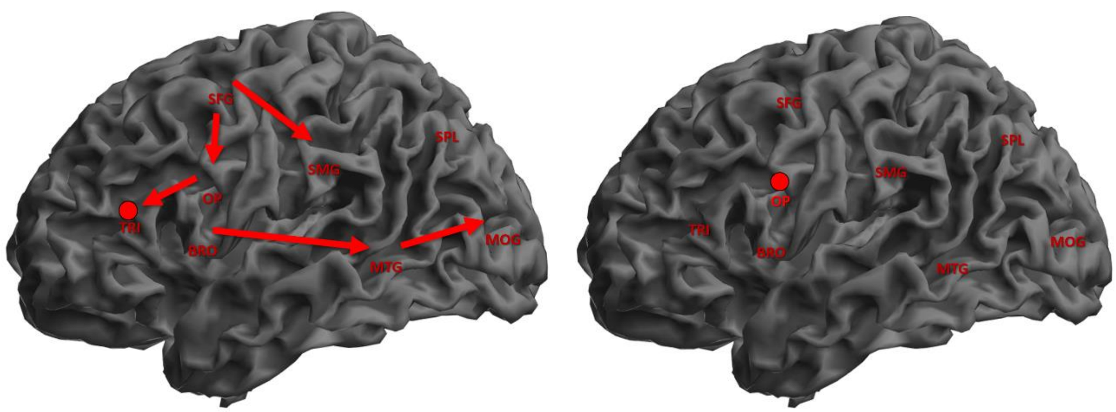
PEB Parameter estimates for all parameters with a P>.99 of being non-zero (significantly contributing towards the model evidence). A dash (-) signifies a connection not present in the model. Commonalities: the underlying connection strengths shared by both tasks, that is then modulated by the parameters for the covariates included in the model. S: Spatial task connection strengths (A parameters) I: modification of connection strength when switching to the iterative task (B parameters)



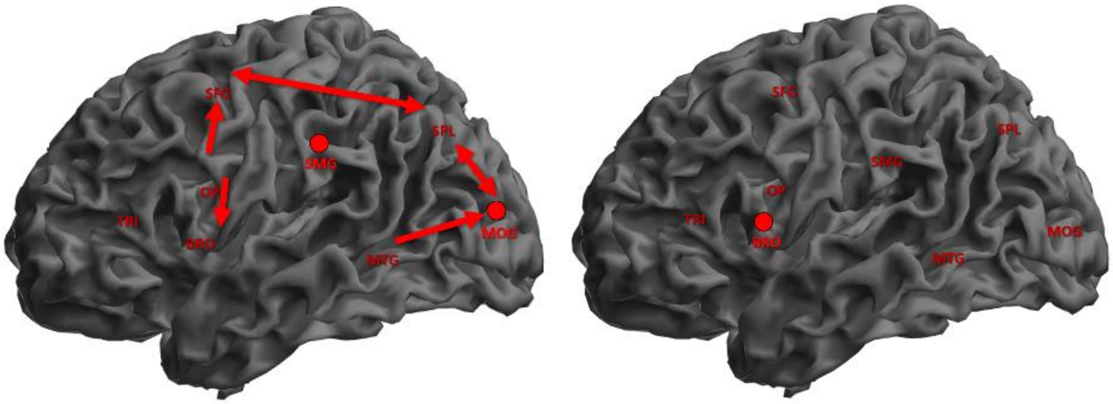
Commonalities



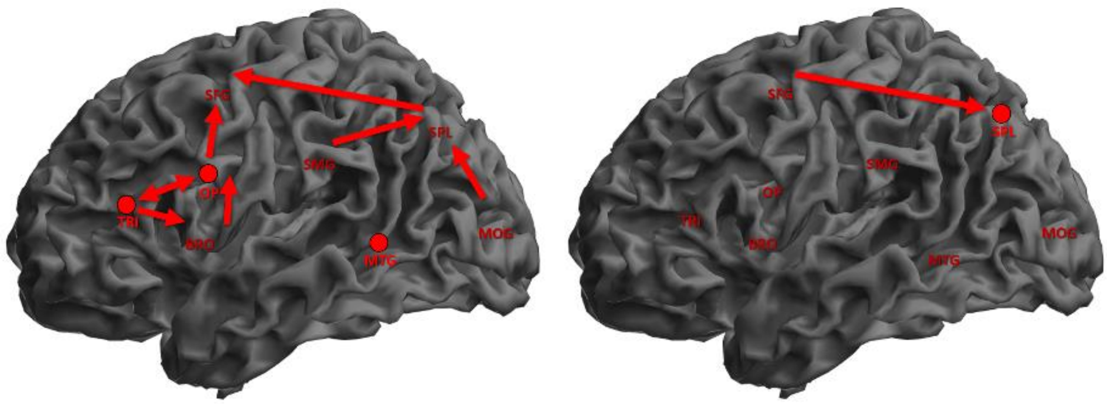
Hands-on



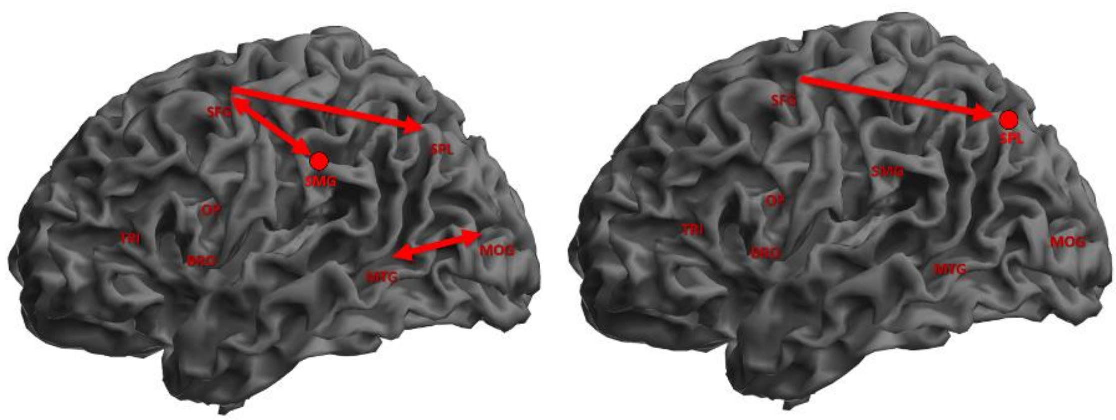
LTM



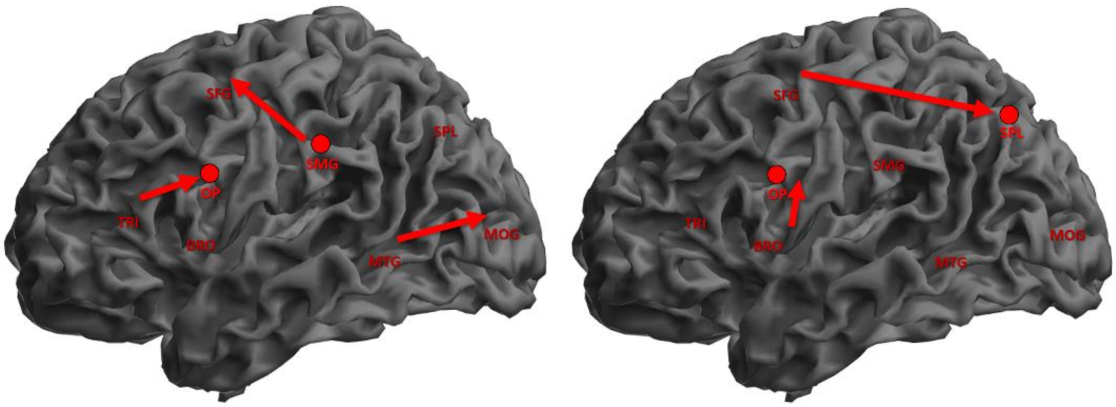
Stress after lesson



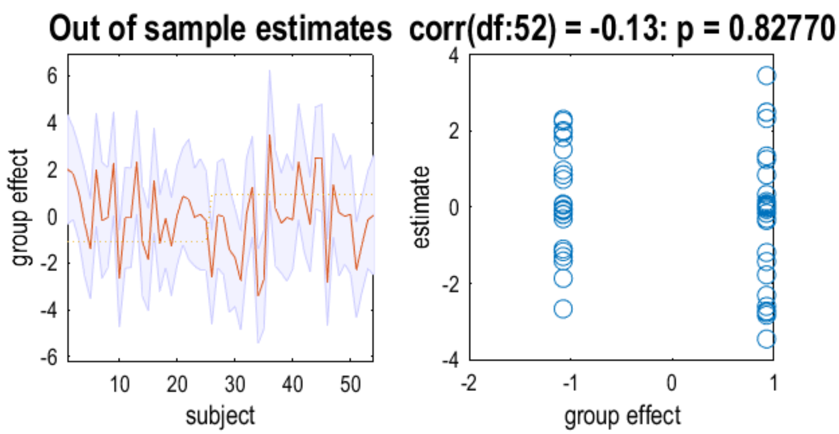
Confidence decrease



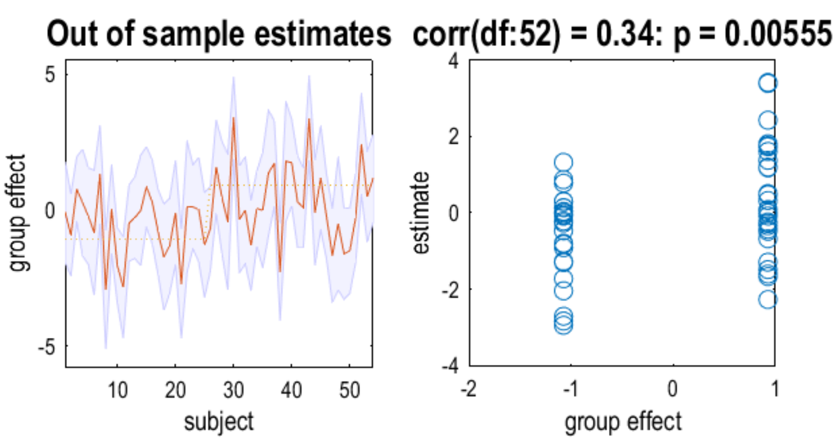
Math grade



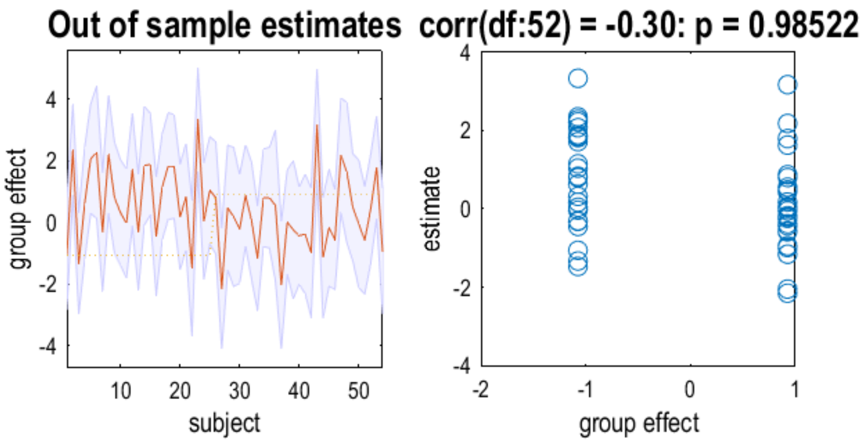
Average grade



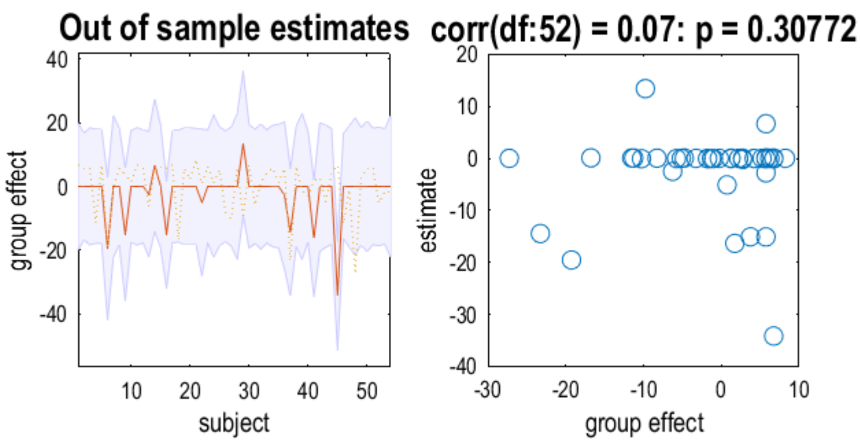
Hands-on A



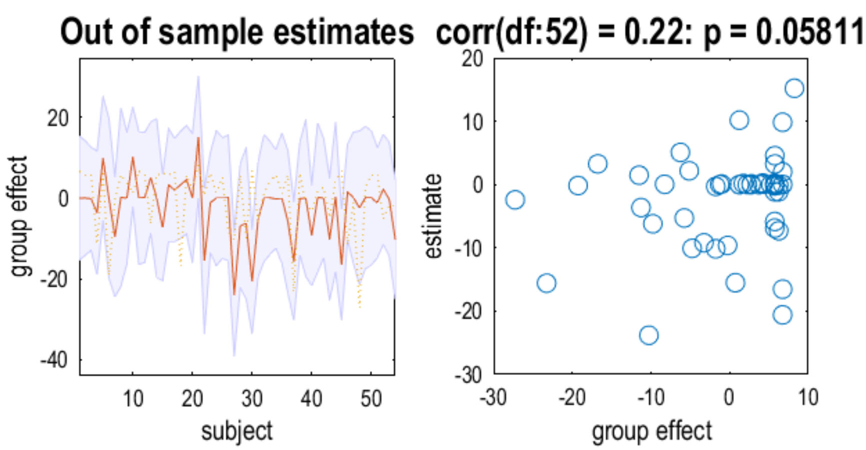
Hands-on B



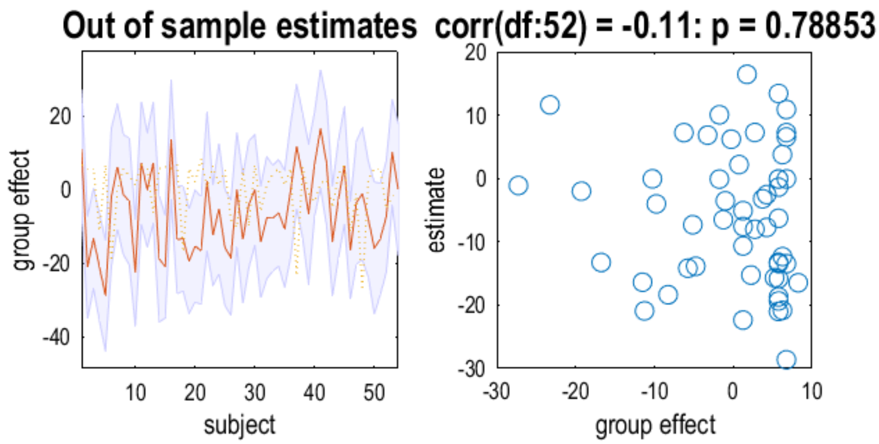
Hands-on AB



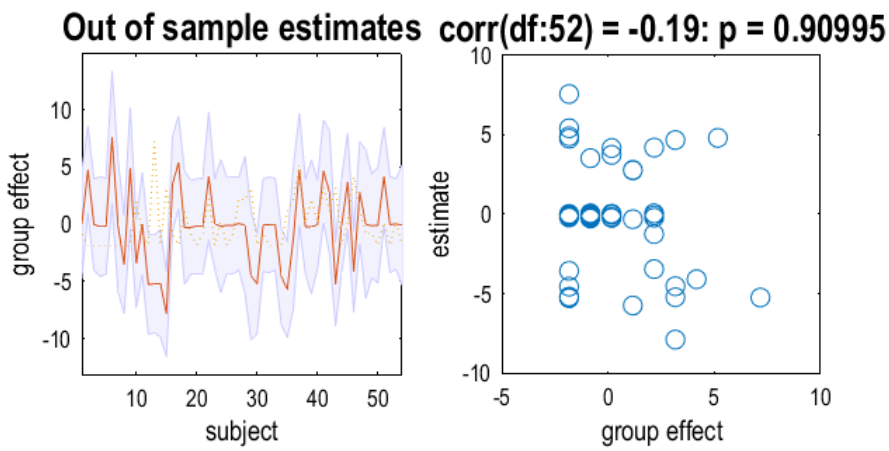
LTM A



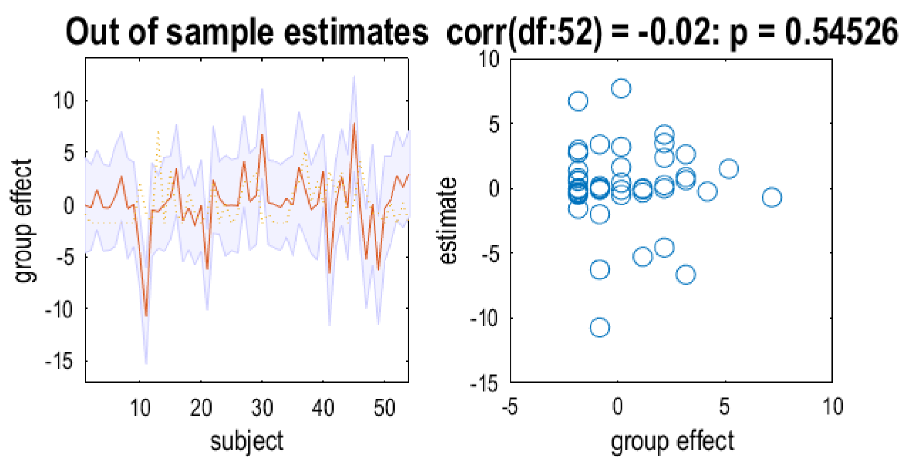
LTM B



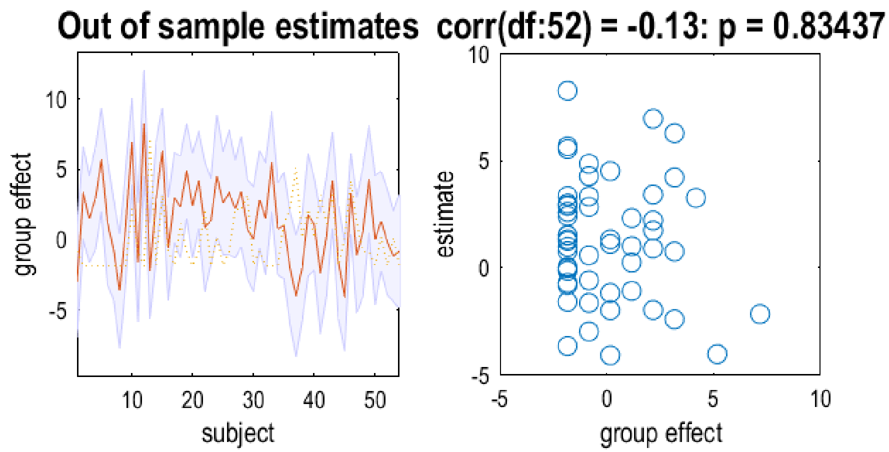
LTM AB



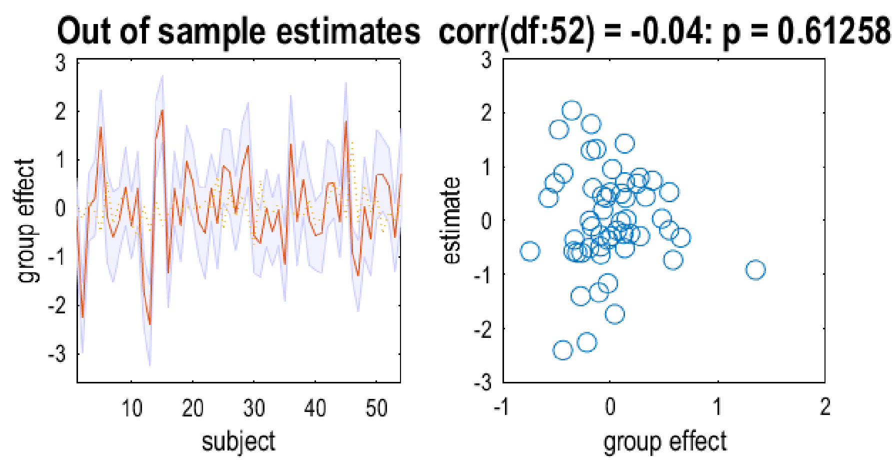
Stress after A



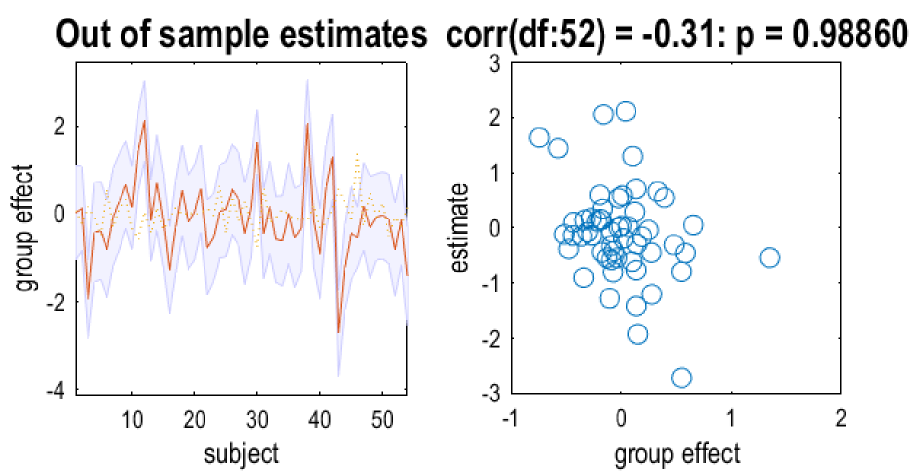
Stress after B



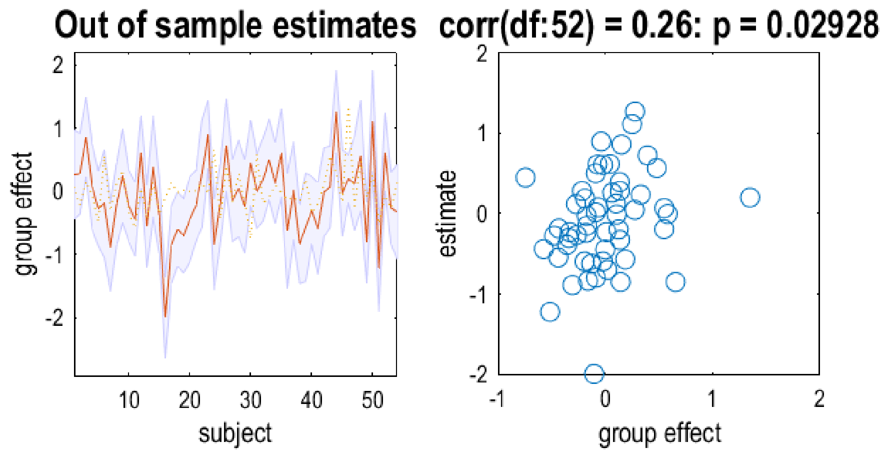
Stress after AB



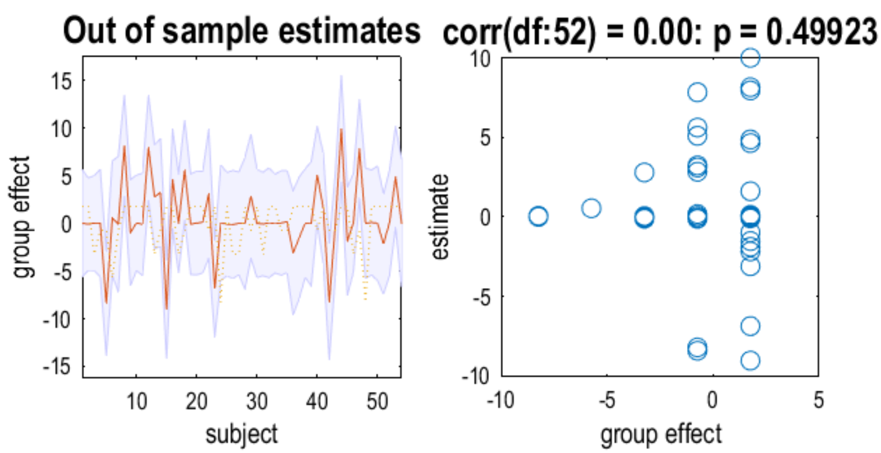
Confidence drop A



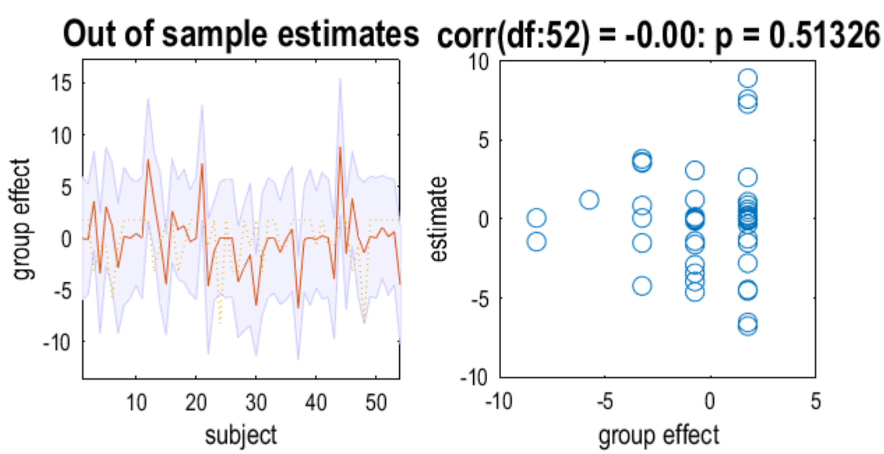
Confidence drop B



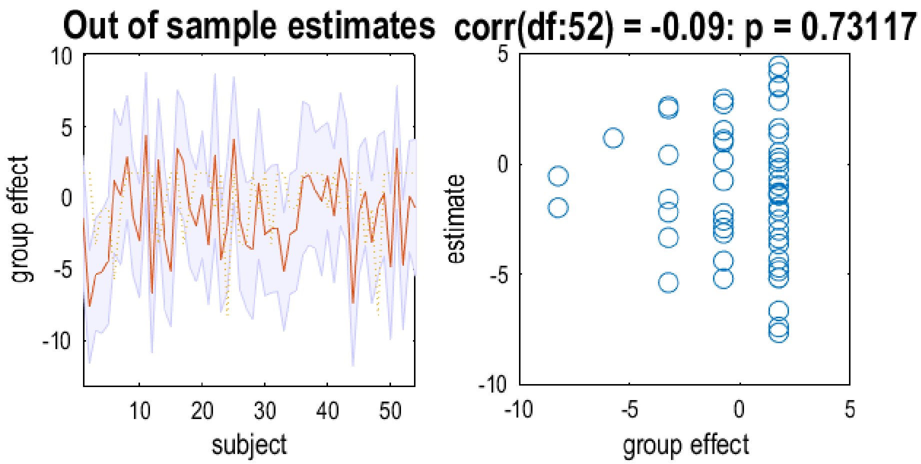
Confidence drop AB



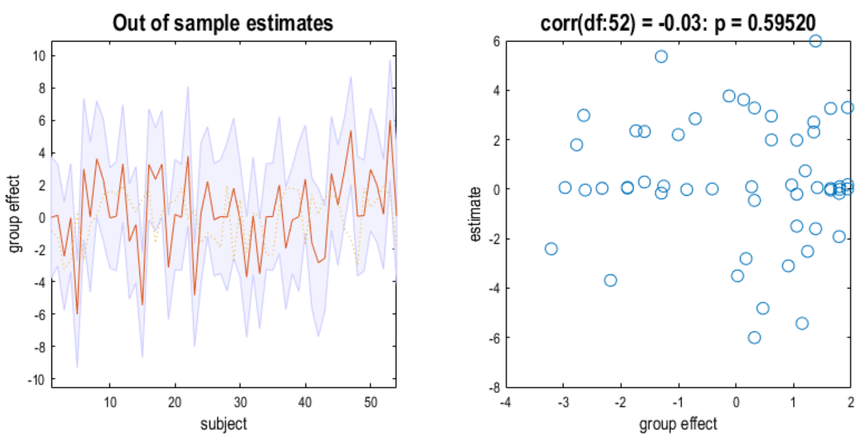
Math A



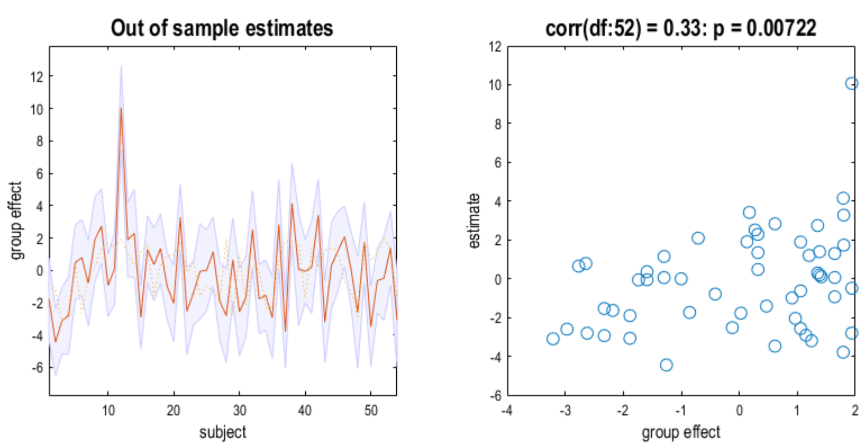
Math B



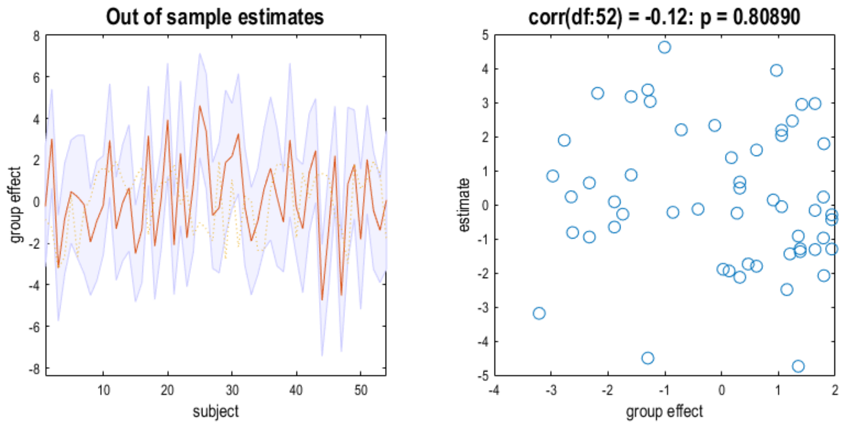
Math AB



Grades A



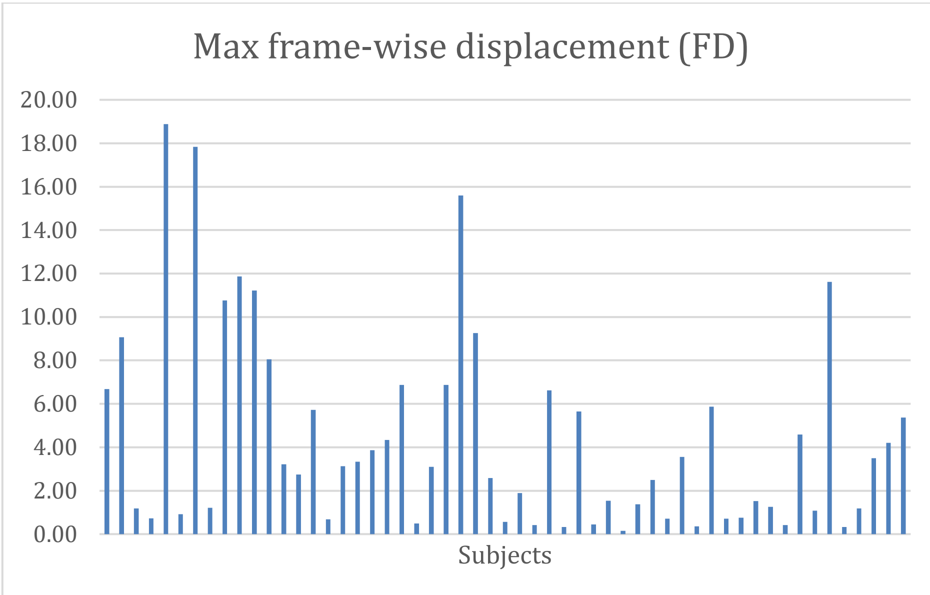
Grades B



Out of sample estimates
group effect
subject
corr(df:52) = -0.12: p = 0.80890
estimate
group effect

Grades AB

Appendix VI - Subject motion analysis



subject	max Δtrans	max Δrot	max FD	# outliers	% of 1836 vols
1	**3.69**	2.56	6.68	**20**	**1.1**
3	**2.26**	**4.21**	9.07	**22**	**1.2**
5	0.88	0.91	1.18	0	0.0
7	0.45	1.96	0.72	0	0.0
8	**4.40**	**8.61**	18.88	**26**	**1.4**
9	1.32	1.17	0.92	0	0.0
10	**4.97**	**9.16**	17.84	**79**	**4.3**
11	0.68	1.10	1.21	0	0.0
12	**4.10**	**7.23**	10.77	19	1.0
13	**2.17**	**4.96**	11.87	**44**	**2.4**
14	**3.46**	2.36	11.22	**16**	**0.9**
15	**3.58**	2.86	8.04	13	0.7
16	**2.20**	0.99	3.22	2	0.1
17	1.33	1.06	2.74	0	0.0
19	1.41	**4.04**	5.73	6	0.3
20	1.15	1.19	0.68	0	0.0
21	1.19	1.05	3.13	1	0.1
22	**4.68**	1.69	3.33	2	0.1

23	**2.28**	1.62	3.87	**19**	**1.0**
24	**2.77**	**4.86**	4.33	3	0.2
25	**2.00**	**4.55**	6.87	**28**	**1.5**
26	0.64	0.35	0.49	0	0.0
27	1.44	1.19	3.11	1	0.1
28	**5.26**	2.78	6.88	5	0.3
29	**3.86**	**6.62**	15.60	**77**	**4.2**
30	**3.11**	**5.20**	9.26	10	0.5
31	*1.43*	*1.62*	*2.58*	*0*	*0.0*
32	*0.43*	*0.27*	*0.56*	*0*	*0.0*
33	*0.71*	*1.51*	*1.89*	*0*	*0.0*
34	*0.40*	*1.25*	*0.42*	*0*	*0.0*
35	*4.36*	*1.51*	*6.62*	*3*	*0.2*
36	*1.56*	*0.95*	*0.33*	*0*	*0.0*
37	*2.00*	*2.96*	*5.64*	*6*	*0.3*
38	*1.86*	*1.70*	*0.45*	*0*	*0.0*
39	*0.83*	*0.56*	*1.54*	*0*	*0.0*
40	0.91	0.54	0.15	0	0.0
41	0.84	0.76	1.37	0	0.0
42	3.66	1.91	2.49	0	0.0
43	0.58	1.03	0.71	0	0.0
44	1.41	1.47	3.56	1	0.1
45	0.88	1.41	0.35	0	0.0
46	3.51	3.21	5.88	10	0.5
47	0.86	0.64	0.71	0	0.0
48	0.53	0.91	0.75	0	0.0
49	1.18	1.72	1.52	0	0.0
50	2.03	0.67	1.26	0	0.0
51	1.16	0.68	0.42	0	0.0
52	2.00	1.78	4.58	2	0.1
53	1.46	1.91	1.08	0	0.0
54	5.21	5.36	11.62	20	1.1
55	0.75	1.57	0.33	0	0.0
56	0.65	0.74	1.19	0	0.0
57	1.80	1.88	3.49	2	0.1
58	2.37	1.58	4.20	3	0.2
59	3.63	4.52	5.36	5	0.3

Maximum rotation(mm at the surface of the brain) and translation(mm) sizes across all three fMRI sessions for each subject. Maximum Frame-wise displacement and number of outliers was calculated using the "Motion

Fingerprint" SPM extension by Marko Wilke (http://www.medizin.uni-tuebingen.de/kinder/en/research/neuroimaging/software/) (M. Wilke, 2012; Marko Wilke, 2014). Last column shows number of outlier volumes as a percentage of the total number of volumes.

Appendix VII – Full One-way ANOVA table

	Sum of Squares	df	Mean Square	F	Sig.	
Test Score	77.812	1.000	77.812	.770	.192	
fMRI Sore	130.809	1.000	130.809	.651	.212	
fMRI Confidence	1.108	1.000	1.108	3.586	.032	*
LTM Score	174.918	1.000	174.918	2.683	.054	*
Stress after lesson	15.324	1.000	15.324	3.347	.037	*
Motivation after lesson	1.822	1.000	1.822	.538	.233	
Stress decrease	3.459	1.000	3.459	1.253	.134	
Motivation decrease	1.170	1.000	1.170	.287	.297	
Test syntax score	5.367	1.000	5.367	1.366	.124	
Test meaning score	.455	1.000	.455	.114	.368	
Test loop score	33.986	1.000	33.986	.737	.197	
Test multiple choice score	.956	1.000	.956	.029	.433	
Test written score	61.515	1.000	61.515	2.586	.057	*
fMRI meaning score	9.012	1.000	9.012	.698	.204	
fMRI syntax score	16.022	1.000	16.022	.827	.184	
fMRI spatial score	.380	1.000	.380	.016	.450	
fMRI iterative(I) spatial(S) score	14.560	1.000	14.560	.997	.161	
fMRI confidence S	1.450	1.000	1.450	4.115	.024	*
fMRI confidence IS	.869	1.000	.869	2.612	.056	*
fMRI confidence decrease S vs IS	.074	1.000	.074	.580	.225	
LTM syntax score	2.886	1.000	2.886	2.176	.073	
LTM meaning score	7.497	1.000	7.497	3.049	.043	*
LTM loop score	77.244	1.000	77.244	2.100	.077	
LTM multiple choice score	51.160	1.000	51.160	3.002	.045	*
LTM written score	36.881	1.000	36.881	1.589	.107	
difference in score Test vs LTM	19.400	1.000	19.400	.718	.200	
NFC	1.465	1.000	1.465	.005	.471	
Stress before lesson	3.669	1.000	3.669	.775	.191	
Motivation before lesson	1.484	1.000	1.484	.445	.254	
Grade average	.070	1.000	.070	.025	.437	
Grade math	.922	1.000	.922	.139	.355	
Grade art	6.579	1.000	6.579	1.215	.138	
Number of languages studied	.472	1.000	.472	1.768	.095	
Grade Sum of Languages	153.008	1.000	153.008	.965	.165	
Grade average of Languages	.010	1.000	.010	.002	.480	

ANOVA comparing our variables for subjects in the Hands-on versus the Hands-off conditions. (one-tailed sig. values reported)