ARN
ASSOCIATION POUR LA RECHERCHE EN NEUROÉDUCATION / ASSOCIATION FOR RESEARCH IN NEUROEDUCATION

Revue NEUROEDUCATION Journal

2018 – Vol. 5, No. 2

EXPERIMENTAL RESEARCH

Executive function as a predictor of physics-related conceptual change

François Thibault[1*] & Patrice Potvin[1]

ABSTRACT

While expert vs. novice comparisons using fMRI revealed a link between executive function and expertise in scientific fields such as electricity and mechanics, it is still uncertain whether these functions are active components of the learning process, rather than its by-product. Using the Force Concept Inventory to track conceptual change, the aim of this study was to confirm that executive function is indeed relevant to this process. Data suggests that participants with higher executive function abilities, as approximated by the Wisconsin Card Sorting Test, are likely to make more progress toward conceptual changes in a 15-week physics course than participants with lesser abilities. This implies that executive function is solicited throughout the process of conceptual change, and not solely being used as a consequence of expertise.

[1] Université du Québec à Montréal, Faculty of Education Sciences, Department of Didactics, Montreal, Canada

[*] Author email address: thibault.francois.2@uqam.ca

To cite this article: Thibault, F., & Potvin, P. (2018). Executive function as a predictor of physics-related conceptual change. *Neuroeducation*, 5(2), 119-126.

DOI: https://doi.org/10.24046/neuroed.20180502.119

Received on November 30th, 2017. Received in revised form on May 30th, 2018.
Accepted on June 9th, 2018. Available online on September 19th, 2018.
Neuroeducation, 5(2), 119-126
ISSN: 1929-1833

1. Introduction

The development of modern imaging technology, combined with its ever-increasing availability, has allowed researchers to study the human brain for purposes that are not medical in nature. This led to interesting discoveries in educational research, bringing valuable insights to fields such as language and mathematics education. Scientific education, on the other hand, has yet to see similar progress.

Up to now, a significant part of imaging research in this field has focused on comparisons between novices and experts in different areas of science, a trend largely influenced by Dunbar and his team's work on this topic (Dunbar & Stein, 2007). In their initial publication, they describe an experiment comparing experts and novices in mechanics, where it was observed that for experts to answer successfully to items related to a common misconception (heavier objects fall faster than lighter ones), increased activation of the anterior cingulate cortex (ACC) was required.

In the past, many functions have been attributed to the ACC: target selection (Frith *et al.*, 1991; Posner *et al.*, 1988), motivational valence assessment (Mesulam, 1990), motor response selection (Badgaiyan & Posner, 1998; Paus *et al.*, 1993; Turken & Swick, 1999), error detection (Gehring & Knight, 2000; Luu, Flaisch & Tucker, 2000), competition monitoring (Carter *et al.*, 1998), working memory (Petit *et al.*, 2000) and reward assessment (Knutson *et al.*, 2000). More recently, its activation has been described as some sort of "teaching signal" indicating that the situation leading to its activation requires increased cognitive resources to process, and thus should preferably be avoided in the future (Botvinick, 2007). Activation of the ACC is also observed during tasks such as the Stroop Task (Adleman *et al.*, 2002; Bush *et al.*, 1998) or the Wisconsin Card Sorting Test (Lie *et al.*, 2006) suggesting that it could be linked to executive functions such as inhibition and shifting, and possibly also updating, considering its previously observed ties to working memory.

Using a nearly identical task involving the common physics misconception that heavier objects fall faster, Brault Foisy, Masson, Potvin, & Riopel (2013) replicated Dunbar & Stein's (2007) findings and effectively observed an increased activation of the ACC for experts viewing non-scientific events. Furthermore, activation of the left dorsolateral prefrontal cortex (DLPFC) and the right ventrolateral prefrontal cortex (VLPFC), regions commonly associated with executive functions such as inhibition, working memory and set-shifting (Bush *et al.*, 1998; Lie *et al.*, 2006; Menon *et al.*, 2001; Monchi *et al.*, 2001; Nathaniel-James, Fletcher, & Frith, 1997) were recorded alongside the ACC, which could be interpreted as a sign that non-scientific events trigger misconceptions even in experts, who are then able to overcome them and consistently produce correct answers during the task. Identical activation patterns were also observed by Masson, Potvin, Riopel, & Brault Foisy (2014) using a similar methodology, but with regards to a common misconception in electricity according to which a single wire connecting a light bulb to a power source is required for it to produce light.

However, such experiments can only account for what happens during task-specific activities. In other words, while they can tell us about the differences between experts' and novices' performances during specific problem-solving tasks, they tell us nothing about how said experts developed their expertise in the first place. As such, before these findings can significantly impact scientific education in general, additional information is required. Since these experiments focus on specific misconceptions, the "experts" that are being studied are generally described as having completed a successful conceptual change, meaning that they have permanently overcome the interference caused by the misconception, and can be reliably expected to answer correctly during the task. This leads to an interesting question: are executive functions part of the process of conceptual change, or is their increased use by experts merely a consequence of the conceptual changes they have completed?

Insights towards an answer to this question can be found in Kwon & Lawson's (2000) study of the effect of brain growth on scientific reasoning and conceptual change. Using a modified version of Lawson's Classroom Test of Scientific Reasoning (Lawson, 1978, 1987, 1992), they found that inhibition capacity, as measured by the Wisconsin Card Sorting Test (WCST; Berg, 1948; Heaton *et al.*, 1993) was, of all other variables considered, the best predictor of scientific reasoning. Furthermore, scientific reasoning was also revealed as the best predictor of post-test performance on a test of air pressure concepts. These results coincide with the recent fMRI studies described earlier, in that they confirm that thinking like a scientist, and thus successfully solving scientific problems, seems to involve the use of executive functions such as inhibition.

The primary focus of their experiment, however, was to learn more about conceptual change. More specifically, it was to learn which functions of a teenager's developing brain were more important in improving conceptual understanding regarding various air pressure concepts. Defining conceptual gain as the difference between pre- and post-test results on a test of air pressure concepts (i.e., students' scores improve because they progress toward conceptual change regarding some of the concepts studied over the 14 weeks of the experiment), the authors found that inhibition capacity was also the best predictor of conceptual change, more so than other variables such as scientific reasoning, mental capacity or planning ability. As such, inhibition appears to not only be important for problem solving in science-related domains, but also for achieving conceptual change in these domains.

But while these results were very promising at the time, more recent knowledge raises doubts as to what was really observed during Kwon & Lawson's (2000) experiment. Knowing that performance to science-related tasks is strongly linked to executive functions, and that these functions, as demonstrated by Kwon and Lawson themselves using the WCST, vary greatly with age, the extended duration of their experiment can be problematic. Indeed, it is not impossible that participants achieved better results on the post-test, not because they learned the targeted concepts, but rather because their executive function abilities improved during the experiment, allowing them to perform better. Furthermore, the lessons that covered air-pressure concepts in this study specifically used the concept acquisition test questions as examples, giving partial credit for answers that were conceptually insufficient, but used the correct terminology. As such, it is difficult to conclude with certainty that conceptual change truly occurred, as opposed to simple memorization of previously unknown key terms.

This distinction between conceptual change and memorization is deeply tied to the Piagetian paradigm (Piaget, 1952), which describes two main forms of learning: assimilation and accommodation. In Posner, Strike, Hewson, & Gertzog's (1982) conceptual change model, assimilation is described as a "simple" accretion of knowledge, while accommodation requires profound changes in the learner's conceptions, a consequence of the incongruities between their prior conceptions and newly acquired knowledge. Assimilation and accommodation have also respectively been described as requiring weak and strong restructuring (Carey, 1985; Tyson *et al.*, 1997), highlighting the inherent difficulties associated with acquiring new knowledge when prior conceptions act as a source of interference.

Such interference also explains the previously described observations, where executive functions such as inhibition were required to perform science-related tasks involving common misconceptions. Indeed, it could be argued that such observations are required to conclude that the strong restructuring associated with conceptual change has occurred or, perhaps more importantly, is in the process of occurring. In this context, a confirmation that executive functions are solicited during the process of conceptual change, and not simply manifested during problem resolution once conceptual change has been completed, is required before pursuing other objectives, such as identifying which executive functions are in play and the extent of their influence on this process. To do so, the current study will use a methodology very similar to Kwon & Lawson's (2000), but controlling for variations in executive function capacities, as well as using stricter instruments to measure conceptual change over the experimental period.

2. Method

2.1 Sample

Participants ($n = 59$) were recruited among first-year students at *École de Technologie Supérieure* (ÉTS), one of Montreal's two engineering universities. To be selected, participants had to be registered for the ING150 course during the semester in which the study took place. This is an introductory level course on Newtonian mechanics and is mandatory for first-year students of all engineering specializations at ÉTS. It is also important to note that ÉTS students are selected from technical college programs rather than pre-university college programs, and that the participants were mostly enrolled in software and electrical engineering programs. Consequently, their only prior Newtonian mechanics training took place roughly three to four years prior to this study, while most of them were still in high school.

2.2 Design

Prior to instruction, students were administered pre-test measures of executive function (Wisconsin Card Sorting Test) and force-related knowledge (Force Concept Inventory). These tests were administered using the *Laboratoire mobile pour l'étude des cheminements d'apprentissage* (LabMÉCAS), a portable laboratory consisting of 35 laptop computers pre-loaded with testing software. This allowed the students to take the tests during their scheduled directed exercise period and required a total of approximately 45 minutes for both tests.

After pre-testing, students were taught a series of 15 three-hour lessons in Newtonian mechanics at a rate of one lesson per week during a 15-week semester, supplemented by another 15 three-hour sessions of directed exercises. Although students were selected in three different groups, all three lecturers responsible for their instruction were required by ÉTS to adhere strictly to the standard curriculum for the ING150 course. After completion of the lessons, both the Wisconsin Card Sorting Test and the Force Concept Inventory were re-administered to participants in the last regularly scheduled directed exercise period prior to the final examination for the course.

2.3 Materials

2.3.1 Wisconsin Card Sorting Test

As per Kwon and Lawson's (2000) research, the WCST was used to measure participants' executive function. Although this test was initially designed to measure mental flexibility (Berg, 1948), it is also known to require increased activation from the dorsolateral prefrontal cortex and the anterior cingulate cortex (Lie *et al.*, 2006; Monchi *et al.*, 2001), brain regions that are often associated with the inhibition of incorrect responses (Konishi *et al.*, 1999; Taylor *et al.*, 1997), detection of dissonance between observations and expectations (Berns, Cohen, & Mintun, 1997; Carter *et al.*, 1998), as well as task-set switching (Aron, Robbins, & Poldrack, 2004). As such, it should serve as a good approximation of executive function in general.

This test consists of a total of 132 cards: 4 stimulus cards and 128 response cards. Each of these cards can be described using three different characteristics: number, shape and colour. During the test, the participant is asked to successively match each of the response cards to one of the stimulus cards. After each attempt, the participant is told whether the match is correct or not, but not the matching principle. Once 10 consecutive correct matches are made, the matching principle is changed. While the participant has been warned that the matching principle will change during the test, the required conditions, as well as the algorithm used to select matching principles, remain unknown to them. When a participant, after receiving negative feedback, persists in using a discarded principle for their attempts, they are said to have committed a perseveration error and are assigned -1 points. Thus, participants with higher executive function capabilities will also have higher scores (i.e., closer to 0).

It should also be noted that the WCST measures executive function in a context where visual stimulation is preponderant, which could be problematic if the other variable studied is not visual in nature. Indeed, it has been observed in the past that executive functions such as inhibition can be highly task-dependent (Simmonds, Pekar, & Mostofsky, 2008). As such, a general approximation of executive function should be more accurate if it uses stimuli that are similar in nature to the other elements being studied. This should not be an issue in the current experiment, however, as visual elements such as graphs or other schematic representations are very often used in mechanics, but also in scientific education in general. There is thus no reason to believe that the WCST would perform differently in experiments involving other scientific domains.

Furthermore, in concordance with the need to ensure that variations in performance were not due to unexpected variations in executive function, steps were taken to control for pre- to post-test variations in executive function abilities. Specifically, executive function abilities were calculated by taking each participant's average score on the WCST, and then disqualifying participants whose deviation (the difference between their actual WCST scores and their calculated average) were more than two standard deviations removed from the average deviation of the studied sample.

Indeed, while a certain increase in WCST scores could be expected as students learn to inhibit their misconceptions, it would be very unusual for a participant to be among the best during one iteration of the test and among the worse during the other. This had the unfortunate side effect of considerably reducing the number of eligible participants (from 59 to 52), negatively impacting the precision of the intended statistical analyses, but also ensuring that any correlation observed between executive function abilities and conceptual change could not be attributed to extreme variations in WCST scores. The typical scenario for these rejected participants was to have a nearly perfect score (fewer than 10 perseveration errors) in either the pre- or the post-test, and abnormally low scores in the other (over 30 perseveration errors). In contrast, the average participant saw their WCST score improve by 1.36 points ($\sigma = 3.07$).

2.3.2 Force Concept Inventory

To measure participants' comprehension of force-related concepts, the Force Concept Inventory (FCI) test was used. Specifically designed by physics professors based on their observations of common students' misconceptions (Halloun & Hestenes, 1985; Hestenes, Wells & Swackhamer, 1992), the FCI is a multiple-choice questionnaire consisting of 30 items. For each item, a single correct answer and four distractors are presented to participants. One of the key points of this test is that each of the distractors is based on one of the most popular misconceptions about forces. There are no calculations required to answer the FCI's items. Rather, participants are asked to interpret or describe a specific phenomenon by selecting the appropriate answer.

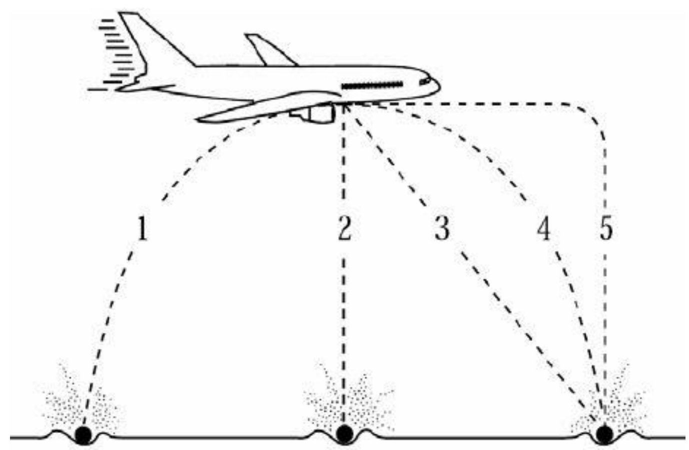

Figure 1. Sample item from the FCI (Hestenes *et al.*, 1992)

Figure 1 shows one such item, where the participant is asked to pick which trajectory would best describe the path of a bowling ball dropped from a plane that is moving toward the right at a constant velocity. Option #1 is based on a misconception that can arise after watching video footage of parachute jumpers. In videos taken from a point of view inside the plane, the jumper seems to fall backwards, causing the misconception. Option #2 is based on more mundane sources: in their daily lives, people often drop items, which apparently fall in a straight line toward the ground. This is because the slow speed of a walking human, combined with the relatively low height at which the object is held, makes it very hard to notice the small forward movement. Option #3 is caused by partial comprehension of the phenomenon. While proponents of this answer understand that the plane imparts a certain velocity toward the right to the bowling ball, they fail to consider that gravity imparts an acceleration, and not a fixed velocity, toward the ground. Similarly, option #5 describes the Aristotelian theory of impetus, intuitively held by many, where the forward momentum must be "spent" before the ball can fall to the ground. Thus, only participants who truly understand the difference between velocity and acceleration, as well as being able to correctly identify the forces at play and the initial conditions of the bowling ball, are able to answer correctly using option #4.

A similar design was used for each of the 30 items of the Force Concept Inventory. Since every distractor is designed to appeal to participants who hold specific misconceptions, only those who have completed conceptual change toward the selected concepts should be able to consistently answer correctly.

While Kwon & Lawson (2000) also correlated post-test scores to executive function, the current study will focus solely on pre- to post-test gains on the FCI. Indeed, correlation between post-test results and executive function could be explained, as was previously described during the short review of more recent fMRI studies, by the fact that executive function correlates with performance in science-related tasks. As such, pre- to post-test gains constitute the best measure of conceptual change during the experimental period, especially given the specific nature of the FCI's items.

2.4 Instrumentation

For this study, a computerized version of the WCST was used. It is available on the Psychology Experiment Building Language (PEBL) platform and conforms to Berg's (1948) original implementation of the test in every way, except for the fact that feedback is given automatically by the program and not by a human agent. As for the FCI, a French version of the test was administered using an interactive Portable Document Format file. Thanks to the LabMÉCAS infrastructure, it was possible for up to 34 participants to take the tests simultaneously, a process requiring approximately one hour.

3. Results

Table 1 presents descriptive statistics of participants' results on the Force Concept Inventory, as well as the average class grade for the ING150 course. As it can be observed, despite having achieved generally good grades on the ING150 course, participants show a limited understanding of concepts related to force, velocity and acceleration. In fact, their conceptual understanding only seems to have minimally improved during the semester, the average gain on the FCI only being 1.96 out of 30 items.

Table 1. Descriptive statistics for the FCI and ING150 course

	FCI pre-test (/30)	FCI post-test (/30)	FCI Gain
Average	12.83	14.79	1.96
Minimum	4	3	-5
Q1	9	11	1
Median	11	14	2
Q3	17	18	3
Maximum	25	28	9
ING150 course average	81 %		

Figure 2 shows a graph of pre- to post-test gains on the FCI as a function of inhibiting ability. Regression analysis shows that a moderate (.30 < R = .427 < .50 [Cohen, 1988]) positive correlation ($p < .01$) exists between participants' inhibiting ability and their ability to successfully achieve conceptual change, as measured by their FCI gains over one semester. Furthermore, although there seems to be little variation in inhibiting ability for the majority of participants, this variable explains 18.2 % of the variance observed for pre- to post-test gains, as can be observed in Table 2.

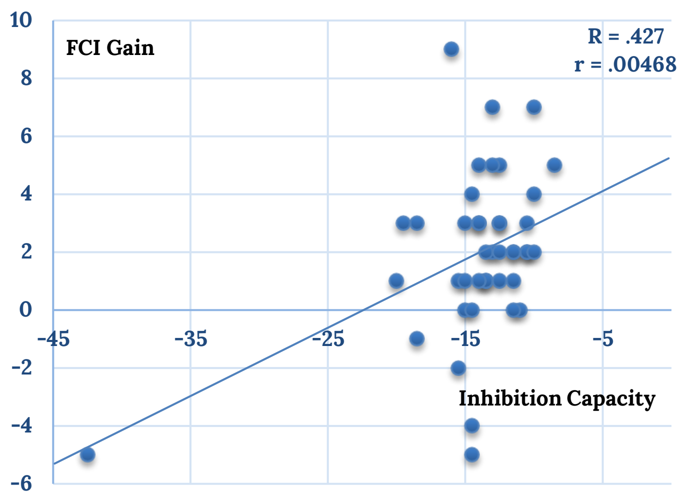

Figure 2. FCI Gain vs. Inhibition Capacity

However, the presence of several outliers on the graph warrants the use of non-parametric measures, such as Spearman's and Kendall's rank correlation coefficients. This is due to the fact that the calculations involved in statistical analyses such as Pearson's coefficient use the data itself, meaning that extreme outliers will have a larger impact on the end result than other points of data. Spearman's and Kendall's coefficients do not suffer from this bias, as they use each participant's rank for each variable rather than the actual test scores. Using these measures, a positive correlation ($p < .05$) can still be observed, although its effect seems to be slightly weaker than what Pearson's coefficient seems to indicate ($\rho = .308$, $\tau = .228$).

Table 2. Correlation Analysis

Pearson's correlation coefficient (R)	.427
Squared Pearson (R^2)	.182
$p(R^2)$	.00468
Spearman's rank correlation coefficient (ρ)	.308
$p(\rho)$	.0469
Kendall rank correlation coefficient (τ)	.228
$p(\tau)$	.0474

4. Discussion

Interestingly, participants seem to have become adept at solving problems related to forces, as shown by their class' generally strong results on the ING150 course, despite their limited understanding of the related concepts, as evidenced by their results on the FCI (as shown in Table 1). This could be tied to the ÉTS's vocation as an engineering school, where more emphasis is placed on practical problem solving over conceptual understanding. It would have been interesting to compare participants' grades to the class average, but the ÉTS decided not to disclose individual grades for ethical reasons.

As can be seen in Figure 2, most of the participants seem to possess very similar inhibiting abilities, which can most likely be explained by the selection process used by ÉTS during student admission. Indeed, similarly to medical programs, engineering programs in Quebec require college grades that are well above average, and enrollment is limited. As such, it is not impossible that participants in the current sample, despite holding common misconceptions in force-related physics, are on the stronger side of the executive function spectrum.

This could possibly explain the large difference between the present correlation and the one observed by Kwon and Lawson (2000), for whom inhibiting ability could explain nearly 30 % of the variance in conceptual gains. There are also large differences in the tests administered to the participants in both studies to measure conceptual gains. While the 30 FCI items all contained "traps" based on popular misconceptions and were meant to lure participants into selecting the wrong answers, the test used by Kwon and Lawson consisted of only six open-ended short answer questions where partial points could be awarded to participants who correctly used certain keywords while still answering incorrectly overall. This has the effect of allowing students to score better on the test despite not necessarily having completed conceptual change, allowing for a wider distribution of scores in the sample. Furthermore, the lessons given to participants in Kwon and Lawson's study specifically included the different test items and the corresponding answers as examples which, combined with the scoring method used, could have allowed students to achieve better gains than expected.

Additional correlation analysis using Spearman's rank correlation coefficient reveals that, while the outliers had an obvious impact on the observed correlation ($\rho = .308$ vs. $R = .427$), the correlation still exists ($p < .05$). Unlike Pearson's coefficient, Spearman's Rho does not assume that the relationship between the two variables is linear in nature (Hauke & Kossowski, 2011). As such, its value should not be interpreted as a strength of the correlation, but rather as an indication of how well an arbitrary monotonic function could describe the association between the variables. Using this non-parametric measure, it is thus possible to conclude that a modest, but significant, relationship exists between measured executive function and conceptual change as measured by the FCI, whereas participants with higher scores on the WCST also tended to learn more during the same physics course.

Another interesting interpretation of the data stems from the use of Kendall's rank correlation coefficient, which is a measure of the concordance between the subjects' rankings according to both studied variables. Using the calculated coefficient, it can be inferred that, when comparing two students, the one possessing higher executive function abilities also has a 61.4 % chance of achieving more conceptual gains during the semester. These odds, while not overwhelming, are consistent with the moderate effect observed with Pearson's and Spearman's coefficients.

5. Conclusion

Novice vs. Expert comparisons using fMRI initially revealed a link between executive function and performance in science-related tasks. In these studies, experts showed increased brain activation in areas linked to executive function such as the anterior cingulate cortex, the left ventrolateral prefrontal cortex or the right dorsolateral prefrontal cortex when successfully solving tasks involving common misconceptions. This link has also been observed in studies using reaction times, where participants required more time to process similar items, which authors explained by the solicitation of increased cognitive resources required by common misconceptions portrayed in the task (Potvin *et al.*, 2014; Shtulman & Valcarcel, 2012).

The data collected during this research extends these findings, supporting the existence of a relation between executive function and conceptual change. While the sample size is small, significant correlations were found between the two variables using three different correlation coefficients. In short, these results imply that students with stronger executive function abilities will tend to be more efficient at achieving conceptual change in domains rich in misconceptions. These results should in no way be interpreted as meaning that these students should be favoured or that science programs be restricted to them as they are most likely to succeed, especially since the correlation is only of moderate strength. Rather, teachers and educators might be interested in helping students develop their executive function alongside the traditional curriculum, as this could potentially make them more efficient at successfully completing conceptual change in the future. Furthermore, tests of executive function could be used as a diagnostic tool to identify the students that are more likely to struggle with the standard curriculum and adjust their instruction accordingly.

But before researching pedagogical practices that favour the development of executive function, it would be important to establish which specific executive functions are being solicited. Indeed, many executive functions, such as inhibition, working memory and set-shifting, are linked to similar brain activity when tested using fMRI (Aron *et al.*, 2004; Berns *et al.*, 1997; Carter *et al.*, 1998; Konishi *et al.*, 1999;

Taylor *et al.*, 1997), meaning that distinguishing one from the other using only fMRI data is not easy, especially if the task was not specifically designed to measure executive function. In turn, knowing which specific executive functions are at play during the process of conceptual change would give researchers and educators new insights regarding the effectiveness of various teaching methods, ideally leading to better learning opportunities for students.

References

Adleman, N. E., Menon, V., Blasey, C. M., White, C. D., Warsofsky, I. S., Glover, G. H., & Reiss, A. L. (2002). A developmental fMRI study of the Stroop color-word task. *NeuroImage*, 16(1), 61-75. https://doi.org/10.1006/nimg.2001.1046

Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2004). Inhibition and the right inferior frontal cortex. *Trends in Cognitive Sciences*, 8(4), 170-177. https://doi.org/10.1016/j.tics.2004.02.010

Badgaiyan, R. D., & Posner, M. I. (1998). Mapping the cingulate cortex in response selection and monitoring. *NeuroImage*, 7(3), 255-260. https://doi.org/10.1006/nimg.1998.0326

Berg, E. A. (1948). A simple objective technique for measuring flexibility in thinking. *The Journal of General Psychology*, 39(1), 15-22. https://doi.org/10.1080/00221309.1948.9918159

Berns, G. S., Cohen, J. D., & Mintun, M. A. (1997). Brain regions responsive to novelty in the absence of awareness. *Science*, 276(5316), 1272-1275. https://doi.org/10.1126/science.276.5316.1272

Botvinick, M. M. (2007). Conflict monitoring and decision making: Reconciling two perspectives on anterior cingulate function. *Cognitive, Affective, & Behavioral Neuroscience*, 7(4), 356-366. https://doi.org/10.3758/cabn.7.4.356

Brault Foisy, L.-M., Masson, S., Potvin, P., & Riopel, M. (2013, June). *Using fMRI to compare cerebral activations between novices and experts in science during a task in mechanics involving a common misconception*. Paper presented at the Canada-Israel Symposium on Brain Plasticity, Learning and Education, Western University, London, Canada.

Bush, G., Whalen, P. J., Rosen, B. R., Jenike, M. A., McInerney, S. C., & Rauch, S. L. (1998). The counting Stroop: an interference task specialized for functional neuroimaging–validation study with functional MRI. *Human Brain Mapping*, 6(4), 270-282. https://doi.org/10.1002/(sici)1097-0193(1998)6:4%3C270::aid-hbm6%3E3.3.co;2-h

Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: MIT Press.

Carter, C. S., Braver, T. S., Barch, D. M., Botvinick, M. M., Noll, D., & Cohen, J. D. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science*, 280(5364), 747-749. https://doi.org/10.1126/science.280.5364.747

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Dunbar, K. N., & Stein, C. (2007). Do naïve theories ever go away? In Lovett (Ed.), *Thinking with data* (pp.193-205), Mahwah, NJ: Lawrence Erlbaum Associates.

Frith, C. D., Friston, K. J., Liddle, P. F., & Frackowiak, R. S. (1991). Willed action and the prefrontal cortex in man: A study with PET. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 244(1311), 241-246. https://doi.org/10.1098/rspb.1991.0077

Gehring, W. J., & Knight, R. T. (2000). Prefrontal–cingulate interactions in action monitoring. *Nature Neuroscience*, 3(5), 516-520. https://doi.org/10.1038/74899

Halloun, I. A., & Hestenes, D. (1985). The initial knowledge state of college physics students. *American Journal of Physics*, 53(11), 1043-1055. https://doi.org/10.1119/1.14030

Hauke, J., & Kossowski, T. (2011). Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data. *Quaestiones Geographicae*, 30(2), 87-93. https://doi.org/10.2478/v10117-011-0021-1

Heaton, R. K., Chelune, G. J., Tally, J. L., Kay, G. G., & Curtiss, G. (1993). *Wisconsin Card Sorting Test manual: Revised and expanded.* Odessa, FL: Psychological Assessment Resources, Inc.

Hestenes, D., Wells, M. & Swackhamer, G. (1992). Force Concept Inventory. *The Physics Teacher*, 30(3), 141-158. https://doi.org/10.1119/1.2343497

Konishi, S., Nakajima, K., Uchida, I., Kikyo, H., Kameyama, M. & Miyashita, Y. (1999). Common inhibitory mechanism in human inferior prefrontal cortex revealed by event-related functional MRI. *Brain*, 122(5), 981-991. https://doi.org/10.1093/brain/122.5.981

Knutson, B., Westdorp, A., Kaiser, E., & Hommer, D. (2000). FMRI visualization of brain activity during a monetary incentive delay task. *NeuroImage*, 12(1), 20-27. https://doi.org/10.1006/nimg.2000.0593

Kwon, Y.-J., Lawson, A. E. (2000). Linking brain growth with the development of scientific reasoning ability and conceptual change during adolescence. *Journal of Research on Science Teaching*, 37(1), 44-62. https://doi.org/10.1002/(sici)1098-2736(200001)37:1%3C44::aid-tea4%3E3.0.co;2-j

Lawson, A. E. (1992). What do tests of formal reasoning actually measure? *Journal of Research in Science Teaching*, 29(9), 965–983. https://doi.org/10.1002/tea.3660290906

Lawson, A. E. (1987). *Classroom Test of scientific reasoning*. Unpublished manuscript, Arizona State University, Tempe, Arizona.

Lawson, A. E. (1978). The development and validation of a classroom test of formal reasoning. *Journal of Research in Science Teaching*, 15(1), 11–24. https://doi.org/10.1002/tea.3660150103

Lie, C. H., Specht, K., Marshall, J. C., & Fink, G. R. (2006). Using fMRI to decompose the neural processes underlying the Wisconsin Card Sorting Test. *NeuroImage*, 30(3), 1038-1049. https://doi.org/10.1016/j.neuroimage.2005.10.031

Luu, P., Flaisch, T., & Tucker, D. M. (2000). Medial frontal cortex in action monitoring. *Journal of Neuroscience*, 20(1), 464-469. https://doi.org/10.1523/jneurosci.20-01-00464.2000

Masson, S., Potvin, P., Riopel, M., & Brault-Foisy, L.-M. (2014). Differences in brain activation between novices and experts in science during a task involving a common misconception in electricity. *Mind, Brain, and Education*, 8(1), 44-55. https://doi.org/10.1111/mbe.12043

Menon, V., Adleman, N. E., White, C. D., Glover, G. H., & Reiss, A. L. (2001). Error-related brain activation during a Go/NoGo response inhibition task. *Human Brain Mapping*, 12(3), 131-143. https://doi.org/10.1002/1097-0193(200103)12:3%3C131::aid-hbm1010%3E3.0.co;2-c

Mesulam, M. (1990). Large-scale neurocognitive networks and distributed processing for attention, language, and memory. *Annals of Neurology*, 28(5), 597-613. https://doi.org/10.1002/ana.410280502

Monchi, O., Petrides, M., Petre, V., Worsley, K., & Dagher, A. (2001). Wisconsin Card Sorting revisited: distinct neural circuits participating in different stages of the task identified by event-related functional magnetic resonance imaging. *The Journal of Neuroscience*, 21(19), 7733-7741. https://doi.org/10.1523/jneurosci.21-19-07733.2001

Nathaniel-James, D. A., Fletcher, P., & Frith, C. D. (1997). The functional anatomy of verbal initiation and suppression using the Hayling Test. *Neuropsychologia*, 35(4), 559-566. https://doi.org/10.1016/s0028-3932(96)00104-2

Paus, T., Petrides, M., Evans, A. C., & Meyer, E. (1993). Role of the human anterior cingulate cortex in the control of oculomotor, manual, and speech responses: a positron emission tomography study. *Journal of Neurophysiology*, 70(2), 453-469. https://doi.org/10.1152/jn.1993.70.2.453

Petit, L., Courtney, S. M., Ungerleider, L. G., & Haxby, J. V. (1998). Sustained activity in the medial wall during working memory delays. *The Journal of Neuroscience*, 18(22), 9429-9437. https://doi.org/10.1523/jneurosci.18-22-09429.1998

Piaget, J. (1952). *The origin of intelligence in children*. New York, NY: International Universities Press.

Posner, M. I., Petersen, S., Fox, P., & Raichle, M. (1988). Localization of cognitive operations in the human brain. *Science*, 240(4859), 1627-1631. https://doi.org/10.1126/science.3289116

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education*, 66(2), 211-227. https://doi.org/10.1002/sce.3730660207

Potvin, P., Masson, S., Lafortune, S., & Cyr, G. (2014). Persistence of the intuitive conception that heavier objects sink more: A reaction time study with different levels of interference. *International Journal of Science and Mathematics Education*, 13(1), 21-43. https://doi.org/10.1007/s10763-014-9520-6

Simmonds, D. J., Pekar, J. J., & Mostofsky, S. H. (2008). Meta-analysis of Go/No-go tasks demonstrating that fMRI activation associated with response inhibition is task-dependent. *Neuropsychologia*, 46(1), 224-232. https://doi.org/10.1016/j.neuropsychologia.2007.07.015

Shtulman, A., & Valcarcel, J. (2012). Scientific knowledge suppresses but does not supplant earlier intuitions. *Cognition*, 124(2), 209-215. https://doi.org/10.1016/j.cognition.2012.04.005

Taylor, S. F., Kornblum, S., Lauber, E. J., Minoshima, S., & Koeppe, R. A. (1997). Isolation of specific interference processing in the Stroop task: PET activation studies. *NeuroImage*, 6(2), 81-92. https://doi.org/10.1006/nimg.1997.0285

Turken, U., & Swick, D. (1999). Response selection in the human anterior cingulate cortex. *Nature Neuroscience*, 2(10), 920-924. https://doi.org/10.1038/13224

Tyson, L. M., Venville, G. J., Harrison, A. G., & Treagust, D. F. (1997). A multidimensional framework for interpreting conceptual change events in the classroom. *Science Education*, 81(4), 387-404. https://doi.org/10.1002/(sici)1098-237x(199707)81:4%3C387::aid-sce2%3E3.0.co;2-8