

City University of New York (CUNY)

## CUNY Academic Works

---

Dissertations, Theses, and Capstone Projects

CUNY Graduate Center

---

6-2021

### **Rethinking Thinking About Thinking: Against a Pedagogical Imperative to Cultivate Metacognitive Skills**

Lauren R. Alpert

*The Graduate Center, City University of New York*

[How does access to this work benefit you? Let us know!](#)

More information about this work at: [https://academicworks.cuny.edu/gc\\_etds/4301](https://academicworks.cuny.edu/gc_etds/4301)

Discover additional works at: <https://academicworks.cuny.edu>

---

This work is made publicly available by the City University of New York (CUNY).

Contact: [AcademicWorks@cuny.edu](mailto:AcademicWorks@cuny.edu)

RETHINKING THINKING ABOUT THINKING:  
AGAINST A PEDAGOGICAL IMPERATIVE TO CULTIVATE METACOGNITIVE SKILLS

by

LAUREN R. ALPERT

A dissertation submitted to the Graduate Faculty in Philosophy in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2021

©2021

LAUREN R. ALPERT

All Rights Reserved

Rethinking Thinking About Thinking:  
Against a Pedagogical Imperative to Cultivate Metacognitive Skills

By

Lauren R. Alpert

This manuscript has been read and accepted for the Graduate Faculty in Philosophy in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

---

Date

---

Michael Brownstein

Chair of Examining Committee

---

Date

---

Nickolas Pappas

Executive Officer

Supervisory Committee:

Michael Brownstein

Elizabeth Edenberg

Eric Mandelbaum

Barbara G. Montero

THE CITY UNIVERSITY OF NEW YORK

## ABSTRACT

Rethinking Thinking About Thinking:  
Against a Pedagogical Imperative to Cultivate Metacognitive Skills

by

Lauren R. Alpert

Advisor: Eric Mandelbaum

In summaries of “best practices” for pedagogy, one typically encounters enthusiastic advocacy for *metacognition*. Some researchers assert that the body of evidence supplied by decades of education studies indicates a clear pedagogical imperative: that if one wants their students to learn well, *one must implement teaching practices that cultivate students’ metacognitive skills*.

In this dissertation, I counter that education research does not impose such a mandate upon instructors. We lack sufficient and reliable evidence from studies that use the appropriate research design to validate the efficacy of metacognitive skill-building interventions (not just evaluate their relationship to learning outcomes). I argue that improved academic outcomes following these interventions aren’t necessarily mediated by increased metacognitive skills; rather, enhanced student performance can be accounted for by other factors that accompany metacognitive training, particularly the explicit provision of domain-specific knowledge.

On the way to this conclusion, I elaborate some complications and controversies surrounding “metacognition.” This is a sprawling and nebulous construct, which makes generalizations about its pedagogical value dubious from the outset. Moreover, it is unclear the extent to which the end goal of metacognitive skill-building (cognitive self-mastery, involving knowledge of and control over our own minds) is even possible, given our mental architecture;

some cognitive scientists allege that we are subject to phenomenological illusions which make this seem more achievable than it actually is.

I also attempt to provide an account of how metacognitive skill-building could receive glowing endorsements from educators and education researchers, even though its conceptual and empirical underpinnings are flimsy. The error theory I offer identifies two major factors that may foster belief in the efficacy of metacognitive training: *goalpost-shifting* around the objective of such efforts, and *motivated reasoning* in defense of the desirable conclusion that educators can significantly reshape students' minds and unlock their intellectual potential through simple pedagogical interventions.

## ACKNOWLEDGEMENTS

I just can't thank Eric Mandelbaum enough, and whatever I manage to express here is bound to be inadequate. I am not exaggerating in saying that we'd have many fewer CUNY Philosophy graduates were it not for his willingness to step up with a much-needed pep talk (and sometimes more serious forms of rescue). Eric, I'm glad to have made you laugh with the jokes I peppered into this manuscript to keep myself entertained while writing it, as some small consolation for the thanklessness of advising.

Thanks to the rest of my committee – Michael Brownstein, Barbara Montero, and Elizabeth Edenberg – for their help in refining the scope, organization, and trajectory of what was in early stages quite an unruly project. I'm so appreciative of their enthusiasm and their encouragement to adapt this research into a public-facing book.

I thank these authors, from whose writing on education I've drawn heavy inspiration: Mark Alfano, Greg Ashman, Jason Baehr, Catherine Elgin, Emily Hanford, E.D. Hirsch, and Daniel Willingham.

I thank Iakovos Vasiliou, John Greenwood, and Noël Carroll for holistic support as a graduate student, opportunities for professional development, and nonjudgment about my unorthodox career ambitions.

I owe many thanks to my highly supportive colleagues at Baruch College: Thomas Teufel, Hagop Sarkissian, Jonathan Gilmore, Eric Mandelbaum, Doug Lackey, and Sandeep Sreekumar. In particular, some of these colleagues secured a two-year full-time lectureship for me, during which I gained invaluable experience (particularly with time management!) and was challenged in ways that forced me to think more innovatively about how to teach. It was through teaching at Baruch that I got interested in the literature on best practices for pedagogy,

wherein I encountered mentions of metacognition that I bristled at for reasons I couldn't quite put into words initially, and now here we are.

I'm grateful for the commiseration and camaraderie with my classmates, in no particular order and not limited to: Jennifer Ware, Joseph Bendaña, Amanda Huminski, John Dyck, Jessie McCormack, David Neely, Jonathan Kwan, Richard Stillman, Kate Pendoley, and Chloë Cooper-Jones. Thanks to them and the availability bias, I can look back at grad school and recall hours of raucous laughter more readily than the innumerable, indistinguishable hours of solitary writing, studying, and teaching prep.

Lastly, I'd like to thank the people in my life who gave me the time and space I needed to invest enough effort into this project to make it worth the sunk costs. They include my very patient startup colleagues, nuclear family, and most of all my husband, Jeff. If I can offer any unsolicited advice, it's to marry someone who believes in you more than you believe in yourself.



## TABLE OF CONTENTS

<b>List of figures</b>	<b>x</b>
<b>0. Introduction</b>	<b>1</b>
§0.1   A preview of what's ahead.....	5
<b>1. A working glossary on metacognition in education</b>	<b>7</b>
§1.1   Metacognition qua pedagogical concept .....	7
§1.2   Cultivating metacognitive skills .....	11
1.2a   <i>Metacognitive skills</i> .....	12
1.2b   <i>Metacognitive training interventions</i> .....	13
§1.3   Genealogy of "metacognition" .....	17
1.3a   <i>"Historical roots"</i> .....	18
1.3b   <i>Coinage and early problems</i> .....	23
1.3c   <i>Early critiques of the construct</i> .....	31
§1.4   Summary and implications.....	34
<b>2. Complications and controversies from philosophy of mind &amp; cognitive science</b>	<b>38</b>
§2.1   Some historical objections to metacognition.....	38
§2.2   Illusions of introspection: Infallibility and transparent access to our mental states and processes.....	43
2.2a   <i>The inscrutability of cognitive processing</i> .....	45
2.2b   <i>Introspective inaccuracy</i> .....	49
2.2c   <i>Reconceiving introspection as motivated self-interpretation</i> .....	52
2.2d   <i>Implications for metacognition in education</i> .....	56
§2.3   Competing psychological models of metacognition.....	57
2.3a   <i>Functionalist higher-order cognition</i> .....	58
2.3b   <i>Control (self-evaluative) theories</i> .....	60
2.3c   <i>Metarepresentational (self-attributive) views</i> .....	62
2.3d   <i>Two-level model</i> .....	64
2.3e   <i>Implications for metacognition in education</i> .....	65
§2.4   Higher-order thinking dethroned.....	68
2.4a   <i>Implications for metacognition in education</i> .....	74
§2.5   Summary.....	75

**3. Scrutinizing metacognitive training interventions** **77**

---

§3.1 | Methodological issues with metacognitive training.....80  
    3.1a | *Explicit modeling: Displaying more than we can know?* . . . . . 80  
    3.1b | *Strategy instruction: Skill-building or “bag of tricks”?* . . . . . 85  
    3.1c | *Reflection revisited* . . . . . 88

§3.2 | Limitations of empirical findings on metacognitive training.....97  
    3.2a | *Assessing the effectiveness of metacognitive training* . . . . . 98  
    3.2b | *Analyzing meta-analyses* . . . . . 103

§3.3 | Explaining the impact of metacognitive training on learning outcomes..109  
    3.3a | *Changes in domain-specific content knowledge* . . . . . 114  
    3.3b | *Learning as a moving target* . . . . . 119

§3.4 | Summary.....122

**4. An error theory for metacognitivism** **126**

---

§4.1 | Shifting the goalposts for metacognitive skill-building.....125  
    4.1a | *Precedents for goalpost-shifting around “metacognition”* . . . . 127  
    4.1b | *Substituting metacognition for metacognitive skill-building* . . 134

§4.2 | Motivated reasoning in defense of metacognitive training.....138  
    4.2a | *Precedents for motivated reasoning in education* . . . . . 140  
    4.2b | *Why metacognitivism is attractive to educators* . . . . . 149

§4.3 | Concluding thoughts.....161

**Appendix: Disambiguating “metacognition”** **166**

---

§A1 | Abstracted cognition.....166  
§A2 | Higher-level cognition.....166  
§A3 | Self-referential cognition, or “thinking about thinking” .....170

**References** **175**

---

## LIST OF FIGURES

FIGURE 1: Example of homuncular imagery in visualization of metacognition.....	41
FIGURE 2: A student’s misinterpretation of the instruction to “show your thinking” .....	68
FIGURE 3: Example of explicit modeling of metacognition, in the form of annotations on a practice exam .....	112
FIGURE 4: Putative evidence that improved student performance resulted from metacognitive skill gain.....	113
FIGURE 5: Example in which “higher order” describes cognition at an advanced skill level....	169

## 0. INTRODUCTION

Over the past half century, the concept of *metacognition* – very roughly, *thinking about thinking* – has cemented its status as a staple of pedagogical discourse across all levels of education. It is a fixture of texts summarizing “best practices” for instruction, and a frequent topic of teacher training workshops. The institution at which I have taught identifies metacognition as “a core pedagogical concept” and encourages instructors to “promote metacognition in your classes by weaving it into every assignment” (Bernard L. Schwartz Communication Center, n.d.). If you peruse the website for any institution of higher education’s “Center for Teaching and Learning” (however it may be named), you are likely to come across an introduction to metacognition, coupled with encouragement to implement pedagogical practices that will prompt students to engage in metacognition.

I’ll use the term “metacognitivism”<sup>1</sup> to capture the seemingly omnipresent view that *metacognition is prompted by good pedagogy*. Minimally, metacognitivists endorse the claim that it is good for teachers to prompt students to exercise metacognition during course activities. But many educators and education researchers have adopted a stronger stance: that teachers should adopt pedagogical methods that will effect lasting enhancements in students’ metacognitive skills. This *strong metacognitivism*, as I’ll call it, often stems from the observation that students with superior metacognitive skills tend to perform better academically (Brown et al. 1983, Bransford et al. 1986, Wang et al. 1990). Especially within the first decade after “metacognition”

---

<sup>1</sup>This term is not widely used, but it’s not exactly a neologism, either. A smattering of thinkers have converged on “metacognitivist” as a natural descriptor for advocates of metacognition in education; e.g., Braten (1991) and Sutherland (1992, p. 98) use it to describe Lev Vygotsky (1896-1934) as an educational theorist who would have recommended that teachers engage students in pedagogy, had “metacognition” been coined in his era. Haynes (2014, p. 138) uses it once in the same way I will throughout this project.

“Metacognitivism” has acquired some alternative meanings in other areas of philosophy. That this term can be adopted for such diverse semantic purposes is an early illustration of how invoking “metacognition” and its linguistic derivatives can be communicatively inefficient and/or obfuscatory, since one nearly always has to clarify what one means by these terms if they don’t want to be misunderstood.

was coined (Flavell 1976), educational psychologists theorized that metacognitive abilities could account for individual differences in aptitude for learning among students with the same level of knowledge about a subject. From there, a further inference is made that metacognition could be a key *leverage point* for pedagogical interventions: that teaching practices which aim at metacognitive skill gain (which I'll call "metacognitive training," as a shorthand) could manifest previously-untapped academic potential in underperforming students. This passage from an influential summary text on the "science of learning" captures the reasoning behind a strong metacognitivist approach to pedagogy:

Everyday experience suggests that there are *intelligent novices*: some novices learn new domains more quickly than other novices. Research tells us that one thing that makes some novices more intelligent than others is their metacognitive skills. . . . Some people develop these skills naturally; others do not. Those who do can become intelligent novices; those who don't may have difficulty learning.

. . . The importance of metacognition for education is that a child is, in effect, a universal novice, constantly confronted with novel learning tasks. In such a situation it would be most beneficial to be an intelligent novice. What is encouraging is that the research also shows that it is possible to teach children metacognitive skills and when to use them. If we can do this, we will be able to help children become intelligent novices; we will be able to teach them how to learn. (Bruer 1993, pp. 70, 72)

Within the past several decades, evangelists of metacognitive training have emphasized the relationship between metacognitive training and measurable gains in student achievement: e.g., enhanced reading comprehension,<sup>2</sup> more successful problem-solving,<sup>3</sup> and elevated exam scores.<sup>4</sup> Some have argued that cultivating metacognitive skills in the classroom is an underappreciated solution to a wide range of scholastic woes including low student motivation,

---

<sup>2</sup> E.g., Brown et al. (1983), Baker & Brown (1984), Paris et al. (1984) Paris & Jacobs (1984), Paris & Oka (1986), Palinscar & Brown (1989), Pressley et al (1992), Artelt et al. (2001)

<sup>3</sup> Including Flavell (1976), Kitchener (1983), Bransford et al. (1986), Whimbey & Lochhead (1986), Schoenfeld (1989), Pressley & Woloshyn (1995), Mayer (1998).

<sup>4</sup> Schraw (1994), Vrugt & Oort (2008), Phillips et al (2016), Casselman & Atwood (2017)

poor transfer of learning to novel situations, and debilitating performance anxiety during assessments (Wilson and Conyers 2014).

Philosopher David Concepción has summed up strong metacognitivism as a simple pedagogical imperative: “If we want students to learn as much as possible, then we should help them improve their metacognitive skills” (2004, p. 356). It’s easy to feel oneself drawn in by this call to action. Who wouldn’t want to at least try out an approach to teaching that is so widely lauded and promises such extensive benefits? Who wouldn’t want to encourage their students – especially in philosophy courses – to become more reflective and self-critical in their thinking? The burden of proof appears to fall upon whoever would challenge the apparent consensus that instructors ought to help their students to improve their metacognition.

In this dissertation, I take up that argumentative burden, and demonstrate that all that glitters is not gold with respect to the impressive prospects of metacognitive training. I argue that the empirical backing for strong metacognitivism just isn’t robust enough to justify the high esteem this approach to pedagogy has garnered.

This project spun out of a deep dive into the segment of the scholarship of teaching and learning<sup>5</sup> focused on teaching *philosophy*. In this small but zealous niche of literature, I came across many mentions of metacognition, but no acknowledgement that this is an area of ongoing debate among cognitive scientists. Moreover, not one of these texts presented direct evidence from an empirical study demonstrating that using teaching methods that target students’ metacognitive skills can enact lasting changes in students’ metacognitive skills.

---

<sup>5</sup> Often abbreviated SoTL, this is an area of research in which teaching faculty (mostly within higher education) investigate and offer insights about the efficacy of teaching practices within their discipline. In SoTL, the instructor is the researcher, testing methods and/or gathering and synthesizing observations from their own teaching practice. This contrasts with standard methods of education research, in which what instructors do is manipulated or observed by a third party. For a comprehensive introduction to the history and methods of this field, see Chick (n.d.).

When a handful of philosophers have endorsed pedagogical methods that aim at enhancing students' metacognitive abilities,<sup>6</sup> they've pointed to general conclusions purportedly drawn from the primary empirical literature — but not that literature itself. Some *reference* primary literature,<sup>7</sup> but none offer any exposition of this research which would illustrate what kind and what magnitude of effects have been observed as a consequence of instructors' concerted efforts to increase students' metacognitive skills. Lastly, since the secondary literature does not differentiate between approaches to cultivating metacognitive skill, authors who derive their impressions of metacognitive from this secondary literature tend to ignore the possibility that there could be discrepancies in effectiveness among various methods of metacognitive training.

On the one hand, it's understandable and defensible to not belabor the results of any individual study on metacognition as a stand-in for a whole body of research. On the other, it is odd to endorse pedagogical interventions wholly on the basis of broad generalities and summary claims, instead of invoking concrete evidence of how specific metacognitive training interventions operate in real classrooms. In my eyes, the omission of primary research findings in discussions applauding the prospects of pedagogical interventions that target metacognitive skills was a *prima facie* reason to at least *wonder* if the empirical record for these interventions is being presented in an unrealistically flattering light. As it turns out, abstracting away from the details of particular studies conceals serious limitations and weaknesses of research in this area

---

<sup>6</sup> Concepción (2004), Crooks (2009), Intellectual Virtues Academy (2012/2016), Stokes (2012), Maynes (2014), Wright (2015), Smith (2017).

<sup>7</sup> Individual empirical studies are cited in Crooks (2009), Intellectual Virtues Academy (2012/2016), Maynes (2014), and Smith (2017) — but the content of these studies is not described. Nor is it addressed that these studies often have small sample sizes, or that they only concern metacognitive interventions within particular domains of instruction (e.g. reading comprehension, math, science) or populations (e.g., students who qualify for remedial courses), which raises questions regarding the extent to which findings in one domain are generalizable to teaching in other disciplines or skill areas. For example, one of the empirical papers cited by Maynes is a qualitative case study on remedial reading instruction for five college freshmen (Nash-Ditzel 2010).

– awareness of which ought to diminish educators’ confidence in the efficacy of metacognitive training.

To be clear, I am *not* arguing that educators must forswear all laudatory talk of metacognition, or should immediately abandon any teaching practices aimed at improving students’ metacognitive skills. There may be valid reasons to position metacognition as a pedagogical ideal, independent of the empirical record of metacognitive training interventions. What I’d hope to accomplish is to stimulate conversation around why strong metacognitivism has been so rapidly embraced, and to give educators the resources to reconsider their attraction to or endorsement of teaching methods that target “metacognition.”<sup>8</sup> At the very least, I think it is valuable for educators to analyze the grounds on which they select pedagogical practices, in order to consider how various motives and preferences might influence judgments about which teaching practices are worth enacting.

## **§0.1 | A preview of what’s ahead**

This project is divided into four chapters. Here’s a quick preview of what appears in each one:

**Chapter 1 (“A working glossary on metacognition in education”)** lays out an account of “metacognition” as used in pedagogical discourse, to set the reader up to follow the critiques offered in Chapters 2 and 3. I make no pretense toward offering an exhaustive summary of extant claims about metacognition in education — the literature is too vast to attempt such a feat. Instead, I suggest that the best way to make sense of what “metacognition” refers to is to

---

<sup>8</sup> I will frequently apply scare quotes to this term, so as to clarify that I’m discussing definitions and properties that various speakers have attributed to *that which they call “metacognition.”* I’m not interested in sussing out which account(s) accurately capture what metacognition truly is; nor do I want to endorse some construals over others. I’m trying to get a grasp on *what it is that users of this term have in mind* when they advance it as a pedagogical ideal and a target of pedagogical interventions, so I can analyze the implications of metacognitivist discourse.



understand this as an umbrella term that brings together an array of phenomena that all *seem* conducive to better learning outcomes, in virtue of their potential to foster knowledge of and/or control over the workings of one's own mind.

**Chapter 2 ("Complications and controversies from philosophy of mind & cognitive science")** catalogs misgivings about an assumption underlying metacognitivism – that humans can come to know our own cognitive operations and/or gain control over them. I recount historical objections and contemporary challenges to this premise, and suggest that belief in our capacity to know and control our own minds may stem from some highly convincing illusions about our minds that manifest in first-person experience.

**Chapter 3 ("Scrutinizing metacognitive training interventions")** lays out my arguments as to why belief in the efficacy of metacognitive training is unwarranted. I draw out problems with each of the major approaches to metacognitive skill-building, lay out methodological issues surrounding the assessment of metacognitive abilities, and explain the limitations of the meta-analyses on metacognitive training that have been conducted to date. I also offer explanations as to why metacognitive training may be strongly associated with improved learning outcomes, even if this training doesn't actually manipulate the psychological quantities that it purports to change.

**Chapter 4 ("An error theory for metacognitivism")** contains my speculations as to why so many educators and education researchers might persevere in the belief that metacognitive training works in the absence of solid empirical evidence. I suggest that *goalpost-shifting* (adjusting one's standards for success) and *motivated reasoning* each facilitate the uptake and persistence of beliefs in the efficacy of metacognitive training.

## 1. A WORKING GLOSSARY ON METACOGNITION IN EDUCATION

This chapter provides working characterizations for items of educational jargon to which I'll refer throughout the remainder of this project. I'll begin by stipulating a working conception of *metacognition qua pedagogical concept* (§1.1). This is meant to capture how *educators* conceive of this construct, judging by what they collectively ostend to as means of engaging students in metacognition during instruction. (Discussion of metacognition as an object of inquiry in cognitive science will be provided in Chapter 2.)

In subsequent subsections, I provide primers on:

- specific abilities that are characterized as *metacognitive skills*, and the major classes of *metacognitive training interventions* through which educators purport to enhance students' metacognitive skills (§1.2)
- *the origins of metacognition* (§1.3), including:
  - *psychological concepts predating "metacognition"* that were woven together to engender this new construct (1.3a)
  - *the circumstances of its coinage* and some presuppositions thereby built into the construct (1.3b)
  - *early critiques of the construct*, centering on its ambiguity and open-endedness (1.3c)

### **§1.1 | Metacognition qua pedagogical concept**

A drumbeat of this project is the lack of consensus regarding the criteria by which a phenomenon qualifies as "metacognitive." In the absence of clarity about its boundaries, the term has undergone an uncontrolled sprawl, encompassing a vast array of referents.<sup>1</sup> This is a

---

<sup>1</sup> An additional layer of complexity comes from unclarity about the relationship of "metacognition" to other pedagogical ideals with which it is often packaged, including *higher-order skills*, *reflectivity*, *self-regulated learning*, *autonomous* or *self-directed learning*, and *constructivism*. I've decided it's beyond the scope of what I can reasonably tackle to try to give an adequate account of the conceptual linkages between metacognition and these other constructs.

long-standing and often-discussed problem among educational psychologists,<sup>2</sup> but the contours of the debate over how metacognition *ought* to be understood have by and large been cordoned off from conversations about how to apply the concept of metacognition in the classroom.

As tempting as it is to offer up an ameliorative conception of “metacognition” (that is, an account that would constrain the extension of the term in an effort to eliminate confusion that arises from its overabundance of possible referents), I don’t think doing so here would serve my purposes. Were I to endorse a specific way that I think the term *should* be used by educators, then I’d no longer be analyzing whether educators are succeeding at achieving *what they are aiming at* by implementing metacognitive training interventions: instead, I’d be assessing whether these interventions succeed in achieving *what I think they ought to be aiming at*. So what follows should be taken as an attempt to accurately describe the construct as it is actually put to use by educators – warts and all.

To convey the breadth of “metacognition” as an umbrella term, here are some examples of teaching practices recommended as means of engaging students in metacognition, roughly organized by theme:

- ***teaching metacognition as a concept***: Positioning the concept of metacognition as an explicit topic of instruction within one’s courses – for example, by including a unit or recurring thread in one’s curriculum on “how to learn,” in which students learn what metacognition is, why it is valuable to them, and what they ought to do to actualize this in their own learning (Crooks 2009, Wilson & Conyers 2016)
- ***planning and project management***
  - **pre-writing**: Having students identify the purpose of an assignment and lay out a concrete plan for how they will approach the project (Sweetland Center for Writing n.d.)

---

<sup>2</sup> The following publications all express concern about the apparent polysemy of the construct and the unconstrained proliferation of its referents: Cavanaugh & Perlmutter (1982), Brown et al. (1983), Wellman (1983), Gleitman (1985), Paris & Winograd (1990), Borkowski et al. (2000), Schraw (2000), Efklides (2001), Veenman et al. (2006), Dinsmore et al. (2008), Kaplan (2008), Schunk (2008).

- **goal-setting:** Prompting students to define the objectives of their learning activities, so they can measure their progress against these yardsticks they have established (Schunk 1990; Paulson & Bauer 2011, Tanner 2012)
- *reflection and self-assessment activities*
  - **exit tickets:** A question or directive which prompts students to reflect upon or demonstrate their comprehension of the material covered in a class session before leaving (Wakeford n.d.)
  - **assignment/exam wrappers:** A brief assignment appended to a graded assessment, wherein students answer questions meant to help them reflect upon why they scored as they did on the assessment and how this should inform their approach to future assessments (Eberly Center for Teaching Excellence and Educational Innovation n.d.; Lovett 2013)
  - **retrospective reflective writing:** An assignment in which students are asked to describe and evaluate *their learning* (e.g., reporting processes by which they approached a learning challenge and what they learned from this approach) and/or *themselves* as a learner (e.g., analyzing one's intellectual habits, strengths, and weaknesses) (Tanner 2012, Harriet W. Sheridan Center for Teaching and Learning, n.d.)
- *cognitive monitoring*
  - **think-aloud:** Prompting students to verbalize their thinking during learning activities, so that students' inner monologue can become the topic of analysis and interpersonal dialogue in the classroom (Ritchhart, Church, & Morrison 2011)
  - **self-questioning protocols:** Providing students a list of questions they ought to ask themselves during learning activities, in order to prompt concurrent assessment of their own performance, comprehension level, and so on (Williamson 1996, Concepción 2004, Tanner 2012)
  - **confidence judgments:** During assessments, prompting students to report on their subjective level of confidence in the accuracy or appropriateness of their responses (Nietfield, Cao, & Osborne 2005)
- *study tactics*
  - **dual-coding:** Assigning learning tasks that require students to encode information both verbally and visually (Clark & Paivio 1991, Cavigioli 2019)
  - **retrieval practice:** Conducting low-stakes or ungraded mini-quizzes to prompt students to reflect the extent to which they have internalized key information (Agarwal et al. 2020)
  - **spaced practice:** Encouraging short study sessions spread across multiple days instead of cramming all study into a single session (Carpenter & Agarwal 2020)

- **mind maps (a.k.a. concept maps or semantic webs):** Having students generate a visualization of the interconnections between recently acquired concepts (Buzan 1993; Sweet, Blythe, & Carpenter 2017; Centre for Teaching Excellence, University of Waterloo n.d.)

Allegedly, each of these diverse activities (which certainly don't run the gamut of pedagogical practices recommended by metacognitivists) is a vehicle through which students will exercise *metacognition*. Yet it's hard to divine what the psychological common denominator of all of these activities could be. (Consider, for example, what mental activity the student would necessarily be engaged in both while setting goals for her performance and while spacing out study over several days, or both while learning what "metacognition" means and reporting how confident she is that she got each answer right on a quiz.)

My view is that there's no psychological common denominator to be found here – not even if one resorts to maximally vague terms like "thinking about thinking" – because "metacognition" does *not* denote a mental kind stimulated by each one of these teaching activities. Rather, what these things have in common with each other which leads them to garner the shared label of "metacognition" is that they've all been assigned instrumental value for eking out better learning outcomes.

For those who appreciate some effort at formalization, I stipulate the following description of how "metacognition" is used by educators and educational psychologists, taken as a whole:<sup>3</sup>

**Working conception of metacognition qua pedagogical concept**

When used as a pedagogical concept, *metacognition* describes any mental activities, attributes, or behaviors...

---

<sup>3</sup> I make no promises that this will capture every last usage of "metacognition" within the domain of educational discourse, nor that it will succeed in excluding phenomena which do not get labelled as "metacognitive." I only allege that this account captures the functional extension of "metacognition" more effectively than an account that tries to stick to criteria meant to pin down the relevant phenomena on the basis of psychological features alone.

- I. possession or performance of which have been theorized to produce superior learning than what one would achieve in their absence,
- II. where the learning benefits of these activities is presumed to derive from their provision of insight into and/or control over the operations of one's own mind.

In this working conception, then, the key features by which we may identify manifestations of “metacognition” are suppositions that [condition I] one would learn more or learn better with these than without them, and [condition II] these facilitate learning by stimulating mental changes conducive to greater mastery over one's own mind. The exact psychological profile of metacognitive phenomena is not defined; the key qualifier isn't belonging to a particular mental kind, but rather the expected educational outcome of possessing or performing it.

This working conception is nebulous by design, since it's supposed to include all possible referents of an extensively polysemous term. The remaining sections of this chapter will motivate this account and provide more specificity about what kind of phenomena speakers may be referring to when they invoke “metacognition” in relation to education.

## **§1.2 | Cultivating metacognitive skills**

As stated in the introduction, many educators and education researchers endorse a form of metacognitivism on which teachers ought to endeavor to produce lasting enhancements in students' *metacognitive skills*. In doing so I take them to posit the existence of an arena of knowledge-how in which students' standing can be expected to dictate their academic performance.<sup>4</sup> Confusingly, authors may enthusiastically champion the pedagogical value of

---

<sup>4</sup> Here my conceptualization draws from Veenman et al's suggestion that we can understand “metacognitive skills” as describing domains of procedural knowledge, encompassing everything under the heading of “metacognition”

cultivating metacognitive skills – and even present this as an imperative for good teaching – without clearly articulating what these metacognitive skills *are*: that is, exactly which abilities students are meant to develop.

### [1.2a] Metacognitive skills

As should be expected given the conceptual murkiness of “metacognition,” there’s no authoritative list of “metacognitive skills” to be found.<sup>5</sup> What follows is my best effort to round up a (non-exhaustive) list of abilities that educators often claim to be aiming at when they position the cultivation of metacognitive skills as a pedagogical objective. I’ve suggested a loose organizational scheme of three broad clusters into which these fall; these aren’t meant to imply ontological or functional divisions within the mind.

- **Cognitive monitoring skills.** Abilities that facilitate the acquisition of valuable information about one’s own cognition. These are often described as being mediated by *cognitive self-reflection*, which I think it’s safe to translate as *introspection directed at one’s cognitive states and processes*.
  - **cognitive self-awareness:** The ability to attain conscious awareness of (and, thereby, the capacity to verbally report the contents of) one’s own occurrent cognitive states.
  - **comprehension monitoring:** The ability to detect the extent to which one is grasping material with which one is interacting – ideally coupled with an impulse to take corrective action to fill in any gaps or misunderstanding in one’s knowledge when one is lost or confused.

---

which is *not* declarative knowledge about navigating learning challenges (2006, p. 4). However, I would emphasize that the exercise of procedural knowledge about how to accomplish certain objectives is likely dependent upon the possession and activation of relevant declarative knowledge. See Bailin et al. (1999, pp. 271-273) for an argument as to why one’s understanding of “critical thinking skills” suffers when these are treated as though they are isolatable from declarative knowledge; I suspect the same applies for metacognition.

<sup>5</sup> Neither is there much consensus on how “skills” ought to be distinguished from other nouns to which “metacognitive” is often appended. For example, “metacognitive skills” are sometimes treated as though they are interchangeable with “metacognitive strategies” (e.g., Centre for Teaching Excellence, U. of Waterloo n.d.); I resist this, since doing so can create circularity in discussions of metacognitive training, in which *strategy* instruction is among the methods used to promote metacognitive *skills*.

- **error detection & correction:** The ability to tell when one has made a mistake in one's cognition (e.g., in memory recall or deduction) and to take appropriate steps to fix these errors.
- **Executive control skills.** Abilities that enable deliberate management of one's cognitive activities and accompanying behaviors.
  - **planning:** The ability to lay out an auspicious approach before tackling an intellectually-demanding task or project. One could also include here the ability to follow through on this plan, as well as the ability to devise a new plan when one's initial approach proves ineffective.
  - **goal-setting:** The ability to specify one's objectives, so that one has a standard against which to judge one's progress or performance.
  - **strategy selection:** The ability to determine which of the tactics in one's mental arsenal is appropriate to deploy in order to succeed in navigating a particular learning challenge.
  - **policing adherence to intellectual norms:** The ability to take corrective action to align one's intellectual conduct with desirable standards & ideals.
- **Self-assessment skills.** Abilities involving the formation of accurate judgments about oneself as a mental agent.
  - **intellectual trait analysis:** The ability to characterize one's own knowledge, intellectual abilities (including strengths & weaknesses), and other mental traits in line with third-party observers and/or objective assessments.
  - **predicting one's performance:** The ability to anticipate how well one will do on a particular intellectual task, on the basis of what one knows about oneself and the demands of the task.

### **[1.2b] Metacognitive training interventions**

Now that we have some sense of the abilities exhibited by someone who could be described as metacognitively-skilled, let's take a quick look at three broad approaches by which instructors purport to enhance students' metacognitive skills. Though I have some a priori concerns with each of these methods, I'll withhold commentary on their prospects for achieving their aims until Chapter 3, at which point I will have laid the groundwork to integrate these concerns with additional empirically-informed objections.



### **(1.2bI) Explicit modeling of metacognition**

Metacognitivists may recommend that teachers adjust their pedagogical approach to put their own exercise of metacognition on display for students' benefit (Quigley et al. 2018, pp. 16-17). For example, this would mean that a logic professor teaching natural deduction should make sure that whenever she demonstrates proofs in class, she goes out of her way to verbalize the thought processes that motivate each successive inference she makes.

A presumption here is that instructors are exemplars of proficient learning (having attained enough knowledge to be in a position to teach), which entails that they themselves are metacognitively adept; as such, they can serve as role models whose metacognitive activity students can emulate. To this end, Collins, Brown, and Newman (1987) recommend that educators think of themselves as engaging learners in "cognitive apprenticeship," wherein they are obligated to model any thinking skills they want students to acquire, in order to provide ample opportunities for students to observe instances of behaviors they are supposed to master.

### **(1.2bII) Explicit strategy instruction**

Another common metacognitivist recommendation is to teach students "metacognitive strategies," which I'll characterize as *methods for monitoring and regulating one's own thinking and behavior that can be deployed to produce desired results in one's learning*. The reasoning behind this method is that if students not only know what these strategies are but also understand when each strategy is useful, they will be metacognitively skilled in the sense of being disposed to use these strategies to their advantage under appropriate circumstances while navigating learning challenges (Quigley et al. 2018, pp. 12-15). Teaching these strategies is often presented as *helping students learn how to learn*: imbuing them with a sort of practical wisdom that makes them adept at managing whatever tasks teachers throw at them.

As should come as little surprise given the semantic flexibility of “metacognition,” there’s no consensus on what qualifies as a “metacognitive strategy”; in practice, that label has been applied generously to accommodate many topics of instruction that may enhance students’ academic performance, including helpful tactics for reading comprehension, writing, problem-solving, studying, and test-taking (for example, *checking all of the options on a multiple choice question before selecting an answer*; Flavell 1976, p. 232). Those who make an effort to maintain some specificity for the label restrict “metacognitive strategies” to *methods of planning, monitoring, and modifying one’s procedure for tackling a learning task* (Dignath et al. 2008, Iris Center, 2020).<sup>6</sup> Examples of this more constrained usage include *self-talk* (elaborating an inner monologue about how one is approaching a learning task throughout task performance) and *self-monitoring* (leading oneself through a mental checklist to assess one’s progress or the products of one’s efforts) (Iris Center). Often both of these activities are merged together in *self-questioning protocols*, in which one poses questions to oneself within one’s inner monologue to manage one’s approach and assess how well one is doing throughout a learning task. To teach this strategy, instructors standardly provide students with a bank of questions to use as a framework for inquiry into their own mental activity; ideally, students will gradually internalize these questions and deploy them as part of an automatic self-monitoring routine deployed whenever the student attempts relevant tasks.

---

<sup>6</sup> “Cognitive strategies,” in contrast, are specific patterns of thinking and behavior that one chooses to implement on the basis of the information one acquires through planning and monitoring one’s learning. Though some cognitive strategies are applicable in many circumstances (e.g., active repetition and chunking of information to aid memorization, breaking a large problem into manageable problems; Dignath et al. 2008), cognitive strategies are often *domain-specific* (pertaining to a specific academic subject) and *situational* (to be deployed when solving particular types of problems or performing specific tasks). The Iris Center (2020) offers this example: if reading a word problem in a high school math class yields the information that this is a trigonometry problem involving a right triangle, 1) drawing a diagram and 2) using the “SohCahToa” mnemonic are two math-specific cognitive strategies the student can use to determine what she needs to plug in to solve for the unknown quantity (length or angle).

### (1.2bIII) Reflective activities

Perhaps the most frequently mentioned intervention for enacting metacognitive skill gain is to prompt students to engage in reflection “on their own performance, their own learning and how they feel they have changed” (Wright 2015, p. 225).<sup>7</sup> The conceptual relationship between *reflection* and *metacognition* is, in my eyes, chronically underspecified – but I gather that metacognitivists typically take reflection to be either a) an instantiation of metacognition or b) a mental mechanism which enables the exercise of metacognition.

At any rate, this pedagogical approach seems to assume a straightforward “practice makes perfect” model of skill-building: to improve metacognitive skill, one must engage in reflection regularly. All of the following are examples of exercises that purportedly invoke reflection in one’s students:

- administering questionnaires about students’ beliefs pre-/post-instruction (Stokes 2012)
- assigning essays in which students are required to report on how their beliefs have changed; e.g., to “articulate their baseline beliefs, the critical analysis opportunities that challenged or reinforced these beliefs, and then discuss their current examined perspectives” (Hornsby 2016, p. 55)
- requiring students to critically assess themselves as a thinker, identifying their intellectual proclivities and judging themselves against standards for intellectual excellence (Crooks 2009)
- prompting students to writing a note about a point of confusion that arises for them during a course session, sometimes called “the muddiest point” (Concepción 2004, Harriet W. Sheridan Center n.d.)
- having students submit a written report recounting how and what they were thinking while executing a particular learning task (Concepción 2004)
- having students complete assignment “wrappers” after getting grades back, where they discuss how they prepared for the assignment and what they learned about the relationship between their approach and the outcome of their efforts (Eberly Center n.d.; Lovett 2013)
- assigning essays where students must report what they learned and how their choices

---

<sup>7</sup> See also Sweetland Center for Writing (n.d.), Tanner (2012), Kaplan et al. (2013).

enabled them to acquire said knowledge or skills (Peterson 2001)

- an end-of-semester exercise where students write a letter to a hypothetical student enrolled in the upcoming semester's iteration of the course, explaining a) which learning strategies lead to their success in the course and b) what they would have done differently to achieve better outcomes (Hornsby 2016)

These activities are a diverse lot, which shouldn't be all that surprising if we consider how open-ended "reflection" is. On the whole, we can classify them into two clusters: reflective activities that aim at *generating self-knowledge* and those that serve *self-assessment*, involving the formation of judgments about oneself or one's actions. (These aren't mutually exclusive, because making judgments about oneself involves generating a conception of what kind of person one is, intellectually speaking.) Moreover, activities that aim at *generating self-knowledge* can be further subdivided according to the type of knowledge aimed at: e.g. awareness of one's *occurrent mental states or cognitive processes* in the midst of a learning task, familiarity with one's *intellectual dispositions* (concerning patterns of mental activity over time), recognition of *facts and concepts one has assimilated* into one's mental stores, or acknowledgment of one's *standing beliefs and attitudes*.

### §1.3 | Genealogy of "metacognition"

One approach to wrapping one's head around the breadth and variance of "metacognition" is to view the construct as a conglomeration of multiple research traditions (Brown et al. 1983, Brown 1987, Hacker 1998, Dunlosky & Metcalfe 2009). In this section, I'll relay enough backstory for "metacognition" to clue readers into some historical contingencies that have contributed to its ambiguity, hewing closely to the account offered by Brown (1987). This isn't a comprehensive genealogy of the concept, but rather a selective examination of some specific aspects of how the term came to be and how it has since been applied.

Retracing the coinage of “metacognition” will go a long way towards explaining why the term is so polysemous. In short, its christening event in published text (Flavell 1976) was not intended to constrain its referential scope; rather, Flavell embraced the potential for the term to encompass a huge array of learning-related phenomena that had previously not been brought together under a common heading. Brown deems the construct “problematic from its inception”:

Metacognition refers loosely to one’s knowledge and control of [one’s] own cognitive system. Two primary problems with the term are: it is often difficult to distinguish between what is meta and what is cognitive; and there are many different historical roots from which this area of inquiry developed. The confusion that follows the use of a single term for a multifaceted problem is the inevitable outcome of mixing metaphors. (1987, p. 66)

Subsequent calls to draw the borders of the term more sharply and reign in its proliferating applications in the ensuing years haven’t succeeded in reversing the trend Flavell set in motion, of using “metacognition” to refer to a “loose confederation of topics” (ibid., p. 106).

### **[1.3a] “Historical roots”**

In Brown’s telling, “metacognition” merges ideas arising from four traditions of psychological research, concerning *reflection*, *executive control*, *self-regulation*, and *other-regulation* (i.e., inducing self-regulation skills in others).<sup>8</sup>

#### **(1.3aI) Reflection**

The first thread of research Brown identifies carries forward a millennia-old inquiry into the apparent human capacity to “consider one’s own cognitive operations as objects of thought; [or] to *reflect* on one’s own thinking” (p. 69, my italics). Brown identifies this as a recurring

---

<sup>8</sup> Other histories of metacognition (e.g. Dunlosky & Metcalfe 2009, Tarricone 2014) largely agree on the significance of prior research in these areas, though they emphasize these threads to different degrees in accordance with their respective emphases regarding what metacognition is.

theme of Western philosophy, most notably in the works of Plato, Aristotle, Spinoza, and Locke (p. 70); it underwent a resurgence around the turn of the twentieth century, with the professionalization of psychology as a field of science, particularly among researchers of the Wurzburg School (ibid.).

The most recent strands of this thread focus on the ability to verbalize thoughts about one's own cognitive operations. Central concerns are the extent to which humans have access to their own cognitive activity through introspection, and the extent to which verbal reports accurately convey the nature and content of the cognitive operations one can access through introspection.<sup>9</sup> Of particular interest to educators is the relationship between the ability to verbalize one's own cognitive processes and cognitive performance: i.e., whether "thinking aloud" (verbally reporting one's occurrent thoughts synchronously during a learning activity) enhances or hinders execution of learning tasks.<sup>10</sup>

### **(1.3aII) Executive control**

Another contributor to the genesis of "metacognition" is the concept of *executive control*, posited in information-processing theories of cognition to designate "a central processor, interpreter [sic], supervisor, or executive system capable of performing an intelligent evaluation of its own operations" (1987, p. 79, my italics). These theories assert that cognitive processes are organized into a hierarchy, in which higher-level processes receive inputs from lower-level ones; within such systems some form of "executive power . . . must be attributed" to account for the system's ability to actively manage activity occurring within the system – for example, shifting a task from automatic to controlled processing, upon detection of some kind of

---

<sup>9</sup>See especially Nisbett & Wilson (1977), Smith & Miller (1978), Ericsson & Simon (1980), White (1980).

<sup>10</sup>Extant research (at the time of publication) suggested that thinking aloud can be beneficial if students are asked to introspect about cognition available to consciousness, but "asking subjects to report on internal events that are not readily available to such [conscious] inspection may significantly impair the processes on which they must report" (p. 77).

disruption to processing fluency (pp. 80-81). This executive has the ability to generate plans and orchestrate them, using input from monitoring functions to implement regulatory control over subcomponents of these plans; however, positing such an executive introduces the challenge of how to make sense of this function without anthropomorphizing a subcomponent of the mind as a homuncular “inner man” (Skinner 1971, p. 14; quoted by Brown 1987, pp. 81-82).

At any rate, Brown reports that “by adopting the notion of a central processor or executive system imbued with very fancy powers, developmental psychologists gained a powerful analogy through which to consider the development of efficient learning” (p. 79): More efficient learning could be understood as the result of having a more skillful executive at the helm of the learning enterprise.

### **(1.3aIII) Self-regulation**

This next thread of research concerns the ability to detect and correct errors in one’s own thinking, originally termed *autocriticism* (Binet 1909) but later rebranded as *self-regulation* (pp. 88-89). To this day metacognition is frequently packaged into a broader educational ideal of *self-regulated learning* – the upshot of which is that educators should facilitate students’ development of a foundational set of skills and dispositions that enable students to autonomously manage their own education.<sup>11</sup>

A central figure in this area is Jean Piaget (1976), who postulated a developmental trajectory in which children move from automatic, unconscious adjustment of motor actions towards progressively more cerebral, conscious forms of cognitive self-regulation (where conscious access to one’s thoughts is indicated through verbalization). Brown attests that Piaget’s trajectory is borne out in research that suggests age-related emergence of successive

---

<sup>11</sup> Accordingly, one can find many pedagogical tactics marketed both as “metacognition” and as methods of cultivating *self-regulation*. See Zimmerman (1990, 2002, 2008), Winne (1995), Boekaerts (1999), Schraw, Crippen, & Hartley (2006), Quigley et al. (2019).

tiers of self-regulatory capacity.<sup>12</sup> At early ages, children work through intellectual challenges (such as the Tower of Hanoi problem)<sup>13</sup> through trial-and-error, without “conscious surveillance” of their thought processes nor the ability to verbally report these processes (pp. 90, 99); with age, children begin to demonstrate the ability to plan ahead to pursue successful strategies and can start to articulate the procedures they used. At the final stage of development, children can not only articulate their procedures but describe the principles underlying these procedures; moreover, conscious awareness and/or verbalization of these principles seems to play a causal role in enabling successful completion of a challenge.

Piaget attributes skillful adult self-regulation to the ability to engage in *reflected abstractions*: roughly, the ability to subject one’s plans and principles to conscious scrutiny. Taking one’s own problem-solving strategies as objects of thought unlocks a purely mental form of “theory testing”: wherein one can determine “on the mental plane” (presumably through reasoning, mental imagery, and whatnot) whether or not a plan or method is hypothetically feasible, without having to put these into action to witness the consequences (Brown 1987, p. 91). Piaget advanced the notion that control of one’s behaviors is directly enabled by conscious reflection upon one’s plan of action.

#### **(1.3aIV) Other-regulation**

This final antecedent to “metacognition” addresses what Brown calls “*other-regulation*”: the ability to induce the development of self-regulation in children through social interaction (p. 100). She associates this work primarily with Lev Vygotsky (1978), for whom “the fundamental process of development is the gradual *internalization* and *personalization* of what was originally a

---

<sup>12</sup> Piaget’s theories have lost much of their luster in the ensuing decades, as experimental studies have superseded observation as the prevailing paradigm in developmental psychology research. See Gopnik (1996).

<sup>13</sup> This is a puzzle where a stack of progressively smaller rings must be moved from one vertical pole to another, moving one ring at a time such that a larger ring is never placed on top of a smaller one. See <https://www.mathsisfun.com/games/towerofhanoi.html>.



social activity” (Brown 1987, *ibid.*; my italics). That is, the ability to self-regulate doesn’t emerge as a result of hard-wiring; rather, it unfolds in a predictable pattern and time course as a function of the mediating function that parents, teachers, older siblings and so on play by providing overt feedback and guidance to the child as they work through intellectual challenges. “If for some reason the child is deprived of a constant history of such interaction, the development of a battery of self-regulatory skills is unlikely to occur,” Brown attests, asserting the need for children to witness explicit models of self-regulation (p. 103).

### **(1.3aV) The throughline of these “roots”**

What exactly do these research traditions have in common that warrants weaving them into a novel construct of “metacognition”? On my reading of Brown’s discussion, what these have in common is that theorists within each of these research traditions posited their respective topics as the psychological basis for *intelligence*, the pinnacle of human cognitive development. (Technically, theorists of other-regulation position their topic as the means by which educators cultivate human intelligence — but this reinforces the message that intelligence stems from self-regulation.) The reason why these threads of psychological research fed into the construction of “metacognition,” then, is *not* because they had all been examining a common mental kind (or set thereof) from different angles. Rather, they were woven together to engender a new *unified theory of intelligence*, on which what we call “intelligence” is underpinned by *metacognition* — to which reflection, executive control, and self-regulation all belong.

Brown attests that researchers would benefit from reimagining intelligence — that which purportedly accounts (at least in part) for why different students arrive at different performance outcomes when they receive the same instruction — as a matter of *metacognitive* ability. Moreover, Brown and colleagues (1983) appear to have originated the phrase “intelligent novices” to denote students who, in virtue of superior metacognitive skills,

outperform their peers in attaining domain-specific knowledge and skills, without having a leg up in prior knowledge or experience in that domain.<sup>14</sup>

### **[1.3b] Coinage & early problems**

“Metacognition” was coined by developmental psychologist John H. Flavell, as a novel umbrella term for a disparate array of phenomena that were collectively thought to account for age-related development of intellectual aptitude, as well as interpersonal differences in academic achievement among same-age peers (1976). The term is an extrapolation from *metamemory*<sup>15</sup> (Dunlosky & Metcalfe 2009, p. 31), which Flavell and colleagues began using to describe processes by which children monitor and assess facets of their memory (e.g., in order to predict how well they would be able to recall vocabulary terms on a memory test; Flavell, Friedrichs & Hoyt 1970).

It isn't entirely clear what prompted Flavell to shift towards the more general construct of “metacognition,”<sup>16</sup> but by the mid-'70s he seems to have become convinced that metamemory

---

<sup>14</sup> This phrase was later taken up by John Bruer (1993, p. 77-78), in a book from which many educators appear to have derived their understanding of the construct. Within the literature on teaching philosophy alone, Bruer's characterization of metacognition crops up in Concepción (2004), Stokes (2012), Maynes (2013), and Wright (2015).

<sup>15</sup> I haven't encountered any testimony from Flavell on why he chose the “meta-” prefix to establish this new locus of psychological research. At any rate, it has proven a consequential choice, as its ambiguity invites diverse interpretations of what metacognition *really* is, judging from its etymology. Plumbing various dictionaries dredges up laundry list of translations for “meta-”; Thomas (1984, p. 16) relays that it has been used to signify at least as many distinct senses as *changed, after, behind, beyond, alternating, later, transcending, reversed, higher, between, over, and with*. Yet none of these senses necessarily captures the content of “meta-” intended in “metacognition”.

Thomas suggests that “metacognition” is instead best understood as one of many implementations of an “assumption that the prefix *meta* . . . would elevate the original meaning of . . . a word to a higher, more abstract philosophical level.” (Amusingly, this assumption seems to stem from a “bibliographic accident,” where Andronicus of Rhodes' choice to label Aristotle's book on “the nature of being” *meta ta phusica* – intending *after [the book on] physics* – resulted in *metaphysics* being “misapprehended as meaning ‘the science of that which transcends the physical’” (ibid.)) Nelson & Narens (1990) also endorse the suggestion that “metacognition” is meant to recapitulate the pattern established with *metamathematics* and *metalanguage* (attributed to Hilbert (1927) and Carnap (1934), respectively), which established the objects of these fields of study as transcending or existing beyond the typical purview of their base subjects (i.e., math and language).

<sup>16</sup> Piaget's account of cognitive development, and related empirical studies on the development of problem-solving ability, seem to have played some role in leading Flavell to conclude that memory wasn't the only arena in which children acquire the ability to engage in monitoring and self-assessment. See Dunlosky & Metcalfe (2009, pp. 31-33).

research had unearthed just one facet of a broader species of mental activity. Here's the first characterization of "metacognition" that Flavell offers in print; I've italicized vague language for emphasis:

'Metacognition' refers to one's knowledge concerning one's own cognitive processes and products *or anything related to them*, e.g. the *learning-relevant properties* of information or data. For example, I am engaging in metacognition (metamemory, metalearning, metattention [sic], metalanguage, *or whatever*) if I notice that I am having more trouble learning A than B; if it strikes me that I should double-check C before accepting it as a fact; if it occurs to me that I had better scrutinize each and every alternative in a multiple-choice type task situation before deciding which is the best one; if I become aware that I am not sure what the experimenter really wants me to do; if I sense that I had better make a note of D because I may forget it; if I think to ask someone about E to see if I have it right. *Such examples could be multiplied endlessly.* In any kind of cognitive transaction with the human or nonhuman environment, a variety of information processing activities may go on. Metacognition refers, *among other things*, to the active monitoring and consequent regulation and orchestration of these processes in relation to the cognitive objects or data on which they bear, *usually* in the service of some concrete goal or objective. (Flavell 1976, p. 232)

Flavell left "metacognition" open-ended by identifying it with knowledge of "anything *related* to [one's cognitive properties or processes]," without specifying the nature of said relationship. Then, he relies upon ostension to examples (without spelling out how each of these qualifies as knowledge of one's own cognitive processes, cognitive products, or "related" ephemera), instead of strictly defining the boundaries of the construct. Finally, he redescribes metacognition as referring (non-exclusively!) to "monitoring," "regulation," and "orchestration" of cognitive processes, "usually" in a goal-directed manner.

### **(1.3bI) Knowledge about cognition & regulation of cognition**

In a retrospective on the status of metacognition research, Ann Brown alleges that what Flavell created in coining *metacognition* — "refer[ring] loosely to one's knowledge and control of [one's] own cognitive system" — is a Frankensteinian mishmash of previously discrete categories of psychological phenomena (1987, p. 66). She elaborates that *knowledge of cognition* and *regulation of cognition* had previously been approached via separate research programs,

because phenomena belonging to each of these categories had been attributed distinct properties.<sup>17</sup> Brown suggests that Flavell just so happened to address both categories in his “metamemory” studies, and hence to him, it made sense to subsume them under a common heading (*ibid.*, pp. 68-69). This was a fateful decision, which has ever since made it difficult to offer any simple answers about the properties of metacognition:

The tension generated by the use of the same term, metacognition, for the two types of behavior [knowledge and regulation of cognition] is illustrated by the fact that the leading proponents in the field tend to answer questions about the nature of metacognition with: ‘It depends.’ *Is metacognition late developing?* It depends on the type of process referred to. *Is metacognition conscious[?]*: It depends. . . (1987, p. 68; italics added to denote rhetorical questions)

Despite the inconvenience it has created, the bifurcation of metacognition into two primary components – knowledge and regulation of cognition – is often replicated in summaries of the construct.<sup>18</sup> Yet even this isn’t a matter of consensus; some authors instead advance a tripartite conception of metacognition which draws out *monitoring of cognition* – roughly, introspective awareness of one’s occurrent cognitive states or processes – as a third element (Pintrich, Wolters, & Baxter 2000; Tobias & Everson 2000; Dunlosky & Metcalfe 2009). Otherwise, monitoring tends to be characterized as an integral component of *regulation* of cognition, insofar as it serves the purpose of informing regulatory adjustment of one’s cognitive activities (e.g. Quigley et al. 2018); note, however, that a case can be made for lumping monitoring into *knowledge* of cognition, if one emphasizes monitoring as a means of accessing or generating self-knowledge about one’s mental states.

---

<sup>17</sup> *Knowledge* of cognition mostly captured phenomena which were “stable” across time (i.e., belonging to long-term memory, as “a permanent part of one’s naïve theory of the topic”), “statable” (i.e., verbalizable), and “late-developing” on an age-related basis. In contrast, *regulation* of cognition tended to be understood as “relatively unstable, not necessarily stable, and relatively age-independent” (p. 68).

<sup>18</sup> Each of the following texts affirms that “metacognition” consists of two major clusters of phenomena – one involving knowledge, awareness, and/or monitoring of cognition, and the other involving regulation and/or control of cognition: Wellman (1985), Nelson & Narens (1990), Paris & Winograd (1990), Livingston (1997/2003), Schraw (2001), Lai (2011).

There's even less alignment regarding how to divide a "domain . . . that lacks coherence" into measurable subcomponents for research purposes (Veenman et al. 2006). For decades, researchers have been lamenting that there isn't enough clarity on the sub-elements of metacognition to be sure whether distinct research teams are accumulating evidence on the *same* metacognitive phenomena or tackling *distinct* components.<sup>19</sup> While early researchers were able to write off this lack of consensus as an expected attribute of an inchoate area of inquiry (Wellman 1983, Brown et al. 1983), the field has yet to achieve anything close to a unanimous conceptual model for metacognition.

### **(1.3bII) Metacognition vs. mere cognition**

Brown (1987) also points out how unsuccessful Flavell's non-definition is at distinguishing "metacognition" from cognition *simpliciter*. Here I'll begin using "mere cognition" to refer to cognition which is *not* meta-; "cognition" can then encompass both mere cognition *and* metacognition – assuming for now that metacognition is a subtype of cognition writ large.<sup>20</sup>

One *prima facie* plausible method of drawing the distinction between metacognition and mere cognition is to posit that the signature feature of *metacognition* is taking *one's own* cognitive activity as an intentional object, whereas *mere* cognition adopts intentional objects outside of this scope. Many quick glosses of "metacognition" endorse this division by defining metacognition as "thinking about *your own* thinking." To elaborate: on this account, *mere* cognition would consist of first-order mental states which represent phenomena out in the world (such as the occurrent thought that "It is raining"), whereas *metacognition* consists of

---

<sup>19</sup> See Tarricone 2014, pp. 3-6 for many citations to this effect.

<sup>20</sup> Some phenomena that have fallen under the referential umbrella of "metacognition" are not obviously cognitive themselves, but might be more accurately considered *behaviors* with consequences for cognitive activities. See Appendix §A3 for discussion of how disagreement about the scope of "cognition" yields unclarity about the scope of "metacognition."

*higher-order* mental states involving propositional attitudes about one's own lower-order mental states ("I believe that it is raining").<sup>21</sup>

However, when Flavell allows that metacognition can include knowledge or awareness of *information in the world that could be relevant* to learning, he opts not to assign "metacognition" even the modicum of specificity that would have come from limiting it to cognition which is *internally* directed (concerning states within one's own mind) rather than *externally* directed (at objects out in the world). At this point, anything that might have previously been deemed cognitive (because it is directed out into the world) suddenly qualifies as *metacognitive*, because this term ostensibly refers to cognition wherein the content is "relevant to learning."

I, for one, am hard-pressed to tell what content out in the world *isn't* "relevant to learning." I'm being somewhat facetious, because I can grasp how the particular learning task one is currently undertaking might set boundaries on which phenomena out in the world it would be appropriate to attend to and process. E.g., if one's current learning task is to decipher the meaning of a text, the text itself (and its subcomponents) would be a much more relevant intentional object to aid completion of that task, compared to other features of one's environment (such as the texture of the couch cushions one is sitting on while reading the text). But the point is that educators would have benefitted from more clarification from Flavell on how to cash out "relevance to learning," given the vagueness of "relevance" as a relational property.<sup>22</sup>

---

<sup>21</sup> This interpretation of metacognition will be revisited in §2.3. Appendix §A2 addresses how some of the semantic breath of "metacognition" reflects loose interpretations of "higher-order" or "higher-level" as *supererogatory* or *advanced*, instead of denoting these states' positioning within a hierarchical cognitive architecture.

<sup>22</sup> An accompanying issue concerns how to understand "learning". This can be a *success term*, wherein one only qualifies as "learning" if one is undergoing enduring changes (in beliefs, memory, habits, and so on) which can be detected through an appropriately designed assessment; beyond this, "learning" could be reserved to undergoing changes that occur in a desirable direction (i.e., *good learning*). But educators often use "learning" in a broader and looser sense, as a stand-in for "participating in an educational enterprise," not specifying the outcome of such participation. For example, the phrase "learning activities" is used to designate *anything* students do during a lesson or independent study (taking notes, speaking to each other in small group discussions, "free writing" in response to a prompt, preparing a study guide, etc.); it is not guaranteed that these activities will actually produce enduring

To their credit, other authors have since been more stringent in constraining “metacognition” to higher-order cognition, including *planning, monitoring, evaluating, and modifying* one’s own cognitive activity (e.g., Papaleontiou-Louca 2003, Dignath et al. 2008, Iris Center 2020). But more often than not, characterizations of metacognition aren’t accompanied by clarifications on in what way this construct should be differentiated from cognition *simpliciter*.

### **(1.3bIII) Presumed benefits to learners**

It’s also important to note that the specific examples of “metacognition” that Flavell offered are all descriptions of thoughts and actions that would be *beneficial* to a learner:

For example, I am engaging in metacognition . . . if I notice that I am having more trouble learning A than B; if it strikes me that I should double-check C before accepting it as a fact; if it occurs to me that I had better scrutinize each and every alternative in a multiple-choice type task situation before deciding which is the best one; if I become aware that I am not sure what the experimenter really wants me to do; if I sense that I had better make a note of D because I may forget it; if I think to ask someone about E to see if I have it right. (1976, p. 232)

For a few of these examples, the benefit to the learner has to do with how the action conduces to the learner’s fulfillment of *epistemic norms* (e.g., that truth-commitments should be based upon careful consideration of the evidence, or that one should subject one’s beliefs to verification procedures). From these examples, we can glean how some thinkers have come to see “metacognition” and “critical thinking” as synonyms, or that metacognition is the mental mechanism by which the intellectual ideal we call “critical thinking” operates.<sup>23</sup> But most of the examples Flavell provides have *practical* benefits to the learner, allowing them to take corrective action to optimize their performance in their current task. This applies to the examples of *observing while studying that one item is more difficult to commit to memory than another, realizing*

---

changes in individual learners, let alone changes that match an instructor’s intentions or any other normative standards.

<sup>23</sup> Scriven & Paul (2008); see also Possin (2008) and Mulnix (2012) for discussion.

during an exam that one should examine all options of a multiple-choice question, recognizing that one needs clearer instructions in order to complete an experimental task correctly, and realizing that one might forget an item unless one records it somewhere for later reference. In all of these cases, the learner is recognizing the salience of information that would allow them to adjust their approach in order to achieve better outcomes (e.g., devoting more study time to the item that is harder to absorb, switching from choosing the first correct option one sees to using a process-of-elimination approach to select the best answer, asking for clarification on the task instructions, and taking a physical note of the item that might otherwise be forgotten).

Some important questions are left unanswered by Flavell's account of metacognition, regarding the normative standards (or lack thereof) for the construct. First off, do observations, beliefs, or self-attributions of mental states need to be *veridical* in order to count as "metacognition"? For example, the practical benefit of noticing that one is "having more trouble learning A than B" would seem to be contingent upon whether one really *is* struggling more with committing A to memory (or developing A as a habit, if it's knowledge how rather than knowledge that) compared to B. If one adjusts one's study time or practice to focus on A when they would really stand to devote more effort to B, then it's actually counterproductive to "notice that [one is] having more trouble learning A than B" – or at least, counterproductive to act upon this knowledge.

Relatedly, *is there such a thing as bad metacognition?* Flavell's examples all describe realizations that lead learners to refine their activity in ways that are conducive to improved cognitive performance (e.g. being able to remember what is important, answering test questions so as to reflect one's full understanding of a topic) or satisfaction of epistemic ideals. Flavell notes that these examples are far from exhaustive, but his statement that "such examples could be multiplied endlessly" provides no indication of whether the unstated instances of



metacognition he has in mind are *all* good for the learner (1976, p. 232). It's not obvious whether instances where learners latch onto ineffective strategies (say, that one should select whichever multiple choice option first strikes them as correct, instead of considering each of the response options methodically) qualify as *bad* metacognition, or whether the status of "metacognition" is reserved for cognitive activity that satisfies particular criteria of desirability and/or conduciveness to success.

Whatever Flavell intended with his initial characterization, his efforts to popularize the concept and put it into pedagogical practice have created an impression that "metacognition" refers to a cluster of phenomena which all account for gains in learning. In fact, his most-cited explanation of metacognition advances unsubstantiated suspicions about what could be gained by positioning metacognition as a target of educational interventions:

*Lack of hard evidence notwithstanding, however, I am absolutely convinced that there is, overall, far too little rather than enough or too much cognitive monitoring in this world. This is true for adults as well as for children, but it is especially true for children. For example, I find it hard to believe that children who do more cognitive monitoring would not learn better both in and out of school than children who do less. I also think that increasing the quantity and quality of children's metacognitive knowledge and monitoring skills through systematic training may be feasible as well as desirable. (Flavell 1979, p. 910; my italics)*

It would be one thing if Flavell advanced a normative claim that people ought to do much more cognitive monitoring, because that is his vision of what comprises competent and responsible thinking. But here he is admitting that he is antecedently committed to an *empirical* claim about the efficacy of metacognitive monitoring as a tool for improving learning outcomes within formal education.<sup>24</sup>

In this passage, Flavell discloses that he "find[s] it hard to believe" that metacognitive monitoring *isn't* a key to unlocking greater levels of achievement. Despite acknowledging some

---

<sup>24</sup>His basis for this impression is a single study by Brown et al., on a sample of about 60 developmentally disabled children (Brown, Campione, & Barclay 1978).

potential externalities of pursuing metacognitive training,<sup>25</sup> Flavell advances the message that enhancing students' metacognitive abilities "may be feasible as well as desirable," with the expectation that confirmatory experimental evidence will follow (1979, p. 910).

### **[1.3c] Early critiques of the construct**

Upon recapping Henry M. Wellman's (1983) unsuccessful effort to pin down definitive features of the concept, Brown questions whether "metacognition" is unified enough to be served by a single heading:

Of some concern, however, is whether the associated phenomena are linked closely enough to warrant the use of a single family name; that is, does it refer to family resemblances within an ill-defined, natural, or fuzzy category, or many categories? And, would it not be better, at this stage, to abandon the global term and work at the level of subordinate concepts, which are themselves fuzzy?

. . . [T]here are nontrivial problems associated with the current blanket usage of the term. . . . At present it is difficult to answer critical questions about metacognition, such as: "Is it late developing?" "Is it general or domain specific?" "Is it conscious?" without pausing to ascertain which type of knowledge or process is in question. Although metacognition may turn out to be a fuzzy concept with indistinct boundaries, this degree of imprecision is not acceptable as a basis of scientific inquiry. (pp. 106-107)

After airing these grievances about the confusion and hindrance to scientific progress engendered by this term, she writes that "for clarity and communicative efficiency, a case could be made that the term *metacognition* should be pensioned-off, or at least severely restricted in its extensional reference" (p. 106, my italics) – but doesn't offer a method of how this pensioning-off should be achieved.<sup>26</sup>

---

<sup>25</sup> Immediately before disclosing that he is "absolutely convinced" that the possible benefits of metacognition in pedagogy warrant implementation in the absence of "hard evidence," Flavell considers some counterpoints:

For example, how much good does cognitive monitoring actually do us in various types of cognitive enterprises? Also, might it not even do more harm than good, especially if used in excess or nonselectively? Think of the feckless obsessive, paralyzed by incessant critical evaluation of his own judgments and decisions. Such questions suggest legitimate caveats about educational interventions in this area. (1979, p. 910)

<sup>26</sup> She does draw attention to two proposals to constrain the usage of the term. Wellman (1983) suggests that metacognition should only refer to "knowledge of cognition" – but Brown observes that this would continue to

A few years later, Paris & Winograd (1990) elaborated on the practical consequences of the lack of clear boundaries for “metacognition”:

Any cognition that one might have relevant to knowledge and thinking might be classified as a metacognition, and thus inclusive definitions are impossible. . . . Flavell and others have used prototypical examples of metacognition to illustrate metacognition rather than operational definitions that may constrain the construct. Thus, metacognition remains open-ended and definitions of metacognition almost become projective tests. Researchers bemoan the imprecision of the term and attribute it to those things that they feel are important about teaching and learning. (1990, p. 19)

As I understand it, these authors are suggesting that “metacognition” has taken on the status of a catch-all category for whatever educators want to designate as important factors in improving students’ performance outcomes. That’s because the clarity that *has* accompanied use of the concept has been around its normative valence: Flavell and other early advocates for the concept advanced consistent messaging that *metacognition is beneficial to learners*. Meanwhile, because the boundaries of the construct are so diffuse, there’s nothing stopping particular educators from classifying borderline phenomena as “metacognition,” so as to make the case that these phenomena are, by extension, *also* beneficial to learners.

Admirably, Brown admits to her own earlier participation in the unprincipled use of “metacognition” to designate *that which is good for learning*. Baker & Brown (1981) characterized many well-established strategies for effective reading as “metacognitive,” but in retrospect, Brown acknowledges that this labeling was hasty: A particular activity — such as asking oneself questions while reading — “can be seen as the [cognitive] strategy itself (looking for main points), its monitoring function (a metacognitive ability), and a reflection of the knowledge (also

---

group disparate types of knowledge together in ways that would produce confusion (1987, p. 106). She seems to see some promise in Gleitman’s (1985) suggestion to limit the term’s usage to “stable and storable” knowledge about cognition (*ibid.*); doing so would mean disaffiliating cognitive *processes* (“such as planning-ahead, monitoring, resource allocation, self-questioning, self-directing, etc.”) from metacognition. Another effort in this vein (which postdates Brown’s chapter) is Paris & Winograd’s recommendation to use “metacognition” only for “knowledge about cognitive states and abilities that can be shared among people” (1990, p. 21).

metacognitive) that it is an appropriate strategy to employ in a given situation” (1987, p. 66).<sup>27</sup>

In short, Baker & Brown could have elaborated their discussion of reading strategies and explained their effectiveness *without* positioning these strategies within the trendy new theoretical framework of “metacognition.” To that end, Brown admits that these “strategic reading activities” were effectively discussed and used as experimental target by educational psychologists in the early twentieth century without recourse to the label of “metacognition” (p. 67). She sees reason to emphasize that the introduction of “metacognition” into educators’ lexicon is far from a revolutionary, paradigm-shifting discovery;<sup>28</sup> instead, the term and its accompanying theory of successful learning turns out to be “a new name for some old ways of thinking.”<sup>29</sup> In earlier writing, Brown acknowledged that “there has been some serious concern that ‘metacognition’ is an epiphenomenon recently elevated and dignified with a new title, but really the stuff that the problem-solving literature has been concerned with all along” (1977, p. 3). She adds that skeptics about this apparently new construct “will be pleased to note the similarity between [metacognitive] activities and the activities traditionally considered under the heading ‘study skills’ . . . The area of metacognition is not as new as it would appear” (*ibid.*, pp. 4-5).

---

<sup>27</sup> Across more recent literature on metacognition, one still encounters inconsistency as to whether a particular learning activity qualifies as the use of a *cognitive* strategy or a *metacognitive* strategy. For example, the use of mnemonic devices is described as a cognitive strategy by Quigley et al. (2018), but as an instance of metacognition by Dunlosky & Metcalfe (2009). Strategies in particular seem difficult to neatly categorize into this dichotomy, perhaps because a strategy can be active or latent. When a strategy is conceptualized as stored knowledge of how one *could* modify one’s cognition, it seems to be at a remove from ongoing cognitive activity (and thereby better thought of as metacognition); however, when the same strategy is being implemented in one’s own cognition, it seems odd to separate it from the cognition which enacts the strategy by assigning it a higher-level status.

<sup>28</sup> Perhaps in doing so, she is conveying a subtle pushback against the tendency for educators to latch onto novel findings, so as to portray one’s pedagogical approach as “cutting edge” (and other theories as comparatively stodgy). Brown does suggest that at least some early work in developmental psychology leading to was conducted by researchers “attracted by the lure of a new-sounding concept” (p. 105), which to my ears sounds like an admission that novelty holds sway in this area. Alternatively, I might be projecting this attitude of mine onto Brown.

<sup>29</sup> . . .to steal William James’ pithy subtitle about pragmatism (1907).

Ultimately, though, Brown ends up defending the use of a new term to bind together existing concepts — strategic reading activities, the hidden target of problem-solving training, “study skills” — by arguing that it marks a meaningful change of emphasis in educational psychologists’ understanding of what is most important for optimizing student outcomes. That is, bundling all of these things together is worthwhile insofar as it serves to draw educators’ attention to the collective significance of these phenomena in accounting for students’ academic performance.

#### **§1.4 | Summary & implications**

In this chapter, I’ve assembled a case that “metacognition” was introduced to lash together a confederation of diverse phenomena, resulting in a construct that is fundamentally ambiguous in reference and indeterminate in scope. I’ll draw out two implications of this account.

Firstly, when educators invoke “metacognition” in discussions about how one ought to teach, this language actually obscures more than it clarifies. It’s hard to tell whether they want students to diagram their thinking, to make a judgment about how they’re faring on objective performance measures, or to just think really carefully about course content — to name just a few options of what they might be referring to. It’s possible for individual speakers to vacillate on how they’re using “metacognition” at different times and in different contexts, but also — more damningly — makes it possible for hearers to misread what speakers are recommending when they advocate for “metacognition” in classroom practice.

What ought we to do with the knowledge that “metacognition” means too many distinct things to enable clear communication among educators, and clarity of purpose for individual educators? This has been well-known for decades; stating the problem and suggesting particular ways of restricting the referential scope of the term hasn’t done anything to change the situation.

One step we can take is to regard the ambiguity of “metacognition” as grounds to weaken credence in any generalizations about the pedagogical value of the construct. How can we be confident in the truth of a claim like *metacognition boosts learning* when “metacognition” captures everything from doing philosophy of mind to taking breaks while studying? More importantly, how can we be sure that *metacognition boosts learning* justifies the implementation of any particular pedagogical method that aims to promote metacognition in the classroom or improve students’ metacognitive abilities?

Let me explain what I mean with an analogy. Consider the generalization that *physical exercise is good for our health*. We’d probably be willing to assent to this as something which is true, and which can inform decision making about what we ought to do if we care about being healthy. But nevertheless we are likely to recognize that “physical exercise” is a vast category, encompassing everything from *walking* to *weightlifting*, *Crossfit* to *Zumba*. These may all qualify as “good for our health” broadly speaking, but whether any particular instantiation of physical exercise is “good for our health” cannot be determined from this generalization alone. In order to know whether we are warranted in recommending that someone act upon the message that *physical exercise is good for our health*, we’d have to get clear on what kind of exercise they might be attempting, in what way, under what circumstances. We would not be warranted in recommending *any* kind of exercise to someone who is supposed to be on bedrest, or recommending that someone try weightlifting half-heartedly with bad form – either of these would actually be counterproductive to health in those instances. In short, the plausibility of the generalization doesn’t mean that we’re automatically justified in carrying out specific instantiations of the implied recommendation that we ought to all be engaged in some form of physical exercise.

Likewise, no matter how consistently we encounter the metacognitivist message that

*metacognition boosts learning* or *is beneficial to learning* or is a “principle [sic] asset in learning” (Concepción 2004, p. 356), the semantic ambiguity of “metacognition” undermines the normative force of the dictum. Even if we were to make a determination that enough investigations of individual constituents of “metacognition” have revealed undeniable benefits to learning that the generalization is worth asserting, this is *not* necessarily a reason to incorporate any particular manifestation of “metacognition” into one’s pedagogy.

Long story short, blanket statements about the pedagogical value of metacognition shouldn’t be treated as a mandate for educators to implement anything that has been or can be called “metacognition.” It doesn’t behoove us to treat these generalizations as a blank check to be cashed out with any translation of “metacognition” in any instructional setting. If we care about determining whether particular instructional practices are warranted or not, we need to examine them in much more granularity than just checking whether or not they can be counted as “metacognitive.”

Another way in which this discussion undermines the credibility of the pedagogical claim that *metacognition boosts learning* is by exposing this statement as a tautology, instead of an empirical finding. In pedagogical discourse, “metacognition” is typically presented as an emanation of research on “the science of learning” (Issitt 2007). When the construct’s scientific origins are brandished within metacognitivist rhetoric, the oft-encountered claim that *metacognition boosts learning* seems to convey a substantive discovery: that engaging in metacognition has a causal effect in producing superior learning outcomes. It may also produce a deleterious misconception that the pedagogical efficacy of *everything* referred to as “metacognition” has been thoroughly validated by education researchers.

But I’m alleging that “metacognition,” as used by educators, does not capture a proscribed set of mental operations demystified by scientific inquiry, so much as it describes a

collection of factors assigned explanatory value in accounting for interpersonal variance in academic aptitude. What binds together instances of “metacognition” under one umbrella is a supposition that these processes collectively conduce to better learning outcomes. As a result, the claim that *metacognition boosts learning* is circular: it restates the qualifications for “metacognition,” rather than articulating the revelation of a causal relationship between metacognition and desired educational outcomes.

If this is correct, then to say that *metacognition boosts learning* is not, as many advocates would have it, necessarily a statement of scientific fact encapsulating a critical mass of concordant empirical findings. Rather, it can be understood as an expression of *choice* about how to understand what learning involves, consistent with an effort to overhaul our understanding of “intelligence.” That is, testifying that *metacognition boosts learning* is to endorse a redefinition of the underlying mental activities and capacities we attribute to someone who learns successfully.

Chapter 2 will carry forward this theme that “metacognition” may be not quite as it seems at first glance or from the standpoint of conventional wisdom. I’ll catalog some *a priori* arguments why we should question the idea of a capacity to know and control our own minds. Then I’ll recount some empirically-informed, revisionist takes on metacognition, according to which our we can hardly help but believe that that capacity for cognitive self-knowledge and self-control is much more expansive than it actually is.



## 2. COMPLICATIONS AND CONTROVERSIES FROM PHILOSOPHY OF MIND AND COGNITIVE SCIENCE

In this chapter, I'll dispute a presupposition underlying metacognitivism: that it's possible for any of us to truly know our own cognitive operations and/or to gain control over them (let alone that we can enact this in others through pedagogy). After recounting some historical objections to the theoretical precursors of "metacognition" (§2.1), I'll elaborate some more contemporary challenges raised against the presumption that we can gain knowledge of and control over our own minds. These include views according to which that transparent access to our mental states through introspection is a highly convincing illusion (§2.2), recently elaborated psychological and philosophical accounts of metacognition (§2.3), and an argument against singling out higher-order cognition as "somehow special" (§2.4; Kornblith 2012, p. 155).

Collectively, the perspectives I'll relay in this chapter back the conclusion that we are disposed to develop inflated confidence in our capacity to a) acquire insight into our own minds and b) modify them at will. The metacognitivist conviction that we can deploy pedagogical interventions to enhance *students'* knowledge of and control over their own minds may, I submit, ultimately turn out to be an undue extrapolation from an unwarranted belief in our own capacity to achieve mental self-mastery.

### **§2.1 | Some historical objections to metacognition**

Some metacognition researchers like to emphasize that "metacognition" isn't so much a novel 20th-century discovery as it is a new name for a concept that philosophers and psychologists have been grappling with for centuries (Brown 1987, Nelson 1996, Veenman et al. 2006, Dunlosky & Metcalfe 2009, Tarricone 2014). As such, plenty of theoretical and empirical

objections to the concept have emerged over the years, most of which are not acknowledged in the laudatory discourse around metacognition in pedagogy. Here I'll relay some historical challenges to the notion that we are capable of attaining accurate insight into our own minds and developing strategic control over our own cognitive operations.

The presumption that a thinker can subject their own thoughts to examination and/or autonomously regulate their own thinking evokes a collection of ongoing debates, centered on an unresolved question: *How exactly can a mind consider and change itself?* Philosophical dialogue in this area has addressed how and to what extent we can truly know ourselves as cognitive agents, where the agency required for self-regulation would come from, and how minds can subject their own activity to normative standards.

One crucial question in this arena is: *Can we trust our own impressions about how our own minds work?* For example, discussions around free will and moral responsibility force us to wrestle with the possibility that we could be deceived about fundamental impressions of ourselves — e.g., that feeling as though our decisions are the sole causes of our actions is *not* a guarantee that we truly are autonomous agents, somehow conjuring causality out of thin air. If we can be mistaken about the degree of control we have over our own behavior, couldn't we be similarly misled about the extent to which we can direct our own mental operations? Might the same mechanisms that provide a convincing impression of unfettered control over our physical actions also deliver up an inflated sense of control over our mental activity?

We could also be mistaken about the internal coherence of our minds, attributing properties to we would expect of a well-designed mechanical system. Since at least as early as Hume's allegation that "the self" is nothing more than a bundle of disparate thoughts, feelings, perceptions and so on (1739–40/1978: 252), we've been contending with the possibility that our minds are actually quite fragmentary, despite the unity & fluidity of first-person conscious

experience. Hume also suggests that we standardly misconstrue our own mental activity – for example, ascribing to ourselves the ability to perceive causation where we are instead *inferring* causation from “constant conjunction” of events (1748/2007). This argument advances the view that some of our most basic cognitive operations (such as inference-making from accumulated sensory data) are opaque to us in everyday experience.

In a similar vein, an evolutionary perspective on brain development suggests that although most of us have a sense of oneself as a singular seat of experience and font of agency, we’re more like a haphazard assembly of various systems that just so happen to deliver up a powerful illusion of seamless coordination, most of the time (Dennett 1992). From this outlook, it seems a minor miracle that some subcomponent of the mind could attain the ability to oversee and impose strategic order over the rest.

A closely-related philosophical objection to the possibility of a mind analyzing or exerting control over itself is *Comte’s Paradox*, the problem of “how one organ [i.e., a unified mind] could both do the observing and be the thing observed” (Nelson 1996, p. 104). Comte was sure that a thinker “cannot divide himself into two, of whom one reasons whilst the other observes him reason” (quoted in James 1890, p. 188). While it may seem as though an easy solution to the paradox is to deny that the mind cannot be subdivided (a premise which admittedly seems naïve in an era in which dual-process models of the mind<sup>1</sup> have become the new orthodoxy), asserting that the mind has distinct parts still leaves open several questions, including how the mind could have evolved the capacity to take its own activity as an object of scrutiny and control, and where to locate the limits on this ability.

---

<sup>1</sup> a.k.a. “dual systems theories”; see, for starters, Stanovich & West (2000), Evans (2003), Kahneman (2003), Evans & Frankish (2009), Evans & Stanovich (2013).

Endowing some portion of the mind with the power to oversee and manage the rest invokes the *homunculus problem*. First raised by behaviorist B.F. Skinner (1964, 1971) and popularized by Daniel Dennett (1978), this is a critique of explanations of psychological phenomena that pass the buck for anything that remains mysterious about the inner workings of the mind by attributing said mystery to the activity of a subcomponent of the mind – an inner *homunculus* who “provide[s] an explanation which will not be explained in turn” (1971, p. 14). In fact, metacognitivist rhetoric in education is often unironically accompanied by homuncular imagery (*Figure 1*).<sup>2</sup>



**FIGURE 1: EXAMPLE OF HOMUNCULAR IMAGERY IN VISUALIZATION OF METACOGNITION.** Cover illustration for the Education Endowment Foundation’s Guidance Report on Metacognition and Self-Regulated Learning (Quigley et al. 2018).

---

<sup>2</sup> Some advocates of metacognition in education lean into homuncular models of the mind in order to make the concept legible to children; e.g., Wilson & Conyers (2016) encourage upper-elementary teachers to present students with images of “brain cars” (picture a homunculus seated atop a brain on wheels) so that they may imagine themselves “driving their brain cars” in order to control their own thinking.

The idea that a mind can regulate its own activity in accordance with some normative standard also invokes an argument sometimes referred to as “*Ryle’s Regress*”:

. . . if, for any operation to be intelligently executed, a prior theoretical operation had first to be performed and performed intelligently, it would be a logical impossibility for anyone ever to break into the circle.

According to the [intellectualist] legend, whenever an agent does anything intelligently, his act is preceded and steered by another internal act of considering a regulative proposition appropriate to his practical problem. But what makes him consider the one maxim which is appropriate rather than any of the thousands which are not? Why does the hero not find himself calling to mind a cooking recipe, or a rule of Formal Logic? . . . Intelligently reflecting how to act is, among other things, considering what is pertinent and disregarding what is inappropriate. Must we then say that for the hero's reflections how to act to be intelligent he must first reflect how best to reflect how to act?

. . . To put it quite generally, the absurd assumption made by the intellectualist legend is this, that a performance of any sort inherits all its title to intelligence from some anterior internal operation of planning what to do. . . . The regress is infinite, and this reduces to absurdity the theory that for an operation to be intelligent it must be steered by a prior intellectual operation. (1949, pp. 30-31)

Ryle is responding to the proposal that good cognition is the result of a mind applying regulations to its own operations. He points out that this explanation creates a chain of causal influence, where each mental action involves choosing the correct regulation which ought to be applied to the next. But among those who defend this view, no effort is made to locate the very beginning of this long chain, or account for the *origin* of the ability to intelligently regulate other processes downstream of itself. To Ryle, the inability to explain where this chain of regulatory control begins invalidates the assumption that such a chain of internal mental processes accounts for our ability to think in accordance with intellectual norms. Any appeal to our apparent ability to engage in cognitive self-regulation has the potential to run up against this regress problem.

To summarize, the concept of “metacognition” is fraught insofar as it rests upon first-person impressions of our mental architecture and capacities, which can be deceptive. It attributes executive control and self-optimization functions to our minds in the absence of solid

explanations as to how such a capacity could have arisen organically in systems that are — unlike computers, from which psychologists derived the notion of executive control — evolved rather than designed. Lastly, some critics contend that attributing puzzling human capacities like intelligence, intellectual aptitude, and self-optimization to “metacognition” merely passes the buck by naming a subsystem of the mind that purportedly actualizes these ideals.

## **§2.2 | Illusions of introspection: Infallibility and transparent access to our mental states and processes**

In recent years, we have seen an upsurge in *illusionist* theories, according to which much of what we think we know about our minds through first-person experience is false. The general thrust of these views — each of which typically targets a specific apparent feature of the mind, like free will or the infallibility of knowledge about our own mental states — is that the mind is not as it seems to us on an everyday basis. (Note that not all of the views I am summarizing here describe themselves as illusionist; I am suggesting that it’s appropriate to extend the label of “illusionism” to describe a general attitude towards how the mind works, cutting across specific topics within cognitive science.)

Illusionists typically argue that human minds work in such a way as to generate first-person experiences which make it inevitable that we will each come to embrace a particular misrepresentation of our minds. In other words, our minds construct impressions that predictably dispose us to arrive at false beliefs about how our own minds work. These false beliefs are self-reinforcing, as they get crystallized into a theory of how minds work, which then governs how we make sense of the data we get about our own minds through first-person experience.

For example, Daniel Wegner generalizes that “the mind is a system that produces

*appearances* for its owner” (2002, p. 28). Among these appearances is the “feel[ing] that we cause ourselves to behave”: a highly convincing sensation of *free agency* in selecting between various possible actions; however, “the experience of consciously willing an action is not a direct indication that the conscious thought has caused the action” (2002, p. 2). In other words, the feeling of conscious will is an illusion that reinforces a false belief about ourselves as unmoved movers. Wegner builds upon Daniel Dennett’s suggestion (1987) that our impression that our minds work differently from other objects in the world (e.g., that they defy patterns of mechanistic causation by generating actions without cause) is a result of adopting a distinctive lens – the “intentional stance” – when we perceive minds. According to Wegner, “the illusion of conscious will may be a misapprehension of the mechanistic causal relations underlying our own behavior that comes from looking at ourselves by means of a mental explanatory system” (2002, p. 26).

Another strand of illusionism concerns *our sense of being a coherent self*: a unified subject of experience and agent of bodily action. Following Dennett’s assertion that the self is merely “an abstraction one uses as part of a theoretical apparatus to understand, and predict, and make sense of, the behavior of some very complicated things” (1992),<sup>3</sup> Thomas Metzinger suggests that our belief that we possess “a self” is the consequence of how the selective filter of consciousness constructs an illusion of unity among our phenomenal experiences (2009).

Perhaps most controversially, illusionism about *phenomenal consciousness* denies that conscious experiences possess properties that introspection seems to make undeniably apparent to us. Keith Frankish distinguishes between weak and strong illusionism about phenomenal properties (i.e. *qualia*, something it is like to have an experience): According to the former, “there really are phenomenal properties, though it is an illusion to think they are ineffable,

---

<sup>3</sup> ...which arguably pays homage to Hume’s famous bundle theory of the self (1738/2007, Book I, part IV, sec. 6).

intrinsic, and so on,” while the later asserts that “it is an illusion to think there are phenomenal properties at all” (2016a, p. 4). But proponents of both variants of illusionism about phenomenal consciousness attest that the ongoing operation of introspection leads us to systematically misrepresent what consciousness involves.<sup>4</sup>

This project is long enough without making forays into the muck of free will, the self, and consciousness — so I won’t expand upon these views further. But I *will* go into more detail about particular accounts in the illusionist vein that bear on claims about “metacognition” in education. Specifically, I’ll present findings which cast doubt on common assumptions about first-personal mental transparency: the idea that each of us can gain insight into the hidden contents of our own minds through privileged channels of access.

### **[2.2a] The inscrutability of cognitive processing**

Synthesizing strands of an accumulating consensus that “it is the *result* of thinking, not the process of thinking, that appears spontaneously in consciousness” (Miller, 1962, p. 56), Nisbett & Wilson (1977) contend that we have little to no direct introspective access to “the cognitive *processes* underlying our choices, evaluations, judgments, and behavior” (p. 231). They argue that when people are prompted to verbally report upon the cognitive processes that led to their visible behavior, the responses they offer are not disclosures of first-personal knowledge of how they got to some particular result; rather, their responses reflect “implicit, *a priori* theories about the causal connection between stimulus and response” — i.e., internalized beliefs about how mental causality works (p. 233).

Particularly relevant for my purposes are Nisbett & Wilson’s comments on the cognitive processes involved in *problem solving*, since educational psychologists have proposed that

---

<sup>4</sup> See also Blackmore (2002, 2016).



students can metacognitively monitor and exert control over problem-solving activities during learning. These authors suggest that the cognitive operations leading to successful solutions are “almost completely hidden from conscious view” (p. 240). When people are asked to explain how they arrived at a solution, they’ll either present the solution as a sudden, all-at-once realization (“It just dawned on me”) or they’ll spin up a plausible explanation *post hoc*. In a classic experiment by Maier (1931), the experimenter prompted subjects to come up with a particular solution to a problem by setting a cord in motion; instead of recognizing that the experimenter’s motion had inspired a solution wherein the cord was swung like a pendulum, a participant famously answered: “Having exhausted everything else, the next thing was to swing it. I thought of the situation of swinging across a river. I had imagery of monkeys swinging from trees. This imagery appeared simultaneously with the solution. The idea appeared complete” (Nisbett & Wilson 1977, p. 241).

This is an example of what psychologists have come to call *confabulation*: the invention of an explanatory story in situations where one cannot access or recall a memory of what actually took place, or “the creation of false memories in the absence of intentions of deception” (Brown et al. 2017). As Nisbett & Wilson put it, “we sometimes tell more than we can know. More formally, people sometimes make assertions about mental events to which they may have no access and these assertions may bear little resemblance to the actual events” (1977, p. 247). Their contention that ordinary people — not just those with clinical memory deficits — confabulate in order to explain their cognitive outputs created a bit of a methodological kerfuffle among psychologists in that era, since it calls into question the validity of treating verbal self-reports from subjects as data about the inner workings of those subjects’ minds.<sup>5</sup>

Nisbett & Wilson’s view qualifies as a form of illusionism (in my eyes) because these

---

<sup>5</sup> See Ch. 1, footnote 9.

authors emphasize the divergence between what introspection *seems to be* from the first person perspective, and what it actually is from the mechanistic standpoint of cognitive psychology. These authors emphasize that “it is often the case that we *feel as though* we have direct access to cognitive processes” (p. 255), and they endeavor to explain why we are susceptible to that illusion of direct access, when in reality we are making “judgments of plausibility” linking a likely cause to our observed behavior. Their account of verbal reports about our own cognitive processing is as follows:

When subjects were asked about their cognitive processes, therefore, they did something that may have felt like introspection but which in fact may have been only a simple judgment of the extent to which input was a representative or plausible cause of output. It seems likely, in fact, that the subjects in the present studies, and ordinary people in their daily lives, do not even attempt to interrogate their memories about their cognitive processes when they are asked questions about them. Rather, they may resort in the first instance to a pool of culturally supplied explanations for behavior of the sort in question or, failing in that, begin a search through a network of connotative relations until they find an explanation that may be adduced as psychologically implying the behavior. (1977, p. 249)

So, in the absence of direct knowledge about what cognitive processes led to a particular result, people offer answers that correspond with socially-acquired theories of how or why they did what they did, or make an inference to the best explanation available.

As for why people *think* they are accurately reporting on their cognitive processes when they’re actually confabulating, Nisbett & Wilson hypothesize four interrelated factors. First, individuals are *more* accurate in explaining the results of their cognitive processes than third-party observers are – not because of direct access to these processes, but rather because they have access to a store of “private knowledge” about ourselves that leads to better-informed inferences as to the causes of our own behavior, compared to the causes of others’ behavior (p. 255). Visible instances of this asymmetry, “when the individual is correct or at least not provably wrong, and the observer is manifestly wrong . . . sustain the individual's sense of

privileged access to process,” even though the actual cause of this asymmetry has to do with privileged access to relevant mental contents (i.e., episodic memories from one’s personal history; p. 256).

Another explanation for the appearance of privileged access to our cognitive processes is that “we are often capable of describing intermediate results of a series of mental operations in such a way as to promote the feeling that we are describing the operations themselves” (p. 255). Nisbett & Wilson suggest that the subject in Maier’s experiment who reported “imagery of monkeys swinging from trees” may very well have experienced this imagery, but that this does not describe the *process* this subject used to complete the task. In other words, Nisbett & Wilson doubt that consciously experiencing this imagery would be causally efficacious in enabling the inference that swinging the cord like a pendulum is necessary to solve this puzzle; they submit it’s more likely that the subject, having made this inference unconsciously, happened to trigger a conscious visualization concordant with the solution.

A third explanation is that misconceptions about our own introspection persist uncorrected due to “a lack of feedback” (p. 256). While first-person experience seems to constantly reconfirm the impression of direct access to our cognitive processing, “disconfirmation of hypotheses about the workings of our minds is hard to come by” unless you’re looking for it, which most people aren’t (ibid.). Nisbett & Wilson’s point here is that the illusion of introspective accuracy into our cognitive processes is well-entrenched, which makes it potentially perseverant even in the face of convincing arguments to the contrary. That we know the workings of our own minds and can report on them accurately is something we practically can’t help but believe and will be hard-pressed to abandon.

Fourth and finally, Nisbett & Wilson note that motivational factors “may serve to sustain our belief in direct introspective awareness”:

. . . It is naturally preferable, from the standpoint of prediction and subjective feelings of control, to believe that we have such access. It is frightening to believe that one has no more certain knowledge of the workings of one's own mind than would an outsider with intimate knowledge of one's history and of the stimuli present at the time the cognitive process occurred. (p. 257)

In other words, we *want to believe* that we know our own minds — that we can understand ourselves in ways that other people can't. On this explanation, we believe in direct access to our cognitive processing for much the same reason most people cannot abandon belief in their own free will in the face of philosophical arguments to the contrary: we'd rather maintain the illusion that our self-knowledge is infallible than to accept a reality in which we are often mysteries to ourselves.

### **[2.2b] Introspective inaccuracy**

Recent years have seen an upswell of a more comprehensive introspection-critical position, on which cognitive processing is *not* the only facet of our minds about which introspection may mislead us. Consider this proposal by psychologist Nick Chater:

The very idea of 'looking' into our own minds embodies the mistake: we talk as if we have a faculty of introspection, to scrutinize the contents of our inner world, just as we have faculties of perception, to inform us about the external world. But introspection is a process not of *perception* but of *invention*: the real-time generation of interpretations and explanations to make sense of our own words and actions. The inner world is a mirage. (Chater 2018, p. 5)

Chater suggests that we are mistaken in assigning *hidden depths* to our minds. He adopts a highly eliminativist picture of mental contents, on which stable, seemingly causally-efficacious features of our minds — “beliefs, desires, hopes and fears,” among other standing attitudes — are merely fictions we turn to in order to tell ourselves cohesive stories, connecting our present actions and experiences to our past.

Chater's view is an extreme one, on which introspection yields us an illusion that we are unearthing buried mental states, but these buried states simply *do not exist*.<sup>6</sup> He suggests that whenever it seems to us that we've brought a previously unconscious thought, belief, or so on to conscious awareness via introspection, we are actually improvisationally generating a conscious thought, the subliminal possession of which in our past could explain our current feelings, behavior, and so on.

More common forms of revisionism about introspection will claim that *some*, but not all forms of introspection involve inventing claims about mental processes to which we have no access. On these views, our minds *do* possess hidden depths (prosaically, unconscious states and processes), but we are mistaken in believing that we can gain *accurate* awareness of what's going on at these deeper levels of the mind. To this end, a small but passionate cadre of philosophers and psychologists has emphasized how *bad* we are at introspection.<sup>7</sup> For example, Eric Schwitzgebel offers this spirited expression of the fundamental unreliability of our introspection, in order to combat the view that the issuances of introspection can provide a justificatory foundation for knowledge:

Most people are poor introspectors of their own ongoing conscious experience. We fail not just in assessing the *causes* of our mental states or the processes underwriting them; and not just in our judgments about nonphenomenal mental states like traits, motives, and skills; and not only when we are distracted, or passionate, or inattentive, or self-deceived, or pathologically deluded, or when we're reflecting about minor matters, or about the past, or only for a moment, or where fine discrimination is required. We are both ignorant and prone to error. There are major lacunae in our self-knowledge that are not easily filled in, and we make gross, enduring mistakes about even the most basic features of our currently ongoing conscious experience (or 'phenomenology'), even in favorable circumstances of careful reflection, with distressing regularity. (2008, p. 247)

---

<sup>6</sup> Chater's view about stable mental contents is thereby comparable to strong illusionism about phenomenal consciousness: it fully denies the existence of that which introspection appears to deliver up to us, as opposed to suggesting that introspection misleads us about the nature of that which introspection brings to our attention.

<sup>7</sup> Titchener (1899), Dennett (1991), O'Regan (1992), Rensink et al. (2000), Blackmore (2002), Schwitzgebel (2008).

Schwitzgebel contends that we're pervasively wrong about what it's like to see, to think, and so on: that our subjective impressions about all of these things are not at all secure or infallible as many philosophers have historically contended.

One motivation for Schwitzgebel's view is our failure to arrive at any kind of consensus about cognitive phenomenology. Here's a question you'd expect we'd be able to answer, given how ubiquitous thought is to our lives, especially as philosophers: *What is it like to have a thought?*<sup>8</sup> As he sees it, introspection ought to make this matter crystal clear for us, but instead, impressions are fuzzy and various. He suggests that we don't notice how shaky we are on this very basic and ever-present feature of being conscious adult humans, because these errors go unnoticed and uncorrected by our peers:

Because no one ever scolds us for getting it wrong about our experience and we never see decisive evidence of error, we become cavalier. This lack of corrective feedback encourages a hypertrophy of confidence. Who doesn't enjoy being the sole expert in the room whose word has unchallengeable weight? In such situations, we tend to take up the mantle of authority, exude a blustery confidence – and genuinely feel that confidence (what professor doesn't know this feeling?) until we imagine possibly being proven wrong later by another authority or by unfolding events. About our own stream of experience, however, there appears to be no such humbling danger. (ibid., p. 260)

This explanation dovetails with Nisbett & Wilson's comment that people confuse their confabulations for genuine self-insight because, being fundamentally fond of the feeling of being right, we'll persist in unwarranted confidence about knowledge of our own mind in the absence of testimony to the contrary.

---

<sup>8</sup> Schwitzgebel unpacks the query as follows:

Does the phenomenology of thinking consist entirely of imagery experiences of this sort, perhaps accompanied by feelings (emotions?) such as discomfort, familiarity, confidence? Or does it go beyond such images and feelings? Is there some distinctive phenomenology specifically of thought, additional to, or conjoined with, the images, perhaps even capable of transpiring without them? (ibid., p. 257)

### **[2.2c] Reconceiving introspection as motivated self-interpretation**

As Chater's comments about the tendency to treat introspection as a facet of perception attest, a traditional view construes introspection as a type of "inner sense." In David Rosenthal's words, the *inner sense theory* of introspection posits an analogy between introspection and our outward- and bodily-oriented perceptual capacities: "Just as exteroceptive perception monitors the external environment and proprioception monitors the positions and movements of our limbs, so introspection performs a parallel monitoring function for the contents of our minds" (2000c, p. 203). On the alternative view Rosenthal defends, however, "introspection is not the perceiving of our mental states, but the having of thoughts about them" (p. 206): in other words, introspection consists of the formation of conscious higher-order representations of our own lower-order states, such that one becomes conscious of the lower-order states in a "an attentive, deliberate, focused way" as opposed to being only fleetingly aware of them (pp. 207-209).

Importantly (for my purposes), Rosenthal contends that these higher-order representations reflect motivation-informed inferential interpretations of our lower-order states (including cognitive states), but the process by which this occurs is concealed from our view:

People interpret themselves in the light of their situation and past experience, and some of these self-interpretations have to do with what mental states they are in. As long as one remains unaware of whatever inference and motivation leads to these self-interpretations about one's mental states, the self-interpretations will seem, from a first-person point of view, to be spontaneous and unmediated. They will seem to arise from just asking oneself what mental states one is in, from a deliberate decision to focus on the states in question by casting one's mental eye inward. But it is likely that such introspective awareness results in substantial measure from desires to see ourselves in a certain light. Introspection is often, if not always, a process of conscious self-interpretation. (Rosenthal 2000c, p. 224)

Rosenthal's view of introspection, then, qualifies as a form of illusionism, insofar as what introspection *seems* to be from the first-person perspective is not revelatory of what's actually

happening in one's mind. Introspection seems to be a means by which we gain sudden insight into which mental states we are experiencing, but in actuality it involves arriving at a judgment about what mental states we are in, and this judgment is informed (and in some cases, distorted) by our desires.

Rosenthal's backing for this view leans on evidence for confabulation about one's mental states, some of which I presented in the previous section.<sup>9</sup> A key takeaway from this body of research is that the subjective impression that one is introspecting does *not* warrant the conclusion that one is in fact accessing and reporting one's own mental states. "[S]ubjects themselves can't tell when they are introspecting and when they are interpreting or confabulating," as Peter Carruthers attests (2009a, p. 127). This leaves open the possibility that *all* instances of purported introspective insight into our own minds are fabricated: as Johansson et al. put it, "confabulation could be seen to be the norm and truthful reporting something that needs to be argued for" (2005: 118).

Another insight from Rosenthal's summary is that the content of confabulated accounts of our mental states can be explained by appeal to normative influences on how we understand our own minds and explain our minds to others. We face substantial pressures to describe our mental operations in ways that satisfy a) our own desires and/or self-image (Bem 1972), b) a drive for coherence and consistency in our self-understanding (Chater 2018), and/or c) cultural narratives about how our minds ought to work (Nisbett and Wilson 1977). This point actually reinforces the worry stated above by Johansson et al., that we have no reason to believe that *any* instances of purported introspection involve direct access to and veridical reporting of our mental states. If these normative pressures clearly force us to provide false reports of our own

---

<sup>9</sup> Particularly Nisbett & Wilson (1977), P. White (1988), Wilson, Hodges & LaFleur (1995), and Bar-Anan et al. (2010). See also Gazzaniga (1995).



mental states under *some* circumstances, why wouldn't *all* reports of workings of our own minds be subject to the same distorting normative pressures?

### **(2.2c1) Rejecting transparent access to cognitive states**

Peter Carruthers (2009a, 2011) incorporates the preceding observations about the fallibility and interpretive nature of introspection into an unorthodox account of what it means to attain self-knowledge. He draws out an assumption of "transparent access" to our own mental states (that is, a route to immediate knowledge about our own mental states, not involving any kind of inference, judgment, or interpretation; 2011, p. 9) as the underpinning of the default view that we discover our own mental states via introspection. As Rosenthal emphasized, some of those who take transparent access for granted further posit that human minds contain a special faculty devoted to monitoring our own mental operations (an "inner sense"); Carruthers adds that those who defend this picture assign this faculty a *raison d'être* of *optimizing our own cognition by exerting direct control over these processes*.<sup>10</sup>

Carruthers argues that these views are dominant because they square with our "common-sense intuitions of mental transparency" (p. 12). It seems to the vast majority of us, consistently and cross-culturally (pp. 25-32), that our mental states are self-presenting (i.e., that "if one is undergoing a given mental state, then one knows or is in a position to know that one is; . . . if one believes that one is *not* undergoing a given mental state, then one isn't"; p. 13), and that knowledge that one is or isn't in a mental state is infallible. But Carruthers argues that these widespread intuitions can be explained as the outputs of an inferential system through which we arrive at judgments about our own minds.

He proposes that knowledge about our own mental states is generated (not discovered)

---

<sup>10</sup> Carruthers' arguments against this aspect of inner sense views of introspection are addressed in §2.3c and §2.3e, after elaborating a bit more on how his unorthodox views on introspection feed into a revisionist account of metacognition.

via “mindreading” – the very same mechanism by which we make judgments about *others’* mental states, consisting of abductive interpretation of whatever sensory data is available to us. Similarly to how Nisbett & Wilson posited that our judgments about our own mental states and processes are driven by *a priori* theories about how minds work, Carruthers suggests that the impression that mental states are self-presenting arises from “tacit processing rules” embedded in the mindreading faculty (p. 34).

To bolster the claim that these rules are responsible for our impressions of mental transparency, Carruthers presents evidence (largely drawn from Gazzaniga 1995) that patients who have undergone a commissurotomy – surgical severing of the corpus callosum, which connects the brain’s two hemispheres – have the subjective impression of immediate, transparent access to their mental states, even when such access is *physically impossible* (because the subject is being asked to report on mental states operative in the opposite hemisphere from the one which is capable of speech production). He observes that split-brain patients “can report plainly-confabulated explanations with all of the same sense of obviousness and immediacy as normal people” (p. 43); their lack of transparent access to states in the opposite hemisphere doesn’t hinder them at all in offering up claims about their mental states, or produce the sensation that any necessary information is *missing*. From this Carruthers concludes that normal subjects probably are just as incapable of detecting “whether they have transparent access, on the one hand, or whether they are interpreting or confabulating, on the other” (ibid.).

In short, Carruthers asserts that mental transparency is a highly convincing, near-universally experienced *illusion*. Despite what our phenomenology offers up, the counterintuitive reality is that our impressions about our own mental states are inferentially constructed from whatever bits and pieces of sensory data we have on hand. Like Nisbett & Wilson, his explanation for why knowledge about our own minds is often more accurate than

knowledge of other agents' minds is that the data set from which we can make inferences about our own mental states is much more substantial than that available from perceptual observations of others' behavior. To make inferences about our own minds, we can draw not only from evidence of our behavior, but also from intramental sources of sensory data, like inner speech and imagistic representations. (It is assumed here that inner speech and mental imagery are sensible representations of cognitive outputs – not cognitive states in themselves.) So the difference between our knowledge of ourselves and knowledge of others is a matter of *degree*, rather than a distinction between kinds.<sup>11</sup>

#### **[2.2d] Implications for metacognition in education**

I've summarized a smattering of views according to which subjective experience misleads us about some fundamental features of our minds. On each of these illusionist views, there are serious discrepancies between what seems to be the case (namely, transparent access to our mental states (including cognitive processing) and introspective infallibility) and what empirical research has revealed about our powers to know and control our own minds.

I'm not asserting that illusionist views are necessarily true (though I am sympathetic to these views, as you likely divined). What I'm highlighting is that there is a countercurrent among philosophers of cognitive science according to which the default, intuitive understandings of how our minds work are unreliable, and our minds are not as they seem. Even if you don't subscribe to illusionism, the considerations presented by proponents of these views suggest that some humility is in order when it comes to our impressions of our own minds.

A recurring theme throughout this discussion has been the idea that we aren't indifferent

---

<sup>11</sup> I credit Arango-Muñoz (2011, p. 73, footnote 3) for making this clear.

to competing characterizations of our mental architecture and capacities. Quite the contrary: we have strong (and, as far as I can tell, interpersonally-consistent) preferences for views on which we *can* know and exercise control over our own minds. Given these biases, it would behoove us to recognize that the fact that “the idea of metacognition maps to our experience of our selves” doesn’t cement the concept (and the accompanying powers of self-awareness and self-mastery it implies) as irrefutable (Issitt 2007, p. 389). Rather, that the idea of being able to monitor and control our mental operations resembles so closely what we would *like* to be true about ourselves should activate our skepticism.

In other words, perhaps “metacognition” has caught on as a pedagogical concept due to a powerful cocktail of persuasive phenomenal illusions about our own minds *and* desires that just so happen to line up with what these illusions suggest to us. Notions taken for granted by metacognitivists – that insight into our minds is readily achievable, and that we gain control over our minds simply by entering into a reflective mode of thinking – are easy to believe not only because they match up with how things seem to us, but because they promise that cognitive optimization is possible for us as long as we build up the requisite skills. I’ll expand upon this in Chapter 4, in an effort to explain the embrace of metacognitivism in the absence of adequate empirical support.

### **§2.3 | Competing psychological models of metacognition**

Now I’ll turn my attention to various proposals as to how to understand metacognition *qua* target of inquiry in cognitive science. These views happen to converge in identifying metacognition with *higher-order cognition* – but diverge as to what “higher-order” entails.<sup>12</sup> One

---

<sup>12</sup> To complicate matters further, “higher-order” has additional senses beyond these outside of the technical language of cognitive science. See Appendix: Disambiguating “metacognition.”

primary difference between them has to do with whether this amounts to *representing* a lower-order state or process as an intentional object (i.e., that which the representation is about), or whether cognition can qualify as higher-order *without representing* other states, in virtue of serving the function of enabling informational flow to and from other cognitive states or processes (Browne 2004, p. 636). These views also vary as to whether metacognition is canonically *conscious* or *unconscious* (or – as in one conciliatory model – *both*, depending on which of “two levels” one is referring to). One final dimension on which they vary is in the *purpose* ascribed to higher-order cognition.

My purpose in outlining these models is to bring attention to incompatibilities between these proposals and assumptions about metacognition that are baked into metacognitive training methods. Despite their differences, the leading theories dominating debates in cognitive science – self-evaluative and self-attributive views – characterize “metacognition” in ways which contradict suppositions of a) transparent access to our mental states and b) direct conscious control of our own cognition.

### **[2.3a] Functionalist higher-order cognition**

Nelson & Narens (1990) proposed a basic theoretical framework on which metacognition consists of higher-order cognitive processes wherein information is transferred between two levels of cognitive operations: an “object-level” and a “meta-level” (pp. 125-7; see *Figure 2*). The *meta-level* contains a “dynamic model” or “mental simulation” of how a cognitive task should be executed (p. 126); the basic cognitive processes necessary to execute the task take place at the *object-level*. I’ll replicate Browne’s (2004) terminology in deeming this a *functionalist* take on “higher-order cognition,” since the processes that qualify – monitoring and control – do so on the basis of the roles they serve in optimizing the unfolding of cognitive processes.

According to Nelson & Narens, optimized performance on a particular learning task (e.g., memorizing a list of vocabulary terms for an upcoming quiz) is enabled by two complementary facets of metacognition, distinguished by the direction of information transfer between the object- and meta-level. Metacognitive *monitoring* describes when cognitive processes occurring at the object-level provide inputs to the meta-level, to inform and “update the model [of performance on the current task] based on what is happening at the object level” (Dunlosky & Metcalfe 2009, p. 5). Metacognitive *control* describes when the model of the task stored at the meta-level is used to modify cognitive activity occurring at the object-level, causing the initiation, continuation, or termination of specific object-level processes.

Nelson & Narens write that metacognitive monitoring and control will often take place automatically and unconsciously, but one *can* become conscious of monitoring of lower-order cognitive states and processes through introspection (1990, pp. 127-128). Consequently, “the main methodological tool for generating data about metacognitive monitoring consists of the person's subjective reports about his or her introspections” – but following Nisbett & Wilson (1977), these authors urge that introspection should not be treated “as a conduit to the mind but rather as a source of data to be accounted for by postulated internal processes” (ibid.). They do, however, emphasize that the faults of introspection don’t invalidate the proposal that metacognitive monitoring plays an important role in enabling control of lower-order cognition: “A system that monitors itself (even imperfectly) may use its own introspections as input to alter the system's behavior” (ibid., p. 128).

This account describes the phenomenon at a high level of abstraction and doesn’t take a stand on many items of controversy among cognitive scientists (like whether metacognition is conscious and unconscious), so it readily accommodates much of what gets described as “metacognition,” at the cost of granularity. As such, it doesn’t challenge anything that would

influence our outlook on the prospects of metacognitive training – so let’s move on to some more recent and more contentious deviations from this foundational model.

### **[2.3b] Control (self-evaluative) theories**

A *self-evaluative* or *control* theory of metacognition plays up the capacity for a subset of our cognitive apparatus to provide strategic direction for one’s mental agency, selecting some paths or strategies over others in order to optimize the use of one’s current cognitive resources in light of various aims and situational constraints (Proust 2010). In short, this view equates metacognition with action-oriented executive control of the cognitive domain. It characterizes metacognition as a fairly primitive mental capacity shared among humans and some nonhuman animals, by which animals automatically and implicitly regulate their own cognitive processes in accordance with internalized standards for successful task performance.<sup>13</sup>

Control theorists often describe metacognition as utilizing the same kind of mechanisms that are standardly postulated for *motor* control. Whereas a motor system produces an internal model of a plan for bodily action, against which sensory feedback can be compared in order to initiate corrective adjustments in the motor plan, a cognitive system generates “an off-line simulation of the cognitive process in question[,] which permits predicting and adjusting future cognitive performance on a given task” (Arango-Muñoz 2011, p. 74).

For those who see metacognition as fundamentally a mechanism for controlling one’s own cognition, the only aspect of metacognition that belongs to our conscious experience is the occurrence of *epistemic feelings*: instances of cognitive phenomenology such as *feelings of knowing*, *feelings of processing fluency*, or *tip of the tongue states* – all of which convey information about

---

<sup>13</sup> A huge swath of the ongoing debates about metacognition among cognitive scientists – specifically between proponents of the self-evaluative and self-attributive views – concerns whether or not it’s appropriate to ascribe metacognitive ability to non-human animals.

one's cognitive operations in a nonlinguistic and nonrepresentational format. Here is one way of summarizing the role of epistemic feelings:

. . . [M]etacognition is a "criticism system" . . . with a list of heuristics in the form of conditionals prescribing the production of a given epistemic feeling for a given situation: if P-event then Q- feeling, if R-event then S-feeling, and so on. . . . The feeling itself is metacognitive in the sense of being directed towards a mental disposition (knowledge, uncertainty, ignorance, etc.), but the content of the epistemic feeling that determines decision-making is non-conceptual and thus not metarepresentational. The feeling points or is directed to a mental property which is not necessarily within the gaze of the subject. For instance, a feeling of uncertainty points to a lack of knowledge or indicates that something is wrong with our perceptual or mnemonic activity, allowing a subsequent correction or improvement, without the need for an introspective effort by the subject. In a nutshell: . . . *metacognition is the capacity of a being a) to entertain epistemic feelings that nonconceptually point to mental dispositions and b) to be able to exploit such feelings in order to control its cognitive activities.* (Arango-Muñoz 2011, p. 77; summarizing Proust 2009; my italics highlighting the gist of this excerpt)

On Arango-Muñoz's interpretation of the self-evaluative view, a metacognitive agent does *not* become aware of their own cognitive processes, but rather becomes aware of an epistemic feeling. The occurrence of the epistemic feeling instigates (or alternatively, signals the instigation of) some subpersonal cognitive procedure to respond with an appropriate corrective to whatever that feeling signals.

Returning to the suggestion of an analogy between cognitive control and *motor* control, we might think of an epistemic feeling as equivalent to the point at which a mismatch is detected between the internal model of a bodily action and the incoming sensory feedback from the body: the motor system uses this mismatch to inform what kind of adjustments need to be made to the motor program, and enacts these adjustments automatically. An epistemic feeling registers a discrepancy between how a cognitive task would ideally proceed and how it's actually proceeding, and that discrepancy triggers subpersonal adjustments to within one's cognitive processing, to bring one's cognitive operations into line with one's plans or ideals. In short, control theorists propose that "our cognitive behavior is often caused not by reflexive



thinking and second-order thoughts but by emotional states“ (Arango-Muñoz 2011, p. 76; characterizing de Sousa 2008).

### **[2.3c] Metarepresentational (self-attributive) views**

On a metarepresentational view of metacognition, a state is designated “metacognitive” only if it involves *higher-order intentionality*, i.e., forming a “propositional attitude [which] takes another propositional attitude for its content” (Browne 2004, p. 634; utilizing Dennett’s concept of “intentional ascent,” 1987). These views typically ascribe metacognition exclusively to humans, and designate it as a process wherein people deliberately and explicitly form higher-order mental representations of their own lower-order mental states. For example, on this view a canonical instance of metacognition is to have a mental state in which you represent yourself as the possessor of a cognitive state: e.g., “I think that I’m finding this discussion to be tedious.” Metarepresentational theorists propose either that one is necessarily conscious of a metacognitive state (taking higher-order intentionality to be synonymous with consciousness), or that metarepresentations are accessible to consciousness via introspection.

Proponents of metarepresentational theories assert that the primary function of metacognition is to enable *understanding* and *rationalization* of one’s own behavior. When people attribute higher-order thoughts to themselves (e.g., “I believe that I know that P”), they utilize “psychological concepts . . . referring to propositional attitudes such as perceptions, feelings, intentions, knowledge, beliefs and expectations, among others” in order to characterize their own mental states. They deploy these concepts in accordance with an internalized theory of mind, which consists of beliefs about causal relationships between mental states and observed behavior or inner speech (Arango-Muñoz 2011, p. 72).

The metarepresentational account rejects the idea of an “inner sense”: an endogenous capacity to access our own mental states for purposes of cognitive optimization. To the extent that we can be described as engaged in “conscious monitoring” of our own cognitive states and processes, this must be understood to describe the attribution of propositional attitudes to ourselves on the basis of inner speech and visual imagery. Through this process we may become aware of representations we’ve generated, and these representations may subjectively seem like a simple readout of our cognitive states — but this is only because the inferential process by which the representation was generated does *not* feature in our cognitive phenomenology.

I’ve already introduced the basic tenets of a prominent metarepresentational account of metacognition, often described as the *self-attributive* view, defended by Peter Carruthers (in §2.2cI). According to Carruthers, humans lack “native metacognitive capacities . . . to monitor and control their own reasoning” (2011, p. 261). By this, he’s not denying that we can engage in metacognition, but rather is rejecting the standard account of where our metacognitive capacities come from: He denies that we have any mental faculty of the sort posited by inner-sense theorists, which purportedly “enables executive guidance and control of some of our own cognitive processes . . . to facilitate cognitive flexibility and improvement” (ibid.). So what we call “metacognition,” he contends, is not a built-in faculty that *monitors* our own occurrent mental states, but rather the *attribution* of mental states to ourselves, in order to explain our own behavior.

He describes self-attribution of mental states as an evolutionary outgrowth of social cognition. That is, we first acquired an ability to make determinations about other individuals’ mental states on the basis of observable signals in their speech, appearance, and behavior (alternately called *theory of mind* or *mindreading*; later, this capacity was exapted to allow us to

make inferences about *our own* mental states. As Carruthers puts this view: “metacognition is merely the result of us turning our mindreading capacities upon ourselves” (2009, p. 123).

### **[2.3d] Two-level model**

Building upon an early proposal in Koriat (2000), Santiago Arango-Muñoz has proposed that we can reconcile the disagreements among control theorists and metarepresentational theorists by determining that both views are correct. “[T]here is no real disagreement between both theories because they are trying to explain different phenomena,” he writes (2011, p. 75). We can understand each theory as describing one of *two levels* at which metacognition occurs within our cognitive architecture, corresponding to the two systems described by dual process models of the mind.

In the two-level scheme proposed by Arango-Muñoz, control theories describe the *low-level* of metacognition: an automatic, implicit, and heuristic-driven procedure for cognitive self-regulation belonging to *System 1*. In contrast, metarepresentational theories capture *high-level* metacognition: a deliberate, explicit, and theory-driven procedure for rationalizing and justifying our own behavior via psychological explanations concocted by *System 2*. Arango-Muñoz postulates a couple of mechanisms by which the two levels of metacognition could interact, but these remain highly indirect relationships: on this scheme, conscious monitoring and cognitive control certainly do *not* comprise a dyadic system enabling cognitive self-improvement.<sup>14</sup> In other words, there is no mental route by which a conscious representation of

---

<sup>14</sup> Some examples he gives are as follows: 1) “a feeling of error or uncertainty . . . elicited by the low-level may trigger an inferential process of verification carried out by the high-level;” 2) “thinking about the fallibility and unreliability of your memory may trigger a strong feeling of uncertainty that might even interfere with your normal cognitive performance;” and 3) “the second-order belief that you are going to make a mistake activates the high-level as the controlling level and then inhibits low-level responses” (pp. 77-78).

one's own cognitive states enables instant revision or correction of those states.

### **[2.3e] Implications for metacognition in education**

Setting Nelson and Narens' unprovocative model aside, let's zero in on the implications of the more recently-defended takes on the psychological basis for metacognition.

#### **(2.3eI) The disconnect between conscious metacognition and cognitive control**

For starters, these views contradict the idea that conscious awareness of our own mental states is either necessary for or conducive to deliberate control over our cognitive operations.

Let's start with the self-evaluative view. Though "it is typically assumed that metacognition is the one aspect of cognition that is mostly likely to be conscious,"<sup>15</sup> some psychologists defend the counterintuitive view that cognitive control occurs primarily through automatic processes at the subpersonal level (e.g. Reder & Schunn 1996, p. 70; see also Reber 1989). Accordingly, those who identify metacognition with these subpersonal control processes, like Proust, argue that conscious surveillance and reporting of one's cognitive processes simply isn't required in order to engage protocols to adjust and correct one's own cognition.

Interestingly, this is a point on which the self-evaluative and self-attributive views agree. Carruthers argues that the metarepresentations in which we attribute mental states to ourselves have no *direct* causal influence over our subsequent cognitive operations, he asserts; exerting control over our own learning requires modifying our behavior (e.g., adopting new study tactics, creating physical visualizations, and so on) instead of being able to initiate changes to our

---

<sup>15</sup> That *conscious monitoring* of one's own mental processes is required to exert control over those processes is a central theme in the developmental theory advanced by Jean Piaget (recounted in Brown 1987). He theorized that being able to engage in *reflective abstraction* — conscious, verbalizable analysis of one's own cognitive processes — is a necessary milestone in one's progression toward the sophisticated form of cognitive self-regulation exhibited by human adults. On his scheme, reflecting upon one's own thoughts and plans yields conscious thoughts containing information about what one ought to do (strategically speaking), which purportedly directs our thinking and actions accordingly.

cognitive activities intramentally (2011, p. 268). So “metacognitive control” is not really a *cognitive skill*, so much as it is a set of behavioral interventions we impose upon ourselves in order to modify how inputs are delivered to our cognitive system. On his view, we lack the power to orchestrate our own cognitive activity, just as we lack transparent access to our cognitive states.

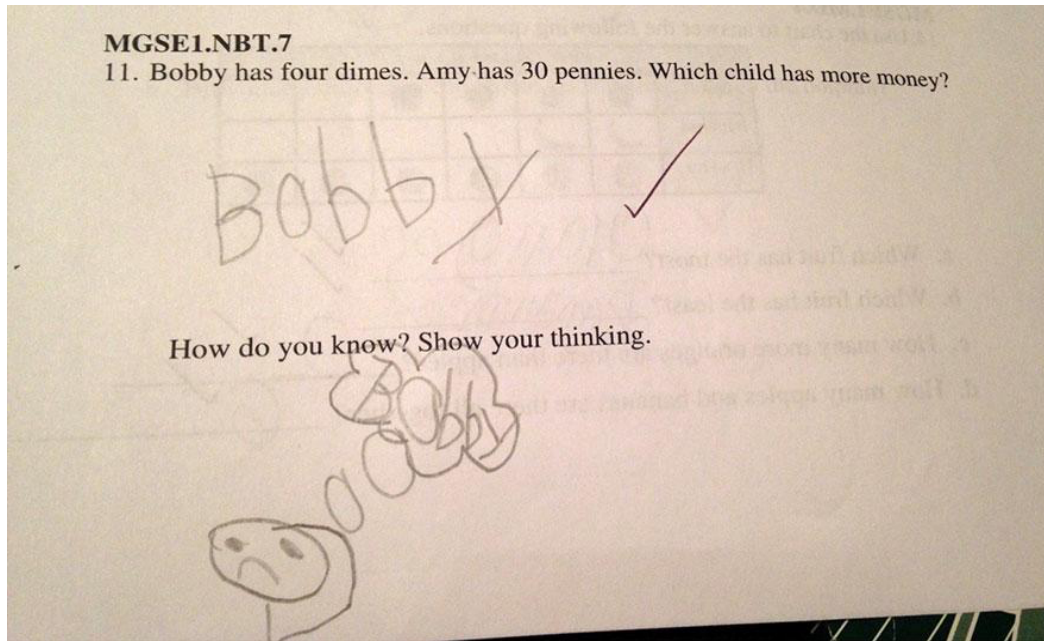
Lastly, the two-level model also happens to deny that the purpose of conscious awareness of one’s own mental states is to direct corrective cognitive activity. Although this account encompasses conscious metarepresentations as a form of metacognition, it does not imbue these states with any influence on subsequent cognitive processing. On this depiction of metacognition, control of one’s cognitive activity occurs either via a) modifications to behavior (such as switching from blocked to interleaved practice), or b) subpersonal mechanisms signaled by epistemic feelings.

To summarize, psychological models of metacognition converge in characterizing conscious representations of one’s own thoughts as *epiphenomenal*, in that they feature in subjective experience but aren’t actually used to change or adjust ongoing cognitive activity. If conscious representations of one’s own mental state are epiphenomenal with respect to cognitive self-regulation, this paints metacognitive training methods centered on conscious representations of one’s own mental states – namely, those involving reflection and self-questioning – in an inauspicious light. The stated purpose of these activities is standardly to allow students to become “self-regulating learners” – but if there’s no necessary linkage between verbalizable expressions of one’s mental states and the ability to enact corrective changes to one’s cognitive activity, then reflection is a lot of talk and no action on the self-regulatory front.

### **(2.3eII) Confabulation in the classroom**

On top of denying that conscious representations of our mental states have causal influence over our cognitive operations, the metarepresentational account emphasizes that because these representations arise through inferential processes, they can be *confabulated*. As a result, we should be concerned that invitations to reflect on one's cognitive activity may elicit confabulatory responses from students, influenced by how students want to portray their own minds, along with their desire to present themselves in ways that align with instructors' expressed ideals regarding *how* students are supposed to think.

More generally, the metarepresentational account supports a cynical picture on which we're regularly deceiving ourselves into believing claims about the contents of our own minds that we are, unbeknownst to ourselves, assembling on the fly. In fact, it reminds me of a meme that circulated a couple of years ago (*Figure 2*). It's a photo of a child's math worksheet; after the student is asked to solve a simple computational problem, they're asked to explain how they figured out the correct answer. Presumably the teacher is looking for the student to write out the reasoning process they used to solve the problem, but what the student cheekily delivers is an illustration of themselves thinking the correct answer.



**FIGURE 2: A STUDENT'S MISINTERPRETATION OF THE INSTRUCTION TO "SHOW YOUR THINKING."** Photograph of a child's math worksheet, posted to Reddit by their parent (irishchck14 2015).

For me, this is an illustration of the mismatch between what the typical idea of what metacognition involves and what it might *actually* be, if Carruthers' metarepresentational view is correct. In asking the student to "show [their] thinking," the instructor is trying to prompt the student to disclose the cognitive processes by which they arrived at an answer to this math problem. Perhaps what students actually do to fulfill this prompt is to offer a representation of themselves as a thinker: a *portrayal* of cognitive activity that can *explain* their mental outputs to themselves and to others. As long as this portrayal is sufficiently convincing, instructors can pride themselves on their success in invoking "metacognition," all the while reinforcing students' habits of performing whatever kind of cognitive activity their instructors seem to be looking for.

## **§2.4 Higher-order thinking dethroned**

In the previous section, I established that metacognition is often cashed out as *higher-order* cognition; here I'll discuss a view according to which higher-order thinking has been undeservedly elevated above other forms of cognition in status and importance. Hilary Kornblith argues that our impression that *reflection* (his verbiage for thinking that qualifies as "higher-order" through intentional ascent<sup>16</sup>) is a superlative species of cognition is unfounded (2002, 2012). I'll consider how his view challenges the idea that having students engage in reflective thinking should be put on a pedestal as a pedagogical achievement.

Kornblith points out that philosophers tend to "attribut[e] magical powers to reflection," invoking it in explanations of an array of problems across multiple areas (2012., p. 7). For example, reflection is often called upon as the means which true beliefs may become properly justified and thereby earn the title of knowledge.<sup>17</sup> A common narrative in normative epistemology identifies "reflective scrutiny" as the means by which one can and should rescue one's beliefs from the doldrums of errors, biases, and other mechanisms that make our unreflective thinking highly unreliable as a guide to truth (*ibid.*, p. 1). Reflection also is invoked the popular Frankfurtian accounts of freedom of the will and personhood, as the mechanism which enables the formation of second-order volitions (Frankfurt 1971), and is proposed to create the normative force which drives action, transforming standing beliefs, desires, and so on into *reasons* to act (Korsgaard 1996). To this list of examples I'll suggest one addition: John Dewey contends that reflection is what allows individuals to draw *meaning* from experience.

---

<sup>16</sup> Contrast this with the colloquial use of "reflection" to capture *thinking deeply or intently* about any subject, not necessarily concerning one's own mental states.

<sup>17</sup> For example, Ernest Sosa names reflection as that which distinguishes human knowledge from that of animals (1991); Lawrence Bonjour defended a coherentist view on which beliefs are justified insofar as one has recognized the coherence of each belief with the others through reflection (1985). See Kornblith 2002, Chapter 4 ("Human Knowledge and Reflection").



Whereas the unreflective individual lets experience wash over them, a reflective thinker funnels this flood of data into the formulation of a theory by which they're able to make sense of experience and attain understanding (Rodgers 2002, pp. 845-849; synthesizing Dewey 1910 and 1916).

Kornblith contends that each of these accounts have to assume, without warrant, that second-order mental states produced by reflection are "self-legitimizing": that they are distinct from first-order states in a way that allows them to escape the need to earn *their* legitimacy by being reflected upon by other higher-order states (p. 135). In other words, Kornblith's critique of the power granted to reflection invokes the same regress problem raised by Gilbert Ryle (see §2.1): If the first-order state earns an elevated status (as knowledge, or as a reason, or so on) in virtue of being the object of reflection, how does the second-order state which arises through reflection meet standards of epistemic legitimacy, other than by themselves becoming objects of even-higher-order reflection, *ad infinitum*?

A second problem, according to Kornblith, is the assumption that reflective processes (including introspection) are more reliable than the first-order processes they analyze (e.g., belief acquisition). He argues that it's a mistake to automatically place credence in second-order states.

To begin with, Kornblith rejects a tendency to conceptually tether second-order cognition to "System 2 reasoning," presuming a dual-process framework.<sup>18</sup> As Kornblith tells it, dual-process models of the mind have coincidentally managed to reify the narrative that "reflective scrutiny" traces a bright line between two distinct strata of mental contents. That narrative reads something like this: Our automatic System 1 processing is fast and untrustworthy; in order to achieve better outcomes (e.g., more accurate judgments and better-

---

<sup>18</sup> See, for example, Evans & Over (1996), Stanovich (2009), and Mercier & Sperber (2009).

informed decision-making), we need to strategically counteract the erroneous outputs of System 1 with the more accurate, controlled, *reflective* thinking of System 2.

In response to this narrative, Kornblith objects that it's misleading to equate System 2 activity with "reflection," as System 2's conscious outputs aren't necessarily second-order:

. . . when psychologists speak of System 2 as involved in reflection, their use of that term better accords with everyday usage, which allows that we may reflect on various features of the world around us and not just on features of our mental life, than it does with the technical usage here which ties reflection to second-order states. (2012, p. 140)

He elaborates that "everyday usage" of reflection marks thought as *reflective* simply on the grounds that one is *conscious* of having these thoughts – but "thoughts which are brought to consciousness in System 2 need not be thoughts about the thinker's mental states" (ibid., p. 142). Meanwhile, a subset of our second-order cognition (namely, whenever one "make[s] inferences about his or her own mental states without focusing attention on these inferences") can be attributed to System 1, since it never crosses into our conscious awareness (p. 143). So, part of the reason why reflective thinking gets treated as a superlative mode of thought might boil down to linguistic error: that speakers who describe System 2 as "reflective" are equivocating between second-order thinking and conscious cognition. Insofar as the latter is taken to constitute a more accurate or more controlled mode of thought, so does "reflection," through this conflation.

Kornblith goes on to question the warrant for treating System 2 outputs as though they are a) immune from error and b) identifiable with one's genuine, *reasoned* beliefs, intentions, and so on. One reason for this is that the border dividing System 1 from System 2 is a porous one: Instead of operating completely in isolation from one another, the two systems interact constantly. For example, inferences made by System 2 are informed by a large stock of "background beliefs" acquired and maintained in System 1; as a result, any faults or limitations

of System 1's operations influence System 2 reasoning upon these foundations (pp. 144-145). For example, an agent may overlook certain options during conscious problem-solving because System 2's conscious explorations can only operate within the confines of what System 1 registers as falling within the space of possibility, based on implicit learning from past experience. So, even when second-order thinking does belong squarely to System 2, it can be faulty (hasty, biased, myopic, desire-driven, and so on).

Another reason not to treat System 2 outputs as an agent's authentic thinking is that we might be prone to falsely equating consciousness with truth. Kornblith writes:

We should be careful to recognize that we cannot simply assume that whenever a series of propositions passes through the conscious mind, even when such a series is accompanied by some feeling of endorsement of the last of the propositions, or endorsement of the last on the basis of the earlier ones, that this amounts to belief on the basis of reasoning. Believing a proposition to be true should not be identified with having it pass before one's mind consciously while accompanied by some feeling of endorsement. . . . [F]eelings of endorsing a proposition as true may, in many cases, come apart from genuine belief. . . . What seems like belief from the first-person perspective, and what seems like reasoning from the first-person perspective, need not actually constitute belief or reasoning. When we examine the workings of System 2, then, it will be important to keep in mind that states which seem like beliefs to the agent who has them . . . should not all be taken at face value. (2012, p. 143)

Although being conscious of System 2's outputs delivers us the *impression* of having exercised agency in endorsing a belief or a judgment as belonging to oneself, Kornblith attests that we can be conscious of states that amount to "self deception[,] . . . after-the-fact rationalization or confabulation" (p. 143). Our sense that our conscious reasoning is a better reflection of our *actual* will or character than our unconscious cognition may be just that: a sense, a phenomenal artifact of how this information presents itself to us.

Lastly, Kornblith argues that we aren't warranted in inflating the *non-automaticity* of System 2 processing into the determination that this form of thinking is fully *controlled*. While we can choose to initiate certain types of thinking (like deliberation), we aren't at liberty to

direct the inferences made therein; rather, these inferences occur passively but present themselves to us with a certain “actish phenomenological quality” . . . the strong feeling, however difficult it may be to describe, that what is going on – whether it is a mental event or a bit of behavior – is one’s own doing” (ibid., p. 151; quoting Ginet 1997, p. 89). In other words, System 2’s outputs are accompanied by an illusory “feeling of agency,” on the basis of which we falsely believe that we are actively orchestrating whatever thinking arises in consciousness. If this is the case, it makes sense that we’d be prone to regarding our conscious second-order thinking as a state in which we’re fully in command of planning, monitoring, and directing what’s going on in our own minds. Nevertheless, being conscious of a mental state isn’t equivalent to causing or being in control of that state.

Kornblith alleges that our impressions that reflection is somehow less mechanical and more agential than the rest of our thinking stem from a “perspectival shift . . . from a third-person perspective on first-order processes to a first-person perspective on reflection” (2012, p. 161). In short, reflective thinking isn’t fundamentally different from first-order thinking (nor conveniently immune from the problems associated with first-order thinking). Rather, we tend to attribute different properties to reflection because we take *ownership* over our reflective thinking, treating it as the outcome of our agency, whereas first-order thinking is treated as the output of a mechanistic system which we happen to possess, but isn’t really *us*. In reality, reflective thinking is also mechanical, and driven by factors that are outside of our ken and beyond our control: It’s only by failing to or choosing not to acknowledge this that we develop the impression that reflection is distinct from and superior to first order thinking. Kornblith labels it a “substantive philosophical error” that “reflection is thus seen as different in kind from processes that go on at the first-order level” (p. 157).

#### **[2.4a] Implications for metacognition in education**

Let's grant Kornblith's account that reflection doesn't merit the plaudits it has been granted by philosophers. What might this mean for the strain of metacognitivism which that advances "reflective activities" as a method of metacognitive enhancement?

Let's start with Kornblith's observation that the colloquial understanding of "reflection" is an inexact one, somewhere in the realm of *thinking that is conscious and deliberate*. I think it's safe to assume that this — not the more technical notion of *higher-order thinking* employed by philosophers — is what most instructors have in mind when they hear metacognitivists recommending reflective activities. As a result, I'd hazard a guess that much of what students actually do while carrying out "reflection"-heavy assignments and activities is merely conscious first-order thinking. And while students may benefit from devoting attention to this form of thinking, I don't see the merit in shoehorning the practice of first-order thinking into one's notion of "metacognitive training." (As Chapter 1.4 attested, being overly permissive about what counts as "metacognition" threatens to dilute whatever utility this concept might have had in marking a sphere of intellectual ability that merits special attention from educators.)

Beyond this, Kornblith's account implies that there's no good reason to treat the products of reflective activities as though they offer a more authentic portrait of students' thinking than what emerges under other circumstances. What students report about their own thinking through reflective exercises isn't necessarily a lens into the depths of their minds; rather, prompting reflection might just create opportunities for students to confabulate about the operations of their minds, unbeknownst to them or their instructors.

As I see it, Kornblith's account diminishes much of the mystique around reflection. While dulling the shine of a concept may not be an especially tangible effect, it could very well make a difference in whether or not educators are compelled to put stock in reflective activities.

Even if you aren't convinced by his arguments flattening the distinction between reflection and first-order thinking, he still offers a non-empirical reason to question the elevated status of "reflection" in pedagogical discourse: Namely, he proposes that engagement in metacognitive reflection is treated as a special pedagogical achievement because we've collectively adopted a stance on which reflection has been put on a pedestal.

What this serves to remind of us is that the value hierarchies on the basis of which we deem some pedagogical practices superior to others are an organizational scheme of our own making — not a fixture of nature uncovered by educational science. If we choose to look at education through the lens of a theory of learning on which certain forms of cognition are more desirable than others (as when "metacognition" is privileged over mere cognition), then we can celebrate the accomplishment of student engagement in "metacognition" as a satisfaction of our desires. This should not, however, be mistaken for attainment of an objectively superior mode of thought.

## §2.5 | Summary

In this chapter, I've provided a round-up of challenges to some key premises underpinning metacognitivist commitments to developing students' metacognitive skills through pedagogical intervention. Some of these are longstanding mysteries about the mind, like how a cognitive system develops the capacity to analyze, evaluate, and regulate itself in the first place. Others are psychological findings that steer us towards a view on which our first-person experiences simply can't be trusted as a source of authority on how our minds work. And some are proposals for how to re-envision mental capacities we have taken for granted (namely, reflection and introspection), which invite us to reconsider the stock we have placed in dogmas like transparent access to our own mental states and the immunity of higher-order cognition

from the errors of lower-order processes.

Having reviewed these challenges, I should hope to have motivated the reader to gravitate toward a view on which our ability to know and exercise control over our own minds is *not* a given which can readily be exploited by savvy education practitioners. Rather, the concept of metacognition might be more like an educational shibboleth, belief in which is bolstered by its congruence with a number of highly convincing illusions about our own minds and the desirability of the notion that we each have the capacity to alter our minds at will, as long as we're provided the right instruction to unlock this ability.

### 3. SCRUTINIZING METACOGNITIVE TRAINING INTERVENTIONS

Beginning in the mid '70s, a critical mass of educational psychologists advanced and backed a proposal that what enables students to learn successfully is *metacognition*: a domain of intellectual ability which constitutes a previously underappreciated form of practical intelligence in academic settings.<sup>19</sup> According to this outlook on education, superior *metacognitive skills* distinguish adept students from struggling ones. From this it would seem to follow that educators, who are occupationally obligated to do what is in their power to help students succeed in school, ought to “treat the development of [metacognitive] skills as the paramount goal of all schooling” (Resnick 1986, p. 43).

A critical assumption of this strong metacognitivist approach to education is that metacognitive abilities are *malleable*. Although it is acknowledged that students enter the educational enterprise with inexplicably varied degrees of metacognitive ability, strong metacognitivism takes for granted that it is within instructors’ power to enhance students’ level of metacognitive skill through explicit instruction.

Many proponents of strong metacognitivism assert that this assumption has been borne out in empirical research on metacognition in education. “What is encouraging is that research . . . shows that it is possible to teach children metacognitive skills and when to use them,” John Bruer reports (1993, p. 72). “Metacognition also improves with appropriate instruction, with empirical evidence supporting the notion that students can be taught to reflect on their own thinking,” writes Emily Lai (2011, p. 2). “Metacognitive skill can be taught,” Craig et al. assert (2020, p. 206). John Perry and colleagues insist:

The available evidence strongly suggests that metacognitive approaches to teaching and learning have the potential to radically improve the outcomes and

---

<sup>19</sup> Flavell (1976, 1979); Flavell & Wellman (1977), Brown & deLoache (1978), Brown et al. (1983), Bransford et al. (1986)



life chances of children, with some evidence suggesting that this is especially the case for disadvantaged children. Such knowledge then places a moral responsibility on policy makers to ensure that schools, school leaders and all those who support them to rapidly develop and implement practical strategies which can deliver, measure and improve metacognitive skills across all schools for all children. (Perry, Lundie & Golder 2019, p. 27)

For Perry et al., studies on metacognition in education demonstrate that we not only can but *must* pursue pedagogy that provides metacognitive training. In this vein, a U.S. Department of Education fact sheet asserts: “Research shows that metacognitive skills can be taught to students to improve their learning,”<sup>20</sup> and “. . . [because] metacognition plays a critical role in successful learning, it is *imperative* that instructors help learners develop metacognitively” (Office of Vocational and Adult Education 2011, my italics).

In this chapter, I’ll explain why I’m not convinced that the available evidence really does impose such a mandate upon educators. I allege that even if the concept of metacognition was entirely sound, our methods of deliberately provoking metacognitive skill gain are not. I’ll begin by revisiting the basic varieties of metacognitive training I catalogued previously (in §1.2b), in light of the challenges to basic assumptions about metacognition that were elaborated in Chapter 2. I’ll note some *a priori* concerns with putting stock in these methods as conduits to metacognitive skill development, before spelling out how the prospects for these methods are threatened by revisionist models of introspection and metacognition elaborated by cognitive scientists (§3.1).

In §3.2a, I’ll explain how most individual studies on metacognitive training have relied on measures with questionable validity (e.g., self-report questionnaires, interviews, analysis of responses to think-aloud prompts, and teacher ratings), all of which are susceptible to problems with social desirability bias and experimenter effects. These methodological issues are layered

---

<sup>20</sup> They attribute these findings to Halpern (1996), Nietfeld & Schraw (2002), and Thiede, Anderson, & Therriault (2003).

on top of concerns about confabulation, self-deception, and desire-driven reporting about one's own mental workings, raised in Chapter 2.2.

In §3.2b, I'll discuss problems with deriving pedagogical justification from meta-analyses that generalize across a wide variety of metacognitive training methods. First off, I emphasize that the vast majority of meta-analyses that concern metacognition in pedagogy do *not* provide evidence for the effectiveness of metacognition-targeting interventions, because they don't actually *measure* students' metacognitive skills pre- and post-intervention. Of meta-analyses that do summarize the right kind of data to bear on the question of whether metacognitive skills respond to pedagogical interventions, there remain problems related to the generalizability of the findings and the aggregation of data concerning many distinct types of interventions that can be classified as "metacognitive."

§3.3 addresses the objection: *If pedagogical interventions that target metacognitive skill don't actually improve students' metacognitive abilities, what else would account for the learning gains that standardly follow these interventions?* I offer two explanations for why learning outcomes could increase even if metacognitive abilities aren't affected by training.

First, I'll demonstrate that in practice, interventions that target metacognition often involve having students put domain-specific knowledge to use in a strategic manner to complete assessment tasks. So, the apparent gains in learning wrought by metacognitive interventions are primarily a consequence of students knowing what standards they're being held to during assessments and behaving accordingly. This *knowledge-driven explanation*, as I'll call it, is a much more parsimonious account of the apparent effectiveness of metacognitive training, compared to the hypothesis that students' minds have been molded so as to allow them to cognitively self-regulate in accordance with standards for intellectual excellence.

Second, those who enact metacognitive training methods often have adopted a view according to which “metacognition” is a constituent part of successful learning, such that the measures they use to affirm achievement of desired learning outcomes will reward students who display uptake of a metacognitive training intervention. In these circumstances it may appear that the metacognitive intervention caused students to perform better, but in reality the goalposts for successful learning have just been shifted such that “being metacognitive” is a trait of accomplished learners.

### **§3.1 | Methodological issues with metacognitive training**

As I see it, efforts to effect lasting enhancements in students’ metacognitive abilities fall into three major classes of interventions – *explicit modeling*, *explicit strategy instruction*, and *reflective activities*. I’ll take a closer look at each method of metacognitive training in the following subsections.

#### **[3.1a] Explicit modeling: Displaying more than we can know?**

This intervention calls upon instructors to regularly “model their thinking as they approach a task to reveal the reflections of an effective learner” (Quigley et al. 2018, p. 16). It assumes that students need to be exposed to explicit, verbalized displays of metacognition from role models in order to develop their own metacognitive capacity.

##### **(3.1aI) Attempting to externalize implicit processes**

A well-established difficulty with this method is that “the reflections of an effective learner” may become automatized and subpersonal as a consequence of the extensive practice it takes to learn a domain well enough to teach it. As Bruer puts it: “One problem with metacognitive skills is that they are usually covert and implicit in expert performance. To teach

these skills and when to use them, the instructor has to make metacognition overt and explicit” (1993, p. 73).

While the unconscious operation of at least some metacognitive processes tends to be presented as a challenge to doing explicit modeling well, some psychologists suggest that this represents a major obstacle for would-be modelers. In particular, Reder and Schunn assembled evidence that a critical component of metacognitive control – selecting a cognitive strategy to enact – cannot be made explicit in the form of a verbal report (1996). These authors argue that even if subjects have explicit declarative knowledge of various strategies they could implement to complete a cognitive task, the process by which the subject selects one strategy as opposed to another is *not* consciously accessible, and is guided by *nondeclarative* knowledge acquired through past experiences:

. . . much of the processing that is called metacognitive typically operates at an implicit level; that is, without conscious awareness. Many of the tasks that are called monitoring are also operating without conscious awareness – people cannot veridically report what they have perceived and acted on. Further-more, when people make efforts to change the nature of the task so that they are conscious and can report what they are doing, they run the risk of fundamentally changing their performance in non-optimal ways. (Reder & Schunn 1996, p. 73)

To bolster this latter point, they refer to evidence that attempting to verbalize the process of strategy selection degrades performance on corresponding tasks. The leading explanation for why this occurs is that “because we do not normally attempt to verbalize these processes, the act of verbalization is especially taxing and disruptive to the relatively automatic processes,” imposing cognitive load the subject is not accustomed to bearing (ibid., p. 71).

Reder & Schunn advise that though cognitive strategies can be explicitly described to students, students attain the desired ability to deploy these strategies under appropriate circumstances through task repetition that establishes tacit strategy-selection rules in implicit memory (pp. 71-72). So, their research suggests that there may be components of metacognitive

ability that just can't be explicitly exposed to learners — but also that explicit modeling is misguided to begin with, since students acquire the ability to automatically select appropriate cognitive strategies through means other than mimicry of visible models of strategy selection.

### **(3.1aII) Rationally reconstructing ideal metacognition**

As far as I can tell, no study has explored the extent to which instructors succeed in externalizing their *actual* metacognitive activity for learners to observe. To accomplish this would require, first, that we could reliably tease out genuine metacognition (i.e., veridical reporting of one's own mental states) from confabulations about one's cognitive activity, which is a methodological doozy in itself.<sup>21</sup> The alternative I'm raising is that instructors who make an effort to verbalize their own metacognitive activity may present a (likely idealized) approximation of their metacognition through rational reconstruction of what kind of metacognitive activity they ought to have engaged in to achieve a particular desired outcome, taking for granted that metacognition is a necessary ingredient in academic proficiency.

For all we know, it might turn out that it doesn't make any practical difference to learners whether their instructors provide a genuine display of their own skilled metacognition, or a hypothetical demonstration of metacognition one ought to engage in. But in order to decipher what impact explicit modeling of metacognition has on students' metacognitive skill gain, we'd need to be able to disentangle the effects of teachers' deliberate display of metacognitive activity for their students from the effects of any accompanying changes in content presentation that might serve as a confound. Let me give an example to explain why I think this disentanglement would be difficult to accomplish.

---

<sup>21</sup> This presumes that there is such a thing as "genuine" metacognition to begin with. Johansson et al. (2005, 2006) allege that confabulation may well be the norm for reporting on our own cognition, and veridical reporting a rarity.

Here's an instance of advocacy for explicit modeling of a particular metacognitive skill, from philosopher David Concepción:

Experts are better than non-experts at grouping related information into a useful, accessible chunk that can be unpacked quickly. For example, experts in moral philosophy easily recognize the conceptual linkages among moral constructivism, subjectivism, intersubjectivism, and objectivism. Experts group these ideas into one intellectually manageable package of related but dissimilar ideas. Novices may not notice the conceptual linkages. Novices may attempt to memorize the meanings of these terms in isolation by rote. Such novices may be able to accurately identify these definitions on a multiple-choice exam. However, they are likely to have difficulty writing a sophisticated essay because they have not discerned the similarities and dissimilarities needed for rich understanding. To help students perform better, professors should do their best to explicitly describe how they chunk information. (2004, pp. 355-356)

Here Concepción is deriving pedagogical conclusions from comments by Bruer (1993) on various distinctions between experts and novices. Specifically, Concepción is reasoning that if experts' superior skills in "chunking" information contribute to their facility in demonstrating "rich understanding" of a domain of knowledge, then professors should be able to eke out more "sophisticated essay[s]" from their novice students by "explicitly describ[ing] how they chunk information." The idea, I gather, is that the explicit description offered by the professor becomes something of an instructional manual for chunking that students can apply strategically, to impose order upon their own newly-acquired knowledge of philosophical concepts.<sup>22</sup>

Presumably, the purpose of having an instructor verbalize how they as an individual chunk information within their domain of expertise is *not* just to allow learners to internalize this same scheme of organizing information. The learner is not supposed to simply take up the instructor's mental frameworks as their own; the point is rather to provide the learner with an illustration of the *kind* of mental framework it is useful to assemble, in order to skillfully make use of the various bits of information one has picked up within a domain.

---

<sup>22</sup> My apologies for the number of instances of "chunk" in this passage. I suspect I'm not alone in finding the word viscerally unpleasant to read.

First off, given that at least some chunking occurs automatically and implicitly (Gobet et al. 2016), it may be quite difficult, if not impossible, for the instructor to unearth the process by which they developed the organizational scheme according to which they have chunked knowledge in the domain of instruction. If an expert tries to report *how* they went about forging these mental linkages, it seems possible that they may very well offer up a narrative of how this *likely* happened or how it might make sense for this to have happened, rather than reproducing the *actual* sequence of cognitive effort by which one chunks like an expert.

But another concern is as follows: How is an instructor supposed to show *how* they chunk information without articulating *what* it occurs to them to chunk together? Perhaps one could deliver an abstract discussion of the criteria according to which one ought to chunk whatever material happens to be chunkable, without providing examples of what one would chunk together on the basis of these criteria. But I suspect that efforts to show how one ought to chunk typically involve the instructor specifying the ways in which items of curricular content are linked together in their own mind. Insofar as this “explicit modeling” of the metacognitive skill of chunking helps students to perform better in a course, my suspicion is that it’s transparency about the instructor’s advanced understanding of the domain which gives novice the information they need – not a breakdown on how to take up the mechanics of chunking into their own cognitive apparatus. Later in this chapter, I’ll argue that much of the apparent benefit of metacognitive training interventions can similarly be explained as a consequence of how these interventions expose students to domain-specific knowledge they can exploit in assessments.

### **[3.1b] Strategy instruction: Skill-building or “bag of tricks”?**

This species of intervention aims not only to make students aware of “metacognitive strategies” they can deploy to *monitor and regulate their own thinking and behavior during learning*, but to cultivate skill in strategically deploying these strategies in appropriate circumstances, in order to optimize their ability to manage scholastic challenges.

One immediate issue with this training approach is how to delineate “metacognitive strategies,” and in turn, how inclusive to be about what qualifies as metacognitive training. As I noted previously (§1.2bII), some speakers restrict “metacognitive strategies” to denote *methods of planning, monitoring, and modifying one’s procedure for tackling a learning task*, examples of which include *self-talk* (elaborating an inner monologue about how one is approaching a learning task throughout task performance) and *self-monitoring* (leading oneself through a mental checklist to assess one’s progress or the products of one’s efforts), or *self-questioning* (posing questions to oneself within one’s inner monologue to manage one’s approach and assess how well one is doing throughout a learning task).

However, other speakers blur the lines between cognitive and metacognitive strategies, seeming to describe *anything one could do that might lead to better learning outcomes* as a “metacognitive strategy” (Florida Center for Instructional Technology n.d.). On expansive interpretations, metacognitive strategy instruction includes practices like *teaching students to annotate text as they read* (Concepción 2004, Nash-Ditzel 2010) or *teaching students to pay attention to chapter titles in order to gain information about what the chapter might contain* (Baker & Brown 1981; see discussion in Brown 1987, p. 66). Similarly,

- *checking all of the options on a multiple-choice question before selecting an answer* (Flavell 1976)
- *reading the whole problem before starting to solve it* (Tomsett 2015, Iris Center 2020)
- *making sure one has underlined key words in a problem before attempting to solve it* (Vula et al. 2017)



- *using signposts and transition phrases in one's writing* (mentioned in Ashman 2016)

...have all been described as “metacognitive strategies,” insofar as knowing and implementing these strategies on key assessments tends to result in better performance in school.

It certainly is *possible* to describe picking up reading comprehension strategies as acquiring a means of regulating one's own learning (say, by taking control over one's uptake of relevant information from text). Same goes with the other strategies listed above: someone who is aware that they could attempt any of these methods is well-positioned to enact auspicious approaches to test-taking, essay-writing, and so on. But implementing these strategies doesn't require conscious, verbalizable thinking about one's own cognitive activity, unlike the self-talk and self-monitoring strategies described above. All of these actions could be thoughtlessly employed by students who aren't able to articulate *why* they are doing these things, who are merely doing as they were instructed.

These instructional choices may very well develop helpful habits in students, and they may each be a boon to academic performance — but to describe the students' acquisition of these habits as “metacognitive skill gain” feels like a stretch. If something like *picking up the habit of reading chapter titles* can count as an enhancement of metacognitive abilities, then this seems to me more a matter of expanding the scope of “metacognitive skills” to include a helpful skill that is easily teachable than it is a matter of cultivating students' practical wisdom about how to learn.

What's going on here, I think, is that there is quite a bit of slipperiness around what is involved in *learning how to learn*. For those who use “metacognitive strategies” narrowly for the likes of self-talk and self-monitoring, “learning how to learn” means cultivating the ability to *engage in an internal monologue about what one is doing while one is tackling a learning challenge, and to deliberately hold one's thinking & behavior to specific standards* (of completeness, accuracy, and so

on). For those who use “metacognitive strategies” more loosely, it would seem that “learning how to learn” can mean *picking up behavioral habits that just so happen to conduce to greater success on academic assessments*. To this end, Willingham & Lovette (2014) have argued that reading comprehension strategies don’t actually promote general, transferable skill-gain, but rather constitute “a bag of tricks” instructors can draw from to yield performance improvements by addressing some low-hanging fruit among obstacles to competent reading.

To put it bluntly, so-called “metacognitive strategies” vary widely in what kind of intellectual achievement their implementation represents. Some strategies, like self-talk and self-monitoring, can only be effectively carried out by students who have developed a robust, consciously-articulable understanding of how to learn well in a particular domain. Having an inner monologue about one’s own learning is only beneficial if one grasps a subject well enough to accurately verbalize what kind of challenge one is facing and how to tackle it; checking the accuracy of one’s outputs only works as a means of error detection and correction if one understands the domain well enough to recognize what would qualify as an error and grasps what they’d need to do different in order to get a different outcome. Succeeding in getting students to use the former type of strategies would indeed represent an impressive cognitive transformation achieved through pedagogical intervention.

But quite a few methods presented as “metacognitive strategies,” like annotating while reading, represent rote uptake of behavioral programs. These could stand to be rhetorically downgraded to a less-impressive designation of *academic hacks*: simple procedures by which one can eke out extra points on an assignment or exam (Ashman 2016).<sup>23</sup> Getting students to use the

---

<sup>23</sup> Drawing on Willingham & Lovette’s observations (2014), Ashman offers this analysis of the apparent effectiveness of reading comprehension strategies billed under the heading of “metacognitive training”:

They are not skills in the sense that a sequence of deliberate practice will make you improve at them. They are useful hacks that, once known, produce a one-off hike in performance. If you take a student who can’t structure a piece of persuasive writing and teach them the ‘firstly... secondly... thirdly...’ hack then you will see an immediate and significant jump against a standardized

latter type of strategies, on the other hand, is only an enhancement of “metacognitive abilities” given a highly permissive sense of “metacognitive.” Unfortunately, the mere uptake of these hacks by students – or even just students’ testimony that they employ these hacks – from students can be misconstrued as a victory of metacognitive training, as I’ll elaborate in §3.2.

### **[3.1c] Reflection revisited**

Metacognitivists often refer to reflection as the basic action which subserves engagement in metacognition, and promote reflective activities in the classroom as a means to enhance metacognitive ability. Previously (§1.2bIII), I characterized these as aiming at *generating self-knowledge* and/or *subservicing self-assessment* (i.e., forming judgments about oneself or one’s actions). Moreover, activities intended to *generate self-knowledge* variously aim at cultivating awareness of one’s *occurrent mental states or cognitive processes* in the midst of a learning task, familiarity with one’s *intellectual dispositions* (concerning patterns of mental activity over time), recognition of *facts and concepts one has assimilated* into one’s mental stores, or acknowledgment of one’s *standing beliefs and attitudes*.

One initial concern with treating reflective activities as “metacognitive training” is that it’s often opaque what it would mean to *get better* or *become more skilled* at reflection. Despite how often reflective activities are recommended, there is very little discussion of what educators are aiming at by implementing these interventions: Are they picturing a student who has mastered reflection as one who is disposed to reflect *often*, or one who reflects *well*, or some

---

persuasive writing test. But how significantly have you improved their writing skill? The slower, much more liminal, curriculum-centered process of building vocabulary is far harder to capture in this way. (2016b)

His allegation is that it’s misleading to portray a practice like the one described as “metacognitive skill-building”. Students are being directed to implement a particular strategy in a setting where this will raise their test scores; to attribute this to a heightened ability to “direct their own learning behavior” is to drastically overascribe self-awareness and self-control to these learners.

combination of these two measures? If the aim is for students to reflect *well*, what exactly does this amount to (e.g., is it *increased accuracy* in their reports about themselves and/or their intellectual activity, *more revelatory outputs* from their efforts to self-reflect, *more action taken* to adjust their cognitive activity in light of what reflection revealed...)? The ultimate objective of these efforts is underarticulated, and as a consequence, it's difficult to assess whether or not practicing reflection yields its intended long-term effects.

### **(3.1cI) Reflection or invention?**

Let's take a look at an example of a use of reflection as a pedagogical intervention to enhance students' metacognitive skills. Shelagh Crooks claims that reflection supports the development of *cognitive self-awareness*, which in turn enables students to find motivation "to find ways to do a better job of thinking," as well as to resist partiality to particular beliefs (pp. 256-7). She attests that her pedagogical approach requires students "to engage in 'metacognitive inquiry' in which their own judgments . . . become a subject matter for reflection and critical evaluation" (p. 247). For instance, she assigned a retrospective self-assessment exercise at the end of the semester, in which students were instructed to write up a document containing:

1. An analysis of the concept of metacognition and its relevance to the study of philosophy.
2. A reflective autobiographical statement describing habits/proclivities of themselves as thinkers engaged in philosophical inquiry.
3. A critical evaluation of these habits/proclivities.
4. Discussions of a minimum of three topics from class identifying what the students learned, failed to learn about each topic.
5. An evaluation of the metacognition as a technique for engaging in philosophical inquiry. (ibid., p. 258)

She asserts that students' output for this assignment "indicate[s] that they did indeed engage in metacognition," because all students reported enhanced awareness of bad cognitive

habits, like holding onto beliefs in spite of contradictory evidence.<sup>24</sup> Crooks considered the intervention a success insofar as it enabled students to make “definite strides in the direction of adopting argumentative strategies in thought” (ibid.).

The concern this method prompts immediately for me, in light of the findings on confabulation discussed in the previous chapter, is as follows: *Why should we trust students' self-reports as a veridical source of data about their thinking?* Let's grant, for a moment, the suggestion advanced by Rosenthal and Carruthers (in §2.2c) that self-knowledge seemingly yielded by introspection turns out to be an interpretative invention of our minds, shaped by normative pressures and prone to confabulation. Accordingly, if we invite students to self-report on their mental processes or examine their own thinking, we're prompting them to generate novel representations of themselves as learners and thinkers – not to plumb the hidden depths of their minds.

Also, though Crooks claims to invite students to arrive at their own judgments how well metacognition fares as a means of developing philosophical sophistication, it's hard for me to believe that this assignment doesn't produce a *demand effect* on student behavior, where students say what the instructor wants to hear instead of reporting their thoughts and views truthfully. If one's pedagogy makes it glaringly obviously that the instructor is convinced of the value of metacognition as the road to successful philosophical learning, we ought to expect that this will result in students singing the praises of metacognition, and insisting that their

---

<sup>24</sup> She discloses that her conclusions about the effectiveness of this strategy are merely tentative, given that the intervention was conducted on a “small sample of students . . . not particularly representative of the university population”: namely, nineteen philosophy majors (p. 261, footnote 25).

It is a relief to hear Crooks admitting that the evidence base for the effectiveness of this intervention is scanty. It's less of a relief that she refers to this report on this particular pedagogical tactic as a “study.” A general problem that concerns the scholarship of teaching and learning (see Chapter 1, footnote 5) is that there is no principled way to distinguish between a “study” and a mere anecdote: in this mode of research, any pedagogical choice (no matter how small or transient) is treated as fodder for discussion with other instructors, and everything that occurs as a result is fair game as “data” on the effects of this choice.

metacognitive training has wrought valuable insights into their own minds. Crooks doesn't acknowledge this could account for the students' apparent enthusiasm about metacognition; instead, she characterizes them as "brutally honest" about their own intellectual faults in "describ[ing] themselves as 'closed-minded,' 'stubborn,' and 'irrationally committed to defending their belief against criticism'," and she takes them at their word when they reported "resolv[ing] to work against this propensity by listening carefully to arguments in favor of alternative beliefs when and if such arguments are articulated in the classroom" (2009, pp 258-9).

More generally, there is a dearth of discussion regarding how educators are supposed to *verify* that the reflective activities they are assigning are actually getting students to acquire self-knowledge and/or form accurate self-assessments. If instructors assign essays in which students are asked to report on what they were thinking during a particular task, or what they are currently confused about, or what they have learned from a course (to name a few examples), said instructors are not positioned to have access to information about what the student is *really* thinking or confused about, or what the student has *really* learned; the student is the only subject to whom those mental states are potentially accessible. As a result, instructors don't have a reliable way to distinguish between veridical self-reports of students' mental states and operations, on the one hand, and (deliberately or inadvertently) fabricated accounts, on the other. Ultimately, whether or not the student has "succeeded" in these activities may be judged on the basis of arbitrary criteria unrelated to the student's metacognitive abilities, like whether students' responses are well-written, or the extent to which written reflections mention concepts specific to the course being taught.

Considering that instructors may not be able to confirm that students' efforts at reflection are actually resulting in heightened access to the inner workings of their minds, it

would seem that many instructors are operating under an implicit premise that reflection is a brute good — that it’s intrinsically worthwhile for students to engage in reflection, whatever its outcomes. In the literature on teaching critical thinking, some critics have (rightly, I think) objected that engaging in reflection doesn’t entail that a student is thinking in accordance with any standards of intellectual rigor; for example, Kevin Possin objects to equating “critical thinking” with metacognition, on the grounds that “metacognition is too easily construed as merely reflecting on what one happens to be thinking or experiencing at the time, with little substantive guidance as to how to do so” (2008, p. 203). His implication is that *reflection* doesn’t automatically or necessarily yield valuable insights, either into one’s own occurrent cognitive states or one’s long-term intellectual character; nor does it always enable the lasting improvements in reasoning we would associate with critical thinking skills.

### **(3.1cII) Dunning-Kruger vs. self-assessment**

Another arena in which reflective activities are supposed to elicit metacognitive skill gain is self-assessment: the ability to form accurate judgments about oneself as a thinker.

Unfortunately, it would seem that those students who would benefit most from improving their abilities in this sphere may be the least capable of benefiting from self-assessment exercises assigned by teachers.

The *Dunning-Kruger effect* describes the tendency for underperformers to be ignorant of their own ignorance. According to the psychologists after whom the effect was named, “the skills that engender competence in a particular domain are often the very same skills necessary to evaluate competence in that domain — one’s own or anyone else’s” (Kruger & Dunning 1999, p. 1121). As a result, individuals with little knowledge in a particular domain aren’t in a position to accurately evaluate their own competence: failing to recognize how little they know or in

what ways their performance is flawed, they tend to systematically overestimate their ability level and predict that they will perform much better than they actually do.

In short, “incompetent individuals lack the metacognitive skills necessary for accurate self-assessment” (ibid., p. 1122; see also Ehrlinger et al. 2008 for confirmatory follow-up research). In my eyes, this poses two serious problems for those methods of metacognitive training that involve prompting students to engage in self-monitoring and self-assessment.

The first is a practical problem: that students who have the least metacognitive skill in the first place will not be able to gather accurate or useful information about themselves through careful self-examination. Regrettably, surmounting some threshold of awareness of one’s own mental attributes might be a prerequisite for making further gains in self-awareness; the skill-building benefits of metacognitive training interventions may be inaccessible to those with the most serious metacognitive deficits.

Take, for example, an exit-ticket reflection exercise in which students are asked to write a few sentences describing what they are confused about at the end of a class session. Presumably, a teacher who assigns this exercise is hoping that students who haven’t been able to make sense of the contents of the lesson will be able to a) recognize this and b) articulate what part of the lesson they’re confused about, so they might have some inkling of what to look up or ask about to resolve their confusion. However, if the Dunning-Kruger effect is real, we should expect the students who have the weakest grasp of the lesson’s material will also be unable to detect that they aren’t seeing connections between the concepts, or have misconstrued a particular point, or so on. Lacking the knowledge necessary to recognize their own errors or deficiencies, they are likely to be left with the impression of understanding, and to report that they aren’t confused at all.<sup>25</sup> Meanwhile, students who were following along closely and are able to make sense of most

---

<sup>25</sup> Anecdotally, this pattern has been borne out in my own teaching experience. When I’ve administered “reading response” assignments wherein students answer comprehension-check questions about an article alongside a



of the content are much better positioned to recognize any small gaps in understanding left to be filled in. In essence, students who are already metacognitively adept are much more capable of gaining useful information from this exercise than students whose metacognitive deficits prevent them from becoming aware of lacunae in their understanding in the first place.

As far as I can tell, the same pattern is likely to follow when instructors implement self-questioning protocols: Students who have the least subject matter knowledge are the least well-positioned to use self-questioning to their advantage as a means to detect and correct their own errors. One can only detect that one has made an error in cases where one has enough command of the normative demands of a domain to be able to register their own performance as deficient. It seems possible that self-questioning might even *backfire* by instilling a false sense of certainty in students whose ignorance prevents them from detecting their errors. Going through an error detection procedure and coming up empty may boost confidence in one's answers, even in cases where one has made glaring errors but lacked the wherewithal to tell.

For what it's worth, Kruger and Dunning's analysis also suggests that the students who have the most to gain from metacognitive training will get the least out of explicit modeling of metacognition by instructors, along with visible demonstrations of metacognition by peers. One of their findings is that "incompetent individuals fail to gain insight into their own incompetence by observing the behavior of other people" (ibid., p. 1126). Students with the least knowledge in a domain are unable to recognize competent performances by peers, and consequently miss out on opportunities to learn how to improve their own performance by emulating more skilled classmates.

---

metacognitive question asking them to specify what portion or aspects of the article they found confusing or unclear, the students who are most likely to report that they didn't find anything confusing or unclear are most often the ones whose answers on the comprehension checks evince the most confusion. Then again, a key third variable in this equation is *time invested in the assignment*: perhaps this pattern is most parsimoniously explained by the fact that spending the bare minimum amount of time necessary to complete the assignment results in both a) miscomprehension of the article and b) reporting no confusion because that's the fastest way to answer that question.

To summarize, the practical problem that the Dunning-Kruger effect poses for some methods of metacognitive training<sup>26</sup> (namely, those that presuppose some minimum quantity of self-awareness in students: self-monitoring, self-questioning, explicit modeling of metacognition, and so on) is that the gains to be accrued from these methods aren't distributed evenly among students of all ability levels. Instead, we see a dynamic wherein the rich get richer, with those who already have metacognitive skill gaining disproportionately while those who are least skilled to begin with gain little if anything from these interventions.

Some empirical support that this dynamic manifests in real educational settings is provided by Hacker et al. (2000), who implemented a set of interventions intended "to help students develop their self-assessment skills" (p. 162). First off, students in an educational psychology course "received instruction about the importance of reflection in learning, particularly its importance concerning accurate self-assessments of one's knowledge and performance"<sup>27</sup> (ibid.); we can think of this as an intervention to get students to take subsequent self-assessment exercises seriously. Next, the experimenters administered practice exams with accompanying answer keys, and encouraged students to "use their performance on the practice exams as a way to identify strengths and weaknesses in their understanding of the material" (ibid.). To assess the effect of this self-assessment exercise on students' accuracy regarding their own performance, students made pre-exam predictions of what percent of answers they'd answer correctly and post-exam judgments of how many they had answered correctly. Lastly,

---

<sup>26</sup> The method of metacognitive training that is most likely to be exempt from this pattern is explicit strategy instruction; students don't necessarily need a baseline level of self-awareness to pick up new habits related to planning, monitoring, and modifying their thinking — though they may need sufficient knowledge in a domain to put these habits to good use to optimize their learning. However, the flip side of strategy instruction *not* hinging on self-awareness is the unflattering conclusion that students who know very little can pick up and implement these strategies because these so-called "metacognitive" strategies don't actually require any thinking about one's own thinking. See discussion in §3.1b.

<sup>27</sup> "Specific lessons that the students received focused on the benefits of accurate self-assessment, including how accuracy in self-assessment could lead to more effective use of feedback, improved time management, and appropriate goal setting" (ibid., p. 162).

when students received their graded exams, they were “urged to reflect and analyze why their predictions and postdictions were or were not accurate and then develop a plan to prepare for the next exam” (ibid.).

Among other measures, Hacker et al. measured how students’ self-assessments changed as the course progressed through three exams. They found that “by the end of the course, high-performing students increased their predictive and especially postdictive accuracy, but low-performing students continued to show little predictive accuracy” (p. 168). In other words, students who performed quite well on exams were able to leverage the feedback they received from their scores on both practice exams and real exams, incrementally ramping up their ability to judge how well they would do or had done; in contrast, students who received feedback indicating that they had substantially overestimated their performance showed no improvement in their ability to accurately judge their performance over the course of the semester. As Al Kuhayli et al. put it, metacognitive training remedies for self-assessment skills “tend to improve self-assessment of students whose work is already good, whereas they have either a negligible or no impact at all on poor performers” (2019, p. 50).

The Dunning-Kruger effect also raises a familiar concern about the extent to which metacognitive skill is isolatable from other quantities of intellectual ability.<sup>28</sup> At least one dimension of metacognitive skill (self-assessment, including *accurate judgment of one’s abilities and error detection & correction*) is tethered to knowledge within a particular domain: The more one knows (about comedy, math, logic, and so on), the more capable one is of assessing oneself accurately.<sup>29</sup> Moreover, psychologists have shown that deficits in self-assessment skill can be

---

<sup>28</sup> Previously alluded to in §3.1a and §3.1b, and to be addressed yet again in §3.3a.

<sup>29</sup> A caveat here is that highly skilled individuals tend to slightly underestimate their ability level or be underconfident in their performance. However, it might be argued that to err in self-assessment in this direction is actually beneficial, as it may motivate skilled students to continue investing effort into improving their performance, instead of fostering complacency. What this suggests is that perfect accuracy in estimating one’s abilities may not

resolved by correcting gaps in subject's knowledge within a domain (Kruger & Dunning 1999, specifically Study 4; pp. 1127-1130). What this suggests is that, counterintuitively, metacognitive training might not actually be the most efficient way to cultivate (at least some types of) metacognitive skill: rather, supplying subject matter knowledge might be all that students need in order to demonstrate greater metacognitive proficiency.

### **§3.2 | Limitations of empirical findings on metacognitive training**

Authors who are in a position to summarize the vast body of research conducted on metacognition in education often claim that the effectiveness of metacognitive training is empirically validated.<sup>30</sup> "Several researchers offer evidence that metacognition is teachable," writes Emily Lai (2011, p. 20). She cites a collection of relatively high-sample size individual studies and a few meta-analyses as evidence that metacognitive skills are in fact teachable.<sup>31</sup>

I'll counter that much of the evidence for the effectiveness of metacognitive interventions is questionable, due to a collection of methodological issues surrounding assessment of metacognitive ability. As with many targets of psychological research, there are inherent challenges with measuring metacognition that have to do with figuring out how to a) expose students' cognitive states and operations to analysis and b) operationalize "metacognition" in terms of quantifiable measures. For "metacognition," there are extra challenges involved with pinning down which of the many translations of this term is under analysis in a particular study.

---

ultimately be that valuable; rather, adaptive underconfidence might be a more suitable target for metacognitive skill development.

<sup>30</sup> See the introduction to Chapter 2 for a cadre of similar claims.

<sup>31</sup> Cross & Paris (1988) Haller et al. (1988), Hennessey (1999), Kremarski & Mevarech (2003), Dignath et al. (2008)

### **[3.2a] Assessing the effectiveness of metacognitive training**

Researchers who want to study metacognitive development (either age-related or as a consequence of instruction) face significant challenges in finding uncontroversial ways to measure metacognitive ability. To state the obvious, “metacognition” refers to a mental phenomenon which is not directly observable; as such, it must be inferred from observable proxies in students’ speech and behavior. “Metacognitive ability” is even more challenging to operationalize, since it’s a construct with continuous variability (i.e., ranging on a spectrum from least to most): researchers must figure out how to quantify individual learners’ positioning within some specified range of possible values for ability relative to their earlier selves (in within-subjects research designs) or to other students (in between-subjects designs).

Accordingly, how to measure metacognitive ability is a long-standing and unresolved topic of debate among education researchers.<sup>32</sup> The two primary areas of uncertainty concern a) how to subdivide the broad category of “metacognitive ability” into discrete measurable skills, and b) how to collect data on each of these subcomponents.

Attempting to measure “metacognitive ability” requires researchers to settle on a working model of what this category comprises. This is an area where disagreement about the extension of “metacognition” introduces serious methodological obstacles for researchers, as Ann Brown noted many years ago (1987; see Chapter 1.3c). Some authors have called for agreement upon “a comprehensive, unified theory of metacognition,” to provide consensus about how many distinct skills are involved in producing metacognitive ability for assessment purposes (Schraw 2000, p. 298). Like many calls to clarify the boundaries and subcomponents of “metacognition” before this, this request has gone unfulfilled; there’s still no agreement on what

---

<sup>32</sup> See Schraw & Impara (2000), Winne & Perry (2000), Veenman (2005), Veenman et al. (2006), Whitebread et al. (2009), Lai (2011), Craig et al. (2020)

quantities researchers ought to assess, so different researchers may conduct assessments in accordance with idiosyncratic visions of what they are striving to measure.

On top of disagreement about what exactly should be assessed, there are disagreements about the reliability and validity of competing assessment methods, each of which involves trade-offs between various desirable features. By far the most popular method of assessing metacognitive ability is to administer *questionnaires*, like Schraw & Dennison's (1994) "Metacognitive Awareness Inventory" (MAI) for adults and Sperling et al.'s (2001) "Jr. MAI," for students in Grades 3-9. Learners are asked to answer survey items concerning their own knowledge, habits, use of learning strategies, and skill at predicting their level of performance; these are often conducted on paper (especially for efficiency of data-collection with large participant samples) or via verbal interview (as in Cross & Paris 1988).

Plenty of criticism has already been lodged against this method on multiple fronts. The MAI's 52 survey items lean very heavily on behavior as an indicator of cognitive activity (e.g., "I ask myself if I have considered all options when solving a problem," "I stop and go back over new information that is not clear," "I ask myself periodically if I am meeting my goals," "I draw pictures or diagrams to help me understand while learning," and so on), and are only marked as "true" or "false" (instead of allowing for gradation of agreement), which limits the total variability of the results across respondents. It also assigns the same scores to a student who a) answers affirmatively to survey items in a truthful report of their skills and a student b) who lacks the self-awareness required to recognize that they don't have the skills and knowledge they are attributing to themselves. In short, the same skill set that the survey is supposed to measure is presupposed by the survey method: students' responses only capture their "metacognitive awareness" accurately if students know their own metacognitive knowledge and behavior accurately enough to answer truthfully.

Especially when administered in settings where instructors have already made it extremely clear that students should strive to be more metacognitive, questionnaires that assess metacognitive ability are highly subject to an experimenter demand effect in which students will tend to overreport behaviors indicating use of metacognition. Basically: it's evident that students who mark more of the MAI questions as "true" will be judged as better students, so even if respondents are strongly encouraged to answer honestly, the social desirability of reporting that one is more metacognitive than one is will tend to produce inflated self-reports of metacognitive knowledge and behavior.

More generally, researchers who rely on questionnaires, as well as those who perform analyses of verbalizations in think-aloud protocols, have been criticized for not heeding the lessons of Nisbett & Wilson (1977), Ericsson & Simon (1980), and others who have pointed out the limitations of using verbal reports from subjects as data about their mental processes. As discussed in Chapter 2, a lack of transparent access to the kind of states one is being asked to report on may force respondents to attribute states to themselves, in order to explain their thinking to a third party; in cases where the evidence base from which to infer one's own mental happenings is lacking, respondents may resort to confabulation. Another objection to self-report-based methods is that, especially when administered on children, they are confounded by individual differences in reading or verbal ability that impact comprehension of survey items and/or production of verbal reports (Whitehead et al. 2009). Lastly, these methods have low validity – i.e., weak alignment with other measures, particularly those that take place in more naturalistic settings. Veenman and colleagues emphasize that "scores on these questionnaires hardly correspond to actual behavioral measures during task performance" (2006, p. 168).

In order to move away from self-report and verbalization, researchers have proposed

alternatives that measure observable behavior. For example, error-detection protocols seek to measure participants' ability to monitor their own comprehension by testing whether participants notice and point out errors deliberately embedded in learning materials; these have largely gone out of favor, since individual differences among subjects impact their likelihood of reporting errors – though some studies evade this issues by measuring nonverbal cues of comprehension checking activity, like eye movements & backtracking (Baker & Cerro 2000, pp. 104-107). Other protocols involve inferring metacognitive ability from portfolios of student work (Valencia & Place 1994) or student discourse about learning (Hennessey 1999). Lastly, observational methods aim to capture exercise of metacognitive ability in the classroom; for example, Whitehead et al. (2009) had teachers log instances of behaviors “theorized to represent metacognitive knowledge, metacognitive regulation, and emotional and motivational regulation” and subsequently rate each student on the frequency of these behaviors (from “always” to “never”).

What these methods gain in bypassing self-report & verbalization is counterbalanced with new problems. Behavioral measures are often time-consuming to administer, and thereby challenging to scale up to large populations. There are also residual issues with ensuring inter-rater consistency on measures involving judgments of student work and behavior, even when scales or rubrics are pre-defined. As a result, many researchers suggest that the best path forward is to conduct mixed-methods studies:

*That we are still far from having adequate tools for measuring metacognition is clear. One solution to the problem of measurement is to use as many methods as possible with each student. . . . Many investigators today do in fact use a combination of measures to obtain converging evidence. As [R.T.] White wrote, "Though each method is weak, the constellation of evidence from them will be more reliable and valid than each alone" ([1988,] p. 74). If different measures are used that do not share the same sources of error, and the same conclusions are reached, we can be more confident that we have measured what we set out to measure. The need for obtaining converging evidence is perhaps even greater in applied settings, where the stakes to the student are higher than it is in basic*



research. Recommendations to collect multiple measures occasionally appear in the literature for teachers and practitioners (e.g., Yochum & Miller, 1990), but not as often as they should. (Baker & Cerro 2000, p. 129; my italics)<sup>33</sup>

To summarize, the problems with extant assessment measures for metacognitive ability are well-documented, and fairly widely agreed upon among those who take these methodological issues seriously. Still, some researchers assert that “any method is superior to no method at all” – that it’s more important to make sure some form of assessment of metacognition is happening than to hold out for problem-free measures (Schraw 2000).

It’s been two decades since these assessments about the inadequacies of measures of metacognition were made; in the intervening years, self-report has remained the dominant method by which metacognition research is conducted (Craig et al. 2020). There seems to be some resignation that there’s no real way around self-report if metacognition research is to be conducted on populations large enough to establish large-scale patterns.<sup>34</sup> If that’s the case, so be it. The problem is that the weaknesses of these measures are only rarely communicated to education practitioners. As Baker & Cerro put it, “Researchers who attempt to translate research into practice have an ethical obligation to frame their recommendations responsibly, providing concrete information on the limitations of the measures;” unfortunately, some theorists persist in “uncritical presentation of measures with questionable reliability and validity” (2000, p. 134).

So, on the whole, it’s quite challenging to determine unequivocally how much metacognitive ability a student has, let alone to make a definitive judgment that they have gained metacognitive skill as a result of a pedagogical intervention. Even if there are quite a few studies that indicate a post-intervention increase in metacognitive ability, a high volume of

---

<sup>33</sup> See also Ericsson & Simon (1980), Garner & Alexander (1989)

<sup>34</sup> Even smaller-scale studies resort to self-report for the sake of efficiency. “[B]ecause of the time-consuming character of interviews, thinking aloud protocols, and simulated tutoring, questionnaires will continue to be used in metacognition research even though the shortcomings are evident,” write de Jager et al. (2005, p. 181-182), before disclosing their own choice to “[use] the questionnaire method for the measurement of metacognition.”

individual affirmative results for the efficacy of metacognitive training doesn't necessarily make for strong evidence that these methods are indeed effective. Due to the psychometric weaknesses of the measures used in these studies, we have much more warrant for the conclusion that students *think* they have increased in metacognitive ability as a result of metacognitive interventions – which we ought to expect anyway, given that students are often in a position to know that their instructor wanted them to gain metacognitive skill.

### **[3.2b] Analyzing meta-analyses**

Now that we've covered the issues surrounding assessment of metacognition in individual studies, we're in a position to consider how these issues can compound when research on metacognitive training interventions is aggregated into meta-analyses. Let's start by distinguishing the set of meta-analyses eligible for consideration – from which conclusions can be drawn about the effectiveness of metacognitive interventions in building metacognitive skill – from other types of research summaries with which they are sometimes confused.

First off, I'm not considering meta-analyses which demonstrate that students' metacognitive ability level is correlated with higher educational achievement, regardless of whether or not they have undergone metacognitive training. For example, Wang et al. (1990) counted metacognitive factors as a variable differentiating individual students, and found that engagement in metacognitive behaviors<sup>35</sup> was among the strongest “influences on learning” (judging by *correlation*, not causal effect) among a huge variety of variables ranging from socioeconomic factors to school culture. Some authors report this finding as if it validates the

---

<sup>35</sup> The metacognitive behaviors counted in the study included “comprehension monitoring (planning; monitoring effectiveness of attempted actions; testing, revising, and evaluating learning strategies), self-regulatory, self-control strategies (e.g., control of attention), and positive strategies to facilitate generalization of concepts” (Wang et al. 1990, p. 36).

implementation of teaching strategies that target metacognitive improvement: e.g., Wilson & Conyers (2016) cite Wang et al. in order to back the claim that “instructional strategies that emphasize metacognition . . . have a solid record of success, according to educational research” (p. 13). However, the studies Wang et al. aggregated examine metacognition as a static characteristic of students — not as an instructional variable. Since Wang et al.’s data doesn’t concern a relationship between metacognitive training *interventions* and improved student outcomes, their meta-analysis does not back the conclusion that metacognition-targeted pedagogy can be expected to produce better learning outcomes.

Secondly, there are plenty of research aggregations which have been used to argue that metacognitive interventions boost students’ *educational achievement*. The overwhelming majority of meta-analyses that have been conducted on metacognition in education fall into this category.<sup>36</sup> An oft-cited example is John Hattie’s (2009, 2018) synthesis of meta-analyses, which ranks various teaching methods categorized as “strategies emphasizing student meta-cognitive / self-regulated learning” as high-impact practices falling into a “zone of desired effects” — meaning that they make enough of a difference on student achievement to actually be worth implementing.<sup>37</sup>

However, meta-analyses of research on the relationship between metacognitive training and student achievement do not necessarily validate the success of training in actually making students “more metacognitive” or “better at metacognition.” In many primary studies on

---

<sup>36</sup> E.g., Haller, Child, & Wallberg (1988), Hattie, Biggs and Purdie (1996), Higgins et al. (2005), Donker et al. (2014), Tongerson et al. (2014), Perry, Lundie, & Golder (2019). Some of these use outcome measures of general academic achievement (GPA, standardized test scores, etc.) while others focus on achievement in specific areas of study (e.g. writing, reading comprehension).

<sup>37</sup> Hattie’s efforts seem to serve educators who want education research findings delivered through the binary filter of “what works” and what doesn’t. I should note that his methodology is quite controversial; see Bergeron & Rivard (2017), Slavin (2018), and Ashman (2018b). Their complaints are various but the general thrust is that meta-meta-analysis is a statistical minefield which Hattie fails to navigate successfully. His “effect size” measurements aggregate data that is so diverse that to average across it gives values that are basically nonsensical; consequently, comparing the effect sizes of various factors against each other doesn’t tell us anything we can reliably act upon.

metacognition in education, “the effects are often measured in terms of improved study results only ([e.g.] GPA and persistence at university) and there is much less attention for the question whether the skills trained have improved” (van Hout-Wolters 2000, p. 83). As Veenman, Van Hout-Wolters, & Afflerbach note, “in order to establish causal relations between metacognitive instruction and learning outcomes, . . . process measures<sup>38</sup> of metacognition need to be assessed in a pretest-posttest design” (2006, p. 9). In other words, researchers need to actually measure how metacognitive ability is impacted by the intervention and detect increased ability (hopefully using reliable measures); otherwise, they cannot conclude that a change in metacognitive ability accounts for improved learning outcomes.

As I will elaborate shortly (§3.3), it is not necessarily the case that improvements in students’ GPAs or standardized test scores following metacognitive training interventions are actually mediated by an increase in metacognitive ability brought about by the intervention; there may be peri-intervention factors (like changes in content presentation and/or subtle effects on how teachers grade student work) that influence or account for apparent gains in learning. At any rate, studies that measure the effects of metacognitive training on learning outcomes simply lack the research design required to defend the hypothesis that this training does in fact garner more or better metacognitive ability.

So, I’m only interested in analyzing meta-analyses that actually test how well pedagogical practices actually fare in accomplishing their stated objective of improving students’ metacognitive skills. Only *one* meta-analysis fits this criterion: technically, it’s just a subset of a larger study that also tracked the effect sizes of metacognitive interventions on student achievement.

---

<sup>38</sup> “Process measures” refer to measures that affirm that an intervention is achieving what it intended to achieve; these are contrasted with “outcome measures,” which assess the effect of the intervention on the high-level goals for the system in which the intervention is taking place.

Dignath et al. (2008) set out to answer two questions: “(1) Are interventions to foster self-regulated learning at the elementary school level effective generally? (2) What types of interventions are most effective?” (p. 104). Obviously, a big limitation here is that the studies included in this analysis only concerned interventions implemented in elementary-school classrooms; the authors explicitly state that “the results cannot be generalized to intervention with older students” (p. 121). Also notice that metacognitive training is being considered here as a subset of a broader category of interventions aimed at enhancing student self-regulation. The studies included in the analysis<sup>39</sup> collectively involved “promotion of cognitive, metacognitive, and/or motivational strategies and focus on strategy application or benefit” (p. 105): studies providing a metacognitive training intervention are a subset of this whole.

41 studies were characterized as providing “metacognitive strategy training;” these are subdivided into two groups based on the type of intervention offered: either *strategy instruction* (i.e., supplying students with explicit knowledge of strategies for planning, monitoring, and evaluating cognition) or “*reflection*” – subcategories of which are labeled “reasoning,” “knowledge about strategies,” and “benefits of strategy use” (p. 115). I’m honestly lost on what “reasoning” instruction is supposed to describe and how this qualifies as a form of metacognitive reflection – no decoding of these category labels is provided in the text. Similarly, instruction that conveys “knowledge about strategies” and “benefits of strategy use” to learners sounds more like an extension of strategy instruction than it sounds like what instructors typically mean when they describe an intervention as involving “reflection.”

Perhaps more importantly, the impact of these instructional methods on students’

---

<sup>39</sup> 48 studies (from 30 articles), spanning years 1992-2006, all of which implemented instructional interventions in classroom settings using longitudinal designs comparing students in the “treatment” condition (receiving the instructional intervention) to students in a control group. These studies were all conducted on “normal” student populations (i.e., studies conducted on students classified as “gifted” or “special education,” as well as those done exclusively on students with low-SES) were all excluded from the sample for the sake of ensuring homogeneity of participant characteristics.

metacognitive ability is measured exclusively in terms of “use of cognitive or metacognitive strategies”:<sup>40</sup> i.e., whether students engaged in (or at least reported engaging in) the use of specific tactics to manage their own learning (p. 106). Other facets of metacognitive ability (e.g., comprehension monitoring or error detection) are excluded from these measures, presumably because they are much more difficult to operationalize than strategy use. Note also that the authors could not pin down how metacognitive strategy instruction impacted the use of “metacognitive strategies” in isolation from cognitive ones:

. . . the use of cognitive or metacognitive strategies was not distinguished. Such a distinction would be problematic as on the one hand the application of cognitive strategies could also be interpreted as an indicator for metacognition. Metacognitive strategy application on the other hand is often not observable and difficult to register (Veenman, Van Hout-Wolters, & Afflerbach, 2006). Strategy application was measured by interviews or questionnaires (self-report) or via simulation tasks or observation. (Dignath et al. 2008, p. 106)

Here these researchers are acknowledging a problem raised in §3.1b: a particular learning strategy (e.g., creating a mnemonic device to enhance memory) can be classified as cognitive *or* metacognitive depending upon whether it is thought of as a lower-order cognitive process or as an orchestration of lower-order cognition (Brown 1987, p. 66). They choose not to litigate the metacognitive/cognitive boundary – but this means that they’re potentially looking at a set of skill measures that extend beyond the skills that were directly being targeted by metacognitive strategy instruction.

The effect size these authors report for the overall impact of metacognitive strategy instruction on (metacognitive *and* cognitive) strategy use is reported as .54 (p. 113). However, most permutations of different types of strategy instruction did *not* produce strategy use in excess of that demonstrated by students who underwent no metacognitive strategy instruction

---

<sup>40</sup> Other outcome measures provided in these studies and tracked in the meta-analysis included *academic achievement* (usually within a particular subject area in which strategy instruction was provided) and *motivational variables* (e.g., students’ levels of self-efficacy and attitudes towards learning) – but these are irrelevant to the determination of whether metacognitive training produces metacognitive skill gain.

at all (effect size .87; p. 115). Only joint instruction of planning and monitoring strategies or of planning and evaluation strategies produced larger effects on strategy use than the null condition (interventions without any metacognitive strategy instruction); confusingly, instruction in individual strategy types and instruction of all three strategies both resulted in a weaker effect on strategy use compared to no metacognitive strategy instruction.

Similarly, instruction in individual metacognitive “reflection” tactics did not produce a larger effect on strategy use than the null condition (no instruction in reflection: effect size .77); only joint instruction in reflection upon knowledge of strategies and the benefits of strategy use produced a stronger effect than interventions not involving any instruction in metacognitive reflection. But recall here that the outcome measure is the *use of cognitive and metacognitive strategies* – it’s not exactly surprising that instruction that reinforces what metacognitive strategies are why one should use them should result in heightened reporting of the use of metacognitive and cognitive strategies. (The authors also warn that these results “have to be interpreted cautiously” due to the small number of effect sizes used in the calculation process; p. 114.)

In their discussion, the authors note that strategy use (the dependent variable being tracked here as a proxy for metacognitive ability) was “mainly registered by the use of questionnaires” (p. 121). They also note that “effect sizes were highest for studies that evaluated the training effect by means of questionnaires,” and that questionnaires “often lack in validity” – implying that the high effect sizes they found for strategy use were potentially inflated due to the reliance on self-report in the primary studies. Had interventions been assessed using more accurate “multi-method analyses” (as they recommend), they’d expect more moderate effect sizes.

To summarize, this meta-analysis only concerns metacognitive training interventions (with a heavy emphasis on instruction in strategy use) among elementary-age children, and only measures the impact of this training on one facet of metacognitive ability — use of cognitive and metacognitive strategies — mostly through students’ self-reports. On top of all of these limitations to the generalizability of the findings, it’s far from obvious that metacognitive training resulted in more robust increases in strategy use than interventions that involved other strategies (cognitive or motivational training).

When this study is reported in Lai’s literature review, a table is provided showing an array of impressive looking effect sizes — all of which fall into the “medium” to “large” range based on Jacob Cohen’s recommendations on how to understand effect size values (1969, p. 23; cited by Coe 2002). When decontextualized from the actual study, these do indeed *appear* to signify that metacognitive interventions really do work. But having considered the mechanics of this meta-analysis, I submit that these values offer a highly misleading characterization of the evidence base for the success of metacognitive training in helping students of all ages develop metacognitive skill.

Long story short, there’s a paucity of high-quality research on how metacognitive training interventions impact metacognitive ability. There are plenty of studies that indicate a relationship between metacognitive training and academic achievement, but it’s a mistake to treat this data as an indication that metacognitive training actually produces an increase in students’ metacognitive skill level.

### **§3.3 | Explaining the impact of metacognitive training on learning outcomes**

I’ve expressed skepticism about the prospects for manipulating metacognitive ability through pedagogical interventions. I offered reasons to adopt a perspective on metacognition



accordingly to which our capacity to know and exert control over our own minds is much more limited than what is assumed by proponents of these methods, and I've contended that apparent increases in metacognitive ability are better understood as performative changes in outward behavior than as cognitive transformations.

Here's a pretty obvious objection: *If pedagogical interventions that target metacognitive skill don't actually improve students' metacognitive abilities, why do they remain so popular? Wouldn't it be obvious to instructors that these interventions aren't achieving their intended objective?*

My response is that these interventions remain popular because they can (and often do) lead to improvements in performance on learning tasks. If these methods are successful in helping students learn better (as divined from metrics like improved test scores, more satisfactory essays, GPA boosts, and so on), instructors may attribute this success to improvements in metacognitive ability that must have arisen from the metacognitive training interventions. Moreover, they may not feel a need to *verify* that students have become "more metacognitive" or "better at metacognition" (however they want to describe the skill-gain): given their commitment to the credo that better metacognitive skills translate to better learning, the proof of metacognitive skill-gain is in the pudding of learning gains.

The trouble with this pattern of inference is that the gains on learning metrics that follow metacognitive training interventions may very well reflect pedagogical changes that *accompany* metacognitive training, instead of resulting from metacognitive skill boosts. Let me illustrate this with an example.

John Tomsett, a high school teacher in the UK, bemoans his students' poor performance on their mock A-level exams, noting that "they know their economics theory, but under examination conditions they do not seem to have a sharp enough grasp of how to respond effectively to score as many marks as possible. Command words are ignored; diagrams are left

unlabeled; answers are expressed carelessly” (2015). He then implements a metacognitive training tactic promoted in the Education Endowment Foundation’s report on “Metacognition and Self-Regulated Learning” (Quigley et al. 2018). The tactic is for teachers to “model [their] own thinking” during instruction while completing the same tasks that students are supposed to master (ibid., pp. 16-17). For Tomsett, this translates to annotating the mock exam document with the thoughts *he’d* have while completing it himself, and then reviewing his annotations with students, “showing them just how alert and alive my brain is when I am being examined, teaching them how to think about their own learning more explicitly” (2015; see *Figure 3*). The annotations highlight key words and data within the case studies, articulate notes about how to compose written answers to guarantee maximal scores (e.g., “Do mention Airfix [the name of the fictional company in the case study] in the answer”), remind students to label the axes on graphs, and urge students “DON’T WAFFLE!” *ad nauseum* – presumably because waffling on one’s answer is anathema to A-level exam graders.

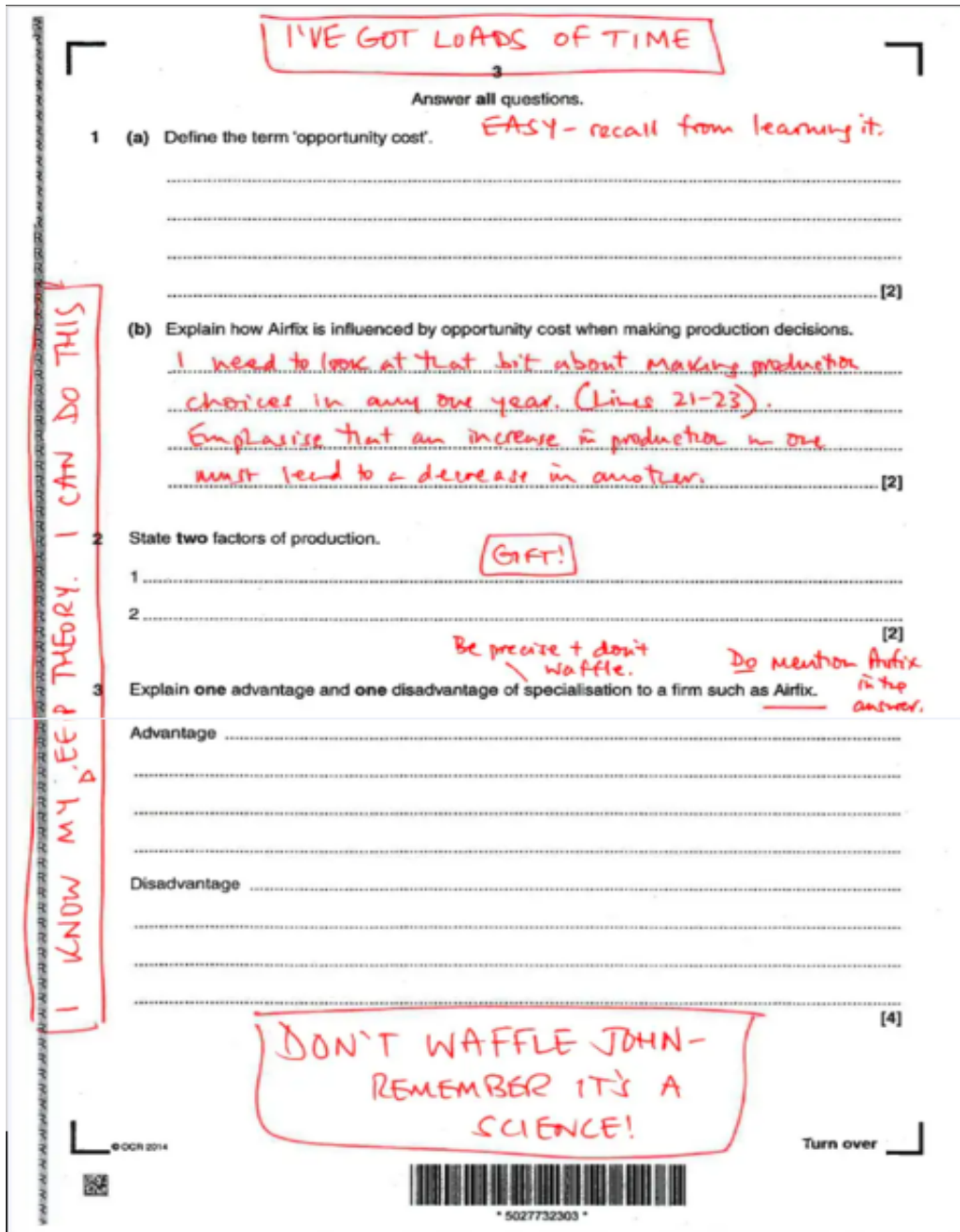


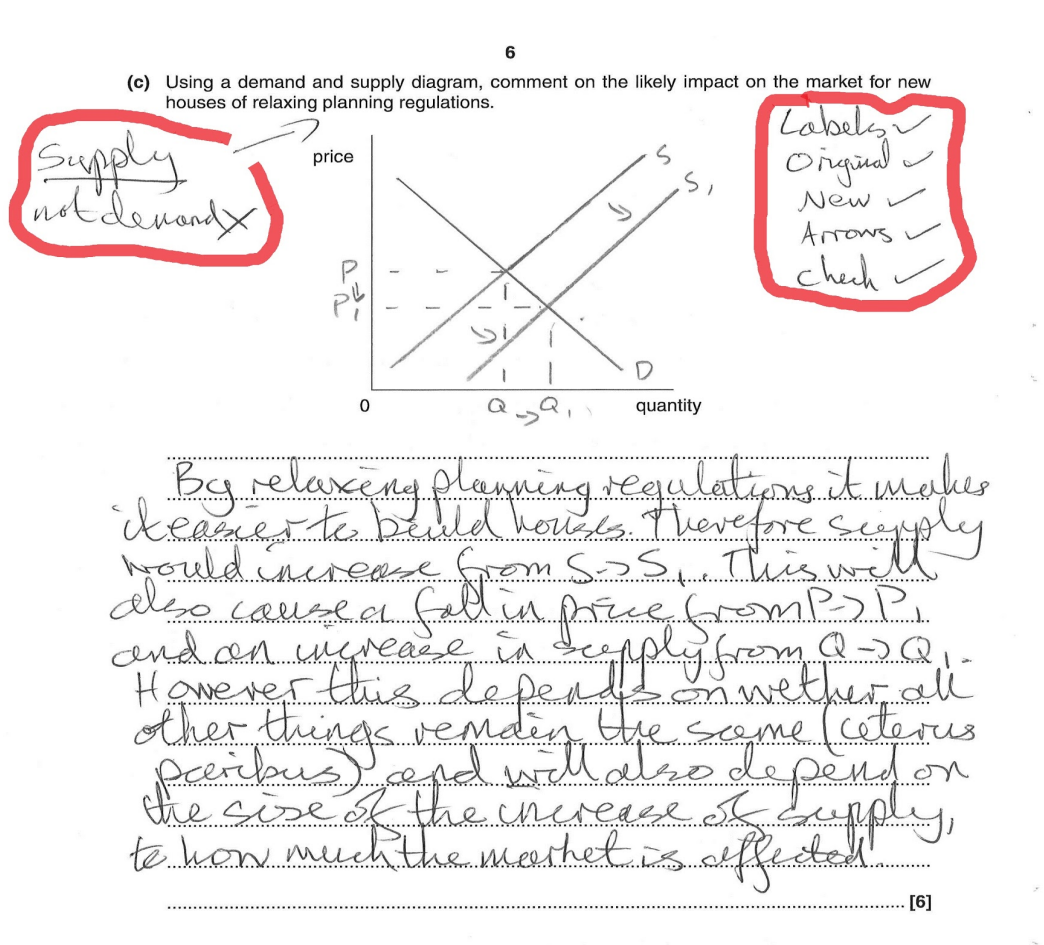
FIGURE 3: EXAMPLE OF EXPLICIT MODELING OF METACOGNITION, IN THE FORM OF ANNOTATIONS ON A PRACTICE EXAM. Shared by Tomsett (2015).

Tomsett reports that the metacognitive training intervention *worked*, because students' exam scores have skyrocketed. He expresses certainty that "explicit modelling/teaching of

examination room meta-cognition and self-regulation skills” was the teaching method responsible for their improvement, singling out one student as proof of principle:

The one student who I know for sure improved precisely because of his use of the meta-cognition and self-regulation intervention I modelled for him is Oliver. He went from getting 24/60 and a grade U in his first paper to getting 51/60 and a grade A in his second paper . . . Why am I so sure it was the intervention which helped Oliver improve? Well look at how he has made explicit on paper his mental checklist for ensuring he completed the question thoroughly . . . (Tomsett 2015)

...at which point Tomsett provides a screenshot of Oliver’s mock exam, demonstrating how Oliver has annotated his exam sheet much like Tomsett did (Figure 4).



**FIGURE 4: PUTATIVE EVIDENCE THAT IMPROVED STUDENT PERFORMANCE RESULTED FROM METACOGNITIVE SKILL-GAIN.** Student Oliver’s grade-A mock exam, purportedly evincing Oliver’s newfound metacognitive prowess (Tomsett 2015).

To me, it seems highly plausible that Tomsett's students' massive improvement on their exams has very little to do with metacognitive ability and very much to do with the fact that Tomsett's "explicit modelling" technique amounts to exposing students to trade secrets regarding *how to do well on this particular type of exam*. He describes this intervention as "teaching them how to think about their own learning," but the much more parsimonious explanation, given a lack of data bearing on how well these students are aware and in control of their own minds, is that Tomsett has offered them a procedure they can mimic in order to complete these exam questions in a way that will satisfy those doing the grading. He's shown them what they must do on specific types of questions to avoid losing points for failing to meet the requirements laid out in the rubric; he's made salient to them details they need to include in their writing in order to make sure their answers stay within examiners' good graces.

I think this example is demonstrative of a broader pattern surrounding "metacognitive training." When educators implement pedagogical practices meant to improve students' metacognitive skills, they're not holding other variables steady. In the upcoming subsection, I'll discuss how metacognitive training interventions tend to come part and parcel with changes in the information delivered to students.

### **[3.3a] Changes in domain-specific content knowledge**

Earlier, I've noted that many metacognitive training interventions often entail alterations in how and what content is delivered to students through instruction. For me, this raises a suspicion: perhaps the apparent effectiveness of some metacognitive training methods may have little to do with engaging students in metacognitive practice, and much more to do with the provision of domain-specific knowledge concurrent with these interventions.

It has long been observed that training in “critical thinking” skills is much more successful when situated within instruction focused on a particular domain of knowledge (e.g. math, science, history) than when it is offered in a domain-general way, abstracted away from any particular subject matter (Bailin et al. 1999, Willingham 2007). In Willingham’s words,

The processes of thinking are intertwined with the content of thought (that is, domain knowledge) . . . You can teach students maxims about how they ought to think, but without background knowledge and practice, they probably will not be able to implement the advice they memorize. Just as it makes no sense to try to teach factual content without giving students opportunities to practice using it, it also makes no sense to try to teach critical thinking devoid of factual content. (2007, pp. 9-10)

Recently, some education researchers have begun to emphasize that the benefits of metacognitive training are similarly context-sensitive. Metacognitive training conducted in the context of one academic domain don’t necessarily translate to gains in other disciplines; instead, “a pupil who shows strong self-regulated learning and metacognitive competence in one task or subject domain may be weak in another” (Quigley et al. 2018, p. 24). The explanation for the lack of interdomain skill transfer is that demonstrations of metacognitive ability depend upon prior acquisition of “sound content knowledge” (ibid., p. 17). These authors assert that “metacognition is specific to the task being undertaken and stronger where learners have a thorough grounding in subject knowledge,” citing as an example the finding that teaching students metacognitive strategies meant to aid reading comprehension only enhances students’ reading performance when students have adequate background knowledge to make comprehension possible in the first place (2018, p. 15; drawing from Willingham & Lovette 2014). The pedagogical upshot of this information is that general “thinking skills” programs, meant to instill cross-disciplinary boosts in metacognition, should be eschewed in favor of discipline-specific instruction concurrent with the construction of domain-specific knowledge (Willingham 2007; re-affirmed by Quigley et al. 2018, p. 14).

Veenman et al. similarly emphasize that the demonstration of metacognitive skill in any domain is contingent upon the possession of an adequate store of knowledge and command of cognitive activities within that domain:

It is very hard to have adequate metacognitive knowledge of one's competencies in a domain without substantial (cognitive) domain-specific knowledge, such as knowledge about relevant concepts and theories in a domain, about intrinsic difficulties of a domain, and about what is irrelevant . . . In terms of metacognitive skills, one cannot engage in planning without carrying out cognitive activities, such as generating problem-solving steps and sequencing those steps. Similarly, one cannot check one's outcome of a calculation without comparing the outcome with an estimation of it, or recalculating the outcome in another way.

If metacognition is conceived as (knowledge of) a set of self-instructions for regulating task performance, then cognition is the vehicle of those self-instructions. These cognitive activities in turn are subject to metacognition, for instance, to ongoing monitoring and evaluation processes. This circular process of metacognitive and cognitive activities makes it hard to disentangle them in the assessment of metacognition. (Veenman et al. 2006, pp. 5-6)

These caveats about metacognitive training may be presented as an affirmation of the “complex interplay” between cognition and metacognition in enabling successful academic performance (ibid., p. 17). But I think we should consider the possibility that the impact of metacognitive training interventions actually lies in the information transfer these interventions entail, rather than in the development of awareness and/or control over one's cognition. If the success of metacognitive training is contingent upon concurrent provision of domain-specific knowledge, we should at least entertain the possibility that domain-specific knowledge is the actual difference-maker in this equation.

### **(3.3aI) Pulling back the curtain on standards of evaluation**

I allege that metacognitive interventions are likely to manifest apparent learning gains when they entail heightened transparency regarding the expectations that student performance is being held to; accordingly, students are able to perform better by meeting these expectations of which they are newly aware. This pattern is blatant in Tomsett's description of the “explicit

modelling” intervention he used: Students underperformed when they didn’t know that failing to label axes would mean losses of points, and miraculously improved after being shown that labeling their axes was an absolute must.

For further illustration of this pattern, I’ll discuss David Concepción’s (2004) intervention to boost students’ metacognitive skills, which is deliberately interwoven with an accompanying intervention of drawing students’ attention to elements of “background knowledge” about philosophy (as well as the specifics of how his courses work).

Concepción accomplishes some provision of background knowledge through straightforward explicit instruction (e.g., assigning students to read a handout which clues them into which phrases often function as signposts for specific components of written arguments). But he also supplies background knowledge through the self-questioning protocol he administers to students. Most of the self-assessment questions students are supposed to ask themselves (in the interest of verifying “How am I doing?” while reading) make salient specific facets of a reading that they are obligated to attend to, as well as tasks that they are beholden to completing in the interest of executing a satisfactory evaluation of the text. Here’s an excerpted list of the questions Concepción encourages students to ask themselves:

[Questions to answer before reading:]

- *Is this a primary or secondary text? Should I expect an argument or a description of an argument?*
- *What is the focal statement of the article? Is there a thesis statement? What is it?*
- *What should I expect to find in the text in light of the title? . . .*

[After a quick full read:]

- *Have I identified the thesis statement and written it down?*
- *Do I know what the conclusion of the author’s argument is and have I marked places in the text where important steps toward that conclusion occur?*

[After re-reading to develop understanding:]



- . . . *Can I connect the dots? Can I explain in my own words why the author concludes what she or he concludes? . . .*

[While evaluating the text:]

- . . . *Do I think the arguments for the conclusions are persuasive? Why or why not?*
- *Can I think of any counter-examples to any assertion made by the author?*
- *Can I put my finger on exactly what bothers me about what the author says? Can I explain where and why I think the author made a mistake?*
- . . . *Is there a conflict between what I believe and what the author says? If so, to avoid being a hypocrite I must ultimately change my mind or show that the author's reasoning fails in some way.<sup>41</sup> Simply identifying a disagreement does not constitute an evaluation.*
- . . . *Have I looked for some point that the author did not consider that might influence what I think is true? (Concepción 2004, pp. 360-363)*

Many of these self-assessment questions function to ensure that students have taken in items of useful information and conformed to discipline-specific norms of engagement with texts.

Concepción characterizes these questions as a means to help students hone their ability to monitor and control their own cognitive processing while reading philosophy. But it's not clear whether any apparent gains in student performance that arise from the use of this intervention stem from the students' examination of their own thoughts through these questions. A simpler explanation for why these questions would help students to acclimate to philosophy quickly is just that these questions foreground and make explicit the standards against which students' performance is being measured, conveying valuable information to students about what their instructor is looking for when they turn in an evaluation of a text

---

<sup>41</sup> As an aside, I find this standard metaphilosophically questionable. He offers a zero-sum image of disagreement, in which "to avoid being a hypocrite I must ultimately change my mind or show that the author's reasoning fails" (2004, p. 362). Unless I am misunderstanding what he means by "chang[ing one's] mind," and he intends this in a way that includes modulating confidence in one's standing beliefs, this seems to be an overly Manichean, on/off-picture of how belief works, and one which could easily turn off those who are new to philosophy. This portrayal of what engagement with philosophical texts involves suggests, for instance, that a student is a hypocrite for continuing to believe that they are morally responsible for their actions after reading a strong defense of incompatibilism and coming to believe that our world is, in fact, deterministic. By this standard, practically all students and most professional philosophers are "hypocrites," for failing to believe in accordance with persuasive argumentation.

they've read. There's really no need to invoke the development of "metacognitive skill" to account for why students are able to turn in work that their instructor deemed more satisfactory after this intervention: Students are better able to meet expectations when those expectations are lucid and publicly known than when the criteria on which they are being evaluated are a mystery.

To summarize, I'm alleging that what instructors may describe as students "being metacognitive" may amount to little more than applying subject-matter knowledge in the right way and at the right time while completing essays, exams, and so on; the extent to which students are monitoring and/or orchestrating their own cognition is not being measured, and so there's no guarantee that these metacognitive processes are underlying students' strategic use of the knowledge they've acquired. In short, the apparent gains in learning wrought by metacognitive interventions can be explained as a consequence of students knowing what standards they're being held to during assessments, and completing these tasks in accordance with their knowledge of what will lead to being rewarded with good grades. What's really making the difference in student performance is knowledge of the norms they're being held to, not their intrinsic ability to self-regulate in accordance with these norms.

### **[3.3b] Learning as a moving target**

John Issitt proposes that when "metacognition" is presented as a revelation from scientific inquiry into "how people learn" (2007), this suggests it's simply a matter of fact that *metacognition is necessary for successful learning* — that this is just how learning works.

Emphasizing the scientific pedigree concept thus tends to obscure the role of our own volition in shaping what we understand "learning" to be: that we have *a choice* about how to understand the normative ideal we are striving for when we advocate for good education.

Educators may not be aware that in endorsing the importance of metacognition in education, they are selecting a particular manner of understanding of what learning *is*, which in turn influences their judgments about when students have succeeded at learning. I contend that instructors who choose to implement metacognitive training methods tend to tacitly adopt a view of learning wherein “being metacognitive” is built into their vision of student success, such that “the ideal student” becomes one who not only performs well but is overtly metacognitive while doing so.

In other words, subscribing to a pedagogical outlook on which metacognition is highly valued will tend to shift one’s standards. An instructor may implement pedagogical methods that aim to improve students’ metacognitive skills, see subsequent gains in student learning, and attribute these gains to students’ enhanced metacognitive ability. But what might actually be happening is that the pedagogical intervention they’ve implemented is accompanied by a shift in their criteria for “learning,” so that students who behave in accordance with the metacognitive training intervention (i.e., being demonstrative about thinking about thinking during class discussions) are rewarded by being viewed (and potentially graded) as more successful learners.

Yet another way to restate the same point is as follows: Teachers who undertake metacognitive training interventions are inadvertently “teaching to the test.” It may seem as though engaging students in metacognition is boosting their performance on independent measures of learning achievement, but in reality, the metrics on which students will be judged are informed by the instructor’s conviction that learning achievement involves metacognition. The manipulation and the measure are tied up with one another in such a way that the intervention is guaranteed to produce the *appearance* of learning gains, but “learning” has been

made artificially achievable by equating it with the *means* chosen to pursue that end (i.e., “being metacognitive”).

This argument can be generalized to any pedagogical intervention – not just those targeting metacognition. I’ll describe this pattern as the *intervention-uptake halo effect*, since it involves developing a positive gestalt impression on the basis of favorably evaluating a person in one domain, which then feeds into further positive evaluations. Whenever an instructor elects to enact a particular intervention, they presumably *want* that intervention to achieve its objective; otherwise, they wouldn’t have bothered to enact the intervention at all. As a result, instructors are motivated to look for evidence that students’ behavior has changed as a result of the intervention, to confirm that their efforts haven’t been for naught. When they do find evidence that students are acting in accordance with the intervention, the positive feelings they get from these signs of student “success” in absorbing the intervention then translate to an impression that the student is *engaged* and doing well as a learner.

For example, let’s return to Figure 4. Tomsett wants to see that his explicit modelling (of his thinking about what a written answer needs to include to be complete and judged positively by A-level examiners) has helped students become aware of these factors while completing their own exams; accordingly, when Oliver writes these factors out in a checklist next to the appropriate exam question, Tomsett is understandably pleased that Oliver has indeed absorbed the message of the modelling exercise. Having successfully conformed to his instructor’s expectations, Oliver is now in a position to be judged as a *good learner*.

This phenomenon recalls an amusing entry into the annals of psychology: the story of Clever Hans (Colman 2008). Retired math teacher Wilhelm von Osten claimed to have taught his horse Hans to do arithmetic, conducting demonstrations where Hans would solve math problems by tapping out the correct number in hoofbeats. By all accounts von Osten had

achieved a pedagogical marvel by manifesting human intelligence in a nonhuman animal — until skeptical psychologist Oskar Pfungst determined that “when no one within its visual field knew the right answer to the question being asked,” Hans couldn’t display his apparent cleverness (ibid.). Pfungst determined that Hans was able to give the appearance of supplying the correct answer to these questions by reading the body language of those watching him: “it began tapping whenever people standing nearby adopted expectant postures and turned their attention to its hoof, and it stopped as soon as the onlookers (probably unwittingly) gave anticipatory head and eye movements when the correct number of taps was reached.”

What von Osten and impressionable spectators attributed to mental operations indicative of remarkable intelligence, we can instead chalk up to an ability to read social cues and conform to other people’s evident desires. Much like Clever Hans did, savvy students can attend closely to an instructor’s signals of what they’d like to see in an ideal student performance, do what is in their power to offer the appearance of satisfying these desires, and thereby earn the instructor’s approval. And this need not be a dishonest ruse from those of the scheming, shortcut-taking set — a drive to earn the goodwill of authority figures could foster this behavior even in the most conscientious of students. When instructors make it clear that they want students to be metacognitive, then they are likely to encounter displays of behavior from students that lead them to affirm that students are indeed being metacognitive — but here they are mistaking students’ efforts to please for the achievement of their pedagogical aspirations.

### §3.4 | Summary

I began this chapter by recounting assertions that the available evidence from education research points to a pedagogical imperative to enact interventions aimed at improving students’

metacognitive skills. I've presented my case against the existence of such an imperative through a multi-pronged approach.

I began by offering critiques of each of the three major approaches to cultivating metacognitive skill through deliberate pedagogical interventions. I've drawn out obstacles and limitations to the efficacy of explicit modeling and reflective activities, as well as the difficulty involved in verifying that these interventions are proceeding as intended (instead of, say, involving or invoking confabulation about one's cognitive activity). I also addressed the ease with which the execution of specific behavioral programs can be shoehorned into "metacognitive training" simply by designating these activities "metacognitive strategies."

Next, I tackled the empirical record for metacognitive training interventions. I outlined unresolved methodological issues with the measurement of metacognitive skills and explained how that difficulty manifests in a tendency to avoid collecting pre-/post-intervention data to assess the direct impact of these interventions on students' metacognitive skills. Even when this data is collected and synthesized in a meta-analysis (as in Dignath et al. 2008), the trustworthiness of this data is suspect, due to its reliance on subject's self-report about their own learning behavior (specifically, how frequently they use metacognitive strategies). There are also open questions about the generalizability of this study's conclusion that metacognitive training is indeed effective: e.g., were these interventions successful only within this specific population, and are all training methods equally efficacious?

Though researchers suggest that subsequent gains in students' academic performance indicate that these interventions have in fact improved metacognitive skills (Perry et al. 2019), I counter that it's inappropriate to conclude that these learning gains must have been mediated by enhanced metacognitive ability. In fact, this recalls a long-standing objection to the construct

of metacognition – that it simply gives a new name to *the X-factor which accounts for why some students perform better than others*:

It is gratuitous to attribute performance variation to levels of metacognition when it can only be inferred from performance itself. Attributions to good or poor metacognition when it is not measured independently are simply attributes [sic] to the “ghost in the machine” and add no psychological explanation. (Paris & Winograd 1990)

In other words, if we don't bother to measure metacognition but allow ourselves to infer its presence from students' academic success, then invoking “metacognitive skill-gain” as the reason why students' scores improve after undergoing these interventions is circular. It also entails positing a mysterious, untouchable quantity as the underpinning of academic achievement – making the role of “metacognition” in education akin to that of phlogiston in early efforts to explain combustion.

Lastly, I expressed doubts about the extent to which one can attempt to manipulate metacognition without impacting confounding pedagogical variables, particularly 1) students' domain-specific knowledge in the area of instruction and 2) instructors' vision of student excellence. I've argued that the well-documented correlations between metacognitive training interventions and improved measures of students' academic performance can be accounted for as a result of the insight these interventions incidentally provide regarding what teachers want to see students do in performances they'll deem proficient. Moreover, the effort to enhance metacognitive skills often stems from a conviction that metacognition is a mark of intellectual achievement – in which case any assessment of student learnings that rest upon instructors' subjective impressions are likely to implicitly enhance instructors' evaluations of students who are demonstrate uptake of metacognitive training interventions. As I see it, these factors offer a much more straightforward explanation for why these interventions “work” to improve

educational outcomes, compared to the alternative of ascribing metacognitive skill gain to students on the basis of unreliable measures that rest heavily on self-report.

I'm not disputing that metacognitive training can do a lot of good for students. Dedicated metacognitive interventions can certainly help students reach higher levels of attainment than they would have otherwise. I just am not convinced that the pathway from metacognitive training to enhanced student outcomes is mediated by the mental transformations that these interventions are claimed to bring about.

In my final chapter, I'll explore reasons why the metacognitivist narrative of the malleability of metacognitive skills and their responsiveness to instructors' efforts would prove compelling to many educators.



#### 4. AN ERROR THEORY FOR METACOGNITIVISM

In this final chapter, I advance a theory as to why it's widely believed that metacognitive training "works," even though the evidence for this is murky. In Chapter 3, I explained that much of the research on metacognition doesn't go through the trouble of verifying that metacognitive training actually modifies metacognitive abilities, but rather infers an increase in metacognitive skill from improved learning outcomes. Certainly this pattern of inference can account for some of the conviction that metacognitive training is effective – but I think there are certainly additional factors that collectively bolster belief that metacognitive skill can indeed be modified through pedagogical interventions.

First off, I'll argue that metacognitivist rhetoric in pedagogical discourse is often hazy on what constitutes successful integration of metacognition into one's teaching practices. When metacognition is introduced to educators as a concept central to good pedagogy, these discussions commonly skirt acknowledgment of the difficulties involved with measuring metacognitive ability, and fail to distinguish between exercise of metacognition and enhancement of metacognitive skills. As a result, educators may believe that they are developing students' metacognitive abilities as long as students show uptake of the interventions they have implemented to promote metacognition during learning. Effectively, educators end up shifting the goalposts for successful metacognitive training to encompass the means by which they attempt to promote metacognitive ability, instead of tracking distinct metrics for metacognitive ability.

Secondly, I'll suggest that motivated reasoning – wherein "people are more likely to arrive at those conclusions that they want to arrive at" (Kunda 1990, p. 495) – provides a useful framework from which to understand the popularity of pedagogical practices that aim to

enhance students' metacognitive skills. In recent years, motivated reasoning has implicated as the key to understanding the formation and maintenance of beliefs about science and politics;<sup>1</sup> I think this explanatory framework is overdue to be applied to decision-making in education.

I submit that educators are likely to *want* it to be true that some simple instructional interventions can set students on a path that culminates in mental self-mastery; as a consequence, they are prone to interpreting available data as sufficient evidence that students' metacognitive abilities can indeed be enhanced through metacognitive training. I'll outline some reasons why, on top of their occupational commitments to believing that minds can be molded through instruction, educators are predisposed to believe that it's possible to cultivate students' metacognitive abilities through simple pedagogical interventions.

#### **§4.1 | Shifting the goalposts for metacognitive skill-building**

In the previous chapter, I noted that it's common for research on metacognition in education to collect data on learning outcomes following metacognitive training interventions, but forego collecting process measures — that is, data that would affirm that these interventions that are supposed to build metacognitive skills actually produce that intended effect (Veenman et al. 2006). But the vast majority of efforts to implement metacognitive training in classrooms occurs outside the scope of formal pedagogical experiments. Much of the time, instructors are free to form their own subjective impressions of whether or not metacognitive training is bearing fruit for their students.

I hypothesize that instructors are likely to treat students' uptake of metacognitive training interventions as a proxy for metacognitive skill-building. That's because what students do in acting in accordance with these interventions — like using dual coding during note-

---

<sup>1</sup> E.g., Kahan, Jenkins-Smith & Braman (2011), Flynn, Nyhan & Reifler (2017)

taking, or going through a mental checklist while taking a test — is typically indistinguishable from what the instructor would expect to see from a metacognitively-adept student. So as long as students are doing what instructors ask them to do in the service of cultivating metacognitive ability, instructors may interpret this as evidence that students have in fact developed metacognitive ability — confusing the intervention with its intended effect.

In forming this hypothesis, I've drawn heavy inspiration from two philosophical critics of metacognition, each of whom has elaborated ways in which the standards for "metacognition" shift in response to pressures for educators to manifest this elusive ideal. I'll explain each of their views briefly, before explaining how I think a similar pattern of goalpost-shifting also leads to the impression that teachers' efforts have resulted in metacognitive skill gain.

#### **[4.1a] Precedents for goalpost-shifting around "metacognition"**

Felicity Haynes (2014) and John Issitt (2007) have suggested that those who champion metacognition as a pedagogical concept tend to exploit the slipperiness of this concept by adjusting their standards as to what "metacognition" is in order to align this with what they are able to accomplish through pedagogical interventions. In short, these authors suggest that instructors' impressions that they have succeeded in cultivating metacognition in their students result from shifting the goalposts to make "metacognition" artificially easy to achieve.

##### **(4.1aI) Substituting low-hanging fruit for metacognition**

Haynes explains that some teaching materials recommended for instructors of Philosophy for Children (P4C)<sup>2</sup> equate critical thinking instruction with "develop[ing] children's

---

<sup>2</sup> "Philosophy for Children" began as an initiative by philosopher Matthew Lipman (with heavy influence from John Dewey), closely linked to the 1974 founding of the Institute for the Advancement of Philosophy for Children at Montclair State University. In the half-century since, P4C has not only expanded to offshoots affiliated with other universities, but has evolved into a decentralized, tentacular, international phenomenon. Some advocates use "P4C"

metacognitive abilities” (2014, p. 131). But she observes that the meaning these materials ascribe to “metacognition” shifts depending on context: specifically, depending upon whether it is invoked as an abstract ideal, versus as an attempt to manifest that ideal through instruction. When it is discussed as the goal of pedagogical intervention, “metacognition” is described as the awareness and control of one’s own cognitive operations, involving abilities “to self-monitor and articulate their daily thinking processes, know when a sub-goal has been obtained, detect errors and recover from errors either by making a quick fix or by retreating to the last known operation” (ibid.). But when instructors try to actually invoke metacognition through pedagogy, they often settle for an inferior substitute for what they were aiming at.

She reports that efforts to help children engage in metacognition typically devolve into boilerplate lessons in content acquisition, where the content to be learned is a taxonomy of cognitive processes. Students will be said to be engaging in “metacognition” if they are memorizing vocabulary words with which they can discuss sub-processes of thinking. While the latter may technically qualify as “thinking about thinking,” it doesn’t amount to the higher-order or self-reflective thinking that this phrase invokes, and it fails to “help children think for *themselves*, creatively as well as critically” (ibid.).

Whereas proponents of metacognitive training assume that “students must understand the language of thinking in order to facilitate their own cognitive growth” (p. 139), Haynes objects that if supplying the “language of thinking” to students amounts to forcing memorization of terminology about cognition, this is a far cry from the method’s objective of empowering students to become self-aware, self-regulating learners. In other words, those who promote metacognitive training don’t uphold their own purported ideal of transferring

---

to refer to a specific pedagogical program centered on “thinking skills,” such as that administered by teachers trained by the UK’s Society for the Advancement of Philosophical Enquiry and Reflection in Education (SAPERRE); others use it more loosely to characterize any philosophical engagement with pre-college students.

responsibility for one's own learning to the learner, so that said student can gain the ability to autonomously self-regulate their cognitive activity.

However, even if metacognitive training is conducted in a way that induces introspection as a means of attaining knowledge of oneself as a thinker, Haynes rejects the idea that "self-reflection about the [cognitive] skills one uses, as is advocated by the metacognitivists, necessarily increases skills by making a given structure [of cognitive architecture] explicit" (p. 138). In short, Haynes suggests that there is a gap between gaining *self-awareness* and gaining *command* over one's cognitive functions, and that the metacognitive training methods popularized as tools for promoting the latter do not necessarily satisfy that objective.

In summary, Haynes provides evidence of a particular pattern of slippage in the referent of "metacognition," where two very different renderings of "thinking about thinking" get conflated with each other, and/or treated as causally related, such that promoting one necessarily promotes the other. Specifically, *absorbing a definition for the mental process of classification* gets treated as if it is a means to *engaging in higher-order analysis of one's own lower-order cognitive activity* – even though the relationship between the two is unclear at best, and nonexistent at worst.

I'll call this the *substitution critique*, because the crux of the complaint is that by labelling a pedagogical method as "metacognition," one can perform a kind of sleight of hand where pedagogical low-hanging fruit (like teaching students vocabulary about cognition) can suddenly be presented as astonishing interventions wherein students gain insight into and control over their own minds. Educators can earn praise for manifesting "metacognition" even when what they're actually enacting is a sorry excuse for achieving the lofty intellectual ideals which that term conjures up.

#### **(4.1aII) Substituting visible thinking for valuable thinking<sup>3</sup>**

Haynes highlighted the interchangeability of different translations of “metacognition” as a means by which one may convince oneself or others that they have pulled off a pedagogical feat. Issitt, on the other hand, proposes that instructors do the same by neglecting any distinction between genuine metacognition and the mere performance of thinking about one’s thinking.

Issitt (2007) proposes that “metacognition” is a covert tool of social control, the function of which is to induce learners to engage in self-surveillance, in order to police their own thinking within narrow bounds of acceptability. (I don’t endorse this assessment, which stems from a Foucauldian analysis of the rhetoric around metacognition in education.<sup>4</sup>) At any rate, he notes that one consequence of positioning “metacognition” as a pedagogical ideal is that students become conditioned to take part in a “contrived performance” of insight into their own minds. Under the value system that metacognitivism advances, the operative factor in students’ academic success is their willingness to “play the game” of engaging in reflection and feigning self-discovery when prompted:

We can imagine two students of, say, physics, one of whom ‘buys into’ education, enjoys the competitive process, realizes that reflection is required as part of this and so reflects upon and adjusts their performance in the light of the requisite criteria and succeeds, while the other comes to see this reflection as but another step in a contrived performance over which they have no control, and so, disillusioned and demoralized, does not ‘succeed’. We may . . . be prompted to

---

<sup>3</sup> “Visible thinking” is a reference to the metacognitivist pedagogical approach defended by Ritchhart, Church & Morrison (2011), according to which instructors ought to prompt students to make their thought processes explicit (in writing, speech, diagrams and so on) in order to treat students’ intellectual activity as the object of educational molding.

<sup>4</sup> Issitt is quite convinced that policy-makers are deliberately using their power to advance a normative agenda under the guise of objective educational science. I’m not comfortable attributing any such deliberate intent to policy-makers. My impression is more that *all* those involved in the institution of education, regardless of their positioning within any strata of power relations, mutually contribute to the generation of a set of normative ideals that tend to go unnoticed because they are taken for granted.

In other words, there need not be ill-intent on behalf of those in power for a situation to emerge in which educators are assuming and advancing particular visions of good education and eschewing others — all the while believing that what they see as “good education” *isn’t* a choice, but rather the simple, scientifically-revealed reality of what education ought to be.

think that the solution [to the second student's failure] is better preparation for the exercise of metacognition. But we may also come to the view that there is something phony about this contrived performance and that an overemphasis on metacognition is part of the problem. (2007, p. 391)

In other words, metacognitivism overvalues *displays* of thinking about thinking, rewarding students for the mere appearance of being metacognitive, while the extent to which this activity genuinely creates self-knowledge or enables self-regulation becomes a mere afterthought.<sup>5</sup>

What Issitt is describing here, is shifting the goalposts for “metacognition” to encompass the *façade* of thinking about thinking, or as I’ll call it, “thought theater.” As I noted in Chapter 1 (§1.3bIII), one axis of ambiguity of “metacognition” concerns whether this construct captures a success term (i.e., thinking about thinking done *well*), or whether one can be “metacognitive” even if one is thinking about thinking inaccurately or platitudinously (see also Fisher 1998). Issitt is suggesting that the pressure placed upon educators to actualize this ideal means that educators will end up trading on this ambiguity, treating the latter as if it represents the intellectual achievement connoted by the former. In other words, if “metacognition” is aggressively marketed as a tenet of quality education, this doesn’t mean that educators will raise their standards for student behavior to meet this high bar; rather, educators are much more likely to brand mediocre or fabricated thinking about thinking as “metacognition” than

---

<sup>5</sup> I have made similar points in Chapters 2 and 3, in arguing that it is naive to assume that students’ self-reports about their own thinking are veridical, especially in cases where instructors place a demand on students to be demonstrative about thinking about their own thinking by communicating to students their belief that successful learning involves metacognition.

See also Greg Ashman’s satirical blog *Making Learners Extraordinary* (<https://makinglearnersextraordinary.com>). In these posts, Ashman presents as the “Extraordinary Learning Foundation™” is a fictional purveyor of pedagogical products and methodology. The basic thesis of the blog is that calling a pedagogical approach extraordinary doesn’t make it so; from his perspective, much of edu-speak is merely hot air – and metacognitivism is a recurring target of his sardonic skewering. He takes aim at those who insist upon placing “metacognition” at the top of a hierarchy of thinking that students must ascend; he demonstrates that it’s certainly not always the case that “thinking about one’s own thinking” means that one is thinking critically, sophisticatedly, or in accordance with any other normative standards to which we collectively desire for learners to conform.

In particular, I recommend the posts entitled “Meta-active thinking,” “Talk-it-out™,” “Mind tools for the whole child: Think-it-out™,” “Actionizing thinkiness,” “Thinkiballs,” “Developing a thinkiness scale,” and “The metaquiry accelerator.”

they are to admit (to themselves or to others) that many of their students fall short of this lofty objective.<sup>6</sup>

For an example of thought theater in action in pedagogy, let's return to David Concepción's defense of metacognitive training in philosophy instruction. Among other interventions, he claims to encourage metacognition with the following activity:

. . . students are *required* to pass notes to each other in class. At the end of class, each student must have contributed at least one question or answer to a written dialogue that took place in note passing during class. To receive credit, students must be on the lookout for material that they do not sufficiently understand and write a question or answer regarding it. In other words, students are given credit for being metacognitive during class. (p. 356)

According to Concepción, this exercise gives students "credit for being metacognitive." I would counter that students are being rewarded for presenting their instructor the *appearance* of thinking about their own thinking, examining their own understanding, and so on: Requiring students to write a question or answer in order to get credit in one's course has no guarantee of prompting students to self-report accurately or to think rigorously; instead, it may prompt students to confabulate about their own cognition — that is, to unknowingly invent narratives about the contents and workings of their own minds.

In defense of this method, Concepción mentions a pair of students who, through note-passing, "independently discovered the 'Euthyphro' question about piety, moved to a discussion of God's attributes in an attempt to resolve it and finally discussed the problem of

---

<sup>6</sup> This is roughly equivalent to Kevin Possin's reasoning as to why it's counterproductive to define "critical thinking" as metacognition (2008, p. 203):

. . . metacognition is too easily construed as merely reflecting on what one happens to be thinking or experiencing at the time, with little substantive guidance as to how to do so, other than with, *e.g.*, open-mindedness, fair-mindedness, clarity, precision, depth, breadth, and logicalness . . . or by means of knowledge acquisition, comprehension, application, analysis, synthesis, and evaluation . . . Using the metacognition account, my university all too easily approved a dance class as a [critical thinking] course, because its students self-reflect on their bodily movements. (p. 203)

In short, critical thinking entails conformity to particular standards of intellectual excellence (regards to the weighing of evidence, evasion of informal fallacies, analysis of arguments, and so on), whereas the many phenomena that can be classified as "metacognition" are not necessarily beholden to standards of rigor, veridicality, etc.



evil”; he contends that “for two students to spontaneously generate an in-depth, on-task conversation in their second week of college is no small achievement” (ibid., p. 357). This is a heartening report of the best-case scenario from this assignment, but it’s also anecdotal – for all we know, this was a rare instance of philosophical insight among a maelstrom of notes completed hastily for mere box-checking. More importantly, this tale of “success” doesn’t actually demonstrate that these students were being *metacognitive*, per se: their spontaneous discovery of the Euthyphro problem seems much more obviously a consequence of thinking *about the course material* than it is an instance of monitoring or controlling their own cognitive processes. To credit this victory to “metacognition” is, at the very least, to stretch the meaning of a term that’s awfully flexible to begin with. But it conforms to the pattern Issitt predicted, where educators prompt students to offer proof that they are thinking about their own thinking while learning, without distinguishing between authentic and inauthentic fulfillment of the demand.

#### **[4.1b] Substituting metacognition for metacognitive skill-building**

I’ve just recounted examples where playing fast and loose with the meaning of “metacognition” allows educators to claim that they have manifested this desired educational ideal. I expect that a similar pattern of goalpost-shifting fosters instructors’ impressions that they have succeeded in enhancing students’ metacognitive skills by implementing new teaching practices aimed at prompting the exercise of metacognition during learning activities.

A contributing factor here is that texts which communicate the importance of metacognition to pedagogy to instructors rarely mention *assessment* of metacognitive skills as something instructors should concern themselves with – and they certainly don’t address the inherent methodological difficulties surrounding measurement of this particular psychological

construct (discussed in §3.2a). Metacognitivist rhetoric doesn't dwell on details about how metacognitive ability should be measured; it's much easier to find recommendations for how to encourage metacognition in the classroom than to encounter discussion around how to tell that these activities are transforming students on a psychological level. I suspect that this means that instructors aren't primed to look for measures of metacognitive ability *beyond or apart from* evidence that the instructor has succeeded in enacting their intervention (e.g., signs that students are making use of a particular metacognitive strategy they taught, or are practicing habits like self-questioning).

Yet another factor that may dispose educators against seeking evidence for the efficacy of metacognitive training interventions is that presenting "metacognition" as emanating from education science bolsters the impression that it has already been extensively verified. An unfortunate side-effect of a climate wherein educators are under considerable pressure to enact pedagogical practices that are "evidence-based" is a tendency to exaggerate the conclusiveness of research from education science, so that this information may be regarded as pedagogically actionable.<sup>7</sup> Summaries of the research on metacognition in education – for example, Bruer's assertion that "the research . . . shows that it is possible to teach children metacognitive skills and when to use them" (1993, p. 72) – are written for educators who want clear answers as to what the "best practices" are and what they are expected to do, and accordingly they tend to convey promising findings as though they are very much settled.

Lastly, confirmation bias might play a role here. Typically, an instructor only goes through the trouble of enacting a pedagogical intervention if they want to realize its expected results. The more they want the intervention to be effective, the less likely it is that they would

---

<sup>7</sup> See Slavin (2002) for an expression of the view that educators must implement research-backed practices in order to bring education in line with fields like medicine and agriculture. More on this to come in §4.2bIII.

go out of their way to seek out or entertain evidence to the contrary. In §4.2, I will dig deeper into exploring the influence of educators' commitments and desires in shaping their engagement with evidence that would bear on the effectiveness of metacognitive training.

Meanwhile, educators have reason to take pride in successful implementation of an intervention as a pedagogical accomplishment in itself. This would be true for any kind of intervention: anyone who has taught anything knows that just seeing that students are doing what you told them to do — an affirmation of the impact of your efforts — is a pedagogical victory, compared to the demoralizing alternative of finding that what you've taught has gone in one ear and out the other, so to speak. Teachers are invested in making a difference for students, and as such their sense of self-worth may hinge upon receiving ongoing validation that students have been receptive to what the instructor has gone through the trouble to teach.<sup>8</sup>

If anything, we should expect that instructors will be even more likely to treat the mere uptake of a pedagogical intervention as a significant accomplishment, in the case where the content of the intervention is the subject of idolatry in educational circles — which is precisely the case for interventions that prompt students to exercise metacognition during learning activities. In other words, since “metacognition” is so celebrated in its own right, educators may shift the goalposts for success in metacognitive training such that merely getting students to “be metacognitive” in the classroom is a win, regardless of its downstream effects on the much less tangible objective of enhancing metacognitive skills.

---

<sup>8</sup> This is, admittedly, likely to be a hard pill for instructors to swallow, since it paints them as easily swayed by the simple flattery involved in seeing students actually do what one has told them to do. In other words, it suggests that instructors are highly sensitive to the degree to which students are complying with instructions, to such an extent that cooperative students who follow directions and play their part in carrying out interventions are better liked by their teachers and judged as better learners. Nevertheless, I stand by it as much more realistic than the alternatives — that teachers are wholly indifferent to whether or not students behave in accordance with their carefully selected interventions, or that they can exercise a saintly level of impartiality when assessing students who cooperate with their pedagogical methods to varying degrees.

That there is so much hype over the mere occurrence of metacognition in the classroom has been a longstanding cause of woe to those who see this as a distraction from the real objective of metacognitive training, which is not to actualize isolated instances of metacognition, but to reshape students' intellectual dispositions in the long term:

It appears that the enthusiasm surrounding metacognition has established the construct as a pinnacle of information processing. It is the most prized, most regulative, top-of-the-hierarchy component in several theories and instructional packages. This appears to us to be an erroneous aggrandizement of decontextualized knowledge. The goal of development and education is not to produce people who reflect, orchestrate, plan, revise, and evaluate their every action. (Paris, Jacobs, & Cross 1987, p. 238; quoted in Baker & Cerro 2000, p. 130)

Here, Paris, Jacobs, & Cross are lamenting how the exercise of metacognition has assumed a reputation of an educational end in itself; Baker & Cerro similarly affirm that "metacognition is important, but it should not serve as an instructional goal in itself but rather as a means to an end" (Baker & Cerro 2000, p. 130).

At any rate, it seems to me that the veneration of metacognition as a good in itself tends to make it more difficult for educators to regard the exercise of metacognition as a means to a further end – i.e., the cultivation of metacognitive skills. As a consequence, they tend to treat student engagement in metacognition-promoting teaching interventions as the end-game. When the distinction between the exercise of metacognition and metacognitive skill gain is flattened, then metacognitive training is suddenly a breeze: someone who exercises metacognition is attaining metacognitive skills, period; top-notch pedagogy is within any educator's grasp.

Finally, educators might be prone to regarding implementation of metacognitive training as an automatic success because the alternative is too unpalatable. It's hard for any of us to admit when we've made choices that weren't warranted. In educational settings, though, the responsibility instructors bear to bring about good results for their students might impose

pressure for success that prohibits serious consideration of the possibility that metacognitive training isn't effective after all. To entertain evidence that metacognitive training didn't actually produce the mental transformation instructors were aiming for means admitting to oneself that they may have let students down by implementing a less-than-stellar approach — or at least wasted valuable time and effort that could have been better spent. As I'll elaborate shortly (§4.2aII), to accept that one has hitched one's wagon to a bogus pedagogical theory is a tall order for educators, as evidenced by the denial and evasion exhibited by those whose teaching methods are challenged.

#### **§4.2 | Motivated reasoning in defense of metacognitive training**

In the previous section, I stated the obvious: that educators are invested in the realization of observable changes in their students. Nevertheless, this very basic fact about educators can be overlooked in discussions around pedagogical decision-making. Here, I'll discuss how instructors' desires and motivations can influence the conviction that metacognitive skills can be enhanced through teaching.

Let's jump all the way back to the origins of "metacognition" as a construct. In Chapter 1, I recounted this statement from John Flavell, in which he admits that he is antecedently committed to the view that students who engage in metacognition will outperform their peers academically:

*Lack of hard evidence notwithstanding, however, I am absolutely convinced that there is, overall, far too little rather than enough or too much cognitive monitoring in this world. This is true for adults as well as for children, but it is especially true for children. For example, I find it hard to believe that children who do more cognitive monitoring would not learn better both in and out of school than children who do less. I also think that increasing the quantity and quality of children's metacognitive knowledge and monitoring skills through systematic training may be feasible as well as desirable. (Flavell 1979, p. 910; my italics)*

It is rare to encounter a scientist admitting this brazenly that his beliefs go beyond what is immediately warranted by available evidence. But whether they disclose it or not, scientists are just as prone as the rest of us to allowing motivational factors to influence our doxastic commitments.<sup>9</sup> As William James put it, “it seems that our passionate and volitional nature lay at the root of all our convictions,” no matter how strongly we try to insist that our beliefs are based upon nothing but the facts (1896, p. 5).

*Motivated reasoning* is one of many constructs with which contemporary psychologists have elaborated upon James’ observation. It refers to *a tendency to attend to and/or process information in a way that supports conclusions one wants to be true or undermines conclusions one hopes to be false* (Kunda 1990; Sinatra, Kienheus, & Hofer 2014; Epley & Gilovich 2016). Thomas Gilovich described this as a bias in how much evidence it takes to earn our assent, where “for propositions we want to believe, we ask only that the evidence not force us to believe otherwise – a rather easy standard to meet” (1993, p. 83-84); for undesired conclusions, on the other hand, we set the threshold for persuasion significantly higher.

Often, motivated reasoning is mediated by selective attention to relevant evidence or selective inference-making from the evidence under consideration. For example, “the assessment of the credibility and weight of scientific evidence may be biased in such a way that

---

<sup>9</sup> His language makes it clear that he would *like* to see people engaged in more monitoring of their own cognition, because this would presumably permit heightened self-awareness and self-control, leading to better decision-making. “Perhaps it is stretching the meanings of metacognition and cognitive monitoring too far,” Flavell acknowledges, “but . . . [it] is at least conceivable that the ideas currently brewing in this area could someday be parlayed into a method of teaching children (and adults) to make wise and thoughtful life decisions as well as to comprehend and learn better in formal educational settings.” He ventures a guess that “trying to teach children and adolescents to monitor their cognition in communication and other social settings” could empower them to resist “the persuasive appeals the young receive from all quarters to smoke, drink, take drugs, commit aggressive or criminal acts, have casual sex without contraceptives, have or not have the casual babies that often result, quit school, and become unthinking followers of this year’s flaky cults, sects, and movements” (*ibid.*, p. 910).

the evidence fits the recipient's preexisting attitudes"<sup>10</sup> (Sinatra, Kienheus, & Hofer 2014, p. 124).

Ziva Kunda's influential account of motivated reasoning specifies that the proposal is *not* that belief formation is a total free-for-all; there are boundaries on what one can manage to convince oneself is true. But within these boundaries, we are able to sway ourselves toward some beliefs (and away from others) by marshalling our cognitive resources in ways that favor desired conclusions:

People do not seem to be at liberty to conclude whatever they want to conclude merely because they want to. Rather, I propose that people motivated to arrive at a particular conclusion attempt to be rational and to construct a justification of their desired conclusion that would persuade a dispassionate observer. They draw the desired conclusion only if they can muster up the evidence necessary to support it. In other words, they maintain an 'illusion of objectivity' . . . To this end, they search memory for those beliefs and rules that could support their desired conclusion. They may also creatively combine accessed knowledge to construct new beliefs that could logically support the desired conclusion. It is this process of memory search and belief construction that is biased by directional goals. (1990, p. 482-3)

To summarize, as long as it doesn't compromise one's self-image as a rational agent, people can construct a case for the truth of a belief that they would like to be true.

I'll offer some examples of how motivated reasoning can manifest in pedagogical discourse and decision-making. Then I'll articulate an array of factors that would lead instructors to want to believe in the efficacy of metacognitive training.

#### **[4.2a] Precedents for motivated reasoning in education**

Here I've assembled a few examples of educational concepts and methodologies that earned widespread and often passionate assent from instructors and other stakeholders, as a

---

<sup>10</sup> For example, subjects who are exposed to alleged scientific evidence are less likely to believe the evidence if they are motivated to disbelieve it, in virtue of some personal investment in the truth of whatever this evidence criticizes. E.g., women who consume high quantities of caffeine were less convinced than low caffeine consumers by an article claiming that "caffeine was risky for women" (Kunda 1990., p. 489; citing Kunda 1987).

result of the operation of motivated reasoning. For both *learning styles* and *whole word reading*, uptake of these approaches to teaching drastically outstripped evidence for their effectiveness; factors beyond the evidence base for these views evidently played a role in their acceptance.

#### **(4.2aI) Learning styles**

Originating in the late '70s, "learning styles" referred to the proposal that each individual a) has a preferred mode of information presentation (e.g. visual, verbal, auditory, kinaesthetic), and b) learns best when information is presented to them in their preferred mode (Dunn & Dunn 1978). This concept was incorporated into education textbooks and taken for granted as the foundation for at least 71 competing variants of learning styles theory (counted in a 2004 review by Coffield et al). It also became the basis for numerous commercial products marketed to schools, which potentially explains why "the learning-styles concept appears to have wide acceptance not only among educators but also among parents and the general public" (Pashler et al. 2008).

However, a team of psychologists commissioned "to assess, as dispassionately as [they] could, the scientific evidence underlying practical application of learning-style assessment in school contexts" found the support for this widely accepted theory to be "weak and unconvincing" (ibid., p. 116). Though many studies had targeted learning-styles theory, only a handful used a research design that tested the key hypothesis underpinning pedagogical applications of the theory: that students perform better when taught in accordance with *their* learning style than when they are taught using methods geared towards a *different* learning style. "The contrast between the enormous popularity of the learning-styles approach within education and the lack of credible evidence for its utility is, in our opinion, striking and disturbing," these authors attest (p. 117). In the years since, following additional confirmation



that the theory's central hypothesis lacks empirical support,<sup>11</sup> this theory has earned the new moniker of "the learning styles *myth*."

In Pashler et al.'s telling, it's possible that learning styles attained such widespread endorsement in part because it was a concept many educators, parents, and students *wanted to believe in*. They surmise "other factors – in addition to, or instead of, actual effectiveness – may [have] play[ed] a role in the popularity of the learning-styles approach." These factors include a) the "eternal and deep appeal" of a theory that sorts individuals into types and offers them insight into which type they belong to, b) the allure (within individualistic cultures, at least) of the idea that children flourish when they receive treatment that is tailored to them as a unique individual, and c) the comfort brought by the suggestion that students who underperform aren't inherently lacking in ability, but rather are the victims of an educational system that fails to cater properly to their learning style (pp. 107-8).

What these authors are suggesting is that education stakeholders may have developed an attachment to the idea that learners will excel when they are instructed in accordance with their personal learning style, on the basis of the intuitiveness of the idea, normative commitments (like valuing individuality), preferences (tending to attribute blame to systems rather than to individuals), and/or raw emotional responses (simply liking the idea that one can understand oneself better by envisioning oneself as an instance of particular type of person). Having developed this attachment, they may retroactively attribute their credence in the effectiveness of this pedagogical method to empirical evidence, and/or tend to evaluate the empirical support for this claim to be more robust than it actually is.

Ironically, Pashler et al. invoke the idea that *metacognition can be misleading* in their

---

<sup>11</sup> Riner & Willingham (2010), Simmonds (2014), Kirschner (2017), Newton & Miah (2017), Nancekivell, Shah & Gelman (2019), Poorvu Center for Teaching and Learning (n.d.).

commentary as to why learning styles took off, despite of a paucity of evidence for the foundational claim of the theory:

Basic research on human learning and memory, especially research on human metacognition, much of it carried out in the last 20 years or so, has demonstrated that our intuitions and beliefs about how we learn are often wrong in serious ways. . . .

As learners, we can also be fooled by subjective impressions, such as the ease or sense of familiarity we gain on reading expository text or how readily some information comes to mind, both of which can be products of factors unrelated to actual comprehension or understanding. There is growing evidence that people hold beliefs about how they learn that are faulty in various ways, which frequently lead people to manage their own learning and teach others in nonoptimal ways. This fact makes it clear that research – not intuition or standard practices – needs to be the foundation for upgrading teaching and learning. (Pashler et al. 2008, p. 117)

In short, that the concept of “learning styles” aligned so smoothly with many stakeholders’ subjective impression of themselves as learning best through a particular mode of information presentation turned out to be a red herring that fostered false confidence in this unverified theory.

#### **(4.2aII) Whole word reading**

In the late ‘60s, education researcher Ken Goodman advanced a theory of reading on which comprehending a text’s gestalt meaning – not its exact lexical contents – is the pinnacle of reading ability (Hanford 2019). On his view, proficient readers have the appearance of many words stored in visual memory, so they can recognize most whole words on sight; however, what makes the reader successful is not that they recognize every word, but that when what they’re reading makes sense to them. From this standpoint, successful reading involves making skillful predictive inferences about what unknown whole words mean, on the basis of three types of cues (graphic, semantic, and syntactic); recognizing each of the individual words comprising the text isn’t necessary, as long as one can fill in the gaps fluidly on the basis of these cues.

Simultaneously, developmental psychologist Marie Clay developed a complementary methodology for reading instruction known as MSV (for meaning, sentence structure, visual information), or “three-cueing.” This method begins with training children to memorize the appearance of many basic vocabulary words. When novice readers encounter words they can’t recognize, instructors prompt them to guess the unknown word by checking for accompanying imagery and filling in whatever could make sense based on the surrounding context.

Proponents of this outlook claim that when instructors focus on distilling meaning out of texts (rather than drilling into the nitty-gritty mechanics), they can help students to associate reading with eye-opening experiences and narratives, which will in turn get them hooked on books from an early age (Hanford 2019). Bolstering the idea that teachers can do more for literacy by doing less in the classroom, Goodman and others (1996; Edelsky et al. 1991) advanced the view that learning to read occurs naturally through mere exposure to texts that contain familiar vocabulary, and that the “joy of reading” will motivate children to resolve comprehension challenges they face when they tackle more complex texts containing previously unencountered words.

The whole-word method to literacy, centered on the instructional technique of three-cueing, is at the foundation of many widely used (and highly profitable) curricular programs.<sup>12</sup> Unfortunately, there’s also ample evidence that three-cueing leads to worse literacy outcomes than instruction based on *phonetic analysis*: teaching students the relationship between letters (or letter clusters) and sounds, so they can decode unknown words by sounding them out).<sup>13</sup>

---

<sup>12</sup> “Units of Study for Teaching Reading” (Calkins et al 2015), “Leveled Literacy Intervention” (Fountas & Pinnell 1996)

<sup>13</sup> An important note is that schools’ choice of literacy instruction method has variable impact on different subsets of students. According to research summarized in Hempenstall (2016) and Young (2020), roughly 40% of students learn to read fluently regardless of their literacy instructional method; this means that “children who learn to read with cueing are succeeding in spite of the instruction, not because of it” (Hanford 2019, paraphrasing David Kilpatrick; see also Kilpatrick 2018). Three-cueing has its most detrimental effects among the roughly 60% of students (about a

Indeed, three-cueing steers students away from attending to “within-word structure” (e.g., how a word is spelled and the sub-parts of compound words; Hempenstall 2002, p. 45), which impedes development of an early habit of attending to this information to decode unknown words by determining their expected pronunciation.

After three-cueing had already taken hold in many educational circles, proponents of phonics found themselves fighting an uphill battle to persuade instructors to change their methods.<sup>14</sup> Marilyn Adams determined that resistant educators were attached to a premise underlying the three-cueing method: that readers use graphic, semantic, and syntactic cues to understand text. The problem was that “teachers understood these cues not just as the way readers construct meaning from text, but as the way readers actually identify the words on the page” (Hanford 2019): they expected the instructional method for reading to emphasize the type of information they expected skilled readers to use, and found it hard to believe that teaching kids to use these cues would *not* lead to the effective use of these cues seen in skilled adult readers.

Another hurdle to ousting three-cueing instruction is that the method seems intuitive to many instructors, as it seems to guide children to “use everything you've got” available to decipher meaning; in contrast, phonics instruction strikes instructors as eschewing a holistic approach to reading comprehension in favor of one narrowly focused on within-word structure. And perhaps most significantly, teachers shied away from phonics instruction because they viewed it as “tedious” and “time-consuming.” It’s true that the method has a steeper initial learning curve for acquiring basic reading fluency with simple texts, compared to three-cueing.

---

quarter of whom qualify for diagnoses of dyslexia) for whom reading proficiency will not be achievable without “code-based explicit, systematic, and sequential instruction” (Young 2020).

<sup>14</sup> Adams (1990), Stanovich (1993, 2000), Hempenstall (2002), Kilpatrick (2018)

But steering clear of phonics because of the initial difficulties it poses for teachers and learners has adverse long-term effects:

Phonics is challenging for many kids. The cueing strategies seem quicker and easier at first. And by using context and memorizing a bunch of words, many children can look like good readers – until they get to about third grade, when their books begin to have more words, longer words, and fewer pictures. Then they're stuck. They haven't developed their sounding-out skills. Their bank of known words is limited. Reading is slow and laborious and they don't like it, so they don't do it if they don't have to.<sup>15</sup> (Hanford 2019)

In other words, three-cueing often produces the appearance of successful reading skill in the early years, where the need to guess unknown words is infrequent, visual cues are ample, and the stakes for miscomprehension are low. It's only after several years that the deficits of this approach start to emerge – which means that the vast majority of instructors who teach literacy through whole-word reading won't see these effects on the children they've instructed, and aren't directly confronted with evidence of the faults of this method.

Lastly, continued defenses of this discredited methodology might be driven by the sunk cost fallacy. Infamously, the lead author of one literacy curriculum laden with three-cueing, *Units of Study* (Calkins et al. 2015) continued to assert that her product aligns with the available science, despite disconfirmation of this claim by an independent review board (Adams et al. 2020; see also Hanford 2020a). School districts that have already purchased and put this curriculum into practice are also disincentivized to acknowledge the charges against *Units of*

---

<sup>15</sup> It gets worse. Cunningham & Stanovich (2001) explain that readers who struggle with text comprehension due to their inability to phonemically decode words quickly lose ground academically compared to their peers, because reading less in the early years withholds access to knowledge that aids future text comprehension:

The combination of deficient decoding skills, lack of practice, and difficult materials results in unrewarding early reading experiences that lead to less involvement in reading-related activities. Lack of exposure and practice on the part of the less-skilled reader delays the development of automaticity and speed at the word recognition level. Slow, capacity-draining word recognition processes require cognitive resources that should be allocated to comprehension. Thus, reading for meaning is hindered; unrewarding reading experiences multiply; and practice is avoided or merely tolerated without real cognitive involvement. (p. 137)

Stanovich (1986) coined this a “Matthew effect,” after the Biblical passage in which the rich get richer and the poor get poorer.

*Study*, since this would require admitting that thousands of taxpayer dollars may have been sunk into a reading instruction method that is ultimately ineffective for many children.

#### **(4.2aIII) Implications of these examples**

Each of these examples highlights specific factors that may facilitate uptake of beliefs in the pedagogical efficacy of approaches to teaching, along with factors that contribute to perseverance in these beliefs in the face of credible challenges.

Let me begin with a throughline linking both examples. For both learning styles and whole word reading, much of the mileage of these approaches seems to be tied to their *intuitiveness* – that it was compatible with many educators’ and stakeholders’ impressions of what leads to success (in learning generally and reading comprehension, respectively).

Learning styles found easy acceptance among those for whom it seemed obvious that students belong to various types, and that they do better when catered to as individuals than when treated as an undifferentiated horde. We may be able to trace these beliefs (that people naturally belong to types, and endorsement of the power of personalization), in turn, to simple and universal cognitive drives and biases: the embrace of types is a predictable consequence of our automatic tendency to group complex phenomena into categories, and belief that each of us requires individual attention is easily fueled by a trenchant conviction in our own singularity or specialness, fostered by our imprisonment in our own limited subjectivity.

Whole word reading, too, is intuitive on several grounds. Its underlying assumption that reading acquisition is a natural facet of development is easy to swallow, thanks to its analogy to the well-observed process of language acquisition. It also seems sensible enough that helping students figure out the meaning of unknown words should result from supplying children with direct instruction in resolving gaps in comprehension the way skilled readers apparently do –

that is, by making use of available context. (Regrettably, this neglects the fact that comprehension strategies don't work equally well for readers at different stages of literacy: Reliance on context is only a viable approach for learners who have adequate conceptual knowledge to make reasonable inferences as to the meaning of words they don't recognize (Hanford 2020b).<sup>16</sup>)

One takeaway from these examples, then, is that some pedagogical proposals have an advantage in the marketplace of educational ideas simply due to their congruence with ideas educators take for granted from personal experience or cognitive predispositions. Counterintuitive ideas, in contrast, may face an uphill battle in earning assent from educators, even if the preponderance of empirical evidence is heavily in their favor (as is the case for reading instruction that gives adequate weight to phonics).

Let's hone in on some specific upshots from each example. The "learning styles" saga is, in my eyes, a compelling cautionary tale about taking the *popularity* of an educational theory as a proxy for its accuracy. We need not take up concepts that emerge from educational psychology as received wisdom about how learning works, even if they are frequently discussed and widely implemented by educational practitioners. "If education is to be transformed into an evidence-based field," write Pashler et. al, "it is important not only to identify teaching techniques that have experimental support but also to identify widely held beliefs that affect the choices made by educational practitioners but that lack empirical support" (2008, p. 117).

I'll co-sign these authors' conviction that *if we want to claim that our educational practices are evidence-based*, then we can't content ourselves with merely trusting that there must be adequate evidence to verify the claims associated with a particular practice that has been

---

<sup>16</sup> Willingham (2006), Lemov et al. (2016), O'Reilly et al. (2019).

labelled as “evidence-based.” We may tend to assume that because a concept is presented (or sold) as the outcome of “the science of learning,” it has already been subjected to this form of testing and run the gauntlet. Unfortunately, it’s apparent that theories may gain considerable traction among educators under the *guise* of educational science, in the absence of the rigorous verification for which “evidence-based” practices are valued in the first place. As education journalist Emily Hanford (2019) emphasizes, educational products won’t sell unless they are marketed as *research-driven* – but this just means that products will be marketed as research-driven regardless of how well they are actually aligned with available evidence.<sup>17</sup>

From whole word reading, we gain a lens into the social and economic forces that can sustain pedagogical methods past their due. For one, once educators have invested a great deal into a particular approach to teaching – and in particular if they have staked their reputation on it – they will be hard-pressed to acknowledge and concede to counterevidence, as this amounts to an admission that they have wasted precious instructional time and school resources on ineffective or counterproductive practices. Add to this the realization that one may have wronged students to whom they applied these methods in the past, abdicating a moral duty to educate students to the best of one’s ability, and you have a potent recipe for denialism.

#### **[4.2b] Why metacognitivism is attractive to educators**

Let’s delve into why educators and education researchers would *want* to believe that metacognitive training is indeed effective. One apparent reason is, of course, that the objective of metacognitive training – to cultivate self-regulating learners with mastery over the workings

---

<sup>17</sup> At present it seems that the only way of enforcing *accurate* descriptions of pedagogical practices as “evidence-based” is to call out false claims after they have been issued, and hope that those who were misled into adopting these practices are listening. Pashler et al. (2008) and Adams et al. (2020) each reflect instances of vigilante verification efforts by nonprofits dedicated to sussing out how well various instructional methods fare against dispassionate investigation.



of their own minds — is a compelling vision of intellectual excellence. To believe that this is readily achievable gives educators an inspiring target to aim at, which rests on an expansive conception of human potential.

I'll outline three additional factors that I suspect play a role in disposing educators to embrace the idea that metacognitive training is indeed effective: I) the influence of Carol Dweck's concept of "growth mindset" on the pedagogical imagination, II) a preference for simple and intuitive accounts of how to teach, and III) the operation of scientism among educators.

#### **(4.2bI) Possible influence of mindset theory**

Back in Chapter 2, I assembled evidence we're susceptible to various illusions and conative influences that dispose us to assign ourselves the power to become masters of our own minds. It's understandable why we'd believe this of ourselves: Who would rather think that we *can't* exert control over one's thinking? But it's even easier to understand why we would adopt this assumption of mental malleability in the context of education. If one's job is to cultivate the minds of the next generation, then one is occupationally committed to the premise that minds can indeed be cultivated — otherwise teaching would be entirely an exercise in futility.

But a strong trend in the educational discourse of the past two decades (give or take) dials belief in the malleability of human minds up to eleven, turning this into an imperative of effective pedagogy. Associated with research by Carol Dweck (1998, 2006; but since taking on a mind of its own<sup>18</sup>), *mindset theory* posits that whether one thinks of their ability as fixed or malleable impacts their potential for future learning by shaping how they respond to learning challenges.

---

<sup>18</sup> See Dweck (2016) on how mindset theory has been misconstrued and misapplied in practice. See also Bandura (1994) for theoretical groundwork on self-efficacy.

One manifestation of this theory's influence on pedagogical strategy is that educators are sometimes exhorted to operate from a worldview according to which ability can *always* be developed.<sup>19</sup> One can find expressions of this outlook in myriad school mission statements asserting belief in the "unlimited potential" of students.<sup>20</sup> The intentions behind this exhortation are admirable, aimed at discouraging educators from giving up on learners who are struggling by pigeonholing them as having limited intellectual potential. Nevertheless, this ardently-defended outlook communicates to instructors that to recognize outer bounds on how much minds can change is detrimental to students' development, and thereby anathema for a responsible pedagogue. One can imagine how the prevalence of the message that an educator *must* believe that their students can achieve *anything* would be disposed to overconfidence in the prospects of pedagogical interventions to mold students' presumptively malleable minds.

#### **(4.2bII) Excessive faith in simple and intuitive teaching methods**

One contributor to the formation of pedagogical myths seems to be the facility with which teaching methods that are simple and intuitive are embraced over alternatives that involve more complexity and nuance. I'll give some examples where the popularity of a pedagogical idea seemed to reflect an assumption that teachers can best facilitate learning through minimal instruction: that is, by merely providing a suitable environment in which students' natural inquisitiveness will draw them steadily toward expertise.<sup>21</sup>

---

<sup>19</sup> E.g., Shenk (2010), Jackson (2011), Delpit (2013)

<sup>20</sup> For example, the website of the influential KIPP Charter School network claims: "Effective teachers set high expectations, believe in every student's unlimited potential, and are equipped with the training, tools, and resources to maximize their success in the classroom" (Knowledge is Power Program n.d.).

<sup>21</sup> E.D. Hirsch traces the origins of this outlook to romantic-era insistence on "education as a natural unfolding" of innate potential (2006, p. 7) – the intellectual equivalent of preformationism in biology. This "naturalistic" vision filtered into the approach championed by American education reformers like Horace Mann and John Dewey, whose theories remain fixtures of education schooling.

Let's return to the now-discredited approach to literacy education, according to which it isn't necessary to teach phonics or the decoding of within-word structure. As I mentioned, an important supposition underlying this approach is the idea that learning to read is a process that comes naturally to children<sup>22</sup> – just like learning language through immersive exposure to language use within one's community. From this standpoint, educators shouldn't try to engineer reading acquisition, but rather ought to get out of the way so that children can pick up literacy on their own through organic encounters with texts.

Unfortunately, this putative analogy between attaining literacy and early-childhood language acquisition just doesn't hold up (Lyon 1998). But once it started to gain ground in educational discourse, it was hard to dispel because its simplicity is highly enticing. I'm sure many early-elementary school teachers really wish it were the case that all one had to do to produce proficient readers is to put books in front of children and leave them to their own devices – which is worlds less demanding than a highly involved “structured literacy” approach (Young 2020).

Moreover, the three-cueing method assumes that teaching children to read requires little more than inducing them to mimic the behavior of adept readers. Accordingly, insofar as it *appears* that skilled readers can rapidly recognize many words on sight and that they make use of contextual cues to fill in gaps, the most efficient way to produce reading proficiency would *seem* to be to help students recognize whole words on sight and to coach them to recruit graphic, semantic, and syntactic cues to infer the meaning of unknown words. Regrettably, this approach rests on a misconception of the cognitive activity involved in fluent reading

---

<sup>22</sup> This is not to be confused with the so-called “Simple View of Reading” (Gough & Tunmer 1986) which despite its choice of adjective, describes a more complex alternative to the view that acquiring literacy just happens if children are exposed to texts. The “simple” descriptor is meant to capture the thesis that reading comprehension emerges from the synthesis of two arenas of skill: graphophonemic decoding (the ability to decipher words on a page) and language comprehension (knowledge of the meaning of vocabulary used in texts).

comprehension: though the speed at which readers recognize words would seem to suggest that reading consists of gestalt processing of whole words as discrete images, skilled readers actually process words on a more granular level, as sequences of letters (Hanford 2019).

Explicitly teaching students to depend upon contextual cues to fill in an unknown word is not to teach them to imitate adept readers, but rather to equip them with a compensatory strategy they can use to make up for an absence of decoding skills.

Moving away from reading instruction, another area in which simplicity has been used a heuristic for identifying an effective pedagogical method is in discussions of “active learning.” This phrase initially emerged as a shorthand for a tenet of *cognitive constructivism*, a theory of learning according to which “learning involves actively building knowledge representations in working memory by applying appropriate cognitive processes” (Mayer 2009, p. 186, p. 188). This theory emphasizes that what we recognize as learning on a behavioral level is underpinned by a requisite pattern of cognitive activity, which involves “attending to relevant incoming information, mentally organizing it into a coherent cognitive structure, and mentally relating it with relevant prior knowledge from long-term memory” (ibid.).

Educational psychologist Richard Mayer coined “the constructivist teaching fallacy” to describe the assumption that “the only way to achieve constructivist learning is through active methods of teaching,” guided by the expectation that *active learning requires active instruction* (2004, p. 15). Many instructors who embraced constructivism as a mechanistic explanation for learning interpreted the theory as backing the pedagogical principle that students must “be active during learning” in order to engage in the appropriate cognitive processing to develop new knowledge. More specifically, they reasoned that students need to be *behaviorally* active during the learning process – engaged in hands-on exploratory activities that allow for “pure discovery” guided by the learner, instead of passively listening to an instructor. However,

learners engaged in “pure discovery” often underperform peers engaged in “guided discovery” structured by an instructor (2004; see also Kirschner et al. 2006). Mayer clarifies that behaviorally-active learning can be cognitively passive, and vice-versa (2009).

A common thread across both of these examples is engagement in wishful thinking: that one can achieve a desired educational end (e.g. “active learning,” literacy) by taking the pedagogical path of least resistance. We see an interest in *not* overcomplicating skill-building, which manifests in beliefs that all it takes to produce the outcome one wants is to get students doing something that resembles where they are supposed to end up after years of progress: e.g., if you want readers to be able to decipher meaning from available cues, teach them how to use cues to guess the semantic content of unfamiliar written terms. This pattern of reasoning is probably intuitive to many, and it’s reinforced in several texts on educational psychology which highlight discrepancies between experts and novices as if these differences present a roadmap for how to transform a novice into an expert<sup>23</sup> (Bruer 1993, Bransford et al. 2000).

The problem with this approach to pedagogy is that it allows our desire to keep things simple (by sticking with a method that is straightforward and intuitive) to generate standing beliefs about how teaching should work that are difficult to oust in favor of less-intuitive views. The simple idea underlying metacognitivist approaches to pedagogy – that *if you want students to be self-aware to self-regulate during learning, you should teach them how to monitor and control their own cognition* – is likely to be sticky, because it suggests that actualizing the objective of promoting metacognitive development in one’s students is no more complicated than having

---

<sup>23</sup> To be fair, Bransford et al. clarify that they draw out differences between these two populations “*not* because all school children are expected to become experts . . . , but because the study of expertise shows what the results of successful learning look like” (p. 31). However, I think emphasizing expert/novice contrasts can be counterproductive insofar as they suggest to instructors that trying to replicate what the experts do should be the goal of their pedagogical choices. At least for the purposes of selecting instructional methods that are appropriate for novices, getting concrete on what *competence* looks like seems to be a more appropriate way of conceptualizing the ideal against which novice performance ought to be compared.

students practice what they're eventually supposed to be able to do on their own without prompting. In contrast, consider the alternative idea that *if you want students to be self-aware to self-regulate during learning, there's not much you can do to guarantee that will happen, because the mechanisms of metacognitive skill gain are opaque to us at present*. It's hard to blame educators for resisting the idea that they're powerless to bring about the change they want to see.

#### **(4.2bIII) Side effects of a demand for "what works"**

Focusing on the U.S. context in particular, historian of education Diane Ravitch notes:

Education in America tends to be akin to religion, with its cycles of stability and instability, its periodic crusades, its movements, and its occasional bouts of zealotry and apostasy. . . .

Just at the point when classroom methods and protocols seem to have grown stale, or at a time when society is experiencing an unusual degree of upheaval, along comes an educational movement to cast out the old and mobilize true believers. Each movement has its prophets, its sacred texts, its peculiar solutions to knotty problems. With each movement there are claims about discovering the Royal Road to Learning, or the key that will unlock the kingdom of student motivation, or the policy innovation that will cure all the ills of the schools at no extra cost. (Ravitch 2003)

Education is certainly not singular in this regard; conformity wields its influence in all sorts of human endeavors. However, there are certain features of education that I suspect render its practitioners especially susceptible to the influence of disciplinary trends and linguistic fashion. That which is *innovative* may seem preferable to education stakeholders insofar as it feeds into the narrative that each generation deserves an education that will prepare them for the unique challenges of the era in which they will become citizens, leaders, and decision-makers. Except for the most traditionalist among us, parents, students, teachers, administrators, and policy-makers typically want to see educational choices cutting-edge enough to equip learners for a rapidly-changing, technologically-driven 21st-century world: in short, *not-your-mother's pedagogy*.

One of the predominant trends in contemporary educational discourse is insistence

upon the use of “evidence-based” teaching practices, often summarized as “*what works*” (Slavin 2002). And not just any “evidence” will do: this honorific is meant to be reserved for findings from quantitative experimental studies using randomized controlled trials, considered the gold standard in education research. This specificity is most prominent in a set of policy announcements around the turn of the millennium, including the 2001 No Child Left Behind Act, 2002 National Research Council report *Scientific Research in Education*, and establishment of the “What Works Clearinghouse” by the US Department of Education (DOE). The DOE’s 2002–2007 Strategic Plan set “Transform[ing] Education into an Evidence-based Field” as one of six macro goals driving education policy-making in the U.S.<sup>24</sup>

This drive to enact “what works” is fueled by a conviction that “our children deserve the best educational programs, based on the most rigorous evidence we can provide” (Slavin 2002, p. 15): that we’d be derelict in our duties to students if we made no effort to leverage available findings from educational science to improve student outcomes, as well as to steer educators away from using well-entrenched but ineffective methods that have no empirical backing. But there are some pitfalls associated with this effort, mostly stemming from the challenge of communicating effectively about what education research does and does not indicate about how we should approach teaching.

In particular, when educators are encouraged or pressured to follow the “science of learning” without accompanying guidance on how to interpret research findings or how to determine which findings warrant changing one’s pedagogical methods, they may engage in *scientism*: indiscriminately embracing the apparent outputs of educational science, instead of using their judgment to assess the extent to which a given pedagogical practice is warranted by

---

<sup>24</sup> Quite a few philosophers of education have expressed reservations about education policy leaders’ adulation of “evidence-based” practices, including that which has trickled down into the discourse of school administrators and teaching practitioners. See Howe (2004, 2005, 2009, 2013); Phillips (2006, 2009), Schwandt (2005), Issitt (2007), Biesta (2007, 2010), Smeyers & Depaepe (2009, 2011, 2012).

the type and quantity of evidence available.

For evidence of how scientism manifests in pedagogical discourse, let's turn to discussions on the prevalence of "neuromyths" in educational rhetoric. Quite a few authors have noted the ease with which commercial programs promising the delivery of "brain-based" pedagogy have been uptaken by schools despite a lack of empirical rigor<sup>25</sup> (and in some cases, utter quackery<sup>26</sup>). Another pernicious problem is the unwitting spread of misinformation about neuroscience among educators.<sup>27</sup>

Elena Pasquinelli (2012) offers a laundry list of factors potentially contributing to the proliferation of (typically distorted) neuroscientific claims in educational discourse: misinterpretations of research findings, faulty inferences from phenomena such as lateralization, innocent miscommunication and irresponsible sensationalization, a general upswell of "neurophilia," and the operation of our ordinary slew of cognitive biases.<sup>28</sup> While all of these factors certainly seem plausible as contributors to the genesis and uptake of neuromyths, I think another factor that is likely at play here is a desire for science to deliver certainty about how one ought to teach.

---

<sup>25</sup> Bruer (1997), Goswami (2006), Geake (2008), Howard-Jones (2008, 2014), Alerink & Farmer-Dougan (2010), Pasquinelli (2012), Dekker et al. (2012)

<sup>26</sup> Goswami describes a product called BrainGym<sup>®</sup>, which "prescribes a series of simple body movements 'to integrate all areas of the brain to enhance learning';" the product's claims are nonsensical, but, chillingly, scientific laypeople would be none the wiser. "Teachers are told that 'in technical terms, information is received by the brainstem as an *impress*, but may be inaccessible to the front brain as an *express*. This ... locks the student into a failure syndrome. Whole-brain learning draws out the potential locked in the body and enables students to access those areas of the brain previously unavailable to them. Improvements in learning ... are often immediate'. It is even claimed that the child can press certain 'brain buttons' under their ribs to focus the visual system for reading and writing." (Goswami 2006, p. 2; quoting Cohen & Goldsmith 2000)

<sup>27</sup> For example, among philosophers of education, Karen Hornsby advances the misconception that instructors can guarantee long-term learning gains by choosing pedagogical practices that catalyze neuroplasticity ("The pedagogical challenge is creating methods to stimulate students' dendrites' developmental activity to thereby create larger, denser neural networks, and brain growth"; 2012, p. 54). As Goswami clarifies, one cannot selectively target neuroplasticity in general or dendrite stimulation in particular: "the effects of any type of training programme that changes behaviour will be reflected in the 'remapping' of neural networks" (2006, p. 3).

<sup>28</sup> For additional discussion on the apparent attractiveness of neuroscientific explanations, see also McCabe & Castel (2008) and Weisberg et al. (2008).



Usha Goswami offers a telling anecdote about an attempt to dispel neuromyths by organizing a conference in which “prominent neuroscientists . . . spoke directly to teachers about the scientific evidence being gathered in scientists’ laboratories” (2006, p. 6). She reports that:

The teachers were amazed by how little was known. Although there was enthusiasm for and appreciation of getting first-hand information, this was coupled with frustration at hearing that many of the brain-based programmes currently in schools had no scientific basis. The frustration arose because the neuroscientists were not telling the teachers ‘what works instead.’ One delegate said that the conference “left teachers feeling [that] they had lots stripped away from them and nothing put in [its] place.” Another commented that “Class teachers will take on new initiatives if they are sold on the benefits for the children. Ultimately this is where brains live!” (Goswami 2006, p. 6)

Instructors did not respond well to the information that the “brain-based” pedagogical methods toward which they had gravitated were empirically unmoored. Having learned that these programs were bunk, the teachers *wanted to be assured* that they could implement alternatives that *are* scientifically bona fide – but this was not something the neuroscientists were prepared to offer, so several educators were left dissatisfied by the encounter.

Plenty of ink has been spilled about if and how we may “bridge the gulf” between neuroscience and education.<sup>29</sup> I’m not sure quite enough emphasis has been placed on the discrepancy between the *modus operandi* of scientists and educators, respectively. Scientists operate under disciplinary norms which favor uncertainty in the absence of overwhelming evidence. (How strongly these norms *actually* curb scientists’ human tendencies to believe without warrant is a separate sociological question.) Educators, on the other hand, are forced to make choices about how they will make learning happen (Chen et al. 2020); moreover, many are under intense pressure from administrators, accreditors and so on, to prove that they’re

---

<sup>29</sup> E.g., Bruer (1997), Willingham & Lloyd (2007), Devonshire & Dommett (2010), Hruby (2012), Horvath & Donaghue (2016).

succeeding in this endeavor. Given these social and occupational pressures, remaining uncommitted to a theory of what pedagogical choices “work” to produce learning isn’t really a live option. They need to believe *something* about how one ought to teach, in order to move forward with decision-making about how they will carry out their responsibility to educate students. Hence, when scientists “are appropriately cautious about saying that something ‘works’,” following norms that urge hedging and withholding full-throated assent, their fidelity to scientific standards of epistemic responsibility may be met with apparent frustration from educators who want definite answers as to “what works” according to science (Goswami 2006, p. 6).

What I’m getting at is that educators may be prone to embracing the issuances of educational science with less caution than would be ideal, because these theories, concepts, and methods offer direction to instructors who are eager for solid answers as to how they can do right by their students. This is, I suspect, part of what has fueled uncritical embrace of metacognitive training as a pedagogical method. When educators hear metacognitivist calls to action, like Concepción’s assertion that “if we want students to learn as much as possible, then we should help them improve their metacognitive skills” (2004, p. 356), this offers a highly satisfying degree of clarity about what is required of an instructor to fulfill their obligations to educate students to the best of their ability. To embrace this message is to ease one’s own mind about what one ought to do to teach well.<sup>30</sup>

---

<sup>30</sup> This might sound like a denigration of the epistemic norms of education. I don’t mean to assert that educators as a whole are careless with their truth commitments, or ignorant about the value of skepticism. I only intend to point out that educators are under occupational pressures that force them to make pedagogical choices in the absence of definitive evidence about what they ought to do, where scientists can remain agnostic until clear answers emerge from accumulated data. Scientists are also better positioned to remain dispassionate about how one ought to teach; most teachers, on the other hand, opted into the profession on the basis of a strong desire to advance learning, and thus have a vested interest in developing convictions as to which methods can best support that objective.

Indeed, education *should* be populated by individuals who are driven by a strong desire to advance learning; any of us who have suffered through passionless instruction are well-aware of the obvious difference that a teacher’s desire for students to succeed makes in their efficacy in fulfilling the role of educator. So I don’t mean to knock educators

Another manifestation of scientism among educators is the positioning of science as the victor in a zero-sum game among sources of knowledge informing pedagogical practice. Some philosophers of education have expressed concern that single-mindedly championing “evidence-based” practices crowds out appreciation of elements of education that undoubtedly make a difference to student outcomes, but are difficult to systematize and/or quantify. For example, Thomas Schwandt highlights how an insistence on “evidence” paints all sorts of non-experimental empirical findings in education – namely, the practitioners’ knowledge that educators acquire through trial-and-error in everyday instructional decision-making<sup>31</sup> – as “makeshift, intuitive, unreliable, and unaccountable”: in all ways inferior to the products of educational science. According to Schwandt, an emphasis on “evidence” has meant the comparative denigration of instructors’ practical judgment and reasoning:

In this scenario, technical, scientific reason becomes the only way to think about what reasoned practice is – and, hence, the messy give-and-take of deliberating ends and means (including asking such questions as what is required to be a good teacher on this occasion, or what is an appropriate and effective way to educate this child in this circumstance) is made to seem deplorable and embarrassing. In the end the essentially normative context of teaching (and administrative) practice, with its understandings of obligations and necessities, is displaced by the scientifically rational order, with its emphasis on calculation and consequence. . . . In other words, expert scientists look upon everyday practice as being in need of salvation. (Schwandt 2005: 295)

In his telling, the rising emphasis on “evidence-based” pedagogy comes at the expense of less technical, vernacular ways of speaking about teaching.<sup>32</sup>

---

for having this desire – only to bring it to our attention, so we may consider what effects such a desire has on their willingness to believe compelling ideas about effective pedagogy.

<sup>31</sup> See Hiebert et al. (2002) for a discussion of how education practitioners’ knowledge could be systematized into a base of professional knowledge, instead of being supplanted completely by scientific research conducted by third-parties.

<sup>32</sup> Another common allegation about the ascent of “evidence-based” practices is that this does *not* reflect a good-faith advocacy for science as the proper font of knowledge from which to guide educational practice, but rather a shallower interest in the *appearance* of scientific backing for educational practice. The grounds for this claim are typically that those who push most fervently for “evidence-based practices” sometimes make choices that evince unsophisticated and reductive views of what “science” amounts to: for example, overemphasizing the definitiveness

If Schwandt is right about how these appeals devalue *unscientific* discourse around educational practice (as “deplorable and embarrassing”), then we can predict that, wherever possible, educators will opt for more scientific language – avoiding the ignominy of seeming as if one is stuck in a defunct, backwoods mode of understanding when it comes to pedagogy. This is part of the reason why, I believe, educators are now disposed to engage with the concept of “metacognition” where they previously might have communicated similar themes in terms of “reflection” or “knowing oneself. “Metacognition” just *sounds* scientific. It involves a technical psychological category (*cognition*) with a prefix that often indicates a more complicated, more theoretical, or more cerebral version of whatever it modifies (*meta-*). It is decidedly *un colloquial*, brandishing its academic origins. It might as well be the poster-child for the type of educational terminology that flourishes in an era when scientism confers a special mystique upon concepts that bring to mind brain scans, laboratories, and other elements of scientific iconography.

#### §4.3 | Concluding thoughts

In this chapter, I’ve sought to explain how educators and education researchers could come to believe that there is a clear pedagogical imperative to develop students’ metacognitive skills, in the absence of adequate evidence to affirm that metacognitive training effects the psychological changes at which it aims.

I began by discussing how educators may come to believe that metacognitive training interventions were effective by shifting the goalposts for success to mark mere uptake of the intervention, instead of collecting independent measures of metacognitive ability to validate the

---

of the results of individual studies (Erickson and Gutierrez 2002), or devaluing qualitative methods as not truly scientific (Feuer et al. 2002).

impact of the intervention. I've demonstrated precedent for pedagogical theories becoming wildly popular despite a dearth of evidence to back their central claims, and I've suggested that motivated reasoning can account for the rapid and widespread embrace of proposals that just so happen to be attractive to various educational stakeholders (for example, because they suggest a one particular intervention could provide quick fix for academic underperformance). Lastly, I've identified some factors that collectively make it easy for instructors to believe that metacognitive training works, the result of which is that they may embrace this claim more readily than they would take up other claims with equally murky empirical records.

In the spirit of not offering critique without supplying suggestions for improvement, I'll close by outlining a possible route away from motivated reasoning in defense of metacognitive training. One step towards a solution is for educators to own up to the normative commitments that lead them to prefer some approaches to teaching over others, instead of concealing these preferences behind assertions that their choices are science-backed. If educators truly want their students to be "metacognitive" (whether they take that to be self-aware, reflective, deliberate in their approach towards learning, or so on), then it makes sense for them to follow through on this desire by engaging students in the exercise of metacognition in their classes. The problem I've harped upon in this project starts to arise when instructors take these classroom exercises to have transformative powers – suddenly granting students insight into their own minds or control over their cognition that they couldn't have had otherwise. So as long as educators are guiding students to engage in "metacognition" because they think this is something intrinsically worthwhile for students, then this pedagogical choice is in good faith. What's suspect is assigning instrumental value to a practice – e.g., by advancing the empirical claims that this type of pedagogy develops students' metacognitive ability – as a justificatory cover-

up for teaching what one values intrinsically, on the supposition that one must only teach “what works.”

Earlier in this chapter, I presented Thomas Schwandt’s argument that when we adopt a culture of justification for pedagogical practice in which all pedagogical choices must be “evidence-based,” we put enormous pressure on educators to prove that that they’re selecting their teaching methods in accordance with one specific mode of decision-making deemed reasonable — anything other than this being “deplorable and embarrassing” (2005, p. 294). I predicted that if Schwandt is right, then educators will suppress admissions that factors *other than* “what works” are influencing their decision-making, even though it’s inevitable that desires and preferences will come into play in shaping educators’ approach to their craft.

A similar point is well-expressed by Gert Biesta (2010), who has contended that even if we’re committed to optimizing educational practice in light of the latest developments in scientific knowledge, it’s folly to pretend that values don’t inform pedagogy. He emphasizes that even the choice to try to transform educational research into practice — what the UK’s Sutton Trust calls following “the Golden Thread from evidence to student outcomes, via deliberate intervention” (Tomsett 2015) — necessarily involves the will to act upon what the evidence indicates: “[If] we wish to use any knowledge about possible relationships between actions and consequences, there is still an important judgment to be made as to whether we wish to apply this knowledge and this, again, is a value judgment” (Biesta 2010, p. 500). We are not compelled by logical necessity or rationality or anything else to do as the “evidence” indicates:

... even if we were able to identify the most effective way of achieving a particular end, we may still want to decide not to act accordingly. There is, for example, important research evidence on the influence of the home environment on educational achievement. Yet in most cases we would find it undesirable to take children away from their parents simply to improve their chances of

educational success somewhere down the line.<sup>33</sup> (ibid., pp. 500-501)

More generally, Biesta's point is that it's "only in light of decisions about the aims and ends of educational practices that questions about evidence and effectiveness begin to have any meaning at all" (ibid.); so, we might as well confront the realities of what it is that we *want*, and get these conversations out in the open instead of banishing them as unfit for a field striving to brand itself as science-driven.

The sooner we get clear on what educators want and how this influences their doxastic commitments, the more insight we gain into what kind of forces are at work in shaping pedagogical doctrine. Should we determine that some dogmas of pedagogical rhetoric are worth revising, I expect we'll have a hard time actually dislodging false, misleading, or counterproductive messages unless we address the desires on the basis of which these messages were attractive in the first place.

Besides being honest with ourselves about the impact of our values on our pedagogical choices, we might consider positioning metacognition as what Mark Alfano calls a *factitious intellectual virtue* (2013, pp. 157-180).<sup>34</sup> Roughly, a factitious virtue is an ideal character trait that isn't attainable by anyone;<sup>35</sup> nevertheless, it's worth discussing this ideal and attributing it to people, on the grounds that the pretense that one *could* attain this virtue motivates people to behave more in line with the ideal than they may have otherwise. In other words, it is worthwhile to speak of the possibilities of becoming open-minded, having intellectual integrity,

---

<sup>33</sup> Similarly, if we really cared about improving student outcomes we'd focus on eliminating poverty, homelessness, and other social conditions which, besides diverting students' focus from their education, impose chronic stress on students. Chronic stress overloads the HPA axis (the endocrine system that, under ordinary conditions allows us to respond appropriately to stress), causing lasting dysregulation which impairs students' capacity to tackle the challenges of education (Tough 2010).

<sup>34</sup> Broncano (2014), Lepock (2014), Green (2019) and Stewart (2019) explore the value of regarding metacognition as an intellectual virtue. For defenses of an "intellectual virtues approach" to education, see Battaly (2006), Baehr (n.d., 2013, 2015, 2017a, 2017b, 2019), Pritchard (2013).

<sup>35</sup> For Alfano this isn't because none of us are actually good enough to become virtuous, but because we don't genuinely possess *character traits* (including virtues); instead, our behavior is dictated by situational factors.

exercising intellectual humility and so on even if no one truly can realize these ideals, because the fiction that they can be attained is more likely to inspire behavior in the direction of open-mindedness, intellectual integrity, and intellectual humility than what we would see had we not articulated these ideals and set students' sights upon them.

Perhaps educators can encourage students to strive to be metacognitive — to know themselves as learners, to take charge of their own thinking, and so on — on the presumption that being aware of this factitious ideal might motivate students to behave in ways that are advantageous to them. The key to making this work is for the instructor to remember that it's pretense, and not let themselves get swept up in the fiction: This is bound to be challenging, since (as I've emphasized repeatedly) we're all motivated at least to some extent to believe that it really is possible to get better at monitoring and controlling our own thinking. But if instructors can pull off the pretense effectively, they'll be encouraging metacognition because they truly believe students are better off when they try to emulate the ideal — not merely because it's fashionable to teach in this way.



## APPENDIX: DISAMBIGUATING “METACOGNITION”

As relayed in Chapter 1 (§1.3b, footnote 20), the sense of “meta-” operative in *metacognition* can be situated among a set of “supra meanings” found in various 20th-century “meta-” terms (Thomas 1984), p. 16). The trouble is that different translations of this *supra-* sense – i.e., of what it means to *transcend*, to elevate to a *higher level*, to *go beyond* – can yield various interpretations of what “metacognition” really *is*.

### §A1 | Abstracted cognition

One possible translation of “metacognition” follows the pattern of “metaethics” and “metaphilosophy”: roughly, *the investigation of the fundamental nature of a subject, as opposed to its practical applications*. This renders “metacognition” akin to fields of study which seek to clarify what cognition *is*. On this construal, all of philosophy of mind and a great deal of epistemology could qualify as “metacognition.” In this vein, Wilson & Conyers’ (2016) characterize *teaching students about brain plasticity* as “metacognition,” because students think about thinking (as an ability mediated by physical processes in the brain) during these lessons.

### §A2 | Higher-level cognition

Another translation of “metacognition” is *cognition which takes place at a higher level*. However, this description is itself ambiguous between at least three senses. One of these, previously discussed in Chapter 2.3, is “higher order cognition,” referring to cognitive states and processes that take other cognitive states or processes as intentional objects. Contrast that sense with each of the following:

*Supererogatory cognition.* “Higher-level” can be taken to indicate *that which goes above and beyond* in a way that entails *exceeding ordinary levels of effort and achievement*. This results in a rendering of “metacognition” as a *supererogatory* form of cognition, where a thinker exceeds basic expectations to engage in a form of thinking that is unusual, extraordinary, and exemplary:

. . . in addition to its cognitive and regulative dimensions there is a certain *aspirational* element to metacognition as well: in seeking to monitor and control my rational activity, I am implicitly or explicitly trying to think *better*, to hold my thinking up to a standard whose validity I acknowledge, but which I am aware I do not yet, or not consistently, reach. (Stokes 2012, p. 145)

On the basis of this translation, to engage in “metacognition” is to not just do well within the normal range of possibilities for cognition, but rather to perform cognitive operations that fall outside of what is typical, in a highly positive way.

*Advanced cognition.* Cognition may be viewed as encompassing a hierarchy of increasingly sophisticated, advanced, and/or desirable intellectual activities. From this standpoint, “higher-level” cognition refers to *intellectual activity occupying the upper echelon of this qualitative hierarchy*.

Many educators have embraced Bloom’s Taxonomy (Bloom 1956) as a useful expression of this hierarchy, for the purposes of analyzing and comparing pedagogical methods. The original purpose of the taxonomy was to systematize assessment of instructional methods, by clarifying the educational objectives to which these methods conduce. However, the taxonomy is presented in terms of “student behaviors . . . related to mental acts or thinking,” which “represent the intended outcomes of the educational process” (ibid., p. 12). In other words, observable behaviors, such as verbalizations, writing, and illustrations, are treated as emblematic of underlying cognitive activity. Bloom and colleagues organized student behaviors into six classes in order of increasing cognitive complexity, where more complex behaviors are

built upon (i.e., require the engagement and/or prior mastery of) the simpler ones: *knowledge, comprehension, application, analysis, synthesis, and evaluation* (ibid., pp. 18-19).<sup>1</sup>

Though the taxonomy was intended as a tool for assessing the objectives of instructional methods, this classification system has been adapted into a means of assessing *the quality of learners' intellectual activity*. For example, a study on agriculture education analyzed recordings of professors' lectures and reports of students' thoughts during said lectures, classified these verbalizations according to their level in Bloom's Taxonomy, and concluded that the "cognitive level" of both professors' and students' intellectual activity heavily favored "the two lowest levels of cognition" (Ewing & Whittington 2009, p. 45). On this basis, the authors express a qualitative judgment about the inadequacy of the instruction relative to the goal of getting students to "think at higher cognitive levels" (ibid.). The understanding of "cognition" operative in this study is a normatively-laden one, in which "higher level" cognition is considered qualitatively superior to "lower level" cognition, and individuals should be engaging in "higher level" cognition as often as possible.<sup>2</sup>

Some authors swap in "higher order cognition" for cognitive activities they attribute to a higher skill level; see for example *Figure 5*, in which Lemons & Lemons (2013) list cognitive

---

<sup>1</sup> A widely-embraced revision to Bloom's taxonomy has been proposed, in which the two top levels are swapped, and all six categories are renamed as *remembering, understanding, applying, analyzing, evaluating, and creating* (Anderson, Krathwohl et al. 2001).

<sup>2</sup> Applications of Bloom's Taxonomy to assess the "level" of students' cognitive activity as opposed to assessing *instructional methods* (like that found in Ewing & Whittington 2009, Aspy & Roebuck 1972, Willson 1973) could be regarded as a misunderstanding or perversion of the classification system. However, some authors seem to think that assessing *students* via the taxonomy is an obvious extension of Bloom's framework. E.g., Forehand (2010) appears to defend the use of the taxonomic categories to evaluate the "level" of student thinking: "Throughout the years, the levels have often been depicted as a stairway, leading many teachers to encourage their students to 'climb to a higher (level of) thought.' . . . One can easily see how this arrangement [of levels] led to natural divisions of lower and higher level thinking." (p. 2)

She also implies that education practitioners require some sort of system for judging students' level of cognitive ability, and that Bloom just so happened to provide a workable system, whether he intended to do so or not: "Out of necessity, teachers must measure their students' ability. Accurately doing so requires a classification of levels of intellectual behavior important in learning. Bloom's Taxonomy provided the measurement tool for thinking" (Forehand 2010, pp. 4-5).

activities at levels 3-6 of Bloom’s Taxonomy as “higher order.”<sup>3</sup> Similarly, “higher order thinking” is the terminology with which some education researchers have chosen to refer to a class of desirable cognitive activities that “enhance the construction of deeper, conceptually-driven understanding” (Schraw & Robinson 2011, p. 2).

Cognitive skill level	Actions required by cognitive skill level
Lower order	<ul style="list-style-type: none"> <li>Recall (memorize) facts, figures, and basic processes</li> <li>Know vocabulary and definitions</li> <li>Understand and illustrate information</li> <li>Includes Bloom’s categories of knowledge and comprehension</li> </ul>
Higher order	<ul style="list-style-type: none"> <li>Use information, methods, concepts, or theories in new situations</li> <li>Predict consequences and outcomes</li> <li>Solve problems in which students must select the approach to use</li> <li>Break down a problem into its parts</li> <li>Identify the critical components of a new problem</li> <li>See patterns and organization of parts (e.g., classify, order)</li> <li>Determine the quality/importance of different pieces of information</li> <li>Discriminate among ideas</li> <li>Weigh the relative value of different pieces of evidence to determine the likelihood of certain outcomes/scenarios</li> <li>Make choices based on reasoned argument</li> <li>Includes Bloom’s categories application, analysis, and evaluation</li> </ul>

**FIGURE 5: EXAMPLE IN WHICH “HIGHER ORDER” DESCRIBES COGNITION AT AN ADVANCED SKILL LEVEL.** Table reprinted from Lemons & Lemons (2013).

In short, since educators sometimes refer to advanced, complex, or otherwise laudable forms of cognitive activity among students as “higher-level” or “higher order,” “metacognition” can be misconstrued as designating *complex forms of cognition*, such as analysis, synthesis, criticism, and forming hypotheses (synonymous with upper levels of the Revised Bloom’s Taxonomy), and/or *cognition which is qualitatively superlative*.

<sup>3</sup> See also Heer (n.d.) and The Learning Center at University of North Carolina Chapel Hill (n.d.).

### §A3 | Self-referential cognition, or “thinking about thinking”

One can also interpret “*meta-X*” to signify a semantic function of *X about X* – replicating the pattern wherein *meta-jokes* refers to jokes about jokes, or *meta-data* refers to data about data. This yields a translation of “metacognition” as *cognition about cognition* – but how to unpack this depends very heavily on what you take “cognition” to be in each of the two positions in that phrase.

One axis of ambiguity embedded in “cognition about cognition” is whether the cognition mentioned in the latter position is *indexed* to any particular individual’s cognitive system, or whether it is intended *generically* to encompass the cognitive activity that is shared across all individual cognitive systems. Those who adopt the former interpretation understand “metacognition” as cognition about *one’s own* cognition, while the latter interpretation allows “metacognition” to “encompass thoughts about the cognitive states and processes of other subjects” (Carruthers 2008, p. 58; my italics).

Beyond this, I can identify two additional axes of ambiguity in “cognition about cognition,” though these two in practice tend to overlap. One is the level of description at which cognitive states or processes are being viewed, ranging on a spectrum from highly *atomistic* analyses (of individual mental states, or instances of processing in service of executing a specific cognitive task) to maximally *systemic* analyses (looking at the function of one’s entire cognitive apparatus). The second is the timeframe on which cognitive states or processes are being analyzed: one can perform *synchronous* analysis of occurrent mental states or processes, or on *asynchronous* cognition – most commonly that which occurred in the past across any stretch of time, but possibly also at future cognitive activity, as occurs in planning. Typically, when cognition is analyzed atomistically, the specific states and processes being considered are

occurrent states being examined synchronously; when cognition is analyzed at the level of the whole system, the target of analysis are past or future cognitions considered asynchronously.

Let's assume for the moment that we are zeroing in on metacognition that consists of cognition about *one's own* cognition. Even with that constraint applied, "metacognition" would include both a propositional attitude one adopts about a specific occurrent first-order mental representation ("I believe that the cat is on the mat") *and* a propositional attitude one adopts about one's large-scale, long-term cognitive attributes ("I believe that I often engage in post-hoc rationalization," or "I believe that I make decisions primarily on the basis of intuition rather than deliberation"). Both of these have some claim to qualify as cognitive *self-awareness* or *self-knowledge*, but the latter seems much closer to what speakers invoke when they claim that developing "metacognition" allows a student to become a "self-regulating learner" capable of exercising strategic control over their cognitive activities.

Yet another layer of complication arises regarding "*what kind of* cognition about cognition" qualifies as "metacognition." For example, does this refer to *implicit* (unconscious, automatic) and *explicit* (conscious, deliberate, controlled) cognition about cognition? This distinction is at the heart of the debate outlined in §2.3, as to whether metacognition is *self-evaluative* (a collection of implicit self-regulatory processes, built into our cognitive architecture to optimize the functioning of our cognitive operations) or *self-attributive* (an explicit process of forming conscious beliefs about one's own cognitive activities).

It's also worth noting that the boundaries of "cognition" are a subject of ongoing debate. The orthodox model of mental architecture – memorably coined "the classical sandwich" by Susan Hurley – positions cognition as "the central core of the mind" (akin to sandwich filling), whereas perception and behavior are peripheral and interface with the world (like the two slices of bread enclosing the filling; 2001, p. 3). Besides being "central" by being enclosed in the

system's interior and partitioned off from the external world (except for the inputs it receives from perception, and the outputs it delivers to action), cognition is "central" within a unidirectional scheme of information processing, from perception to cognition to action. However, this model has been challenged on two separate fronts.

The debate over the perception/cognition border<sup>4</sup> challenges the "cognitive" designation of mental states bearing *phenomenal* content (i.e., having qualitative character belonging to a subject's experience) – and in turn calls into question the classification of *epistemic feelings* as "metacognitive." These labels capture a diverse bunch of mental states ranging from sensations of "fluency, novelty, 'eureka' moments, and the like to more nebulous states like the 'tip of the tongue' state,<sup>5</sup> the 'feeling of knowing', and the 'feeling of forgetting'" – but what all of these share in common is that they are "non-linguistic representations of one's cognition through affective states" (Green 2019, p. 119).

Epistemic feelings occupy a strange territory in our mental landscape. Feelings in general seem to straddle the border between perception and cognition, because they can involve both a subjective *phenomenal* element (something that it is like to be in that state), which likens them to the outputs of perception, and *intentionality* (being about something), which aligns them with cognitive phenomena such as judgments or evaluations (de Sousa 2018). Epistemic feelings thus inherit this ambiguity from the broader class to which they belong, but they add on an additional complication of adopting cognitive states and processes as their intentional

---

<sup>4</sup>Advocates for a border include Carey (2009), Burge (2010), Block (2014), Firestone and Scholl (2014, 2015), Raftapoulos (2015), Mandelbaum (2017), Phillips (2017), and Beck (2018); notable opponents are Clark (2013), Lupyan (2015), Shea (2015).

<sup>5</sup> First discussed in print over a century ago (James 1893), David Rosenthal describes this as "the vivid feeling, familiar to us all, that a word or other piece of information is on the tip of one's tongue" (2000, p. 204): the sensation that one's memory stores contain an item that one is trying to recall, even if one is not able to recall the information in this exact moment. This state presents itself to us a "subjective conviction that one 'knows' the sought-after [item, despite] the actual inability to produce it" (Koriat 2000, p. 151). Moreover, this state seems to be capable of informing behavior, as when a student taking an exam "may decide to dwell more on that question if she feels that the name is on the tip-of-the-tongue than if not" (p. 156).

objects. So, does being *about* phenomena which are canonically cognitive cement the status of epistemic feelings as belonging to cognition? Or does their phenomenal character mark them as something outside of cognition, even though they are intimately related to cognitive activity? I'm not going to resolve this question here (but see Arango-Muñoz 2014 for an interesting discussion); this is just to express that some phenomena labelled as "metacognitive" (like forming explicit, propositionally-expressible beliefs about one's own cognition) more obviously qualify as *cognition* about cognition compared to other phenomena (like epistemic feelings) that exist along the border between cognition and non-cognition.

The other boundary dispute regarding the concept of cognition has to do with the relationship between cognition and action/behavior. A movement variously known as *embodied, enactive, extended, embedded and/or situated cognition* (often summarized as "4E cognition") resists stark divisions between "mental" activity and "bodily" activity.<sup>6</sup> As Clark and Chalmers pithily put it in an early & influential paper, "cognitive processes ain't (all) in the head!" (1998, p. 8); rather, cognition extends out into the physical world to encompass actions which "alter the world so as to aid and augment cognitive processes such as recognition and search" (ibid.). Under these author's "extended" construal of cognition, bodily actions involving manipulation of objects in one's environment – such as using a pen and paper to compute the result of a multiplication problem – belong to cognition, rather than being a physical consequence of cognitive activity. Detractors of this movement argue that cognitive scientists may acknowledge the causal influence of bodily action and environmental manipulation on cognitive processes without regarding an agent's body and environment as constitutive parts of their cognitive system.<sup>7</sup>

---

<sup>6</sup> This list is *far* from exhaustive: Varela, Thompson, and Rosch (1991), Hutchins (1995), Rowlands (1999), Gallagher (2005), Chemero (2009), Shapiro (2011), Hutto & Myin (2013).

<sup>7</sup> Including Rupert (2004), Aizawa (2007), Adams & Aizawa (2008), Adams (2010).



This debate is relevant to adjudicating the “metacognitive” status of phenomena that could alternatively be classified as *behaviors* that inform or are informed by cognition, rather than *cognition* about cognition. Among these phenomena are many classroom activities recommended as means of developing students’ metacognitive skills, like creating visualizations or diagrams of recently-learned concepts (variously referred to as *semantic webs*, *concept charts*, *mind-maps*, and so on); in a similar vein, educators will praise students for “being metacognitive” when they incorporate specific behavioral routines (like spaced practice, retrieval practice, and dual coding<sup>8</sup>) into their study practices. These uses of “metacognition” stretch the meaning of “cognition about cognition” in a way that may be unobjectionable to those of the 4E persuasion, but which represents a large departure from what adherents to the classical model of mental architecture would accept as a subset of “cognition.”<sup>9</sup>

---

<sup>8</sup> See §1.1 for a refresher on these terms.

<sup>9</sup> To complicate matters even further, one can find “metacognitive strategies” included under the banner of “*non-cognitive skills*” – this latter phrase denoting a catch-all category for “attitudes, behaviours, and strategies that are thought to underpin success in school and at work[,] . . . contrasted with the ‘hard skills’ of cognitive ability in areas such as literacy and numeracy” (Gutman & Schoon 2013, p. 2). We could see this as a rejection of the equivalence of “metacognition” and “*cognition* about cognition,” but I expect that this contradiction was an unintentional consequence of lumping metacognitive strategies together with other phenomena (e.g. self-esteem, “grit”) beyond traditional measures of scholastic aptitude which are “increasingly considered to be as important as, or even more important than, cognitive skills or IQ in explaining academic and employment outcomes” (Shechtman et al. 2013).

## REFERENCES

- Adams, F. (2010). Embodied cognition. *Phenomenology and Cognition*, 9(4), 619-628.
- Adams, F., & Aizawa, K. (2009). "Why the mind is still in the head." In P. Robbins & M. Aydede (Eds.), *The Cambridge Handbook of Situated Cognition* (pp. 78-95). Cambridge University Press.
- Adams, M.J. (1990). *Beginning to Read: Thinking & Learning about Print*. MIT Press.
- Agarwal, P.K., Roediger, H.L., McDaniel, M.A., McDermott, K.B. (2020). How to use retrieval practice to improve learning. *RetrievalPractice.org*.  
<http://pdf.retrievalpractice.org/RetrievalPracticeGuide.pdf>
- Aizawa, K. (2007). Understanding the embodiment of perception. *Journal of Philosophy*, 104, 5-25.
- Alfano, M. (2013). *Character as Moral Fiction*. Cambridge University Press.
- Alferink, L.A. & Farmer-Dougan, V. (2010). Brain-(not) based education: dangers of misunderstanding and misapplication of neuroscience research. *Exceptionality*, 18, 42-52.
- Al Kuhayli, H.A., Pilotti, M., El Alaoui, K., Cavazos, S.E., Hassan, S.A.M., Al Ghazo, R. (2019). An exploratory non-experimental design of self-assessment practice. *The International Journal of Assessment and Evaluation.*, 26(1), 49-65.
- Ambrose, S.A., Bridges, M.W., DiPiero, M., Lovett, M.C., & Norman, M.K. (2010). *How Learning Works: 7 Research-Based Principles for Smart Teaching*. Jossey-Bass.
- Anderson, L.W., Krathwohl, D.R., Airasian, P.W., Cruikshank, K.A., Mayer, R.E., Pintrich, P.R., Raths, J., & Wittrock, M.C. (Eds.). (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's 'Taxonomy of Educational Objectives'*. Addison Wesley Longman.
- Artelt, C., Schiefele, U., & Schneider, W. (2001). Predictors of reading literacy. *European Journal of Psychology of Education*, 16(3), 363-383.
- Ashman, G. (2016, Oct 27). Mirror mirror on the web. *Filling the Pail*.  
<https://gregashman.wordpress.com/2016/10/27/mirror-mirror-on-the-web/>
- Ashman, G. (2018a, April 28). A new report on 'metacognition and self-regulated learning': What does it mean? *Filling the Pail*. <https://gregashman.wordpress.com/2018/04/28/a-new-report-on-metacognition-and-self-regulated-learning-what-does-it-mean/>
- Ashman, G. (2018b, June 24). What are the implications of John Hattie being wrong? *Filling the Pail*. <https://gregashman.wordpress.com/2018/06/24/what-are-the-implications-of-john-hattie-being-wrong/>
- Aspy, D.N., & Roebuck, F.N. (1972). An investigation of the relationship between student levels of cognitive functioning and the teacher's classroom behavior. *The Journal of Educational Research*, 65(8), 365-368.

- Austin, R.N. (2006). Writing the teaching statement. *Science Careers*.  
<https://www.sciencemag.org/careers/2006/04/writing-teaching-statement>
- Baehr, J. (n.d.). *What is educating for intellectual virtues?* Educating for Intellectual Virtues.  
<https://intellectualvirtues.org/what-is-educating-for-intellectual-virtues/>
- Baehr, J. (2013). Educating for intellectual virtues: From theory to practice. *Journal of Philosophy of Education*, 47(2), 248-62.
- Baehr, J. (2015) *Educating for Intellectual Virtues: An Introductory Guide for College and University Instructors*. [https://jasonbaehr.files.wordpress.com/2013/12/e4iv\\_baehr.pdf](https://jasonbaehr.files.wordpress.com/2013/12/e4iv_baehr.pdf)
- Baehr, J. (Ed.) (2017a) *Intellectual Virtues and Education: Essays in Applied Virtue Epistemology*. Routledge.
- Baehr, J. (2017b) "The situationist challenge to educating for intellectual virtues." In A. Fairweather & M. Alfano (Eds.), *Epistemic Situationism* (pp. 192-215). Oxford University Press.
- Baehr, J. (2019) Intellectual virtues, critical thinking, and the aims of education. In P. Graham, M. Fricker, D. Henderson, N. Pedersen, and J. Wyatt (eds.), *Routledge Handbook of Social Epistemology*. Routledge.
- Baker, L., & Brown, A.L. (1981). Metacognition and the reading process. In D. Pearson (Ed.), *Understanding Reading Comprehension*. International Reading Association.
- Baker, L., & Brown, A.L. (1984). "Metacognitive skills in reading." In P.D. Pearson (ed.), *Handbook of Reading Research*. Longman.
- Baker, L., & Zimlin, L. (1989). Instructional effects on children's use of two levels of standards for evaluating their comprehension. *Journal of Educational Psychology*, 81(3), 340-346.
- Baker, L. & Cerro, L.C. (2000). "Assessing metacognition in children and adults." In G. Schraw & J.C. Impara (Eds.), *Issues in the Measurement of Metacognition* (pp. 93-145). Buros Institute of Mental Measurements.
- Bailin, S., Case, R., Coombs, J.R., & Daniels, L.B. (1999). Common misconceptions of critical thinking. *Journal of Curriculum Studies*, 31(3), 269-283.
- Bandura, A. (1994). "Self-efficacy." In V.S. Ramachaudran (Ed.), *Encyclopedia of Human Behavior* (Vol. 4, pp. 71-81). New York: Academic Press.
- Bar-Anan, Y., Wilson, T.D., & Hassin, R.R. (2010). Inaccurate self-knowledge formation as a result of automatic behavior. *Journal of Experimental Social Psychology* 46, 884-894.
- Battaly, H. (2006). Teaching intellectual virtues: Applying virtue epistemology in the classroom. *Teaching Philosophy*, 29(3), 191-222.
- Baumeister, R. & Newman, L.S. (1994). Self-regulation of cognitive inference and decision processes. *Personality & Social Psychology Bulletin*, 20(1), 3-19.

- Beck, J. (2018). Marking the perception-cognition boundary: The criterion of stimulus-dependence. *Australasian Journal of Philosophy*, 96(2), 319-334.
- Bem, D.J. (1972). "Self-perception theory." In L. Berkowitz (ed.), *Advances in Experimental Social Psychology*, Vol. 6 (pp. 1-62). Academic Press.
- Bernard L. Schwartz Communication Institute at Baruch College. (n.d.). *Metacognition*. Very Short Guides. <https://blogs.baruch.cuny.edu/blsci/files/2018/06/2-Metacognition.pdf>
- Bialystok, L. (2017). Philosophy across the curriculum and the question of teacher capacity; or, what is philosophy and who can teach it? *Journal of Philosophy of Education*, 51(4), 817-836.
- Blackmore, S. (2002). There is no stream of consciousness. *Journal of Consciousness Studies*, 9, 17-28.
- Blackmore, S. (2016). Delusions of consciousness. *Journal of Consciousness Studies*, 23, 52-64
- Block, N. (2014). Seeing-as in the light of vision science. *Philosophy and Phenomenological Research*, 89(1), 560-572.
- Bloom, B.S., Engelhart, M.D., Furst, E.J., Hill, W.K., & Krathwohl, D.R. (1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook 1: The Cognitive Domain*. David McKay Co., Inc.
- Boekaerts, M. (1999). Self-regulated learning. *International Journal of Educational Research*, 31(6), 445-551.
- Bogdan, R.J. (2003). *Minding Minds: Evolving a Reflexive Mind by Interpreting Others*. MIT Press.
- Bonjour, L. (1985). *The Structure of Empirical Knowledge*. Harvard University Press.
- Borkowski, J., Carr, M., & Pressley, M. (1987). "Spontaneous" strategy use: Perspectives from metacognitive theory. *Intelligence*, 11, 61-75.
- Borkowski, J. G., Chan, L. K. S., & Muthukrishna, N. (2000). "A process-oriented model of metacognition: Links between motivation and executive functioning." In G. Schraw, & J. C. Impara (eds.), *Issues in the Measurement of Metacognition* (pp. 1-43). Buros Institute of Mental Measurements.
- Boström, L., & Lassen, L.M. (2006). Unraveling learning, learning styles, learning strategies and meta-cognition. *Education and Training*, 48 (2-3), 178-189.
- Bransford, J.D., Stein, B.S., Vye, N.J., Franks, J.J., Auble, P.M., Mezynski, K J. & Perfetto, G.A. (1982). Differences in approaches to learning: An overview. *Journal of Experimental Psychology: General* 111, 390-398.
- Bransford, J., Sherwood, R., Vye, N., & Rieser, J. (1986). Teaching thinking and problem solving: Research foundations. *American Psychologist*, 41(10), 1078.

- Bransford, J.D., Brown, A.L., & Cocking, R.R. (eds). (2000). *How People Learn: Brain, Mind, Experience and School*. National Academy Press.
- Braten, I. (1991). Vygotsky as precursor to metacognitive theory: II. Vygotsky as metacognitivist. *Scandinavian Journal of Educational Research*, 35 (4), 305-320.
- Brentano, F. (1874/1995). *Psychology From an Empirical Standpoint*. Routledge.
- Broncano, F. (2014). Daring to believe: Metacognitive, epistemic agency, and reflective knowledge. In A. Fairweather (ed.), *Virtue Epistemology Naturalized: Bridges Between Virtue Epistemology and Philosophy of Science* (49-66). Springer International.
- Brown, A.L., Campione, J.C., & Barclay, C.R. (1979). Training self-checking routines for estimating test readiness: Generalization from list learning to prose recall. *Child Development*, 50(2), 501-512.
- Brown, A.L. & DeLoache, J.S. (1978). "Skills, plans, and self-regulation." In R.S. Siegler (Ed.), *Children's Thinking: What Develops?* Lawrence Erlbaum Associates.
- Brown, A.L., Bransford, J.D., Ferrara, R.A., Campione, J.C. (1983). Learning, remembering, and understanding. In J.H. Flavell & E.M. Markman (eds.), *Carmichael's Manual of Child Psychology*, Vol. 1. Wiley.
- Brown, A.L. (1987). Metacognition, executive control, self-regulation, and other more mysterious mechanisms. In F.E. Weinert & R.H. Kluwe (eds.), *Metacognition, Motivation, and Understanding* (pp. 65-116). Lawrence Erlbaum Associates.
- Brown, A.L. (1990). Domain-specific principles affect learning and transfer in children. *Cognitive Science*, 14, 107-133.
- Brown, G. (1984). Metacognition: New insights into old problems? *British Journal of Educational Studies*, 32(3), 213-219.
- Brown, J., Huntley, D., Morgan, S., Dodson, K.D., Cich, J. (2017). Confabulation: A guide for mental health professionals. *International Journal of Neurology and Neurotherapy* 4(2).
- Browne, D. (2004). Do dolphins know their own minds? *Biology and Philosophy*, 19, 633-653.
- Bruer, J.T. (1993). *Schools for Thought: A Science of Learning in the Classroom*. MIT Press.
- Bruer, J.T. (1997). Education and the brain: A bridge too far. *Educational Researcher*, 26(8), 4-16.
- Burge, T. (2010). *Origins of Objectivity*. Oxford University Press.
- Buzan, T. (1993). *The Mind Map Book: How to Use Radiant Thinking to Maximize Your Brain's Untapped Potential*. Plume Penguin.
- Carpenter, S.K. & Agarwal, P.K. (2020). How to use spaced retrieval practice to improve learning. *RetrievalPractice.org*. <http://pdf.retrievalpractice.org/SpacingGuide.pdf>

- Carruthers, P. (2009a). How we know our own minds: The relationship between mindreading and metacognition. *Behavioral and Brain Sciences*, 32, 121-82.
- Carruthers, P. (2009b). An architecture for dual reasoning. In J.St.B.T. Evans & K. Frankish (Eds.) *In Two Minds: Dual Processes and Beyond*. Oxford University Press.
- Carruthers, P. (2011). *The Opacity of Mind: An Integrated Theory of Self-Knowledge*. Oxford University Press.
- Carruthers, P. & Gennaro, R. (2020). "Higher-Order Theories of Consciousness." In E.N. Zalta (ed.). *The Stanford Encyclopedia of Philosophy* (Fall 2020 Edition).  
<https://plato.stanford.edu/archives/fall2020/entries/consciousness-higher>
- Casselmann, B.L., & Atwood, C.H. (2017). Improving general chemistry course performance through online homework-based metacognitive training. *Journal of Chemical Education*, 94, 181-21.
- Cavanaugh, J.C. & Permuter, M. (1982). Metamemory: A critical examination. *Child Development*, 53(1), 11-28.
- Caviglioli, O. (2019). *Dual Coding With Teachers*. John Catt Educational.
- Centre for Teaching Excellence, University of Waterloo. (n.d.). *Teaching metacognitive skills*.  
<https://uwaterloo.ca/centre-for-teaching-excellence/teaching-resources/teaching-tips/metacognitive>
- Chater, N. (2018). *The Mind is Flat: The Illusion of Mental Depth and the Improvised Mind*. Allen Lane.
- Chemero, T. (2009). *Radical Embodied Cognitive Science*. MIT Press.
- Chen, G.A., Marshall, S.A., Horn, I.S. (2020). 'How do I choose?': Mathematics teachers' sensemaking about pedagogical responsibility. *Pedagogy, Culture & Society*.
- Chick, N. (n.d.). *SoTL Guide*. Vanderbilt Center for Teaching. <https://my.vanderbilt.edu/sotl/>
- Cholbi, M. (2007). Intentional learning as a model for philosophical pedagogy. *Teaching Philosophy*, 30(1), 35-58.
- Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, 58(1), 7-19.
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36, 181-204.
- Coe, R. (2002, Sept. 12-14). *It's the Effect Size, Stupid: What effect size is and why it is important* [Paper presentation]. Annual Conference of the British Educational Research Association, University of Exeter, England.  
<https://www.leeds.ac.uk/educol/documents/00002182.htm>

- Coffield, F., Moseley, D., Hall, E., & Ecclestone, K. (2004). Learning styles and pedagogy in post-16 learning: A systematic and critical review. Learning and Skills Research Centre.
- Cohen, I. & Goldsmith, M. (2000). *Hands On: How to Use Brain Gym® in the Classroom*. Hands On Books.
- Cohen, J. (1969). *Statistical Power Analysis for the Behavioral Sciences*. Academic Press.
- Collins, A., Brown, J.S., & Newman, S.E. (1988). Cognitive apprenticeship: Teaching the craft of reading, writing, and mathematics. *Thinking: The Journal of Philosophy for Children*, 8(1), 2-10.
- Colman, A.M. (2008). "Clever Hans." In his *A Dictionary of Psychology, 3rd Ed.* Oxford University Press.  
<https://www.oxfordreference.com/view/10.1093/oi/authority.20110803095617509>
- Columbia Center for Teaching and Learning. (n.d.). *Metacognition*. Teaching Resources.  
<https://ctl.columbia.edu/resources-and-technology/resources/metacognition/>
- Concepción, D. W. (2004). Reading philosophy with background knowledge and metacognition. *Teaching Philosophy*, 27(4), 351-368.
- Craig, K., Hale, D., Grainger, C., & Stewart, M.E. (2020). Evaluating metacognitive self-reports: systematic reviews of the value of self-report in metacognitive research. *Metacognition and Learning*, 15, 155-213.
- Crooks, S. (2009). Teaching for argumentative thought. *Teaching Philosophy*, 32(3), 247-261.
- Cross, D.R. & Paris, S.G. (1988). Developmental and instructional analyses of children's metacognition and reading comprehension. *Journal of Educational Psychology*, 80(2), 131-142.
- Cunningham, A.E. & Stanovich, K.E. (2001). What reading does for the mind. *Journal of Direct Instruction*, 1(2), 137-149
- Darling-Hammond, L., Austin, K., Cheung, M., & Martin, D. (2003). *Thinking About Thinking: Metacognition*. The Learning Classroom: Theory into Practice.  
[https://www.learner.org/wp-content/uploads/2019/02/The-Learning-Classroom\\_Thinking-About-Thinking.pdf](https://www.learner.org/wp-content/uploads/2019/02/The-Learning-Classroom_Thinking-About-Thinking.pdf)
- De Jager, B., Jansen, M., & Reezigt, G. (2005). The development of metacognition in primary school learning environments. *School Effectiveness and School Improvement*, 16(2), 179-196.
- Delpit, L. (2013). *"Multiplication Is for White People": Raising Expectations for Other People's Children*. The New Press.
- Dennett, D.C. (1969) *Content and Consciousness*. Humanities Press.
- Dennett, D. C. (1978). *Brainstorms: Philosophical essays on mind and psychology*. Montgomery, VT: Bradford Books.

- Dennett, D.C. (1987). *The Intentional Stance*. MIT Press.
- Dennett, D.C. (1991). *Consciousness Explained*. Little, Brown and Co.
- Dennett, D.C. (1992). "The self as a center of narrative gravity." In F. Kessel, P. Cole and D. Johnson (Eds.), *Self and Consciousness: Multiple Perspectives*. Erlbaum.
- de Sousa, R. (2008). Epistemic feelings. In G. Brun, U. Doguoglu, & D. Kuenzle (Eds.), *Epistemology and Emotions* (pp. 185–204). Ashgate Publishing Limited.
- de Sousa, R. (2018). Emotion. In E.N. Zalta (ed.), *Stanford Encyclopedia of Philosophy* (Fall 2018 ed.). <https://plato.stanford.edu/entries/emotion/>
- Devonshire, I.M., & Dommett, E.J. (2010). Neuroscience: viable applications in education? *Neuroscientist*, 16, 349–356.
- Dewey, J. (1910/1933). *How We Think*. Prometheus Books.
- Dewey, J. (1916/1944). *Democracy and Education*. Free Press.
- Dignath, C., Büttner, G., & Langfeldt, H-P. (2008). Components of fostering self-regulated learning among students: A meta-analysis on intervention studies at primary and secondary school level. *Metacognition and Learning*, 3(3), 231–264.
- Dinsmore, D.L., Alexander, P.A., & Loughlin, S.M. (2008). Focusing the conceptual lens on metacognition, self-regulation, and self-regulated learning. *Educational Psychology Review*, 20(4), 391–409.
- Donker, A.S., De Boer, H., Kostons, D., Dignath van Ewijk, C.C., & Van der Werf, M. P.C. (2014). Effectiveness of learning strategy instruction on academic performance: A meta-analysis. *Educational Research Review*, 11, 1–26.
- Dunlosky, J., & Metcalfe, J. (2009). *Metacognition*. SAGE Publications.
- Dunn, R., & Dunn, K. (1978). *Teaching Students Through Their Individual Learning Styles*. Reston Publishing.
- Dweck, C.S. (1998). "The development of early self-conceptions: Their relevance for motivational processes." In J. Heckhausen & C.S. Dweck (Eds.), *Motivation and self-regulation across the life span* (pp. 257–280). Cambridge University Press.
- Dweck, C.S. (2006). *Mindset: The New Psychology of Success*. Ballantine Books.
- Dweck, C.S. (2016, Jan 11). Recognizing and overcoming false growth mindset. *Edutopia*. <https://www.edutopia.org/blog/recognizing-overcoming-false-growth-mindset-carol-dweck>
- Eberly Center for Teaching Excellence and Educational Innovation (n.d.). *Exam Wrappers*. Carnegie Mellon University. <https://www.cmu.edu/teaching/designteach/teach/examwrappers/>



- Edelsky, C., Altwerger, B., & Flores, B. (1991). *Whole Language: What's the Difference?* Heinemann.
- Efklides, A. (2001). Metacognitive experiences in problem solving: Metacognition, motivation, and self-regulation. In A. Efklides, J. Kuhl, & R. M. Sorrentio (Eds.), *Trends and Prospects in Motivation Research* (pp. 297-324). Kluwer Academic Publishers.
- Epley, N. & Gilovitch, T. (2016). The mechanisms of motivated reasoning. *Journal of Economic Perspectives*, 30(3), 133-140.
- Ericsson, K.A. & Simon, H.A. (1980). Verbal reports as data. *Psychological Review*, 87(3), pp. 215-251.
- Erickson, F. & Gutierrez, K. (2002). Comment: Culture, rigor, and science in educational research. *Educational Researcher*, 31: 21-24.
- Erhlinger, J., Johnson, K., Banner, M., Dunning, D., Kruger, J. (2008). Why the unskilled are unaware: Further explorations of (absent) self-insight among the incompetent. *Organizational Behavior and Human Decision Processes*, 105(1), 98-121.
- Evans, J.St.B.T. & Over, D. (1996). *Rationality and Reasoning*. Psychology Press.
- Evans, J.St.B.T. (2003). In two minds: Dual process accounts of reasoning. *Trends in Cognitive Sciences*, 7, 454-459.
- Evans, J.St.B.T. & Frankish, K. (Eds.) (2009). *In Two Minds: Dual Processes and Beyond*. Oxford University Press.
- Evidence for Learning. (2016). Technical Appendix: Meta-cognition and self-regulation. <https://evidenceforlearning.org.au/assets/Oct-2016-Toolkit-Technical-appendix/Toolkit-Meta-cognition-and-self-regulation-E4L-Technical-Appendix.pdf>
- Ewing, J.C., & Whittington, M.S. (2009). Describing the cognitive level of professor discourse and student cognition in college of agriculture class sessions. *Journal of Agriculture Education*, 50(4), 36-49.
- Feuer, M.J., Towne, L., & Shavelson, R.J. (2002). Scientific culture and educational research. *Educational Researcher*, 31(8), 4-14.
- Firestone, C. & Scholl, B. (2016). Cognition does not affect perception: Evaluating the evidence for 'top-down' effects. *Behavioral and Brain Sciences*, 39, 1-72.
- Fisher, R. (1998). Thinking about thinking: Developing metacognition in children. *Early Child Development and Care*, 141, 1-15.
- Flavell, J.H., Friedrichs, A.G., & Hoyt, J.D. (1970). Developmental changes in memorization processes. *Cognitive Psychology*, 1, 324-340.
- Flavell, J.H. (1976). "Metacognitive aspects of problem solving." In L.B. Resnick (ed.), *The Nature of Intelligence* (pp. 231-235). Lawrence Erlbaum.

- Flavell, J.H., & Wellman, H.M. (1977). "Metamemory." In R.V. Kail, Jr. & J.W. Hagen (Eds.), *Perspectives on the Development of Memory and Cognition*. Lawrence Erlbaum Associates.
- Flavell, J.H. (1978). "Metacognitive development." In J.M. Scandure & C. J. Brainerd (Eds.), *Structural Process Theories of Complex Human Behavior*. Sijthoff.
- Flavell, J.H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, 34(10), 906-911.
- Flynn, D.J., Nyhan, B., & Reifler, J. (2017). The nature and origins of misperceptions: Understanding false and unsupported beliefs about politics. *Political Psychology*, 38, 127-150.
- Forehand, M. (2010) Bloom's Taxonomy. In M. Orey (Ed.), *Emerging Perspectives on Learning, Teaching, and Technology*.  
[http://epltt.coe.uga.edu/index.php?title=Bloom%27s\\_Taxonomy](http://epltt.coe.uga.edu/index.php?title=Bloom%27s_Taxonomy)
- Forrest-Pressley, D.L., MacKinnon, G.E., Waller, T.G. (Eds.). (1985). *Metacognition, Cognition, and Human Performance, Vol. 2*. Academic Press.
- Frankfurt, H.G. (1971). Freedom of the will and the concept of a person. *The Journal of Philosophy*, 68(1), 5-20.
- Frankish, K. (2016a). Illusionism as a theory of consciousness. *Journal of Consciousness Studies*, 23(11-12), 11-39.
- Frankish, K. (2016b). Whatever you think, you don't necessarily know your own mind. *Aeon Magazine*. <https://aeon.co/ideas/whatever-you-think-you-don-t-necessarily-know-your-own-mind>
- Gallagher, S. (2005). *How the Body Shapes the Mind*. Oxford University Press.
- Garner, R. & Alexander, P.A. (1989). Metacognition: Answered and unanswered questions. *Educational Psychologist* 24(2), 143-158.
- Gazzaniga, M. (1995). "Consciousness and the cerebral hemispheres." In M. Gazzaniga (Ed.), *The Cognitive Neurosciences*. MIT Press.
- Geake, J. (2008). Neuromythologies in education. *Educational Research*, 50, 123-133.
- Gilovich, T. (1993). *How We Know What Isn't So: The Fallibility of Human Reason in Everyday Life*. Free Press.
- Girash, J. (2014). "Metacognition and Instruction." In V.A. Bennassi, C.E. Overton, & C.M. Hakala (eds.), *Applying Science of Learning: Infusing Psychological Science into the Curriculum* (pp. 152-68). Society for the Teaching of Psychology, Division 2, American Psychological Association.
- Goldman, A.I. (2006). *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*. Oxford University Press.

- Gough, P. & Tunmer, W. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, 7, 6-10.
- Goodman, K.S. (1996). *Ken Goodman on Reading: A Common Sense Look at the Nature of Language and the Science of Reading*. Heinemann.
- Gopnik, A. (1996) The post-Piaget era. *Psychological Science*, 7(4), 221-225.
- Green, J. (2019). Metacognition as an epistemic virtue. *Southwest Philosophy Review*, 35(1), 117-129.
- Gutman, L.M. & Schoon, I. (2013). *The impact of non-cognitive skills on outcomes for young people* [Literature review]. Institute of Education.
- Hacker, D.J. (1998) "Definitions and empirical foundations." In D.J. Hacker, J. Dunlosky, & A.C. Graesser (eds.), *Metacognition in Educational Theory and Practice*. Lawrence Erlbaum.
- Haller, E.P., Child, D.A., & Walberg, H.J. (1988). Can comprehension be taught? A quantitative synthesis of metacognitive studies. *Educational Researcher*, 17(9), 5-8.
- Halpern, D.F. (1996). *Thought and Knowledge: An Introduction to Critical Thinking*. Lawrence Erlbaum Associates.
- Hanford, E. (2019, Aug 22). At a loss for words: How a flawed idea is teaching millions of kids to be poor readers. *APM Reports*.  
<https://www.apmreports.org/episode/2019/08/22/whats-wrong-how-schools-teach-reading>
- Hanford, E. (2020a, Jan 27). Experts say widely used reading curriculum is failing kids. *APM Reports*. <https://www.apmreports.org/episode/2020/01/27/lucy-calkins-reading-materials-review>
- Hanford, E. (2020b, Aug 6). What the words say. *APM Reports*.  
<https://www.apmreports.org/episode/2020/08/06/what-the-words-say>
- Harriet W. Sheridan Center for Teaching and Learning, Brown University. (n.d.). *Promoting Metacognition*. <https://www.brown.edu/sheridan/teaching-learning-resources/teaching-resources/classroom-practices/promoting-metacognition>
- Hartman, H.J. (2001). "Developing students' metacognitive knowledge and skills." In H.J. Hartman (ed.), *Metacognition in Learning and Instruction* (pp. 33-68). Kluwer Academic Publishers.
- Hattie, J. (n.d.). *Hattie Ranking: 252 Influences And Effect Sizes Related To Student Achievement*. Visible Learning. <https://visible-learning.org/hattie-ranking-influences-effect-sizes-learning-achievement/>
- Hattie, J. (2009). *Visible Learning* (1st Ed.). Routledge.

- Hattie, J., Biggs, J., & Purdie, N. (1996). Effects of learning skills interventions on student learning: A meta-analysis. *Review of Educational Research*, 66(2), 99-136.
- Haynes, F. (2014). Teaching children to think for themselves: From questioning to dialogue. *Journal of Philosophy in Schools*, 1(1), 131-146.
- Heer, R. (n.d.). "A model of learning objectives based on *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*." Iowa State University Center for Excellence in Learning and Teaching.
- Hempenstall, K. (2002). The three-cueing system: Help or hindrance? *Direct Instruction News*, 42-51.
- Hempenstall, K. (2016). *Read about it: Scientific evidence for effective teaching of reading*. CIS Research Report 11. Sydney: The Centre for Independent Studies.  
<https://www.cis.org.au/publications/research-reports/read-aboutit-scientific-evidence-for-effective-teaching-of-reading/>
- Hennessey, M.G. (1999). Probing the dimensions of metacognition: Implications for conceptual change teaching-learning. Paper presented at the annual meeting of the National Association for Research in Science Teaching, Boston, MA.
- Hiebert, J., Gallimore, R., & Stigler, J.W. (2002). A knowledge base for the teaching profession: What would it look like and how can we get one? *Education Researcher*, 31(5), 3-15.
- Higgins, S., Hall, E., Baumfield, V., & Moseley, D. (2005). A meta-analysis of the impact of the implementation of thinking skills approaches on pupils. EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.  
[http://eppi.ioe.ac.uk/cms/Portals/0/PDF%20reviews%20and%20summaries/t\\_s\\_rv2.pdf?ver=2006-03-02-125128-393](http://eppi.ioe.ac.uk/cms/Portals/0/PDF%20reviews%20and%20summaries/t_s_rv2.pdf?ver=2006-03-02-125128-393)
- Hirsch, E.D. (2006) *The Knowledge Deficit: Closing the Shocking Education Gap for American Children*. Houghton Mifflin.
- Hornsby, K. L. (2016). The pedagogical imperative: Achieving areté in philosophy graduate programs. *Teaching Philosophy*, 39(1), 51-68.
- Horvath, J.C. & Donaghue, G.M. (2016). A bridge too far – revisited: Reframing Bruer's neuroeducation argument for modern science of learning practitioners. *Frontiers in Psychology*, 7, 1-12.
- Howe, K.R. (2004). A critique of experimentalism. *Creative Inquiry*, 10(4), 42-61.
- Howe, K.R. (2005). The question of educational science: experimentism versus experimentalism. *Educational Theory*, 55(3), 307-321.
- Howe, K.R. (2009). Positivist dogmas, rhetoric, and the education science question. *Educational Researcher*, 38(6), 428-440.

- Howe, K.R. (2013). "Philosophy of social science and educational research." In B.J. Irby, G. Brown, R. Lara-Alecio, S. Jackson (Eds.), *The Handbook of Educational Theories* (pp. 79-86). Information Age Publishing.
- Howhy, J. (2013). *The Predictive Mind*. Oxford University Press.
- Hruby, G.G. (2012). Three requirements for justifying an educational neuroscience. *British Journal of Educational Psychology*, 82, 1-23.
- Hume, D. (1738/2007). *A Treatise of Human Nature*, Vol. 1. Eds. D.F. Norton & M.J. Norton. Oxford University Press.
- Hume, D. (1748/2007). Excerpts from *An Enquiry Concerning Human Understanding*. In (J.E. Adler & C.Z. Elgin (eds.), *Philosophical Inquiry: Classic and Contemporary Readings* (pp. 175-206). Hackett.
- Hurley, S. (2001). Perception and action: Alternative views. *Synthese*, 129, 3-40.
- Hutchins, E. (1995). *Cognition in the Wild*. MIT Press.
- Hutto, D.D., & Myin, E. (2013). *Radicalizing Enactivism: Basic Minds Without Content*. MIT Press.
- Intellectual Virtues Academy. (2012/2016). *Intellectual Virtues Academy Charter Petition*. <http://www.ivalongbeach.org/30-our-charter>
- The Iris Center, Peabody College at Vanderbilt University. (2020). What evidence-based mathematics practices can teachers employ? Page 7: Metacognitive strategies. *High quality mathematics education: What teachers should know*. <https://iris.peabody.vanderbilt.edu/module/math/cresource/q2/p07/>
- [irishchck14]. (2015, November 3). *This is why my kid is going places* [Online forum post]. Reddit. [https://www.reddit.com/r/funny/comments/3rfx5h/this\\_is\\_why\\_my\\_kid\\_is\\_going\\_places/](https://www.reddit.com/r/funny/comments/3rfx5h/this_is_why_my_kid_is_going_places/)
- Issitt, J. (2007). Evidence and metacognition in the new regime of truth: Figures of the autonomous learner on the walls of Plato's cave. *Journal of Philosophy of Education*, 41(3), 381-393.
- Jackson, Y. (2011). *The Pedagogy of Confidence: Inspiring High Intellectual Performance in Urban Schools*. Teachers College Press.
- James, W. (1890). *The Principles of Psychology*. Vol. 1. Holt.
- James, W. (1896). *The Will to Believe and Other Essays in Popular Philosophy*. Project Gutenberg EBook. <https://www.gutenberg.org/files/26659/26659-h/26659-h.htm#P1>
- Johansson, P., Hall, L., Sikström, S., & Olsson, A. (2005). Failure to detect mismatches between intention and outcome in a simple decision task. *Science*, 310(5745), 116-119.

- Johansson, P., Hall, L., Sikström, S., Tärning, B. & Lind, A. (2006). How something can be said about Telling More Than We Can Know: On choice blindness and introspection. *Consciousness and Cognition*, 15, 673-692.
- Kahan, D.M., Jenkins-Smith, H., & Braman, D. (2011). Cultural cognition of scientific consensus. *Journal of Risk Research*, 14(2), 147-174.
- Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin and D. Kahneman (Eds.), *Heuristics of Intuitive Judgment: Extensions and Applications*. Cambridge University Press.
- Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, 58, 697-720.
- Kahneman, D. (2011). *Thinking Fast and Slow*. Farrar, Straus and Giroux.
- Kaplan, A. (2008). Clarifying metacognition, self-regulation, and self-regulated learning: What's the purpose? *Educational Psychology Review* 20, 477-484.
- Kaplan, M., Silver, N., LaVaque-Manty, D. & Meizlish D. (Eds.). (2013). *Using Reflection and Metacognition to Improve Student Learning: Across the Disciplines, Across the Academy*. Stylus Publishing.
- Kilpatrick, D. (2018, March 11). Phonological awareness and intervention. *Cracking the ABC Code*. <https://crackingtheabccode.com/phonological-awareness-intervention/>
- Knowledge is Power Program. (n.d.). *Excellent teaching*. <https://www.kipp.org/approach/highly-effective-teachers-and-leaders/teaching/>
- Kirschner, P.A., Sweller, J. & Clark, R.E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching, *Educational Psychologist*, 41(2), 75-86,
- Kirschner, P.A. (2017). Stop propagating the learning styles myth. *Computers and Education*, 106, 166-171.
- Kitchener, K. S. (1983). Cognition, metacognition, and epistemic cognition: A three-level model of cognitive processing. *Human Development*, 26(4), 222-232.
- Kruger, J. & Dunning, D. (1999). Unskilled and unaware of It: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality & Social Psychology*, 77(6), 1121-1134.
- Koriat, A. (2000). The feeling of knowing: some metatheoretical implications for consciousness and control. *Consciousness and Cognition*, 9, 149-171.
- Kornblith, H. (2002). *Knowledge and Its Place in Nature*. Oxford University Press.
- Kornblith, H. (2012). *On Reflection*. Oxford University Press.

- Korsgaard, C. (1996). *The Sources of Normativity*. Cambridge University Press.
- Kramarski, B. & Mevarech, Z.R. (2003). Enhancing mathematical reasoning in the classroom: The effects of cooperative learning and metacognitive training. *American Educational Research Journal*, 40(1), 281-310.
- Kunda, Z. (1987). Motivation and inference: Self-serving generation and evaluation of evidence. *Journal of Personality and Social Psychology* 53, 636-647.
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108(3), 480-498.
- Lai, E. (2011). Metacognition: A literature review research report. *Pearson Assessments*.
- Lemons, P.P. & Lemons, J.D. (2013). Questions for assessing higher-order cognitive skills: It's not just Bloom's. *CBE – Life Sciences Education*, 12, 47-58.
- Lepock, C. (2014). Metacognition and intellectual virtue. In A. Fairweather (ed.), *Virtue Epistemology Naturalized: Bridges Between Virtue Epistemology and Philosophy of Science* (33-48). Springer International.
- Lehman, B. J., & Crano, W. D. (2002). The pervasive effects of vested interest on attitude-criterion consistency in political judgment. *Journal of Experimental Social Psychology*, 38(2), 101-112.
- Lemov, D., Driggs, C., & Woolway, E. (2016). Why background knowledge is crucial for literacy. *Education Next*. <https://www.educationnext.org/background-knowledge-doug-lemov-reading-reconsidered-excerpt/>
- Livingston, J.A. (1997/2003). Metacognition: An overview. <http://www.gse.buffalo.edu/fas/shuell/cep564/Metacogn.htm>
- Lovett, M.C. (2013). "Make exams worth more than the grade." In M. Kaplan, N. Silver, D. LaVaque-Manty, & D. Meizlish (eds.), *Using Reflection and Metacognition to Improve Student Learning: Across the Disciplines, Across the Academy* (pp. 18-52). Stylus Publishing.
- Lupyan, G. (2015). Reply to Macpherson: Further illustrations of the cognitive penetrability of perception. *Review of Philosophy and Psychology*, 64(4), 585-589.
- Lyon, G.R. (1998). Why reading is not a natural process. *Educational Leadership*, 55(6), 14-18.
- Mandelbaum, E. (2017). Seeing and conceptualizing: Modularity and the shallow contents of perception. *Philosophy and Phenomenological Research*, 94(3), 1-17.
- Markman, E.M. (1985). "Comprehension monitoring: Developmental and educational issues." In S. F. Chipman, J. W. Segal, and R. Glaser (Eds.), *Thinking and Learning Skills, Vol. 2: Research and Open Questions*. Erlbaum.
- Florida Center for Instructional Technology. (n.d.). "Metacognitive strategies." *MathVIDS*. <http://fcit.usf.edu/mathvids/resources/metacognitive.html>

- Mayer, R. (1998). Cognitive, metacognitive, and motivational aspects of problem solving. *Instructional Science*, 26, 49-63.
- Mayer, R.E. (2004). Should there be a three-strikes rule against pure discovery learning? The case for guided methods of instruction. *American Psychologist*, 59, 14-19.
- Mayer, R.E. (2009). Constructivism as a theory of learning versus constructivism as a prescription for instruction. *American Psychologist*, 59, 14-19.
- Maynes, J. (2013). Thinking about critical thinking. *Teaching Philosophy*, 36(4), 337-351.
- McCabe, D.P. & Castel, A.D. (2008). Seeing is believing: The effect of brain images on judgments of scientific reasoning. *Cognition*, 107, 343-352.
- Mercier, H. & Sperber, D. (2009). "Intuitive and reflective inferences." In J.St.B.T. Evans & K. Frankish (Eds.), *In Two Minds: Dual Processes and Beyond*. Oxford University Press.
- Metcalf, J. & Shimamura, A.P. (Eds.). (1994) *Metacognition: Knowing About Knowing*. MIT Press.
- Miller, G.A. (1962). *Psychology: The Science of Mental Life*. Harper & Row.
- Mulnix, J.W. (2012). Thinking critically about critical thinking. *Educational Philosophy and Theory*, 44(5), 464-479.
- Mulnix, J.W. & Liberman, A. (2017). Philosophers folding origami: Illustrating essential strategies for learner-centered teaching. *Teaching Philosophy*, 40(4), 437-462.
- Nancekivell, S.E., Shah, P., & Gelman, S.A. (2019). Maybe they're born with it, or maybe it's experience: Toward a deeper understanding of the learning style myth. *Journal of Educational Psychology*, 112(2), 221-235.
- Nelson, T.O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. *The Psychology of Learning and Motivation*, 26, 125-173.
- Nelson, T.O., & Narens, L. (1994). "Why investigate metacognition?" In J. Metcalfe & A.P. Shimamura (Eds.), *Metacognition: Knowing About Knowing* (pp. 1-25). MIT Press.
- Nelson, T.O. (1996). Consciousness and metacognition. *American Psychologist*, 51(2), 102-116.
- Newton, P.M., & Miah, M. (2017). Evidence-based higher education: Is the learning styles 'myth' important? *Frontiers in Psychology*, 8, 444.
- Nietfeld, J., Cao, L., & Osborne, J. (2005). Metacognitive monitoring accuracy and student performance in the postsecondary classroom. *Journal of Experimental Education*, 74(1), 7-28.
- Nietfeld, J., & Schraw, G. (2002). The effect of knowledge and strategy explanation on monitoring accuracy. *Journal of Educational Research*, 95, 131-142.



- Nisbett, R.E., & Wilson, T.D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84(3), 251-9.
- O'Regan, J.K. (1992). "Solving the 'real' mysteries of visual perception: The world as an outside memory." *Canadian Journal of Psychology*, 46, 461-88.
- O'Reilly, T., Wang, Z., & Sabatini, J. (2019). How much knowledge is too little? When a lack of knowledge becomes a barrier to comprehension. *Psychological Science*, 30(9), 1344-1351.
- Palincsar, A. S., & Brown, A.L. (1989). "Instruction for self-regulated reading." In L. B. Resnick & L. E. Klopfer (Eds.), *Toward The Thinking Curriculum: Current Cognitive Research*. Association for Supervision and Curriculum Development Yearbook.
- Papaleontiou-Louca, E. (2003). The concept and instruction of metacognition. *Teacher Development*, 7(1), 9-30.
- Paris, S.G., Cross, D.R., & Lipson, M.Y. (1984). Informed strategies for learning: A program to improve children's reading awareness and comprehension. *Journal of Educational Psychology*, 76(6), 1239-1252.
- Paris, S.G., & Jacobs, J.E. (1984). The benefits of informed instruction for children's reading awareness and comprehension skills. *Child Development*, 55(6), 2083-93.
- Paris, S. G., Jacobs, J. E., & Cross, D. R. (1987). Toward an individualistic psychology of exceptional children. In J. D. Day & J. G. Borkowski (Eds.), *Intelligence and Exceptionality: New Directions of Theory, assessment, and Instructional Practices* (pp. 215-248). Ablex.
- Paris, S.G., & Oka, E.R. (1986). Children's reading strategies, metacognition, and motivation. *Developmental Review*, 6(1), 25-56.
- Paris, S.G. & Winograd, P. (1990). "How metacognition can promote academic learning and instruction." In B.F. Jones & L. Idol (Eds.), *Dimensions of Thinking and Cognitive Instruction* (pp. 15-51). Lawrence Erlbaum Associates.
- Pashler, H. McDaniel, M., Rohrer, D., Bjork, R. (2008). Learning styles: Concepts and evidence. *Psychological Science in the Public Interest* 9(3), 105-119.
- Paul, R. & Elder, L. (2001). *Critical Thinking: Tools for Taking Charge of Your Learning and Your Life*. Prentice Hall.
- Paulson, E. J., & Bauer, L. (2011). Goal setting as an explicit element of metacognitive reading and study strategies for college reading. *NADE Digest*, 5(3), 41-49.
- Perry, J., Lundie, D., & Golder, G. (2018). Metacognition in schools: what does the literature suggest about the effectiveness of teaching metacognition in schools?. *Educational Review*, 71(4), 483-500.
- Peterson, R.M. (2001). Course participation: An active learning approach employing student documentation. *Journal of Marketing Education*, 23: 187-94.

- Phillips, B. (2017). The shifting border between perception and cognition. *Noûs*, 53(2), 316-346.
- Phillips, D.C. (2006). A guide for the perplexed: Scientific educational research methodolatry, and the gold versus platinum standards. *Educational Research Review*, 1(1), 15-26.
- Phillips, D.C. (2009). "Empirical educational research: Charting philosophical disagreements in an undisciplined field." In H. Siegel (Ed.), *The Oxford Handbook of Philosophy of Education* (pp. 381-407). Oxford University Press.
- Phillips, J.A., McElwain, C., & Clemmer, K.W. (2016). Metacognitive training in professional development can improve and sustain student achievement. *arXiv preprint arXiv:1607.07856*. Web.
- Pintrich, P.R., Wolters, C.A., & Baxter, G.A. (2000). "Assessing metacognitive knowledge monitoring." In G. Schraw & J.C. Impara (Eds.), *Issues in the Measurement of Metacognition* (pp. 43-97). Buros Institute of Mental Measurements.
- Poorvu Center for Teaching and Learning. (n.d.). *Learning Styles as a Myth*. Yale University. <https://poorvucenter.yale.edu/LearningStylesMyth>
- Possin, K. (2008). A field guide to critical thinking assessment. *Teaching Philosophy*, 31(3), 201-228.
- Pressley, M., Borkowski, J.G., Schneider, W. (1989). Good information processing: What it is and how education can promote it. *International Journal of Educational Research*, 13(8), 857-67.
- Pressley, M. & Harris, K. R. (2006). "Cognitive strategy instruction: From basic research to classroom instruction." In P. A. Alexander & P. Winne (Eds.), *Handbook of Educational Psychology* (2nd Ed., pp. 265-286). Erlbaum.
- Pressley, M., & Woloshyn, W. (1995). *Cognitive Strategy Instruction That Really Improves Children's Academic Performance* (2<sup>nd</sup> ed.). Brookline Books.
- Pritchard, D. (2013). Epistemic virtue and the epistemology of education. *Journal of Philosophy of Education*, 47(2), 236-247.
- Proust, J. (2003). Thinking of oneself as the same. *Consciousness and Cognition*, 12(4), 495-509.
- Proust, J. (2007). Metacognition and metarepresentation: is a self-directed theory of mind a precondition for metacognition? *Synthese*, 159, 271-295.
- Proust, J. (2009). "Is there a sense of agency of thought?" In L. O'Brien & M. Soteriou (Eds.), *Mental Actions and Agency* (pp. 253-275). Oxford University Press.
- Proust, J. (2010). Metacognition. *Philosophy Compass*, 5(11), 989-998.
- Proust, J. (2013). *The Philosophy of Metacognition: Mental Agency and Self-Awareness*. Oxford University Press.

- Purdy, N. (2008). Neuroscience and education: How best to filter out the neurononsense from our classrooms? *Irish Educational Studies* 27(3), 197-208.
- Quigley, A., Muijs, D. & Stringer, E. (2018). "Metacognition and self-regulated learning: Guidance report." Education Endowment Foundation.  
<https://educationendowmentfoundation.org.uk/tools/guidance-reports/metacognition-and-self-regulated-learning>
- Raftopoulos, A. (2009). *Cognition and Perception: How Do Psychology and Neural Science Inform Philosophy?* MIT Press.
- Ravitch, D. (2003, April 15). Does education really need more innovation in the age of scientifically based research? [Presentation] Office of Innovation and Improvement, U.S. Department of Education.  
<https://www2.ed.gov/about/offices/list/oii/resources/20030415a.html>
- Reber, A. S. (1989). Implicit learning and tacit knowledge, *Journal of Experimental Psychology: General*, 118, 219-235.
- Reder, L.M. & Schunn, C.D. (1996). "Metacognition does not imply awareness: Strategy choice Is governed by implicit learning and memory." In L.M. Reder (Ed.), *Implicit Memory and Metacognition* (pp. 45-78). Erlbaum.
- Rensink, R.A., O'Regan, J.K., & Clark, J.J. (2000). On the failure to detect changes in scenes across brief interruptions. *Visual Cognition*, 7, 127-45.
- Resnick, L.B. (1986). Education and learning to think: A special report prepared for the Commission on Behavioral and Social Sciences and Education. National Research Council.
- Reiner, C. & Willingham, D.T. (2010). The myth of learning styles. *Change*, 42, 32-35.
- Ritchhart, R., Church, M., & Morrison, K. (2011). *Making Thinking Visible: How to Promote Engagement, Understanding, and Independence for All Learners*. Jossey-Bass.
- Rodgers, Carol. (2002). Defining reflection: Another look at John Dewey and reflective thinking. *Teachers College Record*, 104(4), 842-866.
- Rosenthal, D. (1998). "Consciousness and metacognition." In D. Sperber (ed.), *Metarepresentation: Proceedings of the Tenth Vancouver Cognitive Science Conference* (pp. 265-295). Oxford University Press.
- Rosenthal, D. (2000a). Consciousness, content, and metacognitive judgments. *Consciousness and Cognition*, 9(2), 203-214.
- Rosenthal, D. (2000b). Metacognition and higher order thoughts. *Consciousness and Cognition*, 9(2), 231-242.
- Rosenthal, D. (2000c). Introspection and self-interpretation. *Philosophical Topics*, 28(2), 201-233.

- Rowlands, M.J. (2005). *How the Body Shapes the Mind*. Oxford University Press.
- Rowlands, M.J., & Monsó, S. (2017). "Animals as reflexive thinkers: The Aponian paradigm." In L. Kalof (ed.), *The Oxford Handbook of Animal Studies* (pp. 319-342). Oxford University Press.
- Rupert, R. (2004). Challenges to the hypothesis of extended cognition. *Journal of Philosophy*, 101(8), 389-428.
- Schoenfeld, A.H. (1987). "What's all the fuss about metacognition?" In A.H. Schoenfeld (ed.), *Cognitive Science and Mathematics Education* (pp. 189-215). Lawrence Erlbaum Associates.
- Schoenfeld, A. H. (1989). "Teaching mathematical thinking and problem solving." In L. B. Resnick & L. E. Klopfer (eds.), *Toward The Thinking Curriculum: Current Cognitive Research*. Association for Supervision and Curriculum Development Yearbook.
- Schraw, G. (1994). The effect of metacognitive knowledge on local and global monitoring. *Contemporary Educational Psychology*, 19: 143-154.
- Schraw, G. (2000). "Assessing metacognition: Implications of the Buros Symposium." In G. Schraw & J.C. Impara (Eds.), *Issues in the Measurement of Metacognition* (pp. 297-323). Buros Institute of Mental Measurements.
- Schraw, G. (2001). "Promoting general metacognitive awareness." In H.J. Hartman (ed.), *Metacognition in Learning and Instruction* (pp. 3-16). Kluwer Academic Publishers.
- Schraw, G., Crippen, K., & Hartley, K. (2006). Promoting self-regulation in science education: Metacognition as part of a broader perspective on learning. *Research in Science Education*, 36, 111-39.
- Schraw, G. & Dennison, R.S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology*, 19, 460-475.
- Schraw, G. & Impara, J.C. (Eds.) (2000). *Issues in the Measurement of Metacognition*. Buros Institute of Mental Measurements.
- Schraw, G. & Moshman, D. (1995). Metacognitive theories. *Educational Psychology Review*, 7(4), 351-371.
- Schraw, G., & Robinson, D.R. (2011). Conceptualizing and assessing higher order thinking skills. In Schraw, G., & Robinson, D.R. (eds.), *Assessment of Higher Order Thinking Skills* (pp. 1-15). Information Age.
- Schunk, D.H. (1990). Goal setting and self-efficacy during self-regulated learning. *Educational Psychologist*, 25(1), 71-86.
- Schunk, D.H. (2008). Metacognition, self-regulation, and self-regulated learning: Research recommendations. *Education Psychology Review*, 20(4), 463-467.

- Schwandt, T.A. (2005). A diagnostic reading of scientifically based research for education. *Educational Theory*, 55(3), 285-305.
- Schwitzgebel, E. (2008). The unreliability of naive introspection. *The Philosophical Review*, 117(2), 245-273.
- Scriven, M. & Paul, R. (2008). Defining Critical Thinking, *Foundation for Critical Thinking*.  
<http://www.criticalthinking.org/aboutCT/definingCT.cfm>
- Shapiro, L. (2011). *Embodied Cognition*. Routledge.
- Shea, N. (2015). Distinguishing top-down from bottom-up effects. In D. Stokes, M. Matthen, and S. Biggs (Eds.), *Perception and its Modalities* (pp. 73-91). Oxford University Press.
- Shechtman, N., DeBarger, A.H., Dornsife, C., Rosier, S., & Yarnall, L. (2013). *Promoting grit, tenacity, and perseverance: Critical factors for success in the 21st century*. U.S. Department of Education Office of Educational Technology.
- Shenk, D. (2010). *The Genius in All of Us: Why Everything You've Been Told About Genetics, Talent, and IQ Is Wrong*. Doubleday.
- Simmonds, A. (2014). How neuroscience is affecting education: Report of teacher and parent surveys. *Wellcome Trust*. <https://wellcome.ac.uk/sites/default/files/wtp055240.pdf>
- Sinatra, G.M., Kienhues, D. & Hofer, B.K. (2014). Addressing challenges to public understanding of science: Epistemic cognition, motivated reasoning, and conceptual change, *Educational Psychologist*, 49(2), 123-138.
- Skinner, B. F. (1964). Behaviorism at fifty. In T. W. Wann (Ed.), *Behaviorism and Phe-nomenology*. University of Chicago Press.
- Skinner, B. F. (1971). *Beyond Freedom and Dignity*. Knopf.
- Slavin, R. E. (2002). Evidence-based educational policies: Transforming educational practice and research. *Educational Researcher*, 31(7), 15-21.
- Slavin, R. E. (2018, June 21). John Hattie is wrong. *Robert Slavin's Blog*.  
<https://robertslavinsblog.wordpress.com/2018/06/21/john-hattie-is-wrong/>
- Smeyers, P., & Depaepe, M. (Eds.). (2009). *Education Research: Proofs, Arguments, and Other Reasonings*. Springer.
- Smeyers, P., & Depaepe, M. (Eds.). (2011). *Education Research: Why "What Works" Doesn't Work*. Springer.
- Smeyers, P., & Depaepe, M. (2012). The lure of psychology for education and educational research. *Journal of Philosophy of Education*, 46(3).
- Smith, E. R., & Miller, F. D. (1978). Limits on perception of cognitive processes: A reply to Nisbett and Wilson. *Psychological Review*, 85(4), 355-362.

- Sosa, E. (1991). "Knowledge and intellectual virtue." In his *Knowledge in Perspective: Selected Essays in Epistemology*. Cambridge University Press.
- Sperling, R.A., Howard, B.C., Miller, L.A., & Murphy, C. (2002). Measures of children's knowledge and regulation of cognition. *Contemporary Educational Psychology*, 27, 51-79.
- Stanovich, K.E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, 21, 36-407.
- Stanovich, K.E. (1993). Romance and reality. *The Reading Teacher* 47(4), 280-291.
- Stanovich, K.E. (1999). *Who is Rational? Studies of Individual Differences in Reasoning*. Lawrence Erlbaum.
- Stanovich, K.E. (2000). *Progress in Understanding Reading: Scientific Foundations and New Frontiers*. Guilford Press.
- Stanovich, K.E. (2008). "Individual differences in reasoning and the algorithmic/ intentional level distinction in cognitive science." In J.E. Adler & L.J. Rips (eds.), *Reasoning: Studies of Human Inference and Its Foundations* (pp. 414-436). Cambridge University Press.
- Stanovich, K.E. (2009). "Distinguishing the Reflective, Algorithmic, and Autonomous Minds: Is it Time for Tri-Process Theory?" In J.St.B.T. Evans and K. Frankish (Eds.), *In Two Minds: Dual Processes and Beyond* (Oxford University Press).
- Stanovich, K.E., & West, R.F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, 23, 645-726.
- Stewart, T.M. (2019). Comments on Green's "Metacognition as an epistemic virtue". *Southwest Philosophy Review*, 35(2), 21-22.
- Stokes, P. (2012). Philosophy has consequences!: Developing metacognition and active learning in the ethics classroom. *Teaching Philosophy*, 35(2), 143-169.
- Strawson, G. (1997). "The self." In S. Gallagher & J. Shear (eds.), *Models of the Self* (pp. 1-24). Exeter.
- Sutherland, P.A. (1992). *Cognitive Development Today: Piaget and His Critics*. Sage Publications.
- Sweet, C., Blythe, H., & Carpenter, R. (2017). Mind mapping: A technique for metacognition. *Teaching with Metacognition*, 1-4. <http://www.improvewithmetacognition.com/mind-mapping-technique-metacognition/>
- Sweetland Center for Writing, University of Michigan. (n.d.). *Metacognition - Cultivating reflection to help students become self-directed learners*. <https://lsa.umich.edu/content/dam/sweetland-assets/sweetland-documents/teachingresources/CultivatingReflectionandMetacognition/Metacognition.pdf>

- Tanner, K.D. (2012). Promoting student metacognition. *CBE-Life Sciences Education*, 11(2), 113-120.
- Tarricone, P. (2011). *The Taxonomy of Metacognition*. Psychology Press.
- The Learning Center, University of North Carolina at Chapel Hill. (n.d.). *Higher-Order Thinking: Bloom's Taxonomy*. <https://learningcenter.unc.edu/tips-and-tools/higher-order-thinking/>
- Thiede, K.W., Anderson, M.C., & Theriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology*, 95, 66-73.
- Titchener, E.B. (1899). *A Primer of Psychology*. New York: Macmillan.
- Tobias, S. & Everson, H. (2000). "Assessing metacognitive knowledge monitoring." In G. Schraw & J.C. Impara (Eds.), *Issues in the Measurement of Metacognition* (pp. 147-222). Buros Institute of Mental Measurements.
- Torgerson, D., Torgerson, C., Ainsworth, H., Buckley, H.M., Heaps, C., Hewitt, C. & Mitchell, N. (2014) Improving writing quality: Evaluation report and executive summary. *Education Endowment Foundation*. <https://files.eric.ed.gov/fulltext/ED581140.pdf>
- U.S. Department of Education. (2002). Strategic Plan 2002-2007. <http://www.ed.gov/about/reports/strat/plan2002-07/plan.pdf>
- U.S. Department of Education, Office of Vocational and Adult Education. (2011). *Just Write! Guide*. Teaching Excellence in Adult Literacy Center. <https://lincs.ed.gov/state-resources/federal-initiatives/teal/guide/metacognitive>
- Valencia, S., & Place, N.A (1994). "Literacy portfolios for teaching, learning, and accountability: The Bellevue literacy assessment project." In S.W. Valencia, E.H. Hiebert, & P.P. Afflerbach (Eds.), *Authentic Reading Assessment: Practices and Possibilities* (pp. 134-156). International Reading Association.
- van Hout-Wolters, B.H.A.M. (2000). Assessing active self-directed learning. In P.R.J. Simons, J. van der Linden, & T. Duffy (eds.), *New Learning* (pp. 83-101). Kluwer.
- Varela, F.J., Thompson, E., & Rosch, E. (1991). *The Embodied Mind: Cognitive Science and Human Experience*. MIT Press.
- Veenman, M.V.J. (2005). The assessment of metacognitive skills: What can be learned from multi-method designs? In C. Artelt & B. Moschner (Eds.), *Lernstrategien und Metakognition: Implikationen für Forschung und Praxis* (pp. 77-99). Waxmann.
- Veenman, M.V.J., van Hout-Wolters, B.H.A.M., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. *Metacognition and Learning*, 1(1), 3-14.
- Vrugt, A., & Oort, F.J. (2008). Metacognition, achievement goals, study strategies, and academic achievement: Pathways to achievement. *Metacognition & Learning*, 30, 123-36.

- Vula, E., Avdyli, R., Berisha, V., Saqipi, B., & Elezi, S. (2017). The impact of metacognitive strategies and self-regulating processes of solving math word problems. *International Electronic Journal of Elementary Education*, 10(1), 49-59.
- Wakeford, L. (n.d.). *Sample exit tickets*. The Harriet W. Sheridan Center for Teaching and Learning at Brown University. <https://www.brown.edu/sheridan/teaching-learning-resources/teaching-resources/course-design/classroom-assessment/entrance-and-exit/sample>
- Wang, M. C., Haertel, G. D., & Walberg, H. J. (1990). What influences learning? A content analysis of review literature. *The Journal of Educational Research*, 84(1), 30-43.
- Wegner, D.M. (2002). *The Illusion of Conscious Will*. MIT Press.
- Weisberg, D.S., Keil, F.C., Goodstein, J., Rawson, E., & Gray, J.R. (2008). The seductive allure of neuroscience explanations. *Journal of Cognitive Neuroscience*, 20(3), 470-477.
- Wellman, H.M. (1983). "Metamemory revisited." In M. Chi (Ed.), *Trends in Memory Development Research: Contributions to Human Development* (pp. 31-51). Karger.
- Wellman, H.M. (1985). "The origins of metacognition." In D. L. Forrest-Pressley, G. E. MacKinnon, & T. G. Waller (eds.), *Metacognition, Cognition, and Human Performance* (Vol. 1, pp. 1-31). Academic Press.
- Weinert, F.E. & Kluwe, R.H. (Eds.). (1987). *Metacognition, Motivation, and Understanding*. Lawrence Erlbaum Associates.
- Whimbey, A., & Lochhead, J. (1986). *Problem Solving and Comprehension*. Lawrence Erlbaum Associates.
- White, P. (1980). Limitations on verbal reports of internal events. A refutation of Nisbett and Wilson, and Bern. *Psychological Review*, 87, 105-112.
- White, P. (1988). Knowing more than we can tell: Introspective access and causal report accuracy 10 years later. *British Journal of Psychology*, 79(1), 13-45.
- White, R.T. (1988). "Metacognition." In J. P. Keeves (Ed.), *Educational Research, Methodology, and Measurement: An International Handbook* (pp. 70-75). Pergamon.
- Whitebread, D., Coltman, P., & Pasternak, D. (2009). The development of two observational tools for assessing metacognition and self-regulated learning in young children. *Metacognition and Learning*, 4(1), 63-85.
- Williamson, R.A. (1996). Self-questioning – An aid to metacognition. *Reading Horizons: A Journal of Literacy and Language Arts*, 37 (1).
- Willingham, D.T. (2006). How knowledge helps: It speeds and strengthens reading comprehension, learning, and thinking. *American Educator*, 31, 8-19
- Willingham, D.T. (2007). Critical thinking: Why is it so hard to teach? *American Educator*, 31, 8-19



- Willingham, D. T., & Lloyd, J. W. (2007). How educational theories can use neuroscientific data. *Mind, Brain, and Education*, 1, 140-149.
- Willingham, D.T. & Lovette, G. (2014). Can reading comprehension be taught? *Teachers College Record*.  
[http://www.danielwillingham.com/uploads/5/0/0/7/5007325/willingham&lovette\\_2014\\_can\\_reading\\_comprehension\\_be\\_taught\\_.pdf](http://www.danielwillingham.com/uploads/5/0/0/7/5007325/willingham&lovette_2014_can_reading_comprehension_be_taught_.pdf)
- Willson, I.A. (1973). Changes in mean levels of thinking in grades 1-8 through use of an interaction analysis system based on Bloom's Taxonomy. *The Journal of Educational Research*, 66(9), 423-429.
- Wilson, D., & Conyers, M. (2014). The boss of my brain. *Educational Leadership*, 72(2).
- Wilson, D., & Conyers, M. (2016). *Teaching Students to Drive Their Brains: Metacognitive Strategies, Activities, and Lesson Ideas*. Association for Supervision and Curriculum Development.
- Wilson, T.D. (2002). *Strangers to Ourselves*. Harvard University Press.
- Wilson, T.D., Hodges, S.D., & LaFleur, S.J. (1995). Effects of introspecting about reasons: Inferring attitudes from accessible thoughts. *Journal of Personality and Social Psychology*, 69, 16-28.
- Winne, P.H. (1995). Inherent details in self-regulated learning. *Educational Psychologist*, 30, 173-188.
- Winne, P.H. (1996). A metacognitive view of individual differences in self-regulated learning. *Learning and Individual Differences*, 8, 327-353.
- Winne, P.H., & Perry, N.E. (2000). Measuring self-regulated learning. In P. Pintrich, M. Boekaerts, & M. Zeidner (Eds.), *Handbook of Self-Regulation*. Academic Press.
- Wright, J. (2015). Course participation: Defining and evaluating student participation. *American Association of Philosophy Teachers*, 4(1), 219-232.
- Yochum, N., & Miller, S.D. (1990). Classroom reading assessment: Using students' perceptions. *Reading Psychology: An International Quarterly*, 11, 159-165.
- Young, N. (2020). *The ladder of reading* [Infographic]. <https://www.nancyyoung.ca/research-and-links?lightbox=dataItem-isv6h29t>
- Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist*, 25(1), 3-17.
- Zimmerman, B.J. (2002). Becoming a self-regulated learner: An overview. *Theory into Practice*, 41(2), 64-70.
- Zimmerman, B.J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal*, 45(1), 166-183.

Zohar, A. (1999). Teachers' metacognitive knowledge and the instruction of higher order thinking. *Teaching and Teacher Education*, 15, 413-429.

Zumbrunn, S., Tadlock, J., & Roberts, E.D. (2011). Encouraging self-regulated learning in the classroom: A review of the literature. Metropolitan Educational Research Consortium (MERC).