Why do we learn science better when it looks like a novel?

Sara Alexandre da Palma Soares

PhD in Psychology

Supervisors:
Doctor Rita Jerónimo, Assistant Professor, Iscte – Instituto Universitário de Lisboa
Doctor Régine Kolinsky, Research Director, FRS-FNRS, Université Libre de Bruxelles

December, 2022

Department of Social and Organizational Psychology
Center for Research in Cognition & Neurosciences

Why do we learn science better when it looks like a novel?

Sara Alexandre da Palma Soares

PhD in Psychology

Jury:
Doctor Sven Waldzus, Full Professor, ISCTE – Instituto Universitário de Lisboa (President of the Jury)
Doctor Arthur Jacobs, Full Professor, Freie Universität Berlin
Doctor Matilde Gonçalves, Assistant Professor, Universidade NOVA de Lisboa
Doctor José Morais, Invited Full Professor, Université Libre de Bruxelles
Doctor Gaetane Deliens, Assistant Professor, Université Libre de Bruxelles
Doctor Cecília Aguiar, Assistant Professor, ISCTE – Instituto Universitário de Lisboa

December, 2022

Why do we learn science better when it looks like a novel?

Sara Alexandre da Palma Soares

There have been great societies that did not use the wheel, but there have been no societies that did not tell stories

Ursula K Le Guin, The Language of the Night: Essays on Fantasy and Science Fiction

The scientists walk more slowly, over to the brow of the hill
and down to the water's edge and past the place where the red clay runs

Neil Gaiman, The Mushroom Hunters

A educação libertadora consiste em atos de cognição, não em transferências de informação [Liberating education consists in acts of cognition, not transferals of information]

Paulo Freire, A pedagogia do oprimido [The pedagogy of the oppressed]

Acknowledgements

Many characters were part of this story, so hold on tight and bear with me while I show them my appreciation and gratitude.

. first and foremost, to my supervisors, two gifted storytellers who weaved this web of stories with me. Thank you for accepting the challenge, for *taking a walk on the wild side.* Thank you for sticking up with me even when the rest of the world was not on board with our ideas, for giving me strength and support and sharing your wisdom. It has been a long journey, but here we are at last! Obrigada, Rita. Ours is an old story, with many chapters; thank you for always *being there.* Thank you for being a cooperative raven, always embracing my multidisciplinary cravings and bringing new people on board. Merci, Régine. Our story also began in a previous chapter. Thank you for enthusiastically accepting to co-create this story and always making me feel at home in faraway Bruxelles. Thank you for being a wise owl who always finds a solution, even when the light is very dim and faded.

. to all the other people who collaborated closely in the project. To José, another gifted storyteller, who gave valuable insights that shaped the project since it was an embryo. Sofia, who told stories from the laboratory (our stories crossed again); Matilde, who told stories from linguistics, Raquel, who told stories from natural science, Cláudia, who told stories from statistics, Susana, who told stories from qualitative methods, and Maria do Carmo, who told stories from the science museum.

. to all the other people who helped me along the way by recruiting participants, going over materials and procedures with me, and doing all sorts of tasks. A very special thanks to Ana, and also to João and Ricardo. I could not have done without your help.

. to the participants. Learning is a two-way street and I learned much from them. I hope they also took something with them.

. to the souls of the PhD program and from room 2n6 in particular, especially to Eduardo, Tânia, Patrícia, Sofia, and Sara.

. to Elizabeth Collins, who always provided encouraging feedback.

. to Marco Rego at Iscte and Angélique Bernacki at ULB, who helped me navigate the kafkaesque worlds of Portuguese and Belgian bureaucracy, and to Helena Santos, who always provided much needed flexibility and solutions to problems concerning data collection and scheduling at LAPSO.

. to my grandmother, the first great storyteller I met. She told me so many fables, folk tales, historical episodes, paranormal accounts.. such richness and versatility, and over and over, *on demand*. I was truly lucky. Sometime during this story, she returned to stardust, the thing stories are made of, but her stories still peek through the small details of this dissertation. Obrigada Av'Olenca.

. to my mother, who added expressionism and musicality to storytelling. While we are not always able to share stories, I will forever be grateful for the confidence, support, appreciation and confidence she always put on my creativity and intelligence.

. to my nephew, who was one year old when this story started. With him I ran, jumped, laughed, played, danced, and created a many surrealistic stories involving dinosaurs, cows, snails and Batman.

. to L., an ever-present lantern when walking through subterranean passages.

. to the brave and beautiful people and collectives who defy injustice and power asymmetries by putting politics where it belongs: in the streets and at the heart of community-building.

. to that beautiful place in beautiful Alentejo, and the beautiful people who create it, where I started to tell a new story. Although I have been less present in more recent chapters, they never closed the book on me, and await patiently and lovingly for me to regain my active role as a character.

. to my cats, Kali and Kabuki, who served as comic reliefs, warm blankets, but also as naughty villains in some moments.. anyway, thanks for not stop loving me!

. to my friends. I am blessed that they are too many to mention, but two people have to be mentioned: Sinead who supported me throughout these last weeks with delicious food and empathy, and Lara, who is always here, no matter how far.

. a very warm and tender thanks to Dioma, for the unconditional support, love, and understanding. Words could never express how grateful I am for being a rock during this tumultuous journey.

. and last, but not least, to Art. Where would I be without brushes, hanging notes, alien melodies, poems, coloured pencils, words, ethereal sounds, harsh industrial noises, fantasy worlds, wordswordswords, bright colours, black Sakura pens? I dare not think.

Research reported in this dissertation was financed by the Fundação para a Ciência e Tecnologia (Foundation for Science and Technology; FCT) with a PhD grant to the author (PD/BD/135345/2017 and COVID/BD/152481/2022)

FCT
Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

Resumo

A ciência é ubíqua e aprendê-la é vantajoso. Contudo, esta aprendizagem é desafiante, devido a uma lacuna entre pensamento quotidiano e pensamento científico. O presente trabalho subscreve à proposta de destacar a componente de literacia da educação científica, através da investigação do impacto dos textos narrativos enquanto ferramentas de aprendizagem científica mediadas pela literacia. Propusémos um diálogo entre quadros teóricos multidisciplinares e identificámos duas questões complementares para examinar a questão da aprendizagem com textos narrativos científicos, concernentes, em primeiro lugar, aos resultados de aprendizagem e, em segundo lugar, às condições e processos que os geram (Capítulo 2). Desenvolvemos materiais de aprendizagem para investigar estas questões (Capítulo 3). Relativamente aos resultados de aprendizagem, estudos comportamentais (Capítulos 4 e 5) e de eye tracking (Capítulos 5) mostraram que jovens adultas/os com pouco conhecimento científico prévio aprendem ciência de textos narrativos a vários níveis de compreensão. O primeiro estudo mostrou que esta aprendizagem pode ser superior ou equivalente àquela produzida por textos expositivos, dependendo do tópico científico. Relativamente às condições e processos de aprendizagem, os mesmos estudos mostraram que um conjunto de características das/os aprendizes contribuíram conjuntamente para esta aprendizagem. O segundo estudo mostrou que atenção a e pensamentos sobre acção humana contribuíram independentemente para a aprendizagem (Capítulo 5). Finalmente, um estudo qualitativo revelou como textos narrativos e museus podem preencher lacunas entre pessoas e ciência (Capítulo 6). Globalmente, os resultados sugerem que os textos narrativos podem ser uma ferramenta útil para aprender ciência e aproximar as/os aprendizes da faceta humana da ciência.

Palavras-chave: aprendizagem de ciência; textos narrativos; textos narrativos científicos; educação não-formal; compreensão de texto; características dos aprendizes; acção humana; percepções sobre aprendizagem de ciência; ciência e pessoas

PsycINFO Codes:
3500 Educational & School Psychology
2343 Learning & Memory
2340 Cognitive Processes

Abstract

Science is pervasive, and everyone can benefit from learning it. Yet, it is also challenging, for reasons stemming from a gap between everyday and scientific thinking modes. The present work endorses the proposal of foregrounding the literacy component of science education, by thoroughly examining the impact of narrative texts as literacy-mediated science learning tools. We proposed a dialogue between multidisciplinary theoretical frameworks and identified two complementary questions for tackling the issue of learning from science narrative texts, pertaining, firstly, to learning outcomes and, secondly, to the conditions and processes generating them (Chapter 2). We developed learning materials to investigate these questions (Chapter 3). Regarding learning outcomes, a behavioural study and a combined behavioural and eye tracking study (Chapters 4 and 5) showed that young adults with low prior science knowledge learn from science narrative texts at various comprehension levels. The former study showed that this learning can be superior or equivalent to the one yielded by expository texts, depending on the science topic. Regarding learning conditions and processes, the same studies showed that a set of learner features jointly contribute to this learning. The latter study further showed that attention to and thoughts on human action make independent contributions to learning (Chapter 5). Finally, a qualitative study revealed how narrative texts and museums can help bridge gaps between people and science (Chapter 6). Overall, results suggest that narrative texts can be a useful tool for science learning and can bring learners closer to the human facet of science.

Keywords: science learning; narrative texts; science narrative texts; non-formal education; text comprehension; learner features; human action; perceptions on science learning; science and people

PsycINFO Codes:
3500 Educational & School Psychology
2343 Learning & Memory
2340 Cognitive Processes

Index

Index of Tables

Index of Figures

CHAPTER 1

INTRODUCTION

"I WOULD LIKE TO TRY OUT AN IDEA THAT MAY NOT BE QUITE READY, indeed may not be quite possible. But I have no doubt it is worth a try. It has to do with the nature of thought and with one of its uses. It has been traditional to treat thought, so to speak, as an instrument of reason. Good thought is right reason, and its efficacy is measured against the laws of logic or induction. Indeed, in its most recent computational form, it is a view of thought that has sped some of its enthusiasts to the belief that all thought is reducible to machine computability. But logical thought is not the only or even the most ubiquitous mode of thought. For the last several years, I have been looking at another kind of thought, one that is quite different in form from reasoning: the form of thought that goes into the construction not of logical or inductive arguments but of stories or narratives." (Bruner, 1986, p. 1)

Thirty-six years have passed since Bruner's "Actual Minds, Possible Worlds" was first published. Although we are surrounded by stories, be it in books, films, TV series, or daily exchanges with people around us, and that storytelling approaches seem to be booming a bit everywhere, it will still probably feel weird to equate narration and learning the formal and natural sciences (henceforth, science), such as mathematics or chemistry. In fact, learning science might feel weird in and of itself, even though, similarly to narratives, science permeates a myriad of aspects of our lives: From informally commenting on why we should refrain from picking certain mushrooms, to deciding if we believe the earth to be round, to performing heart surgeries, to building sophisticated devices to explore the cosmos, and, ultimately, making decisions that allow us to deal with a global pandemic. As such, acquiring some level of scientific knowledge can benefit us and those around us, even if we do not wish to become scientists ourselves.

The present work has the goal of examining the potentially symbiotic relationship between science learning and narration. In specific, we aimed at contributing to finding better ways to learn science by doing a thorough examination of the impact and features of the literacy-mediated tool of narrative texts. This idea is, by no means, new. Besides Bruner (e.g., 1986), many educators, authors, and researchers have lauded the features of narratives, such as its familiar structure, ability to evoke emotions, and overall proclivity to human exterior affairs and inner workings (e.g., Abd-El-Khalick, 1999; Egan, 1997; Graesser & Ottati, 1995; Strube, 1994). Paired with empirical evidence of narrative texts' comprehension and retention benefits (e.g., Britton et al., 1983; Kintsch & Young, 1984; Mandler & Johnson, 1977), these idea inspired the idea of a *narrative effect*, thus encouraging researchers and educators to use them to teach science contents. Yet, there is still much about the use of science narrative texts in science education which we do not know.

In theoretical terms, a solid framework seems to be missing, or, more precisely, a dialogue between frameworks that can guide the design of materials and interventions, and the interpretation of the obtained results. Whereas many of the authors who used narrative texts to convey science draw on ideas about what narrative is, as a concept, and what makes up a narrative text (e.g., Arya & Maul, 2012; Flynn & Hardman, 2019; Prins et al., 2017), we believe this to be not enough for at least two reasons. First, narrative texts do not tend to convey science contents, and so some of their assumed features and benefits might not directly translate to science narrative texts, at least on a one-to-one basis. Second, other aspects fundamental for framing and understanding the present issue are not considered in that approach, such as aspects pertaining to literacy. These include a more complete understanding of text features (e.g., structure and content), cognitive text comprehension (levels of comprehension and corresponding processes), and the more encompassing process of learning by text (the elements that are involved, such as the learner and the context).

Understanding these theoretical underpinnings can, in turn, improve our comprehension of the empirical learning outcomes that are produced. Research with science narrative texts has not only provided mixed results (e.g., Arya & Maul, 2012; Hadzigeorgiou et al., 2012; Jetton, 1994; McQuiggan et al. 2008; Negrete & Lartigue, 2010), but has also presented considerable variability in terms of materials, populations, and learning activities, to name a few. The lack of incorporation of literacy-related frameworks makes these results even more difficult to interpret. In addition, we believe that the idea of a general narrative effect, whether explicit or tacit, is misleading from the outset. Besides being, to a great extent, founded in non-science narratives, by thinking about learning outcomes in *absolute* terms, one is more prone to overstate the dimension of the effect when it is found, and to dismiss important information when it is not found.

To avoid this, it is paramount to consider the complexity of the learning process, and to investigate the conditions and processes leading to *specific* learning outcomes. Whereas it is potentially misleading to assume learning outcomes or processes from the features and workings of general narratives without testing them, it can nonetheless inspire hypotheses on the processes that may be triggered by science narratives. Moreover, many studies which did not find a narrative learning advantage found that narrative texts nonetheless had a positive impact on other aspects, such as feeling immersed in the story (e.g., McQuiggan et al., 2008) or connected to its characters (Jetton, 1994). Yet, most studies tend to approach these aspects as yet another outcome of reading a science narrative text, and not as a process underlying the observed learning outcomes.

The present work was rooted on the premise that the multidisciplinarity inherent to the topic at hands should be acknowledged and embraced; with that in mind, there is still much to learn about a long-standing proposal. The overarching aim of this dissertation was to build a more complete picture of the topic of learning science through narrative texts, by focusing on the three dimensions we outlined: theoretical backbone, learning outcomes and learning processes. These three dimensions underlie our exploration of one main topic: how narratives can help highlight the *human facet* of science, that is, how retrieving human elements of the process of doing and learning science can contribute to make science more understandable, appealing, and relatable to participants.

The present dissertation was guided by three questions. Our first question was which theoretical frameworks would be useful to draw from in order to better understand the issue of learning science from narrative texts (**theoretical backbone**). Our second question was whether people, and particularly young adults with little prior knowledge, can learn science from narrative texts, and even learn better from them (**learning outcomes**). Finally, our third question was directed at how the process of learning science through narrative texts occurs (**learning process**). These questions were examined in four studies, corresponding to five chapters.

Chapter 2 addresses the three research questions. To address the first question (theoretical backbone), we propose a dialogue between theoretical frameworks coming from text linguistics, cognitive psychology, and pedagogy, a combination not frequently encountered in the literature. To address the second and third questions, we draw on this theoretical backbone and ask if science narrative texts have consistently produced good memory and learning (learning outcomes), and how the process of learning with science narrative texts takes place (learning process). The resulting theoretical proposal and analysis of previous results and ideas can be used to make sense of previous results, as well as in future conceptualization and designing of interventions.

In Chapter 3, we report how and why we developed a set of materials that would enable us to investigate the second and third questions (i.e., learning outcomes and learning processes, respectively). In addition to clarifying methodological aspects from following chapters, the provided information and guidelines can be used by educators and researchers to build similar materials for research and educational interventions. Chapters 4 and 5 report our empirical examination of the second question (learning outcomes); importantly, we focused on young adults, a rather understudied population as regards learning through science narrative texts. In Chapter 4, this examination involved comparing narrative texts with expository texts (more common in science instruction) with equivalent contents. Besides giving us a control condition, it further allowed us to examine key questions in the relevant literature warranting further investigation. The findings thus contribute to determine not only if young adults can benefit from learning science from narrative texts, but also the extent of this benefit, namely in terms of different science topics, levels of comprehension, and retention intervals.

In Chapter 5, we delved further into one of the three facets of the third question (learning process), pertaining to the impact on science learning of human-related elements in narrative texts, and the human-related processes they may trigger. Besides helping characterize the processes engaged when learning from science narrative texts, this thorough investigation of human-related elements delivers new information, as these aspects have been chiefly investigated using non science narrative texts. This examination included an eye tracking analysis, which is a premiere, to our knowledge. The resulting information can prove useful for future research and for educators to design science education practices that incorporate these human elements. Chapters 5 and 4 also examined the impact on science learning of more general features from the learner. Approaching learners' features as processes underlying learning, instead of just outcomes in themselves, enabled us to address another facet of our third research question (learning process). This facet offers hints of the kind of more general aspects that can be further examined in connection to learning from science narrative texts, as well as integrated and stimulated in science education.

Finally, in Chapter 6, a final facet of the third research question was tackled, this time relating to the perceptions that people have on science learning and literacy, and on the specific role that narrative texts might play in this learning. Whereas the few previous studies investigating perceptions on the role of narrative texts in science learning interviewed school students and teachers, we talked with adults of varying age and contact with science in a non-formal (i.e., non-school) context. As such, we were able to gather perceptions on the kind of difficulties people feel during lifelong science learning, and how narrative texts and other non-formal tools, such as museums, can grapple with these challenges. Importantly, this study was the result of the collaboration with a science museum, Museu de História Natural e da Ciência de Lisboa, further cementing the socially and culturally situated nature of this Chapter and extending the reach of the present dissertation. In addition, examining such perceptions offers a discursive facet of the process of learning science through narrative texts that complements the more cognitive facets of the third research question addressed in Chapters 4 and 5.

Given the pervasiveness and relevance of science in our society, as well and the challenges to its learning, the relevance of the findings of the present dissertation becomes evident. Overall, we hope to produce results that will help finding innovative and more inclusive ways of learning (and doing) science, which can resonate at different levels: for researchers to interpret previous findings and plan future research on this topic; for educators to design and explore educational tools and practices; and for a broader audience to reflect upon the topic of science learning and literacy, especially in connection with their social and cultural context. Although our main focus concerned lifelong learning and a broader sense of scientific literacy, the findings of the present dissertation may be of relevance for various learning settings and contexts.

In the following paragraphs, we will introduce the core ideas pertaining to this problematic, so that the aims and research questions that guided this dissertation, as well as its organising structure, can be more fully understood. The research questions will be fully outlined, following the presentation of the ideas and evidence that sustain them.

1.1. Texts, processing them, and learning from them

At the outset, it is helpful to summarize our view on texts, text processing, and learning from text, here, as these concepts form the theoretical backbone that shapes the remaining concepts and theoretical ideas. This theoretical backbone will be more thoroughly presented in Chapter 2.

In line with proposals from text linguistics (e.g., Adam, 1997), we regard texts as being made up of linguistic features and of pragmatic features, which interact in important ways (e.g., the structure of a text is in part determined by the purpose and the context in which the text will be shared). Additionally, in accordance with cognitive models of text processing (e.g., Kintsch, 1998), we conceive text processing as entailing different levels of representation (e.g., mentally representing the meaning of each individual sentence on a page is at a different level than being able to extract the main idea contained on that page), and consider that these different levels correspond to different comprehension levels (e.g., the distinction between recalling explicit information from the text and being able to interpret and use information from a text). Finally, we endorse the view, posited by several pedagogical frameworks (e.g., Snow, 2002), that learning from text involves different but interacting elements, namely the reader, the text, the learning activity in which the text is being read (and, by extension, listened to or written), and the wider sociocultural context of which these elements are part.

1.2. Why learning science and scientific literacy are important

Science permeates several aspects of daily life, from simple informal remarks to activities and decisions that affect everyone around us. It is therefore concerning that both the appeal and the performance in science start to fall short during school years, with many students displaying lack of interest (e.g., Avraamidou & Osborne, 2009; Hazelkorn et al., 2015; Logan & Skamp, 2008) and low performance levels (e.g., Blotnicky et al., 2018; OECD, 2019).

On the one hand, learning challenges discourage students to pursue studies in scientific fields and develop a fundamental and specialized scientific literacy (Norris & Phillips, 2003), which they might otherwise be interested in doing. This corresponds to formal education, a type of education in which a mandatory curriculum is administered in a systematic and structured way, according to a set of rules and norms, and performance is graded (Dib, 1988). Besides potential losses for individual students' paths, developing specialized scientific knowledge is very crucial for communities and society more generally, as it represents a key resource to solve pressing issues and drive innovation (e.g., Roberts, 2007).

However, regardless of ambitions of specialization and personal interests, it is claimed that the goal of scientific education should be broader, aiming at the development of a scientific literacy from which everyone can benefit (Sadler, 2009). This derived sense of literacy (e.g., Norris & Phillips, 2003) corresponds to non-formal education, a kind of education in which one or more of formal education's features are absent (Dib, 1988). Non-formal education can happen at many different places throughout life and have varying levels of informality, including learning that takes place at home or in students' more specific communities, more organised communities of practice or groups, and institutions such as museums (Bell et al., 2009; Callanan et al., 2011; Lave & Wenger, 1991). In the present dissertation, we will use the terms "science learning", "scientific literacy" and "science education" in a mostly interchangeable way, as we are not concerned with the specificities of these concepts, but instead with the acquisition and use of scientific knowledge, which can be conveyed by any of these terms.

Furthermore, scientific literacy can be used to inform discussions and decisions on science-related matters taking place in the wider sociocultural environment, affecting not only individuals but also communities and societies at large (e.g., Bruner, 2009; Morais & Kolinsky, 2016; 2021; Solomon et al., 1992). This becomes especially crucial, and alarming, when we consider that pseudoscientific beliefs and misconceptions, such as that we only use 10% of our brain (Swami et al., 2012), or that humankind was created by God less than 10 thousand years ago (Silva et al., 2010), are pervasive even among highly schooled people (e.g., Morais & Kolinsky, 2021; Van Prooijen, 2017; Silva et al., 2010) and are often correlated with anti-scientific attitudes (e.g., Swami et al., 2012; Swami et al., 2014). Crucially, the impact of such beliefs is not limited to public discussions, as they can also influence and shape specialized policymaking, which in turn has the potential of affecting the whole planet, as we have recently witnessed during the COVID-19 pandemic and continue to witness in ongoing discussions on climate change (e.g., Ball, 2021; De Pryck & Gemenne, 2017; Marleau & Girling, 2017). Importantly, given that acquiring and using science knowledge takes place within a wider social and cultural context, these two dimensions are, to some extent, inextricable.

1.3. Learning science: challenging aspects with literacy overtones

The reasons pinpointed by educators and researchers for the difficulties experienced in science learning are manifold. The challenging aspects of learning science begin with its own language, considered by many students to be "the tongue of foreigners, equally exotic" (Montgomery, 1996, p. 9). Scientific language is often dense, technical, and jargon-laden (e.g., Avraamidou & Osborne, 2009; Snow, 2002), and some even point out that its readability has been decreasing over time, even among the wider scientific community (Ball, 2017; Plavén-Sigray et al., 2017). A related problem is that of scientific concepts; not only are they unfamiliar to new learners due to concepts' level of abstractness and technical specificity (Best et al., 2005; Graesser et al., 1991), but they can be felt as counter-intuitive (e.g., evolution, Browning & Hohenstein, 2015; projectile motion, Alvermann et al., 1995).

Moreover, scientific discourse is often regarded as authoritative or dogmatic (Kloser, 2013; Negrete & Lartigue, 2004), as it frequently frames science as a monolithic set of unquestionable conclusions (e.g., Avraamidou & Osborne, 2009; Clough, 2011). Yet another problem, partially stemming from the previous ones, is the human and cultural decontextualization that pervades science education (e.g., Sánchez Tapia et al., 2018; Solomon, 2002). Science education often fails to convey the process-like and cultural-bounded nature of science, with all its failures, errors, and frustrations (Arya & Maul, 2021; Hadzigeorgiou et al., 2012; Hong & Lin-Siegler, 2012; Morais & Kolinsky, 2021). This lack of information on the human and cultural side of science creates an incomplete and myopic view of science and can be particularly damaging for students whose social and cultural identities are under-represented in science and society at large, such as those belonging to marginalized and *othered* groups (e.g., Jackson et al., 2016; Johnson, 2008; Visintainer, 2020).

Corollaries of these problems include widespread misconceptions regarding the nature and workings of science and the characteristics of scientists (e.g., Clough, 2011; Matthews 1994), as well as perceptions of narrowness of science as an educational experience (e.g., Rowe et al., 2015; Tobias, 1990). In fact, it has been reported that students tend to give up on pursuing science earlier than other fields perceived to be more interdisciplinary (Avraamidou & Osborne, 2009; Tobias, 1990), and even students who end up pursuing more specialized science-related paths continue to face challenges in higher educational levels (e.g., Frisch, 2010; Jackson et al., 2016; Johnson, 2008).

Despite their diversity and range of application, the main outlined challenges easily and acutely apply to text materials. It is pressing to take a closer look at text materials because they are a widespread and well-established source of scientific information. Texts have been the cornerstone of formal science education (Kloser, 2013; Snow, 2010; Van den Broek, 2010); learning is still mostly based on successful text comprehension (e.g., Mason et al., 2013), which has been found to be correlated with academic achievement (e.g., Cromley et al., 2010). Additionally, to literate people, texts continue to be a resource for acquiring new facts and knowledge on science throughout their lives, and still have a pivotal role to play in an increasingly digitalized world in mediums such as posts in social media and blogs, newspapers and magazines (digital or not), and even museum plates (e.g., Negretti, 2022; Ravelli, 2007; Sullivan et al., 2021).

This role is further cemented by claims for a greater integration of language and literacy processes (i.e., reading and writing) when addressing science education (e.g., Klein, 2006; Morais & Kolinsky, 2016; Norris & Phillips, 2003). As already noted elsewhere (Arya & Maul, 2021; Morais & Kolinsky, 2016), an illustrative example is the fact that the eminent journal *Science* dedicated a whole special issue to the importance of strengthening the ties between science and literacy (e.g., Snow, 2010; Webb, 2010). Under this view, literacy (i.e., reading and writing) processes are therefore not merely instrumental to science, but instead "*constitutive parts* of science" (Norris & Phillips, 2003, p. 226, our emphasis). Indeed, literacy plays a fundamental role in cognition, as attested by its well-established impact on thinking, reasoning, and a series of other widespread processes (e.g., Kolinsky, 2015; J. Morais, 2015; Morais & Kolinsky, 2021). However, this strong impact is often made less apparent, or even heavily overshadowed, by the fact that we can only appreciate it through "literate glasses" (Kolinsky & Morais, 2018, p. 322), as the ways in which we get information, think and communicate are themselves under the influence of literacy.

Furthermore, this focus on literacy has also the potential of retrieving social and cultural dimensions of this processing, beyond cognitive processes. This is crucial, as the impact of literacy (scientific and of other kinds) on cognition is not confined to our individual brains, but extends to several dimensions of our life, notably to social, economic, and political ones (J. Morais, 2015; Morais & Kolinsky, 2021).

The literacy component of science education is thus an essential avenue of research when addressing difficulties in science learning and devising better learning strategies and tools. Moreover, and importantly, the outlined problems seemingly denounce the existence of a sort of gap between science concepts and literacy-mediated tools and processes, as science texts, at least to some extent, fall short of connecting learners to science. Having said that, this is not the only gap relevant to examine to understand the challenging aspects pertaining to science learning.

1.4. "If everyone understands how to think in stories…"

1.4.1. A gap between thinking modes: science made of stories

The challenges to science learning that we have outlined can also be traced to a fundamental difference between everyday and scientific modes of thinking (e.g., Bruner, 1986, Egan, 1997; Klein, 2006; Phillips & Norris, 2009). Much like science texts, cognition was first thought of as consisting of logically manipulated propositions expressed literally through language (e.g., Klein, 2006). However, it went on to be acknowledged by second-generation cognitive scientists as being perceptually based, fuzzy, and contextual, being expressed primarily through metaphoric and narrative language (e.g., Klein, 2006). This later conception of cognition is much more in tune with the notion that our most common and familiar way of thinking is not logically based nor analytically bounded.

Bruner (1986) was one of the main proponents of this notion, positing that people learn to think and make sense of the world through human intentions, reasons, actions, and agents, that is, using a *narrative* mode of thought. He contrasted this mode with a paradigmatic or logico-scientific mode, by which people learn to interpret the world through general causes, proofs, and theories. The latter is developed later, particularly through formal instruction, and, most people do not get to effectively think in that mode (Bruner, 1990). In a related vein, Egan (1997) proposed that children begin to think in terms of storytelling and fantasy, dubbed mythic understanding, followed by a focus on heroic but possible people and events, also known as romantic understanding; only later can they master conceptual or philosophical understanding.

This idea of a gap between modes of thinking helps clarify why scientific language and ideas seem so removed and distant from everyday life. Everyday life communication is generally made up of relatively simple terms, referring to familiar concepts and objects. Therefore, an “anthropocentric-leaning” thinking mode, that is a thinking mode based on reasoning and actions which are highly familiar are centred around people’s actions and intentions, can arguably make it difficult to grasp the abstractness and formality of scientific logic and ideas, rendering them inaccessible and incomprehensible, at times even seemingly irrelevant. Thus, although both narratives and science are pervasive in daily life, the first is part of a thinking mode which is highly familiar and meaningful for us, whereas the second is part of a thinking mode that is for the most part alien and hard to grasp, even though we make extensive use of the outputs of such thinking.

It would thus seem like the scientific and the narrative thinking mode are diametrically opposed and difficult to reconcile. But is this really the case? Or does science have narrative threads in its fabric?

Although science and narrative thinking modes seem to be at odds, many authors argue that the nature and practice of science have solid narrative roots. Many scientific hypotheses started out as stories or myths (Hadzigeorgiou & Schulz, 2019; Popper 1972), and theories rely on metaphors, analogies, and interpretive frameworks (e.g., Bruner, 1996; Fuchs, 2015). Additionally, scientists often engage in storytelling (Hadzigeorgiou, 2016; Leipzig, 2018) and, much like narratives, science is an unfolding process centred on human endeavour (Bruner, 1996; Larison, 2018), where protagonists execute sequences of events with a purpose and interpret its consequences (e.g., Hoffman, 2014; Strube, 1994). The very structure of scientific articles (and of dissertations, for that matter) bears resemblance to the temporal unfolding characteristic of narratives (i.e., beginning, middle and end; Leipzig, 2018; Olson, 2015). Furthermore, and significantly, many authors contend that science and more artistic and aesthetic engagements are wrongly divorced, as they not only complement but also feed each other, to create more versatile and complete ways of thinking and acting (Bruner, 1996; Egan, 2005; Hadzigeorgiou, 2016; Morais & Kolinsky, 2021).

These different claims suggest that science and narrative thinking modes and practices share fundamental traits. Narratives centre on human action and science has human action at its centre; narratives and science therefore share a fundamental humanness. As we are interested in the role of texts, a literacy-mediated tool combining science and narrative features, such as narrative texts, may bring out the complementary aspects of narrative and science to bridge the everyday and the scientific thinking modes.

Consequently, it becomes necessary to better understand why authors like Bruner consider narratives to hold such a special place in human cognition, and what is important to learn about them. As we build a case for this tool, we will identify the areas in need of further conceptualization and investigation and present the research questions we asked to address them.

1.4.2. Stories made of science: narratives to bridge the gap

Klein (2006) contented that the gap between science and everyday thinking modes should be bridged by innovative science education practices. Concomitantly, it has been remarked that "If everyone understands how to think in stories, then why not use stories as a way into other modes of thought?" (Gilbert et al., 2005, p. 3). One innovative tool fitting the claims for a greater focus on literacy processes in science education practices, and complementary to scientific thinking, is thus narrative textual materials, as has been advocated by several educators (e.g., Bruner, 1991; Egan, 2005; Solomon et al., 1992). Narrative texts are therefore proposed to establish a much-needed literacy-mediated bridge between the everyday and scientific modes of thinking, an idea represented in Figure 1.1. Although we are particularly interested in textual materials, we will also mention narrative's presence in other mediums, as to better characterize the pervasiveness of narration in daily life.

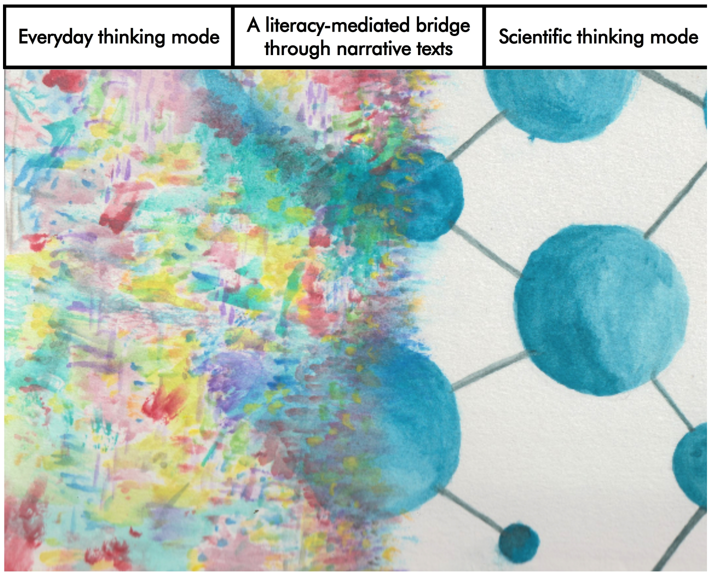


Figure 1.1. Narrative texts as a literacy-mediated bridge between everyday and scientific thinking modes

Narration is generally understood as a sequence of temporally organized actions or events (e.g., Adam, 2011; Norris et al., 2005; Strube, 1994), and is considered to have a privileged status in human cognition (Bruner, 1990; Graesser & Ottati, 1995) for several reasons. Believed to have emerged early in human history (e.g., Donald, 2001), (oral) narration is integral to every culture (e.g., Fisher, 1987; Gottschall, 2013), rendering narrative communication ancient and embedded in evolutionary and cultural significance (e.g., Boyd, 2009; Campbell, 1949; Gazzaniga, 2008; Gottschall, 2013; Sibierska, 2016). The close correspondence between narrative structure and the way the human mind orders experience is also frequently mentioned (Fisher, 1987; Graesser et al., 2002), and there is evidence that people build temporally organized representations from a text even in the absence of such a structure (Claus & Kelter, 2006). This proclivity to the narrative structure seems to begin early on in life, as indicated by newborns' auditory cortex sensitivity to the temporal structure of sounds (Telkemeyer et al., 2009); the pervasiveness of the narrative structure goes on throughout the rest of our lives, starting with bedtime stories and cartoons, and carrying on with books, movies, songs, theatre, among many other cultural artefacts and activities (Schank & Berman, 2002). It should therefore come as unsurprising that narrative's ubiquity is also reflected in the way the human brain processes narrative texts, namely by the engagement of a wide network extending way beyond language processing areas (e.g., Mano et al. 2009; Mar, 2004; Mason & Just, 2009; Xu et al., 2005; Young & Saver, 2001). These additional areas include emotional processing, mental imagery, executive processing (i.e., set of processes necessary for goal-directed behaviour, Diamond, 2013), and theory of mind (i.e., the ability to infer mental states in oneself and in others, Baron-Cohen et al., 1985).

Even though science narrative texts have been used as science learning tools, the total set of studies is somewhat limited (Prins et al., 2017) and there is a great deal of variability among the existing ones. As such, there is still much we do not currently know, or that at least would benefit from further conceptual framing and direct investigation.

A first question pertains to the **theoretical underpinnings** of narrative texts as science education tools. As far as theory goes, it is possible to find shared ideas, generally stemming from the basic idea previously outlined: Science learning presents challenges and narrative texts can be a way to tackle some of those challenges. Some authors emphasize narrative's familiarity (e.g., Cervetti et al., 2009; Flynn & Hardman, 2019; Morais et al., 2019; Prins et al., 2017), and ability to bring learners closer to real-life situations (Alvermann et al., 1995; Frisch, 2010; Maria & Johnson, 1989; Mutonyi, 2016). Other authors focus on the benefits that the imaginative (Browning & Hohenstein, 2015; Corni et al., 2010; Hadzigeorgiou et al., 2012; Negrete & Lartigue, 2010) and/or affective (Banister & Ryan, 2001; Browning & Hohenstein, 2015; Negrete & Lartigue, 2010; Prins et al., 2017) features of narrative may yield. Others still highlight the human and/or historical component of narrative-based materials (Arya & Maul, 2021; Hadzigeorgiou et al., 2012; Hong & Lin-Siegler, 2012; Klassen, 2007). Given that the process of learning science through narrative texts touches upon several aspects, we believe that ideas on narrative as a concept, as the ones described, would benefit from being integrated in a background framework, or set of frameworks, touching on those elements.

Regarding the narrative texts to be used with the purpose of science learning, some authors focused on proposing a set of defining narrative features to be considered when doing studies using science narrative texts (Avraamidou & Osborne, 2009; Norris et al., 2005; Prins et al., 2017; Wilcken, 2008). In addition, a few authors incorporated cognitive frameworks of text processing and/or levels of comprehension (Arya & Maul, 2012; Negrete & Lartigue, 2010; Wolfe & Mienko, 2007; Wolfe & Woodwyk, 2010). Still others have applied constructivist, student-centred, pedagogical notions (e.g., Akarsu et al., 2015; Klassen, 2007; Kokkotas et al., 2010; Morais et al., 2019), highlighting the affinity between narrative-based materials and innovative learning practices (Klein, 2006).

While by no means exhaustive, this selection of examples illustrates the conceptual diversity that permeates this literature, and that authors tend to endorse more than one idea. However, overall, there seems to be a lack of dialogue between theoretical frameworks that focus on different aspects of the process of learning science through narrative texts, having texts and literacy processes at its core. Such a dialogue could, in turned, be used to shape theoretical ideas on the impact of specific narrative features on specific comprehension outcomes, and to better accommodate and interpret any resulting empirical results.

This set of ideas led us to our first, theoretically motivated, research question (RQ):

RQ 1: Better understanding the issue of learning science from narrative texts from a literacy point of view: Which theoretical frameworks would be useful?

A second question pertains to the **memory and learning** benefits afforded by narrative texts. As a matter of fact, narrative texts have been considered to make ideas more coherent, meaningful and memorable (Kintsch, 1998; Schank & Berman, 2002; Strube, 1994). Because creating stories in our minds is a fundamental means of meaning-making, it is considered to “pervade all aspects of learning” (Wells, 1986, p.214), and stories are considered to be inherently didactic (Schank, 2002). These claims have been to some extent corroborated by evidence that participants more easily understood and recalled information from narrative texts, as compared with expository texts (e.g., Britton et al., 1983; Graesser et al., 1980; Kintsch & Young, 1984; Mandler & Johnson, 1977; Zabrucky & Moore, 1999). For instance, Britton et al. (1983) showed that the performance of a secondary task, simultaneously while reading a text, was more affected when participants were reading narrative texts than when they were reading expository texts. However, despite this interference, comprehension results were higher for narrative texts. The authors interpreted this as evidence that narrative texts “fill more cognitive capacity”, producing more meaning and yielding a more complete comprehension. Narrative’s enhanced memorability is also described, with authors arguing that narrative texts provide an organising structure for knowledge to be more effectively built and recalled (e.g., Bruner, 1986; Kintsch, 1998; Mandler, 1984).

However, these comparisons between the comprehension and memory effects of narrative texts and expository texts are problematic for the question at hands, as these narrative texts *did not* convey science contents, but instead daily, and way more familiar, topics. This is crucial, and calls attention to the fact that different text structures or text types, such as narrative texts and expository texts, tend to be associated to specific contents and have different purposes. These differences are rooted in and shaped by wider discursive and sociocultural practices and influence the way texts are perceived and processed (e.g., Adam, 1997; Kintsch, 1998; Snow, 2002). Unlike narrative texts, which follow a temporal structure, in expository texts contents are usually structured according to concepts and their relations (Meyer, 1975). Being the most commonly used text in science education (e.g., Avraamidou & Osborne, 2009), expository texts have gained the reputation of being abstract and difficult (Best et al., 2005; Graesser et al., 1991). Narrative texts, on the other hand, are not usually the chosen type of text when the goal is to instruct students on science contents. Although there are some exceptions, narratives tend to have broader or complementing goals, such as providing historical context or details on the scientific process (e.g., discovery narratives, Curie, 1904), communicating science to a wider and non-specialized audience (e.g., popular science books, e.g., Brown, 2003), or entertaining readers (e.g., science fiction, Dnieprov, 1969; short stories, Levi, 1985).

The notion of a memory, learning, and interest advantage for narrative materials, is sometimes termed the *narrative effect* (e.g., Norris et al., 2005). It was thus important to ascertain what kind of learning outcomes can be generated by science narrative texts, and whether they point to a generalized narrative effect. For instance, previous studies on science narrative texts have observed different results in different conditions, depending for instance on the text’s topic (e.g., Arya & Maul, 2012), the assessment measure (e.g., Hong & Lin-Siegler, 2012), the delay of assessment (e.g., Negrete & Lartigue, 2010), and the participants’ level of prior knowledge (e.g., Wolfe & Mienko, 2007). Given the variability of the reported results, and of the studies themselves (in terms of e.g., paradigm; population and/or materials), it is unclear whether narrative texts are particularly suited for specific science topics, retention intervals (e.g., immediate vs. delayed, e.g., Negrete & Lartigue, 2010), levels of comprehension (e.g., recall vs application of learned ideas, e.g., Arya & Maul, 2012), as measured by different assessment measures. Importantly, due to the strong association between narrative materials and children, particularly in learning contexts (Sanacore, 1991), young adults are an understudied population in what concerns the use of narrative texts for science learning (for exceptions, see Negrete & Lartigue, 2010; Wolfe & Mienko, 2007; Wolfe & Woodwyk, 2010). In addition, the impact of text types on learning can be better ascertained among participants with low levels of prior knowledge on the texts' topics (e.g., Prins et al., 2017).

Drawing on the selected theoretical frameworks, our second research question addressed these various aspects:

RQ 2: Can people, in particular young adults with little prior science knowledge, learn science from narrative texts at different levels of comprehension, and even better than from expository texts?

This second question has both empirical and theoretical implications: If, on the one hand, it concerned empirical evidence on learning outcomes, it could further contribute to determine the extent of the conceptual proposal of a generalized narrative effect. It also partially echoes one of the questions that the educators Norris and colleagues (2005, p. 559) suggested to guide future research, "What are the implications of any narrative effect for teaching science?"

A final question concerns the **process of learning** science through narrative texts. Our third research question was directed at this matter, again drawing on the selected theoretical frameworks:

RQ 3: How does the process of learning science through narrative texts occur?

Importantly, the question touching on the process of learning science through narrative texts was threefold. A first, and more specific, facet of this learning process comes in the form of human-related elements and the processes they trigger. This question is key, as many of the claimed benefits of narrative learning materials are based on the idea that narrative texts more aptly retrieve the human side of science, as they centre on human action and are socially and culturally rooted. It is for instance argued that science narrative texts can offer enhanced contextualization, as they can more easily connect science contents with learners' own experiences and daily situations (Murmann and Avraamidou, 2014; Mutonyi, 2016), or with scientists' actions and feelings and the wider context in which science contents are developed (e.g., Arya & Maul, 2021; Clough, 2011; Kubli, 2005). On the one hand, the incorporation of such elements can be particularly useful to show that science is an ever-changing and culturally-charged process made by people, who struggle to establish the currently best perception of truth amidst setbacks and uncertainty (e.g., Abd-El-Khalick, 1999; Bruner, 1996; Hong and Lin-Siegler, 2012). Encouraging a more human and social frame of the scientific process can humanise scientific meaning (e.g., Egan, 1997; Hadzigeorgiou et al., 2012), providing a more accurate image of science (e.g., Arya & Maul, 2021; Clough, 2011; Kubli, 2005) that contributes to diminish science's authoritative voice (Arya & Maul, 2021; Kloser, 2013; Mutonyi, 2016).

Furthermore, it can reveal a fundamental humanness in science, capturing reader's attention and interest through the inclusion of human affairs, emotions, thoughts, hopes, frustrations, among others (e.g., Arya & Maul, 2012; Hadzigeorgiou et al., 2012; Hong & Lin-Siegler, 2012), making scientists more relatable in learners' eyes (e.g., Hadzigeorgiou et al., 2012; Hong & Lin-Siegler, 2012; Klassen, 2006), and ultimately making science more inclusive (Arya & Maul, 2021; Gilbert et al., 2005; Mutonyi, 2016). There is indeed evidence, mostly qualitative, that science narrative texts can encourage learners to make connections between science contents and human-related information (e.g., Arya & Maul, 2021; Hadzigeorgiou et al., 2012; Lin-Siegler et al., 2016), to see scientists as struggling and hard-working individuals and feel connected to them (e.g., Arya & Maul, 2021; Hadzigeorgiou et al., 2012; Lin-Siegler et al., 2016), and even to adopt the narrative character's point of view (Jetton, 1994; Murmann & Avraamidou, 2014). Still, this aspect has not been closely and more directly investigated. Importantly, although studies with non-science narrative have showed that narrative texts can engage a range of socio-cognitive abilities such as the ones described previously (e.g., perspective thinking; theory of mind) these remain practically unexplored in the context of science narrative texts.

These aspects can easily be related to a consistent body of literature that has investigated how reading non-science narratives texts can impact a series of affective and socio-cognitive processes. Indeed, a fundamental characteristic of narrative is their power to evoke powerful emotions and simulate social processes. Emotional involvement is a fundamental part of the narrative reading experience (e.g., Bruner, 1986; Egan, 2005; Oatley, 2016), and becoming part of narrative worlds can help readers simulate the complexities of the social world through a range of socio-cognitive abilities and gain a deeper understanding of themselves and others (e.g., Djikic et al., 2013; Oatley, 2016), potentially even improving those abilities. Examples of such abilities include theory of mind, perspective-taking (i.e., the capacity to recognize, and potentially adopt, another person's point of view, Baron-Cohen, 2001), and empathy (i.e., the ability to understand or feel what someone else is experiencing from *their* frame of reference; Bellet & Maloney, 1991). Previous studies have established connections between these abilities and reading narrative texts (e.g., Kidd & Castano, 2016; Mar et al., 2006; Mar & Oatley, 2008). However, as these texts did not convey science contents, and thus did not have the aim of instructing people on such matters, it remains to be ascertained whether these processes are also engaged when reading science-conveying narrative texts, and whether this engagement impacts science learning.

These concerns can be related to another of the suggested research questions from Norris et al. (2005, p. 559), namely "What features of narrative prove through empirical research to be most crucial, and how do they operate?". Accordingly, the following, more specific, research question was formulated:

RQ 3a: How does the processing of human-related elements in science narrative texts impact science learning, at different levels of comprehension, and which kind of processes does it engage?

In addition, little is known about the more general processes contributing to the science learning outcomes generated by narrative texts. Whereas assuming learning outcomes and engagement of specific processes based on previous studies that used non-science narrative is potentially misleading, such literature can offer important hints and inspire hypotheses, that should, nonetheless, be directly investigated and tested. At the outset, many authors claim that science narrative texts can catch learners' interest more easily than expository texts (e.g., Arya & Maul, 2012; Hong & Lin-Siegler, 2012), and have a more accessible and familiar language and structure (e.g., Avraamidou & Osborne, 2009; Browning & Hohenstein, 2015). Most studies have focused on how narrative-based materials or interventions can impact specific aspects, such as enhanced interest (e.g., Ritchie et al., 2011), attention (e.g., Hadzigeorgiou et al., 2012), proactivity during the learning process (e.g., Akarsu et al., 2015), or make use of previous knowledge (e.g., Wolfe & Mienko, 2007). Yet, few studies directly investigated how these aspects contributed to the observed learning outcomes. For instance, Reuer (2012) found a positive correlation between interest in the narrative texts and the outcomes they produced, and Wolfe and Mienko (2007) observed that less knowledgeable students (i.e., with lower levels of prior science knowledge) benefited more from narrative texts that more knowledgeable ones. Yet, the latter authors did not find a correlation between working memory capacity and learning scores. Crucially, the impact of literacy-related aspects, such as reading habits and experience, has scarcely been investigated in the context of learning science through narrative texts, even though it is quite common in studies using non-science narrative texts (e.g., Kidd & Castano, 2016; Mar et al., 2006).

The following question addressed the potential impact of these more general processes when learning from science narrative texts:

RQ 3b: What kind of learner features impact the process of learning science through narrative texts?

Lastly, and as a complement to the quantifiable cognitive measures of features and processes impacting learning, it is relevant to know more about the perceptions that people have on the use of narrative texts in science education. Interviews with students and teachers have shed some light on the perceptions these actors hold on the use of narrative texts as science learning tools, both positive (e.g., is enjoyable, Ritchie et al. 2011; aids in recall, Murmann & Avraamidou, 2014) and negative (e.g., not appropriate for learning and hard to tell between fact and fiction, Prins et al., 2017). However, there is still much to learn about this issue outside the classroom. Given the pivotal and socially rooted role of lifelong science learning, recently illustrated by the COVID-19 pandemic and by ongoing debates around climate change, public perceptions of science have become increasingly significant (e.g., Abdool Karim, 2022; A. Costa, 2021; Yang et al., 2020). In this context, non-formal learning settings, such as museums, become key places for communicating science to broader audiences and providing them with opportunities for science learning, where innovative pedagogical practices, such as narrative-based ones, can be more easily applied (Callanan et al., 2011; Murmann & Avraamidou, 2014). Examining the perceptions elicited by non-formal tools, such as narrative texts and museums, can therefore bring vital information, and can even be seen as an additional way of retrieving the humanness of the scientific learning process.

This confluence of ideas led to our final research question:

RQ 3c: What are people's perceptions on science learning and literacy, particularly on the role played by narrative texts and museums?

In conclusion, we expected that the overall findings resulting from the three outlined research questions, concerning a literacy-based theoretical backbone, learning outcomes and the learning process, could provide a multidimensional lens and prove relevant for conceptualization, designing research and educational interventions, interpreting results, and better understanding the process of learning through science narrative texts.

1.5. Overview of the dissertation

The present dissertation is organized in seven chapters (see Figure 1.2. for a synthesis of its structure). In **Chapter 1**, which corresponds to the present chapter, we introduced the problem and the main concepts and ideas that support our research questions and aims. Namely, we outlined the problematic of science learning and the challenges that come with it, how it connects to the perception of a gap between everyday and scientific thinking, and why narrative texts can be a tool to bridge this gap and improve science learning. An overview of the structure of the dissertation is also offered.

In **Chapter 2** we conducted a theoretical review with two overarching aims. The first was to provide a theoretically-grounded mapping that drew on complementary theoretical frameworks, namely from text linguistics, cognitive psychology and pedagogy. The second aim was divided in two complementary questions which mirror the second and third research questions of the dissertation. First, we were interested in examining *if* there was evidence that narrative texts have consistently benefited retention and learning from science at different levels of education, by selecting and examining a set of previous studies. Second, we sought to analyse *how* those educational outcomes might occur, namely the conditions and underlying mechanisms of the learning process, by establishing connections with a broader literature.

In **Chapter 3** we detail how we developed a set of science texts and corresponding learning items. Namely, we conducted a set of pretests to develop the science educational materials that would be used in following empirical studies to answer the research questions of the present dissertation. We aimed at combining the insights from theoretical frameworks described in Chapter 2 with those from coming from a range of linguistics and science experts, ongoing findings, and participants' feedback. Besides building texts of different text types (i.e., narrative and expository) and comparing their impact on learning, we also sought to build texts on different science topics, as this allowed us to further examine the extent of a potential narrative effect in science learning. We expected to build at least two pairs of texts from two science topics, controlled in a set of parameters, and corresponding learning measures.

In **Chapter 4** we conducted the first main empirical study of the dissertation. We investigated whether young adults with little prior knowledge could learn science from narrative texts, particularly when compared with expository texts with equivalent contents. We examined the impact of math and chemistry narrative texts and expository texts at four different levels of comprehension, requiring different levels of elaboration, and two retention delays. We also evaluated the impact of a set of learners' features on this learning, to uncover more about the learning process, namely on what was common to both text types and what was specific to narrative texts. These features were learners' contact with literacy, science background, and evaluative and motivational attitudes. We expected an advantage for at least one of the narrative texts, particularly at more elaborate comprehension levels and on delayed measures, as well as a positive impact of the examined learners' features.

In **Chapter 5**, akin to Chapter 4, we analysed the science learning outcomes of young adults with little prior knowledge, as well as the same set of learners' features. However, this time we focused on narrative texts and on the impact of human-related elements on science learning. We recorded participants' eye movements while they read the science narrative text, to examine how regions depicting specific kinds of human actions were processed. We expected that dedicating attention to these human-related regions would impact science learning. Moreover, we also applied a set of tasks evaluating both the extent to which reading the narrative text prompted human-related thoughts, as well as a more general socio-cognitive ability. We expected to observe inter-individual variability on thoughts on human action, and that these thoughts impacted science learning to some extent.

In **Chapter 6** we aimed at building a more global picture of what science learning means to people, and particularly on the role that narrative texts and museums may play in this learning. We examined the perceptions of a group of people with different levels of contact with science, who participated in an online non-formal learning experience. The learning experience was developed as a science museum activity, and included reading science texts and taking part in a science learning activity facilitated by a museum science educator. The learning activity was followed by focus group discussions on the topic of science learning and literacy, and the role of specific tools.

Lastly, **Chapter 7** presents a summary of the main empirical findings, and discusses their theoretical and applied contributions, particularly regarding the question of whether narrative texts can be a useful tool for tackling challenging aspects of science learning and improving that same learning, taking into consideration theoretical frameworks, learning outcomes, and learning process. The general limitations of the present work are also discussed, and suggestions for future research are given. The chapter ends with concluding remarks.

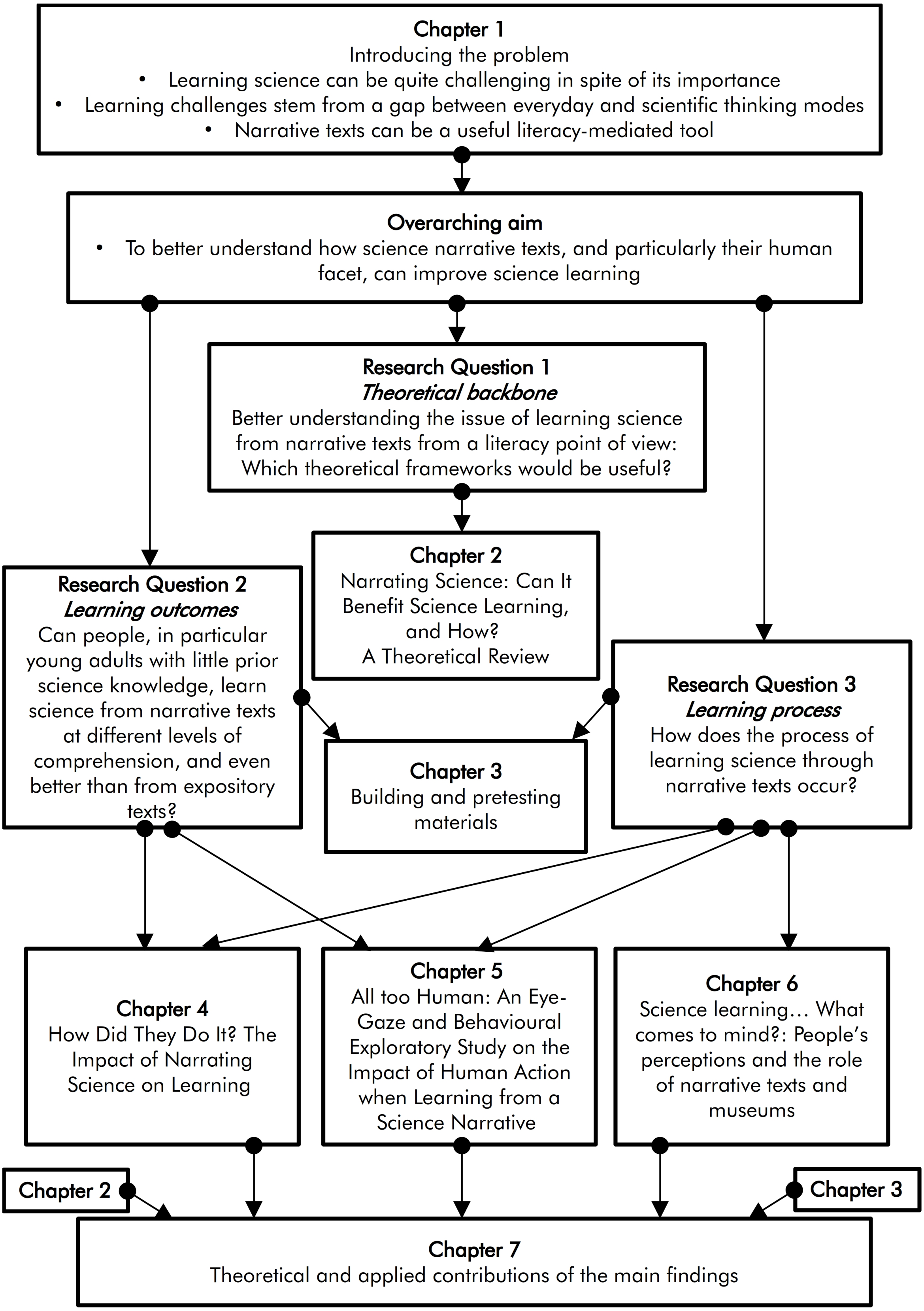


Figure 1.2. Synthesis of the structure of the dissertation. Overview of the problem, research questions, aims, and chapters

CHAPTER 2

Narrating Science: Can It Benefit Science Learning, and How? A Theoretical Review

This chapter is based on the manuscript: Soares, S., Gonçalves, M., Jerónimo, R., & Kolinsky, R. (2022). *Narrating Science: Can It Benefit Science Learning, and How? A Theoretical Review* [Manuscript accepted for publication with pending minor revisions], *Journal of Research in Science Teaching*

2.1. Abstract

Narrative texts have been advocated as tools to tackle science learning challenges, and there is even the proposal of a "narrative effect" on learning. We believe it is necessary to examine previous evidence on this effect, as well as to characterize the process of learning through science narrative texts more broadly. In this paper, we offer a theoretical review drawing on three frameworks, namely on pedagogical aspects of text learning, linguistic features of texts, and cognitive aspects of text comprehension. Based on that, we analyzed two complementary questions. First, we reviewed 36 studies to ask if science narrative texts can benefit learning and memory outcomes at different educational levels (i.e., the "If" question). We found encouraging evidence for the use of science narrative texts at various educational levels, especially in delayed assessments and longer-lasting interventions. Second, we gathered and linked ideas, hints, and evidence on how the process of learning with science narrative texts takes place, namely on conditions and underlying processes (i.e., the "How" question). There are many features from conditions (texts, learners, activities, wider context) and underlying processes (integration with prior knowledge, affective dispositions, cognitive abilities) that can help to account for variability in outcomes; yet ideas and evidence are not always tightly connected. We suggest that education and research should focus on specific narrative effects, that specify with what (texts), with whom (learners), when and where (activities and wider context) these effects occur, as well as the "why" (underlying processes). We believe the proposed framing can help both make sense of previous evidence and inform future educational practices and research and provide some recommendations in this regard.

Keywords: science learning, science narrative texts, narrative effects, text comprehension, learner features

2.2. Introduction

"The universe is made of stories, not of atoms"

(Rukeyser, 1968, p. 111)

Many authors contend that the challenges of science learning should be addressed by improving language and literacy processes (e.g., Morais & Kolinsky, 2016; Klein, 2006; Norris & Phillips, 2003; Webb, 2010). At the same time, these challenges are thought to stem from a fundamental difference between everyday and scientific modes of thinking (e.g., Bruner, 1986, Egan, 1997; Phillips & Norris, 2009). For the latter reason, narrative texts, which are generally viewed as temporally organized actions or events (e.g., Adam, 2011; Norris et al., 2005; Strube, 1994), have been advocated as effective tools to tackle the challenges of science learning, which is commonly based on expository texts (e.g., Arya & Maul, 2021; Olson, 2015; Solomon et al., 1992).

The idea that narrative materials can improve the understanding and retention of information is sometimes termed the narrative effect (e.g., Norris et al., 2005). Yet, it comes from theoretical (e.g., Bruner, 1986) and empirical (Graesser et al., 1980; Kintsch & Young, 1984; Zabrucky & Moore, 1999) works based on non-science narrative texts. It is therefore relevant to ascertain if narrative materials actually consistently benefit science learning.

Additionally, learning occurs through the combination of different elements (Snow, 2002). Namely, texts have specific features (e.g., Adam, 1997) and are cognitively processed by readers in specific ways during learning activities (e.g., Kintsch, 1998), all these aspects interacting within and with a wider context (e.g., Adam, 1997; Snow, 2002). To give a few examples, qualitatively different contents have been used in science narrative texts (e.g., fiction, Banister & Ryan, 2002; non-fiction, Hong & Lin-Siegler, 2012), as well as different activity goals (e.g., studying, Wolfe & Mienko, 2007; evaluating text quality, Arya & Maul, 2012). Science narrative texts have also been claimed to connect to readers' social and cultural identities (e.g., Mutonyi, 2016), and to engage differently processes such as integration with prior knowledge (e.g., Maria & Johnson, 1989), emotions (e.g., Murmann & Avraamidou, 2014), and attention (e.g., Hadzigeorgiou et al., 2012). Thus, it is also important to characterize how the process of learning science through narrative texts takes place.

In short, providing a theoretically grounded examination of whether science narrative texts consistently improve memory and learning outcomes, as well as insights into the characteristics of the learning process that can lead to such outcomes, is an important step toward better understanding this science educational tool.

2.3. A Theoretical Review on Narrating Science for Learning

The goal of the current paper is to analyze by means of a theoretical review two questions pertaining to the topic of learning science through narrative texts, that we believe to be of relevance for educators and researchers in education. To the best of our knowledge, such a review has not yet been provided.

Our first question (henceforth, the “If” question) is whether there is evidence that narrative texts have consistently benefited retention and learning from science at different levels of education. Learning and retention are relevant cognitive and pedagogical outcomes (e.g., Kintsch, 1994) whose conditions may depend on learners’ educational level. Although the use of narrative educational materials is conventionally associated to young children, it has been claimed that these materials may benefit older learners as well (e.g., Klassen, 2006; Olson, 2015). This question will be examined by using a set of studies chosen on the basis of specific criteria (see Method).

A follow-up question (henceforth, the “How” question) concerns the characteristics of the learning process that may lead to the aforementioned educational outcomes. We will explore the conditions involved in this process, as well as the mechanisms that may underlie it, by establishing connections with a broader literature.

To our knowledge, there has been no strong theoretical framework guiding the interpretation of previous studies and the planning of future interventions. In the present theoretical review, we draw on a set of theories from relevant disciplines to accommodate the different aspects that our questions touch on.

As we aim at connecting science learning to literacy processes (e.g., Morais & Kolinsky, 2016; Norris & Phillips, 2003), we draw on pedagogical aspects pertaining to learning through reading. We chose the framework outlined by the Reading for Understanding (RAND) Reading Study Group (Snow, 2002) for several reasons. First, because it subscribes to the idea that improving learners’ literacy skills promotes content learning. Second, because it recognizes the multifaceted nature of learning, providing a backbone of the conditions that should be considered when planning interventions or analysing its outcomes. These elements combine research traditions (e.g., Pearson & Cervetti, 2015) that we see as highly relevant and complementary for the question at hands. Indeed, RAND builds its proposal around three elements that have been very present in cognitive models (i.e., text, reader, and activity), but whose interactions are context-dependent, which has been given more attention in sociocultural models. Finally, RAND summarizes and structures these elements within a policy-context, aiming at providing guidelines for research and development that focus on text-and discipline-specific reading practices. These aims are in agreement with the aims of the current review.

Because textual materials are a key aspect of our theoretical review, we also draw on a framework that describes its features. We chose text linguistics (TL), particularly the francophone line (e.g., Adam, 1997; Bronckart, 1997), because its conception of texts as social objects made up of textual and contextual features (Gonçalves, 2019) provides us with a better grasp of the features that make up science narrative texts, as well as with an important sociocultural lenses, as it brings wider sociodiscursive practices into play.

Finally, we also integrated cognitive aspects of text comprehension, as they provide valuable insights on memory and learning (i.e., the outcomes under examination) and on how readers cognitively process science narrative texts (i.e., the processes underlying the examined outcomes). We selected the Construction-Integration (C-I) model (e.g., Kintsch, 1988, 1998) because it is regarded as the most comprehensive cognitive model of text comprehension (McNamara & Magliano, 2009), dedicates special attention to science texts, and has been the predominant paradigm when conceptualizing basic processes and pedagogical practices for comprehension (e.g., Pearson & Cervetti, 2015).

Drawing on this multidisciplinary effort, we will examine the two questions outlined in the beginning of this section. In the following section, we will present the selected theoretical frameworks, underlining both their shared and specific (and thus complementary) aspects.

2.4. Learning Science from Text: A Proposal Drawing on three Frameworks

For the RAND model (Snow, 2002; Sweet & Snow, 2003), text comprehension is always a specific combination between features from the text, the reader, and the activity. These elements are highly permeable to each other's influences (dashed lines in Figure 1) and interact both within a wider sociocultural context (the wider circle surrounding these elements in Figure 1) and also with it (middle circle in Figure 2.1). Reading and learning by reading occur at the interfaces of these elements, and the process of learning is deemed as important as its content.

TL and RAND conceive literacy as a cultural and historical activity (e.g., Adam, 1997; Snow, 2002). Texts are viewed by TL as social objects that connect to expectations and practices of the wider sociocultural context. For instance, a text with fantastic elements it expected to entertain, hence the fairy tale genre is not commonly found in science education. In other words, texts are governed by textual genres (Adam, 1997; Bakhtin, 1984; Dijk & Kintsch, 1983), such as fable, scientific report, or cookbook, which are abstract models of what is to be expected and adopted in specific communicative situations (Bronckart, 1997). As genres have conventionalised structures, purposes and target audiences, the latter influence how texts are perceived and processed (e.g., Hidi et al., 1982; Rastier, 2001; Zwaan, 1994), connecting the textual and pragmatic components of a text (Adam, 1997).

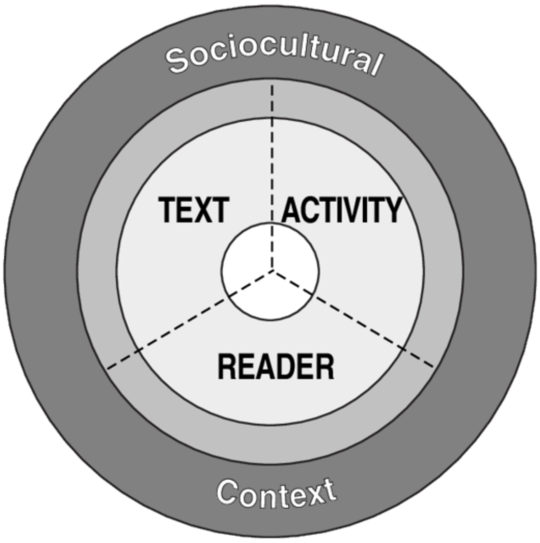


Figure 2.1. The RAND Framework Proposal (Taken from Snow, 2002)

According to TL, the textual component has different levels (see the arrows in Figure 2.2), namely clauses, sequences and text plans. Sequences belong to five prototypical categories: narrative, descriptive, explicative, argumentative, and dialogical. Crucially, these heterogeneous sequences are intermingled in most texts (Adam, 1997; 2011). Sequences are organized by a common text plan, which determines the global configuration of the text. In a narrative text, sequences are temporally organized; in expository texts, they are organized by topics and ideas.

Also according to TL, contextual aspects, or the pragmatic component, influence the configuration and processing of texts as well. Pragmatic aspects include enunciative features, content (or “semantic representation” in the original terminology), and communication aims (or “illocutiveargumentative”). Narrative texts often include fictional contents and have the aim of entertaining; science texts tend to stick to factuality and aim at instructing. The pragmatic and textual components interact constantly (the dashed lines in Figure 2.2).

What makes a text being perceived as a narrative text is therefore not homogeneous narrative sequences, but a combination of textual and pragmatic features. As, overall, texts are highly variable and heterogeneous, by definition all texts are, to some extent, “mixed” or “hybrid” (e.g., Hidi et al., 1982; Norris et al., 2005). Materials only containing prototypical narrative features have likely been extracted from larger heterogeneous texts, or carefully built that way for a specific pedagogical or experimental purpose. They would more aptly be defined as sequences than as texts; yet, for simplicity’s sake, we will refer to all materials as “texts”.

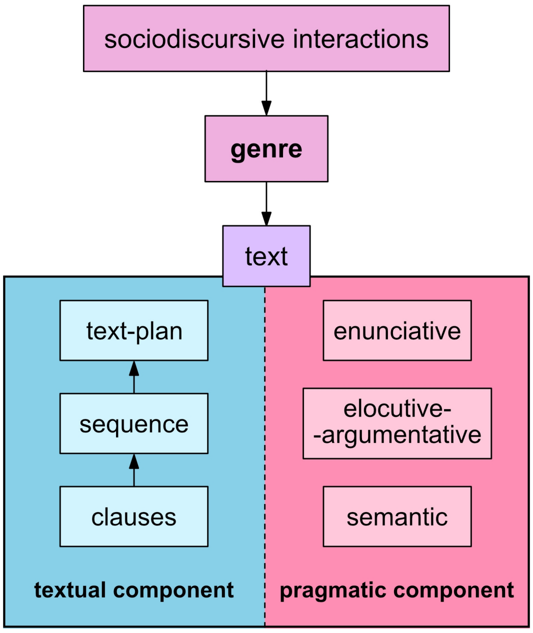


Figure 2.2. A Text Linguistics’ View of Text Features (Adapted from Adam, 1997; 2008)

Both a novel about fictional characters’ lives and a science discovery narrative have mixed features, yet the latter will more likely be perceived as “mixed”. Factual science contents (in contrast with science fiction) are not usually associated to an overall narrative structure, as this combination is present in few textual genres. This difference sets studies with science narrative texts apart from the studies on narrative comprehension they frequently draw on. Nonetheless, some science educational materials incorporate narrative features. To name a few examples, some texts add narrations about scientists thoughts and actions to provide a personal and social context to the science contents (e.g., hybrid adapted primary literature, Shanahan, 2012). Others present science contents as part of a story with characters and events (e.g., secondary literature or popular fiction, Baram-Tsabari & Yarden, 2005), or of a historical narrative that explains how concepts were discovered or developed, as to make more explicit how knowledge is constructed (e.g., epistemologically considerate texts, Kloser, 2013).

The communication aim of the text, whether externally communicated or inferred, has a crucial impact on how the activity is perceived and executed (e.g., to entertain vs. to learn), and hence on its comprehension outcomes. This notion is shared by all the theoretical models, and illustrates an interaction between the text, the activity, and wider sociocultural conventions.

Just as texts have different levels of organization, so do readers' representations of them. Drawing on a similar body of research, the three frameworks acknowledge that readers build various representations of the text's information (e.g., Kintsch & van Dijk, 1975; van Dijk & Kintsch, 1983). However, it is in the C-I model that these representation levels are thoroughly developed (e.g., Kintsch, 1998; Kintsch & Rawson, 2005). This model proposes that readers construct (i.e., represent meanings) and integrate (create a coherent representation) information into representations through an interactive interplay of text-driven and reader-driven processes, and focus on three levels of representation (see Figure 2.3). At the surface level, linguistic information is represented literally (i.e, exact wording and phrasing), which is generally assumed to have little effect on comprehension (McNamara & Magliano, 2009). At the textbase level, the explicit meaning of the text is represented by a propositional network of interrelated idea units, based on words, their syntactic relationships, and inferences generated for text cohesion. These idea units are organized into higher meaning units according to global topics and their interrelationships that link larger portions of the text, and often follow the conventions of familiar schemata. Schemata are mental structures containing knowledge (e.g., elements; rules; strategies) on specific genres or discourses (e.g., fairy tale vs. informational piece), that orient and facilitate information processing and comprehension (Adam, 2011; Kintsch & Rawson, 2005; van Dijk & Kintsch, 1983). This is another illustration of how wider contextual conventions influence activities' aims and readers' cognitive processing of the text.

In addition to deriving relations between information explicitly mentioned in the text, which is fairly shallow, readers elaborate on this propositional network, generating inferences and integrating their own experience and world knowledge. Integration with prior knowledge cannot take place when this knowledge is inadequate or absent, but texts containing only already known information are also useless for learning (Wolfe et al., 1998). This deeper level of representation is called the situation model, and it involves a set of knowledge, affective dispositions and cognitive abilities, including previous experiences, motivation, memory and visual imagery (e.g., Kintsch, 1998; Xu et al., 2005). The extent to which these reader's features are engaged would depend on the interactions with features from the text, the activity, and the wider context. The C-I model acknowledges that the situation model is influenced by contextual features such as the genre, discipline, and goals (Kintsch, 1998; Dijk & Kintsch, 1983). Yet, these features have a more supporting role in the model, which places greater emphasis on reader and text features.

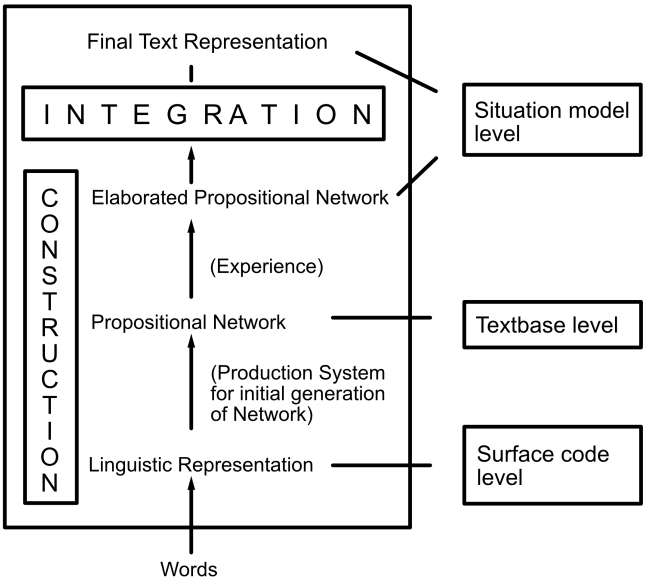


Figure 2.3. Schematic Overview of the C-I model (adapted from Wharton and Kintsch, 1991)

The representation of a text's explicit meaning (i.e., textbase level) and its deeper integration with extra-text information (i.e., situation model) occur simultaneously and interdependently (Kintsch, 1988). Yet, for conceptual and pedagogical purposes, they are treated as memory and learning outcomes, respectively (Kintsch, 1994; 2012).

We believe that these theoretical frameworks provide a complementary multidisciplinary view of the different elements we wish to examine. Using this theoretical background, we will now describe the methods applied to address the two questions addressed in the present review.

2.5. Method

2.5.1. The "If" Question

The aim of the "If" question was to examine evidence on memory and learning outcomes at different educational levels, having the outlined theoretical frameworks as theoretical grounding. We did this by selecting research papers that evaluated these outcomes according to a set of four inclusion criteria listed in Table 1 and analysing their findings.

As we were interested in the role of narrative texts as science education tool, (1) texts had to be the central learning activity, and had to be read, listened to or written, as these activities are integral to literacy (e.g., Morais & Kolinksy, 2021); in addition, (2) text materials must be described as having narrative, story or novel features (the latter two convey central narrative features, e.g., temporal organization), even if other terms are used simultaneously (e.g., "hybrid"); authors' definitions were not called into question.As we wanted to examine comprehension outcomes at different educational levels, (3) participants should be students, i.e., completing a formal educational degree (or in preschool, which in some countries is considered formal education) and (4) outcomes must examine memory and/or learning of contents from the scientific body of knowledge (i.e., explanations intrinsic to science, Norris et al., 2005); other outcomes could be analysed (e.g., affective aspects; understanding the nature of science), insofar as the outcomes of interest were directly examined.

Table 2.1. Inclusion and Exclusion Criteria for Selecting Studies Assessing Memory and Learning

Criteria	Inclusion	Exclusion
1. Reading, writing, or listening to narrative texts	Texts are the central task of the learning activity	Texts are secondary
2. Definition of narrative	Narrative/narrative features, story, or novel	Only other terms are used (e.g., primary vs. secondary literature; history of science)
3. Participants	Students, from preschoolers to pre-service teachers	Others (e.g., teachers)
4. Educational outcomes	Memory or learning is one of the outcomes	Only other outcomes (e.g., engagement; understanding the nature of science)

To include different materials, we searched for papers in the following databases: Web of Science (very encompassing), ERIC (directed at education) and OATD (theses and dissertations). It is worth noting that the main addressed concepts were quite encompassing (e.g., science; learning; narrative), and that the outcomes of interest could be described using a variety of terms (e.g., comprehension, learning, recall, memory). As a result, broader searches would have resulted in many papers that did not fit our aims; yet, narrowing the search terms too much would likely result in missing relevant papers. In an attempt to achieve a balance between these extremes, we opted for flexibly combining the following search terms: “science”, “scientific”, “narrative text”, “learning”, “comprehension”, “memory”, “retention”, “recall”, “teaching”. As effective Boolean queries including various terms are difficult to achieve (e.g., Karimi et al., 2010; Scells & Zuccon; Wang et al., 2022), it proved hard to apply a stricter search methodology that would have allowed to perform an exhaustive search of the retrieved results. Thus, we performed various searches using relevant combinations of the terms, guided by the goal of examining the overall (in)consistency of previous evidence on narrative learning benefits. The search terms were preferentially looked for in the abstracts. The search covered works produced between January 1990 and December 2019, as many of the seminal theoretical and empirical works on the benefits of narrative texts for memory and comprehension dated from the 1980’s. Studies could be written in English, European Portuguese, Brazilian Portuguese, Italian or French, and we also did some searches using the search terms translated in these languages.

We began by checking the titles and abstracts against the inclusion criteria. When papers were related to our question, we inspected them fully for potentially relevant references. We started by inspecting the methods and results sections, to check whether memory and learning outcomes had been directly addressed. These outcomes should have been to some extent learned through a text-based activity of the study. We did not consider papers in which the learning activity was only a means of applying knowledge acquired before/elsewhere. We were interested in both quantitative and qualitative measures. Quantitative measures had to be supported by some numerical data (e.g., means; percentages; test statistics). Qualitative measures had to include some sort of analysis or explanation of the observed comprehension outcomes, connecting them with least one illustrative example. We did not include papers which only presented general statements on learning gains or on the quality/relevance of narrative tools or interventions. When papers fitted our purposes, we also checked the papers that cited them.

We analysed the findings of 36 papers, whose main characteristics are summarized in Table 2.2 (see also Tables 2.3 and 2.4). Among them, 20 compared narrative materials to some control material (narrative vs. control studies; henceforth, N vs. C). In the other 16 studies, narratives were examined as stand-alone tools (narrative-only studies; henceforth, N-O). As using science narrative texts is often presented as an alternative to conventional teaching methods, directly comparing narratives with a control material provides more straightforward evidence on a potential narrative effect in science learning, while also minimizing confounding effects. However, many educators are not interested in comparing different teaching methods, but rather in exploring in more depth the narrative tool.

Table 2.2. Measures, Design and Tasks of the 36 Studies Selected to Examine the "If" Question

	Quantitative measures	Qualitative measures	Mixed-methods	Pre-post design	Delayed assessment	Written tests or questions	Interview, discussion or observation	Recall, retell or rewrite	Writing stories or journaling	Drawings or hands-on-activities
N vs. C	17	0	3	9	7	16	4	7	1	0
N-O	6	6	4	8	4	7	5	5	6	5
Total	24	6	6	17	11	24	9	12	7	5

Note. The first three columns refer to the outcomes we wished to examine, not necessarily to the methodology of the full study.

Studies using delayed assessment could have applied a pre-post design or other (post-test or case study).

A same study could have examined memory and/or learning using more than one type of task.

Recall and retell could have been done orally (e.g, interview) or through writing.

These different paradigms often used different methods and tasks to examine the outcomes of interest. A similar number of N vs. C and N-O studies (3 and 4, respectively) used mixed methods to assess memory and learning. Yet, whereas N vs. C studies used predominantly quantitative measures, the same number of N-O studies (i.e., 6) used either quantitative or qualitative measures. A similar number of N vs. C and N-O studies (i.e., between 5 and 7), examined memory and/or learning using interviews, discussions or observations, and recall, retell or re-rewrite. All but four N vs. C studies used written tests of questions, while N-O studies used more diverse tasks, such as storywriting or journaling and drawings or hands-on-activities. Despite these differences, the general design is often similar. In about half of the studies of each paradigm, participants' knowledge was assessed before and after reading, writing, or listening to the texts (i.e., pre-post design). The remaining studies either only applied post-tests or analysed several measures in the context of case studies. Regardless of paradigm and design, some studies also included delayed assessments.

In the case of quantitative comparisons, we verified whether there was evidence that the interventions produced gains in memory and/or learning (e.g., pre vs. post measures) and whether these gains were stronger in specific conditions (e.g., with N vs. C; with different narrative texts; in delayed vs. immediate assessments). When effect sizes (ES) were not reported and sufficient information was available, we calculated them for between or within effects (Lakens, 2013). For the sake of clarity, when studies included several tasks, conditions (e.g., groups) or factors irrelevant to our purpose, we either made a selection or aggregatedthe results when they were of similar direction and significance (see notes of Tables 2.3 and 2.4). As regards qualitative analyses, we checked which kind of memory and/or learning outcomes were observed, and how they were patent in the illustrative examples presented by the authors.

As we will discuss further in the Results, the C-I model differentiates memory from learning. Yet, memory and learning are intertwined at both the cognitive and design level, with some measures tapping both processes at once (e.g., tests; interviews). When tests were reapplied after a delay, we considered them asassessing both learning and memory. In addition, some studies explicitly stated they included measures tapping different levels of representation but did not present results separately; in this case, we interpreted them as pertaining to both levels.

2.5.1. The "How" Question

The aim of the "How" question was to provide an overview on aspects that, based on the outlined theoretical frameworks, should be relevant to characterize the process of learning science through narrative texts. We did this by gathering and linking ideas, hints, and evidence, establishing connections with a broader literature in a more exploratory mode.

The 36 studies examined in the "If" question were also used as the main basis for the analysis, whenever they contributed to characterize the process leading to the observed learning outcomes. Many of the variables not included in the analysis of the "If" question were relevant and thus were considered here, such as prior knowledge, interest, or readers' social and cultural identities. However, we also drew on and established connections with a broader literature with that same purpose of characterising the underlying process in mind. We included information from theoretical and empirical works on science narratives and on conventional narratives more generally. These were mostly retrieved during the previously described searches, especially by inspecting the full papers for relevant references. When relevant, we also made connections to more general research on learning and science learning, and to literature on specific features from the examined conditions and processes. These literatures were again not exhaustively searched or described, as our goal was to sketch a theoretical overview. The presented evidence is also not exhaustive, but instead illustrative. We included different kinds of evidence: direct, that is, the feature has been investigated in direct connection to learning outcomes stemming from science narrative texts; indirect, that is, the feature has been investigated in a learning intervention that used science narrative texts but was not directly connected to learning outcomes; or more tentative, namely, the feature was investigated using non-science narrative texts.

We organized the analysis of this question under two main sub-questions, each addressing specific features. Namely, drawing on the outlined theoretical frameworks, we mapped a set of conditions (texts, activities, and populations, as well as their interactions with the wider context) and underlying processes (prior knowledge, affective dispositions, and cognitive abilities) that are relevant to characterize the process of learning through science narrative texts.

2.6. Results

2.6.1. If: Can Science Narrative Texts Improve Memory and/or Learning?

The presentation of results will be organized according to the memory vs. learning distinction provided by the C-I model, and the tasks usually used to assess them (e.g., Ferstl, 2001; Kintsch, 2012).According to the C-I model (Kintsch, 1994; 1988), memory and learning correspond to different levels of representation. Memory is related to the reproduction and paraphrasing of information, being strongly associated to the textbase level. Learning is related to changes in knowledge and requires integration of new content within prior knowledge, being more closely associated to the situation model level. As outlined, authors from N vs. C and N-O studies used different tasks to measure these outcomes. Table 2.3 lists all the examined N vs. C studies, along with information on the main variables (educational level and memory and/or learning outcomes) and, whenever possible, the corresponding ES. Table 2.4 presents the same information for the N-O studies. As can be seen in Tables 2.3 and 2.4, both types of studies included a range of educational levels, but whereas N-O studies mainly included preschoolers and only seldom undergraduates, N vs. C studies focused more on the latter and none included preschoolers.

Memory and learning are preferentially measured by specific tasks (e.g., Ferstl, 2001; Kintsch, 2012). Memory is often tapped by free recall (e.g., 71), recognition questions (e.g., multiple-choice; 10), comprehension questions about explicitly mentioned information (e.g., fill-in questions, 10; interviews, 21), questions that probe automatic inferences based on explicitly mentioned information (e.g., 14), and questions that evaluate retention, such as comparing immediate and delayed measures (e.g., 16), or repeating learning assessment sometime later (e.g., test, 15).

N vs. C studies reported mixed results, using mostly written tasks and some interviews (see Table 2.3). Among undergraduates, there was no overall difference between text types in immediate recall (7) but, depending on the recalled items (the total items of the text vs. the items common to both texts) and level of representation (textbase vs. situation model), either a narrative or expository advantage was found. Delayed recall among younger students (primary: 3; first middle school year: 2) did not benefit from one text type specifically. Expository gains in immediate recall were reported among primary and middle school students (e.g., 4, 9, respectively; the latter just for one topic).

Narrative gains, on the other hand, were seldomly observed in immediate recall (14, but only for one topic) but were clear cut in delayed assessments of students from several educational levels (middle school: 1,14, 15; high school: 16; university: 10; medium to large ES). It is interesting to note that when only one of the presented topic yielded a text type advantage, this topic had been deemed as less interesting (14) or more difficult (9) by students. Many ES were medium to large, with some exceptions (1, 2, 3, 7, 9, 11, 14, 16).

Table 2.3. Characterization of the Main Variables (Educational Level, Memory Outcomes, Learning Outcomes) of the N vs. C Studies

No.	Study	Educational Level	Main Memory and/or Learning Findings	Effect Sizes (When Available)
1.	Maria & Johnson (1989)	Gr 5 and 7	Gr 7 scored higher than Gr 5 in Im Misc Post and Im App Post but there we no differences between N and E; Gr 7 scored higher than Gr 5 in Del App Post and N scores were higher than E scores	Im Misc Post Gr 5: *d*= 0.47 *r*= .23; Gr 7: *d*= 0.24, *r*=.12; Im App Post Gr 5: *d*= 0.45 *r*= .22, Gr 7: *d*= 0.47 *r*= .23; Del App Post Gr 5: *d*= 0.76 *r*= .35; Gr 7: *d*= 0.24 *r*= .12
2.	Maria & Junge (1993)	Gr 5	No N vs. C differences in Im Rec and Del test; N Rec were longer; Rec had few Sci ideas	Im Rec (length): *d*= 1.07, *r*= .47; Im Rec (Sci ideas): *d*= 0.20, *r*= .10; Del test: *d*=0.19, *r*=.1
3.	Jetton (1994)	Gr 2	No N vs. C differences in Im free response or Del Rec; Rec included more on N ideas	Im story ideas: *d*= 0.57 *r*= .27; Im Sci ideas: *d*= 0.17 *r*= .08; Del Rec: *d*= 0.13 *r*= .06
4.	Alvermann et al. (1995)	Gr 9	E advantage in Rec & App Post	Rec: *d*= -0.88 *r*= -.40; App *d*= -0.90 *r*= -.41
5.	Hellstrand & Ott (1995)	Gr12	N LRN advantage in Post	*d*= 0.57, *r*= .27
6.	Lamartino (1995)	Gr 3	No N vs. C LRN differences in Post	*d*= 0.50, *r*= .24
7.	Wolfe & Mienko (2007)	UndGr	No significant N vs. C differences in LRN and Rec	LRN: *d*= -0.13 *r*= -.06; Rec: *d*= 0.27 *r*= .14
8.	McQuiggan et al. (2008)	Gr 8	Largest Pre-Post gains for E text, followed by min N and lastly full N	N (full) vs E: *d*= -0.99 *r*= -.44; N (full) vs N (min): *d*= -0.32 *r*= -16; N (min) vs E: *d*= -0.62 *r*= -.30
9.	Cervetti et al. (2009)	Gr 3 and 4	General E advantage in Post LRN and Retell, but only significant for one of the Sci topics	General LRN: *d*= -0.47 *r*= -0.23; General Retell: *d*= -0.46 *r*= -0.22
10.	Negrete & Lartigue (2010)	UndGr	E advantage in Im Post MEM/LRN; N advantage in Del MEM/LRN Post	
11.	Wolfe & Woodwyk (2010)	UndGr	N MEM advantage for total text elements, but E MEM advantage for common text elements; Textbase stronger for N and decreased	MEM (total): *d*= 0.57 *r*= .28; MEM (common): *d*= -0.77 *r*= -.36; Im textbase: *d*=0.44, *r*=.22; Del

No.	Study	Educational Level	Main Memory and/or Learning Findings	Effect Sizes (When Available)
			significantly in Del; situation model marginally stronger for E and no decrease in Del	textbase: d=0.19, r=.10; Im situation model: d=0.41, r=.20; Del situation model: d=0.22, r=.11
12.	Rosa (2010)*	Gr 11	N Pre/Post LRN and CCPT elaboration advantage as measured by tests, DISC and INTVW	
13.	Ritchie et al. (2011)	Gr 6	N Pre/Post LRN advantage as measured by story WRT and INTVW	PreWRT to Story Part A: d= 1.59, r=.69; PreWRT to Story Part B: d= 1.16
14.	Arya & Maul (2012)	Gr 7 and 8	Im N LRN advantage for one Sci topic and Del N LRN advantage for both Sci topics; Gr 8 students did not benefit from Radioactivity N	Im (Radioactivity) d= .17 r= .08; Im (Galilean telescope) d= .43 r= .20; Del (both topics) d= 0.95 r=.43
15.	Hadzigeorgiou et al. (2012)	Gr 9	N advantage in Im and Del LRN Post	Im Post: d= 1.31 r= .55; Del Post: d= 1.72 r= .65
16.	Hong & Lin-Siegler (2012)	Gr 10	No N (struggles) vs. N (achievements) vs. C differences in Im LRN or Rec. N (struggles) advantage (vs. other two texts) in Del LRN and Rec	Im Rec: d= -0.35 r= -.17; Del Rec: d= 0.67 r= .32; Im LRN: d= 0.07 r= .04 , Del LRN: d= 0.90 r= .41
17.	Reuer (2012)*	Gr 12	N LRN advantage in Post and INTVW	Chapter test: d= 0.37, r= .18; Exam: d= 1.21, r= .51
18.	Browning & Hohenstein (2015)	Gr 1, 2 and 3	N LRN advantage in all Gr, but especially in Gr 3	Total Gr: d= 1.06 r= .47; Gr 1: d= 0.83 r= .38; Gr 2 d= 0.73 r= .34; Gr 3 d= 1.72, r= .65
19.	Akarsu et al. (2015)	Gr 7	N Post LRN advantage	Pre: d=-0.15, r= -.08; Post: d= 1.42, r= .58
20.	Dinsmore et al., (2017)	UndGr	Highest increase in LRN complexity with N, followed by E and decrease with persuasive text	$p\eta^2$ = .14

Note. White cells represent better memory and/or learning outcomes for the narrative (N) text, light grey cells represent the absence of difference between N and control (C) texts or mixed results (e.g., N advantage only in delayed measures), and dark grey cells represent better memory and/or learning outcomes for the expository (E) text. App: application; CCPT:

concept/conceptual; DISC: discussion(s); Del: delayed; Gr: grade(s); INTVW: interview(s); Im: immediate; LRN: learning; MEM: memory; Misc: misconception; Pre: pretest; Post: post-test; Rec: recall; Ss: students; Sci: science/scientific; UndGr: undergraduates; WRT: writing/wrote; yr-o: year-old(s). The letter *d* stands for Cohen's effect size and *r* for Pearson's correlation coefficient.

All studies were published articles except for 2 MSc theses (*).

Selection of conditions/tasks: in study 1, we only present the comparisons between the N text and one of the applied E texts (considerate E text) because they had equivalent length and their scores only significantly differed in Del measures. In study 4, we only present part of the applied tasks, one representative of each relevant outcome, and the Control condition was not examined.

Aggregation of results: the following results were collapsed: in study 2, the results from good readers and bad readers; in study 4, the results from the Discussion web and Question/answer conditions; in study 7, the results from the two applied E texts (Topical E text and Sequential E text); in study 14, the results from both Sci topics in the Del assessment.

In study 5, effect sizes were calculated by approximation (approximately 25 students per class).

N-O studies reported encouraging retention outcomes (see Table 2.4). In interviews, primary and preschool students showed good recall of information after three to five months (21, 36). In the latter case (36), drawings made by the students supported recall and there were different levels of performance. In middle-to-high-school students, the retention interval could amount to one year (32). Among high schoolers, after a one-week delay, one study found that the narrative texts with concrete details promoted better retention than their abstract counterparts (22; large ES), and another found that students rewrote the narrative they had read using less factual information (33; large ES). Intriguingly, when comparing primary students' retelling of the narrative with their explicit description of the scientific model it contained (both after three months), more abstract science ideas were included in the former. When directly compared, older students recalled more content than younger ones in immediate measures (29, 33).

Learning can be assessed by problem-solving tasks that demand the transfer or the application of information (e.g., complex problem-solving, 16; implementing an experiment, 25; applying classroom-acquired knowledge in other contexts, 22), by inference questions that cannot be answered with explicitly mentioned information (e.g., 14), and by questions directed at determining knowledge change or improvement, such as pre- vs. post-tests (e.g., 7) and delayed reassessments (e.g., 1).

The pattern of results from N vs. C studies mirrors the previously presented one for memory but contains more findings (see Table 2.3). In some studies, there were no significant differences between text types in immediate measures using written questions or tests (2, 3, 6, 7; primary school, middle school, and university). Other studies reported an expository advantage in immediate measures of mostly the same kind (study 9 used interviews) and among the same educational levels (9, 4, 8, 11). However, more studies reported a narrative advantage. This learning advantage was scarcely observed on immediate measures (primary school: 18; middle school: 14; interview and written test, respectively).

Table 2.4. Characterization of the Main Variables (Educational Level, Memory Outcomes, Learning Outcomes) of the N-O Studies

No.	Study	Educational Level	Main Memory and/or Learning Findings	Effect Sizes (When Available)
21.	Banister & Ryan (2001)	Gr 4	Ss showed CCPT change from Pre/Post questions and reWRT of N; in Del INTVW Ss used more abstract ideas in retell of N than in description of Sci CCPT; some imperfections in CCPT development	
22.	Wilcken (2008)**	High school	Better Post Im and Del LRN for N with concrete details (vs. abstract N)	Im: Cohen's *dav*= 0.45; Del: Cohen's *dav*= 0.79
23.	Ritchie et al. (2008)	Gr 4	LRN was evaluated as functional and fluent use of Sci CCPT through N WRT, INTVW and field observation	
24.	Corni et al. (2010)	Gr 3	Ss drawings and WRT texts showed LRN (from descriptions to interpretation and formulation of hypotheses)	
25.	Frisch (2010)	UndGr (preservice teachers)	Average-scoring Ss (exam) used N they WRT to understand Sci CCPT more than below- or above-average-scoring Ss; teacher guidance helped Ss integrate and LRN SCI CCPT in N they WRT	
26.	Kokkotas et al. (2010)	Gr 6	Ss showed LRN in comprehension questions, classroom DISC and by implementing experiments	
27.	Tomas & Ritchie (2010)	Gr 9	Ss were able to transform Sci knowledge to WRT accurate Sci N but this knowledge was better explained and elaborated in INTVW (which can explain decreases)	Improvements: N PreWRT to Part B: *d*=1.25; N Part A to N Part B: *d*= 0.85; Decreases: N PreWRT to Part C: *d*=0.55; N Part A to N Part C: *d*= 0.89; N Part B to N Part C: *d*= 2.19
28.	Kalogiannakis & Violintzi (2012)	Preschool	LRN improved from Pre to Post as assessed by INTVW and drawings	

No.	Study	Educational Level	Main Memory and/or Learning Findings	Effect Sizes (When Available)
29.	Legare et al. (2013)	5 to 12 yr-ol	Specific N (desire-based, need-based, natural selection) promoted corresponding explanations of evolution in Im LRN and Rec (INTVW) Older Ss recalled more content, used more need-based and evolution-based explanations, and used more evolution CCPT than younger Ss; the later used more desire-based explanations	Older vs younger desire-based explanations d=-2.29; r=-.75; Older vs younger need-based explanations d=3.18, r=.85; Older vs younger evolution-based explanations d=6.08 r=.95; Older vs younger evolution CCPT d=6.55 r=.96
30.	C. Morais (2015)	Primary school (8-10 yr-o)	Ss drawings showed LRN of Sci CCPT	
31.	Lin-Siegler et al. (2016)	Gr 9 and 10	Ss who read intelectual struggle N or life struggle N had better Post LRN (tests) than Ss who read achievement N	Achievement N vs. intelectual struggle N: d=0.04, r=.02; achievement N vs. life struggle N: d=0.09, r=.05; intelectual struggle N vs. life struggle N: d=0.06, r=.03
32.	Mutonyi (2016)	Gr 9-11	Ss journals, focus-groups DISC and INTVW demonstrated CCPT LRN; in many cases it was retained after months/a year	
33.	Prins et al. (2017)	Gr 10 and 11	Both Gr scored high in Im and Del LRN and MEM Post (except Retell); Gr 11 had higher Im Retell; worse Del retell (less Sci information) in both Gr	Retelling between Gr: d= 2.02 r= .71; Retelling between sessions: Cohe's dz= 0.74
34.	Flynn & Hardman (2019)	Gr 12	Ss improved LRN (from first 15 Post questions to last 15 Post questions)	d= 0.71
35.	Morais et al. (2019)	Gr 8	Ss showed LRN of Sci CCPT in N WRT (explanation of Sci ideas) and creation of hands-on-activities (connecting and App of Sci ideas)	
36.	Walan & Enochsson (2019)	Preschool (4-6 yr-o) and primary school (7-8 yr-o)	Ss demonstrated different levels of SCI LRN and Rec in Del INTVW (from no identified LRN to LRN connected to reality); Im drawings supported Rec)	

Note. App: application; CCPT: concept/conceptual; DISC: discussion(s); E: expository; Del: delayed; Gr: grade(s); INTVW: interview(s); Im: immediate; LRN: learning; MEM: memory; N: narrative; Pre: pretest; Post: post-test; Rec: recall; Retell: Retelling; Ss: students; Sci: science/scientific; UndGr: undergraduates; WRT: writing/wrote; yr-o: year-old(s). The letter *d* stands for Cohen's effect size and *r* for Pearson's correlation coefficient. All studies were published articles except for 1 PhD dissertation (**).

Aggregation of results: in study 22, the results from the factors Structure and Gender were collapsed with the results of the factor Concreteness.

However, it was often reported in delayed assessments (1, 10, 14, 16; primary school, middle school, and high school) and in assessments in which the “moment” of learning is harder to categorize (5, 12, 13, 15, 17, 18, 19; primary school, high school, and middle school), as they were developed through the course of several weeks (e.g., school term). Two of these studies also assessed learning through discussions and interviews (e.g., 12, 13, 17) and story writing (13). When directly compared, older students tended to score higher (1, 18), and in one study older students did not benefit from the narrative text in one of the science topics (14). ES were mostly medium to large (exceptions: 1, 2, 3, 9, 14, 17, 20).

Positive outcomes relating to learning were also reported in N-O studies, at different educational levels and using a more varied set of tasks (see Table 2.4). Among preschool, primary school, and middle school students, conceptual appropriation and development was observed with written questions (21, 26, 31), interviews (21, 23, 27, 29, 36), drawings (24, 28, 30), written stories (23, 27, 35), created hands-on-activities (26, 35) and field observations (23). However, in many studies, knowledge demonstration was not without imperfections, as evaluated by incorrections or naive concepts in drawings and interviews (21, 28, 36) and even by a decrease in the demonstration of learning in the last part of the task of story writing (27). Moreover, in one study conducted with participants from 5 to 12 years-old, younger participants endorsed more naive and anthropomorphic explanations of evolution (desire-based), whereas older students endorsed more natural selection explanations and used more evolution concepts (although they also endorsed more need-based explanations as well). Many of the older students (including undergraduates, 25) demonstrated conceptual learning and refinement not only in written tests (25, 31, 33,3 4), but also through journaling (32), discussions and interviews (32), as well as story writing (25). ES were medium to large, with one exception (31).

Overall, although there is encouraging evidence for the use of narrative texts in science learning, it is difficult to build a clear pattern from these results. This can be partly explained by the variability in features, which, beyond educational level, can be found in these studies and impact memory and learning outcomes. We provided a few hints of such features (e.g., narrative text elements, students’ interest, configuration of the activity) that can help characterize this learning process. We will delve further into this question in the next section.

2.6.1. How: Characterizing the Process of Learning from Science Narrative Texts

Under Which Conditions.

Texts Features. Regardless of the narrative educational materials used, it is useful to consider how their features can specifically impact memory and learning, so that the conditions in which they are more effective for learning can be ascertained (e.g., Norris et al., 2005). The way narration is structured (i.e., temporally organized events) is proposed to aid memorization and learning (e.g., Prins et al., 2017; Strube, 1994). This textual feature resembles the way the human mind organizes experiences (Bruner, 1991; Fisher, 1987; Kintsch, 1998), and there is evidence that people build temporally organized representations of texts even when that structure is absent (e.g., Claus & Kelter, 2006). Two studies that used science narratives report results in line with this idea. One of them reports that, compared to the narrative text, the expository text more frequently caused chronological confusions that interfered with comprehension (Browning & Hohenstein, 2015). In another, the expository text presenting events in a temporal order promoted greater knowledge integration than another presenting events by topics and, interestingly, than the narrative text (Wolfe & Mienko, 2007). Yet, although the temporal expository and narrative texts had a similar structure, they differed greatly in another text feature, namely its content, a pragmatic feature.

Narration contents tend to focus on personal and social events (e.g., Arya & Maul, 2012; Corni et al., 2010; Klein, 2016), and science narratives have portrayed these contents differently. They can include fictional elements (e.g., Wolfe & Mienko, 2007), which are strongly associated to sociocultural conceptions and practices of narratives, or stick to factual information, which is more characteristic of science discourse. Examples of the fiction elements used in science narratives are anthropomorphism (e.g., Banister & Ryan, 2002; Cervetti et al., 2009), myths (e.g., Kalogiannakis & Violintzi, 2012), and fantasy and science fiction (e.g., Akarsu et al., 2015; Wolfe & Mienko, 2007). Some authors argue that, because fiction suspends disbelief, it creates unrestricted hypothetical worlds that are useful to illustrate complex (e.g., Negrete, 2005) or counterintuitive (Browning & Howenstein, 2015) science concepts. At the same time, there is the concern that fantastic and anthropomorphic elements may make it difficult to separate fact from fiction, promoting inaccuracies and misconceptions (e.g., Broemmel & Rearden, 2006; Gomez-Zweip & Straits, 2006), or animistic or teleological explanations of science (Klein, 2006).

Many studies that used fictional elements in science narratives report positive memory and/or learning outcomes (e.g., Akarsu et al., 2015; Corni et al., 2010, Kalogiannakis & Violintzi, 2012), and one suggests that anthropomorphic elements enhanced students' recall of ideas (Banister & Ryan, 2001). Yet, others report that fantastic and anthropomorphic elements interfered with recall, with students exhibiting more misconceptions (Cervetti et al., 2009), less scientifically accurate interpretations (Legare et al., 2013), and difficulties in separating facts from fiction (Prins et al., 2017) or in integrating science contents into stories (Frisch, 2010; Tomas et al., 2011). Students also recalled more story than science ideas from the texts (Jetton, 1994; Maria & Junge, 1993; Wolfe & Woodwyk, 2010), and some expository materials benefited knowledge integration further (Wolfe & Mienko, 2007; Wolfe & Woodwyk, 2010). Importantly, the interference of interesting yet irrelevant elements (e.g., Garner et al., 1989), such as fiction, seems to depend on how coherent and intertwined with the text's topic these elements are (Glaser et al., 2009; Lehman et al., 2007; Negrete, 2005). In at least one study, the authors acknowledged that this was not the case in their materials (Wolfe & Miekno, 2007).

Despite their close pragmatic association, narrative materials do not necessarily contain fictional contents. Examples of factual or feasible fictional information used in studies with science narratives are depictions of daily or contemporary events (Dinsmore et al., 2017; Reuer, 2012; Rosa, 2010) and historical/discovery accounts (e.g., Arya & Maul, 2012; Lin-Siegler et al., 2016). There are different proposals on how scientists should be portrayed in the latter. Some educators argue that a romanticized view of scientists, that brings out heroic and wonder-like qualities, can facilitate learning (Egan, 1997; Hadzigeorgiou et al., 2012). Other educators argue that scientists should be portrayed in a realistic and accessible way that highlights their struggles and challenges (e.g., Arya & Maul, 2012; Hong & Lin-Siegler, 2012) and avoids stereotypical images of innate ingenuity and monumentality (e.g., Allchin, 2003; Solomon, 2002). In one study that followed the former approach, students learned better with the narrative than with the expository text, and more than half associated romantic qualities to science knowledge in their journals (Hadzigeorgiou et al., 2012). In other studies, however, narratives focused on scientists' achievements and innate intelligence did not produce the same learning gains as narratives focused on scientists' struggles did (Hong & Lin-Siegler, 2012; Lin-Siegler et al., 2016). Narratives portraying scientists' discoveries also yielded better learning than their expository counterparts (Arya & Maul, 2012).

The use of historical-based accounts is claimed to serve another purpose, which is to connect scientific knowledge with the social and cultural context in which it was discovered or developed (Arya & Maul, 2021; Klassen, 2007). This provides students with a better grasp of what science is and how it works (i.e., the nature of science), averting misconceptions that can be harmful to learning (e.g., Allchin, 2003; Clough, 2011). Some authors further contend that the scientific process is fairly narrative in itself (e.g., Bruner, 1996; Hadzigeorgiou, 2016; Larison, 2018).

Science narratives have been reported to encourage students to challenge their perceptions (e.g., Arya & Maul 2021; Dinsmore et al., 2017; Erten et al., 2013) and hold a more accurate image of science, such as viewing it as a process (Evangelista & Zimmermann, 2008; Leipzig, 2018). One study reported that the narrative text made students use more evidence in their responses and display a more complex learning of the science contents (Dinsmore et al., 2017).

Activity Features. Besides the text's content, it is important to consider situational aspects or circumstances of learning, pertaining to the activity and its interactions with the wider context (Snow, 2002). One feature important to consider is the goal(s) of the activity. Educators and researchers may communicate goals to students, but students generate their own goals, which can be influenced by existing schemata. People often draw on schemata when interpreting texts (e.g., Adam, 2011; van Dijk & Kintsch, 1983), which include genre-specific processing strategies, activated according to the knowledge of what is usually expected from texts with specific features (e.g., Hidi et al., 1982; Rastier, 2001).

Instructional/study goals can favour the activation of expository-processing strategies, as learning is associated with this kind of materials (e.g., Kloser, 2013; Wang, 2009). A text with features associated to the narrative textual genre (e.g., temporal organization; fictional information) may activate an entertaining aim, stemming from socioculturally-based expectations. The overtly communicated goal (e.g., to learn) may thus conflict with activated pragmatic knowledge (e.g., to entertain), and interfere with comprehension (Snow, 2002). Indeed, students exposed to science narratives have expressed that narratives were not adequate to learn science (Prins et al., 2017), or were surprised to have learned from them (Murmann & Avraamidou, 2014). Additionally, students may activate a story or an informational mode depending on the activity they are doing (e.g., short stories and drawings vs. instructed group work, respectively, Murmann & Avraamidou, 2014).

Studies using science narratives have communicated different goals to students. For instance, some goals were related students' own evaluation (understanding how students make sense of difficult information; Alvermann et al., 1995; solving a mystery; McQuiggan et al., 2008; studying; Wolfe & Mienko, 2007; Wolfe & Woodwyk, 2010), and students tended to benefit more from expository texts in these cases. In one study where the goal was not related to students' own evaluation, but instead to the evaluation of the texts, they benefited more from the narrative texts (Arya & Maul, 2012). Some goals were related with entertainment, such as solving a mystery (McQuiggan et al., 2008), where students benefited more from the expository text, or co-writing an ecological mystery (Ritchie et al., 2008, a N-O study), in which case students demonstrated written and spoken fluency of the scientific concepts. One study specifically manipulated the activity's goal (hearing the same text as a "story" or as a "book") but, regardless of the instruction, all students focused more on story elements (Jetton, 1994). As 2nd graders, these students must have been familiar to narrative texts via social practices, yet have little contact with expository materials, leading them to activate story-processing strategies under both conditions.

Another relevant feature of the activity is its duration. As commented in the analysis of the “If” question, studies ran over the course of weeks tended to yield more positive results for narrative than for expository texts. Authors from one study reasoned that the narrative storyline may have overloaded students’ cognition, which was assessed on the same day (McQuiggan et al., 2008). Duration can also provide the opportunity to integrate more and varied activities. Examples of activities included in longer studies with science narratives, but not shorter ones, are fieldtrips (e.g., Ritchie et al., 2008), journaling (e.g., Hadzigeorgiou et al., 2012; Mutonyi, 2016), and preparing hands-on activities and experiments (e.g., Kokkotas et al., 2010; Morais et al., 2019). Reasons for the learning gains afforded by longer and more varied activities may include knowledge consolidation (e.g., Squire et al., 2015) and increased meaning making (e.g., Bruner, 1990). Finally, the duration and variety of learning activities are likely to vary depending on economic and cultural factors of the learning setting (Snow, 2002), such as socioeconomic status (SES) and cultural features of the neighbourhood.

Reader Features. The reader is at the centre of the learning process. In addition of differing by their educational level, readers may have varied social and cultural identities. At the same time, developing text materials and learning from them takes place in a sociocultural context that reflects the interpretations of specific cultural groups (e.g., Adam, 1997; Snow, 2002), particularly the dominant ones (e.g., Arya & Maul, 2021; Phillips Galloway et al., 2020). The underrepresentation of historically/currently marginalized social and cultural groups in mainstream science poses challenges to learning (e.g., Jackson et al., 2016; Harper & Kayumova, 2022; Visintainer, 2020). The need of integrating the thinking and learning dynamics of these marginalized groups has been stressed by some authors (e.g., Harper & Kayumova, 2022; Lee & Grapin, 2022; Mutonyi, 2016), who claimed that cultural background impacts text interpretation and knowledge construction (e.g., Greenfield, 1997, cited by Arya & Maul, 2021; van Dijk, 2001, cited by Arya & Maul, 2012), and that learners feel the need to see people like them doing science (e.g., Arya & Maul, 2021; Bowman et al., 2022; Gilbert et al., 2005).

Being culturally relevant mental models about the world (Bruner, 1986; Kintsch, 1988), narratives are proposed to help readers connect to science by bringing them closer to familiar and relevant contexts (e.g., Avraamidou & Osborne, 2009; Graesser et al., 2002). This issue has been tackled in some studies with science narratives. In one study, preschool Greek children learned about volcanoes through a Greek myth (Kalogiannakis & Violintzi, 2012) and in another, Ugandan students made use of cultural tools like proverbs and stories to learn more about HIV, a very socially relevant issue in their country (Mutonyi, 2016). Both studies report engagement and learning gains. Other studies found through interviews that narratives detailing episodes of discovery were effective in reaching different genders and cultural backgrounds. Because they felt they could also be scientists, students found science more relatable and interesting, which may have boosted their comprehension of the contents (Arya and Maul, 2021; Lin-Siegler et al., 2016). This sense of relatedness may be dependent on the level of match between the text's social and cultural elements and the student's own background (Lin-Siegler et al., 2016).

Additionally, SES has well established effects on brain and cognition (for a review, see e.g., Farah, 2017), including in science learning (e.g., Lee & Luykx, 2007; Yang, 2003). There is some evidence that science narratives work well (e.g., Lin-Siegler et al., 2016; Mutonyi, 2016) or better than expository texts (Arya & Maul, 2012; Hong & Lin-Siegler, 2012) among middle and high-school participants from low and middle-income backgrounds. Importantly, SES is often confounded with ethnicity (e.g., Cheng et al., 2015), and in some studies the students who benefited more from narratives were both from low SES and predominantly Latinx and Black (Arya and Maul, 2012; Lin-Siegler et al., 2016). Yet, another study with mostly Black students found worse results in the narrative condition (Alvermann et al., 1995).

Through Which Underlying Processes. Ascertaining the extent to which reader's features are engaged during the process of learning from science narrative texts can help understand how these texts generate memory and learning outcomes (e.g., Norris et al., 2005). As such, in this section these features will be framed as underlying processes.

Integration with Prior Knowledge. The notion than learning from text requires linking and integrating new information with prior knowledge is central to the C-I model (e.g., Kintsch, 1998). These processes have been vastly investigated using science expository texts, which are more dependent of the integration processes than non-science narrative texts (e.g., Best et al., 2008, McNamara et al., 2011).

Narrative texts are proposed to provide meaningful organizing structures (e.g., Negrete & Lartigue, 2010; Strube, 1994) that help activate prior knowledge (e.g., Leipzig, 2018; Maria & Johnson, 1989) and integrate information (Negrete, 2005; Prins et al., 2017). For these reasons, they can be particularly useful as scaffolding tools for beginner or struggling learners (e.g., Gilbert et al., 2005; Klassen, 2007; Mutonyi, 2016).

Narratives have been successfully used to derive science teaching methodologies for preschool (e.g., Kalogiannakis & Violintzi, 2012; C. Morais, 2015) and primary school (e.g., Corni et al., 2014) students. Additionally, it has been shown that high-school students with very little prior knowledge were able to develop adequate scientific understanding through science narratives (Prins et al., 2017), or were the only ones demonstrating learning gains (Flynn & Hardman, 2019). Yet, significant correlations between prior knowledge and learning through science narratives were not always found (Wilcken, 2008). When compared to expository texts, results are mixed. Even though text type did not have an overall impact in the learning and memory of undergraduates, it interacted with their prior knowledge: students with minimal prior knowledge learned better with the narrative, and students with higher knowledge learned better with the expository text (Wolfe and Mienko, 2007). Prior knowledge did not correlate with narrative text recall, but it correlated with expository text recall (Wolfe & Mienko, 2007; Wolfe & Woodwyk, 2010). Middle-school students who had previous contact with the topic also benefited less from the narrative texts (Arya & Maul, 2012; but prior knowledge was not directly measured). However, younger middle and primary school students all benefited more from the narrative text, regardless of age and prior knowledge (Browning & Hohenstein, 2015; Maria & Johnson, 1989).

There is also some evidence that narratives may be used as scaffolding tools for struggling learners. In Reuer (2012), narratives were on average more effective for learning than textbooks, but especially so for average and low achievers, a pattern matched by the students' own perceptions. Narrative texts also improved the learning of below-average and average achieving students (Hadzigeorgiou et al., 2012; Hong & Lin-Siegler, 2012; Lin-Siegler et al., 2016). Yet, in one study it did not particularly benefit struggling readers (Maria & Junge, 1993).

Affective Dispositions. Even though affective processes have received less attention than cognitive ones in science learning research, their impact has been documented (e.g., engagement, Fredricks et al., 2018; self-efficacy, Britner & Pajares, 2006; emotions, Sinatra et al., 2014).

Narratives have been consistently suggested to have a positive effect on students' affective dispositions (e.g., Avraamidou & Osborne, 2009; Bruner, 1986; Norris et al., 2005). Studies using science narratives report a range of results in support of this idea. Students seem to react with interest to the reading (e.g., Arya & Maul 2021; Hadzigeorgiou et al., 2012) and writing (e.g., Evangelista & Zimmermann, 2008; Tomas et al., 2011) of science narratives. Students' engagement was also manifest by the expressiveness of their drawings about the narrative-based intervention (C. Morais, 2015). Moreover, students described the experience of reading (e.g., Prins et al., 2017; Reuer, 2012) or writing (Ritchie, 2008; Ritchie et al., 2011; Tomas et al., 2011) science narratives as enjoyable and engaging, writing more (Hadzigeorgiou et al., 2012) and more positive (Akarsu et al., 2015) journal entries than students in the expository text condition. High levels of immersion during narrative-based science learning activities were also reported (McQuiggan et al., 2008; Murmann & Avraamidou, 2014). Finally, students with low self-reported levels of interest in science increased this interest by reading narratives about scientists' struggles (Hong & Lin-Siegler, 2012), and one study found a significant positive correlation between students' interest in the narrative texts and their performance on an science exam (Reuer, 2012).

Interventions with science narratives have also reported behaviours suggestive of active learning, an intrinsically motivated type of learning (Deci & Ryan, 1982) expressed through autonomy, initiative, and responsibility for one's learning (Kane, 2004). Students were curious (e.g., Akarsu et al., 2015; Morais et al., 2019), participated actively (e.g., Evangelista, 2008; Kokkotas et al., 2010), engaged in the preparation and execution of tasks (e.g., Klassen, 2007; Kokkotas et al., 2010; Vrasidas et al., 2015) and spontaneously wrote stories (Akarsu et al., 2015) and planning notes (Klassen, 2017). Students also showed interest in learning more about the science topic (Evangelista, 2008; Rosa, 2010) and proactively made additional research on it (Evangelista, 2008; Hadzigeorgiou et al., 2012).

Some studies found evidence that learning with science narratives impacted the willingness and belief in the capacity to achieve by positively affecting the ratings of self-efficacy (McQuiggan et al., 2008; Tomas et al., 2011) and self-confidence (Flynn & Hardmann, 2019).

Finally, many authors contend that the science learning gains prompted by narratives are in part due to its ability to involve readers emotionally, particularly with the thoughts, feelings, and actions of characters (e.g., Banister & Ryan, 2001; Hadzigeorgiou et al., 2012; Lin-Siegler et al., 2016; Murmann & Avraamidou, 2014). Emotions are a fundamental part of the narrative experience (e.g., Bruner, 1986; Egan, 2005; Oatley, 2016) and it is known that understanding non-science narratives recruits a wide brain network (e.g., Mar, 2004; Mason & Just, 2009; Xu et al., 2005) that includes areas related to emotional processing, perspective-taking, and theory of mind.

Studies with science narratives offer some concordant evidence. A range of emotional responses has been observed: enthusiasm and excitement (e.g., Hadzigeorgiou et al., 2012; Vrasidas et al., 2015); laughter (Banister & Ryan, 2001; Klassen, 2007); comments of how enjoyable and fun the intervention was (Murmann & Avraamidou, 2014; Tomas et al., 2011); and other emotionally-charged appraisals (Mutonyi, 2016). Some studies offer evidence of students' emotional involvement with the text's characters, whether they were scientists (e.g., connecting with the scientist's life and work, Hadzigeorgiou et al., 2012; Lin-Siegler et al., 2016) or fictional characters (adopting the character's point of view, Jetton, 1994; Murmann & Avraamidou, 2014).

Cognitive Abilities. Studies with science narrative texts also mention cognitive abilities that have an established role in the comprehension of narratives (e.g., attention: van den Broek et al., 1999; mental imagery; Sadoski et al., 1990) and expository science texts (e.g., attention: van den Broek, 2010; working memory: Linderholm et al., 2002).

Narratives are thought to capture attention because they centre on human action (e.g., Bruner, 1986; Banister & Ryan, 2001; Corni et al., 2010). Studies that used science narratives reported that students were more focused and attentive than in regular classes (Hadzigeorgiou et al., 2012; Morais et al., 2019; Murmann & Avraamidou, 2014).

One study comparing science narrative and expository texts directly examined working memory (Wolfe & Mienko, 2007), namely, the processes involved in the maintenance and manipulation of information during cognitive tasks (Baddeley & Hitch, 1974). Contrary to previous studies cited by Wolfe and Mienko (2007), on both expository texts (e.g., Linderholm & van den Broek, 2002) and narrative texts (e.g., Hambrick & Engle, 2002), working memory did not predict learning or recall from either text.

Some studies with science narrative texts offer evidence on the engagement of mental imagery/visualization abilities. Students claimed to have been able to visualize invisible concepts that were difficult to understand (Akarsu et al., 2015) and that traditional school texts did not promote visualization the same way narrative texts did, making them less memorable (Prins et al., 2017). Students also described elaborate mental images of the stories, seemingly using them as a basis for understanding the learning activity (Murmann & Avraamidou, 2014; Vrasidas et al., 2015). One study (Wilcken, 2008) found that narrative texts with concrete details were more easily understood and remembered than more abstract narratives, possibly due to an enhancement of mental imagery (e.g., Driscoll, 2000, cited by Wilcken, 2008).

Imagination is proposed to encourage processes that can aid the learning of complex and counterintuitive science contents, such as divergent thinking (Browning & Hohenstein, 2015), suspension of disbelief (Alvermann et al., 1995; Browning & Hohenstein, 2015), and the envisioning of different realities (Bruner, 1986; Gilbert et al., 2005), including scientists' own reality (e.g., Arya & Maul, 2012). In some studies, using science narrative texts, teachers (e.g., Klassen, 2007; Vrasidas et al., 2015) and students (Tomas et al., 2011) claimed that these materials allowed students to exercise imagination to a greater extent than traditional activities.

The use of imagination is thought to be closely connected to the interpretative nature of narrative texts, as readers must draw on their imagination to fill parts that are ambiguous or left unanswered (e.g., Bruner, 1986; Klassen, 2007; Negrete & Lartigue, 2004). This interpretative effort is proposed to trigger high-level abilities such as abstraction, thinking and reflection (e.g., Bruner, 1991; Klassen, 2007; Rosa, 2010). In some studies, specific kinds of narrative (personal, Skydsgaard et al., 2016; historical, Evangelista, 2008; science-fiction, Vrasidas et al., 2015) helped students reflect and develop critical thinking skills.

Finally, specific features from these texts should engage abilities not directly discussed in studies using science narratives, such as inference generation and executive functioning. Namely, filling ambiguous or unanswered parts should involve inference generation, processing multiple perspectives (Bruner, 1996) can engage cognitive flexibility, and event sequentiality (e.g., Negrete & Lartigue, 2004; Norris et al., 2005; Reuer, 2012) might trigger planning abilities. These abilities are deemed crucial to learn from text by the C-I model (Kintsch, 1998). They have been often compared in non-scientific narrative and expository texts, but have been reported to be more important for the comprehension of the latter (e.g., Eason et al., 2012; Wu, 2020).

2.7. Discussion

The current theoretical review had a twofold aim: to examine if science narrative texts have consistently benefited learning and/or memory outcomes (the "If" question) at different educational levels, and to provide an overview of aspects that characterize the learning process leading to such outcomes (the "How" question). These aims were grounded on three theoretical frameworks based on concepts from pedagogy, text linguistics, and cognitive psychology.

The "If" question revealed encouraging results for the use of science narrative texts. Students from different educational levels benefited from narrative texts in memory and learning outcomes. However, this advantage was particularly marked in delayed assessments and in longer-lasting interventions. Despite the strong pragmatic association between narrative texts and younger students, these results suggest that narrative texts can be appropriate science education tools to students of diverse educational levels. However, there is a need for more studies with higher level students, such as undergraduates.

Despite this overall pattern, it is not always clear which representation level benefited most from narrative texts, partly because they were not always differentiated in the studies. Importantly, narrative texts did not always provide an advantage to expository texts. This lack of consistency contradicts the idea of a single narrative effect. We further argue that such a narrative effect fails to consider the multifaceted nature of the learning process, which is further highlighted by the unavoidable variability of conditions between different studies.

In the "How" question, we provided an overview of conditions and processes that are part of the learning process and may impact its outcomes, attempting to connect them with evidence on learning outcomes whenever possible. As regards conditions, we discussed how different text features can differ in science narrative texts, how aims and duration can differ in learning activities, and how readers' social and cultural identities can vary. As for processes, we discussed how integration with prior knowledge, affective dispositions and cognitive abilities can be engaged by science narrative texts. This mapping stems from the three theoretical frameworks.

Based on our analysis, we propose that education and research should focus on specific narrative *effects*, that specify with what (texts), with who (learners), when and where (activities and wider context) these effects occur.

We should, however, keep in mind that these conclusions derive from a theoretical review, in which we mainly attempted to understand the narrative effect in relation to science learning. In the "If" question, we made a restricted analysis, focusing on the (in)consistency of previous results, and not a comprehensive analysis of the existing literature. This choice reflects a set of difficulties. First, as already commented on in the Method, developing a unified search strategy proved difficult. In addition, we wanted to include various types of data, stemming from quantitative, qualitative, and mixed methods studies, as well as from masters' theses and PhD dissertations. Satisfactorily integrating quantitative and qualitative data in systematic reviews can prove challenging, and including data from different research levels likely introduces varying levels of rigour. Combined, these issues may arguably increase a false sense of precision. Despite these challenging aspects, future studies may use our theoretical proposal to perform mixed methods syntheses (e.g., Heyvaert et al., 2013), providing systematic reviews that combine qualitative and quantitative evidence and research elements, and hence add systematicity and comprehensiveness to our approach. As regards the "how" question, our goal was even wider in breadth, as we were interested in providing an overall and exploratory overview of the conditions and processes underlying learning science through narrative texts. Again, future studies may benefit from our qualitative suggestions to provide more systematic reviews.

In addition, although we considered theoretical concepts from three different domains, we did not provide a fully unified and integrated theoretical frame, which, although was not the purpose of the current work, could undoubtedly be useful. Our interpretations and the reach of our findings are necessarily limited by these choices.

Nevertheless, in addition to supporting the interpretation of previous results, the theoretically-grounded mapping we provided can also contribute to design future education practices and interventions, as well as research. Embracing these different features of the process of learning science through narrative texts can also help bring together multidisciplinary educators and researchers interested in these educational tools. We provide some recommendations below.

2.7.1. Recommendations for Research

According to RAND, it is important to distinguish what readers take from the activity (i.e., which outcomes) from what they bring to the activity (i.e., which underlying processes). Whereas there are several evidence on the former, which is particularly interesting for educators, the latter is less understood, and researchers can help shed light on it. Examining how features from texts, learners, and activities impact learning from science narrative texts can thus provide valuable insight as to in which conditions and *why* these materials work.

It is clear from the hints we gathered and linked, that ideas and proposals about why narrative texts can improve science learning have not been tightly connected to evidence. For example, is increased mental imagery what makes concrete details improve memory and learning? Is visualization more important when learning from some science topics? And how do these elaborate images specifically promote content learning? Another example is the claim that science narrative texts allow students to exercise imagination to a greater extent than traditional activities. How does the use of imagination promote science learning? Is it by activating divergent thinking, as it has been suggested? At the same time, many abilities that have been extensively investigated with non-science narratives have not been addressed, or even mentioned, in studies using science narrative (e.g., perspective-taking; inference generation; executive functioning). More directed research can help fill in the gaps.

It is also noteworthy that many justifications for the use of science narrative texts as learning tools stem from theoretical or empirical works using more conventional, non-science, narratives. The latter body of literature can certainly provide interesting sources for future research. Yet, science and non-science narrative texts differ in important ways, namely in terms of their connections to wider sociocultural practices such as learners' conceptions of textual genres. Because of this, they can engage abilities very differently. In a related vein, researchers should also consider the difference between sequences and textual genres. It is perfectly fine to develop and investigate highly prototypical narrative materials, but any conclusions drawn should pertain to the level of (narrative) sequences, and not to be generalized to a textual genre as a whole. Instead, more experimental research can benefit from focusing on specific and well-defined text features.

As science narrative texts contradict socially-based expectations, they may activate different processing strategies. It would be interesting to check whether resulting conflicts can be minimized, or even activated in a complementary manner. Another activity-related aspect worth examining is why narrative learning gains are less evident in shorter interventions.

In addition, the effectiveness of any feature will likely vary according to variability in other features, so these interactions are important to keep in mind. It would be important to ascertain what narrative features are more likely to engage specific processes, among which readers and using which activities. For instance, examining which features promote deep thinking and reflection would inform how to build or select science narrative texts that are appropriate for more complex learning and knowledgeable learners. It would also be very relevant to examine which kind of processes are engaged by science narrative texts that bring out social and cultural elements of science, and how they can improve content learning. This would both shed light on how the learning of students from diverse backgrounds benefits from science narratives, and on how understanding the nature of science improves learning.

Finally, even though the C-I model has privileged the investigation of expository texts, the interest in using narrative texts as science education tools should spark more research on these texts based on the C-I model.

2.7.1. Recommendations for Educational Practices

Educators have frequently pointed out a set of challenges to science learning that can easily be connected to text, reader, activity and sociocultural aspects of learning. The language of texts is seen as dense and technical (e.g., Plavén-Sigray et al., 2017; Snow, 2010); readers find many science ideas unfamiliar or even counter-intuitive (e.g., Browning & Hohenstein, 2015; Gilbert et al., 2005); the framing discourse of learning activities is authoritative or even dogmatic (Kloser, 2013; Negrete & Lartigue, 2004); and education is generally decontextualized from human and cultural aspects (e.g., Harper & Kayumova, 2022; Sánchez Tapia et al., 2018; Solomon, 2002). The features tackled in the "How" question can therefore help address these challenges.

When selecting or creating narrative texts for science education it is important to consider the ways in which they connect readers and activities to the wider context. For instance, fiction can be used to address complex or counterintuitive ideas and may engage students as well as prompt positive emotions; however, these elements can also distract students or induce misconceptions. Educators should ensure that fictional elements are well weaved with science contents, in a coherent and contextualized way.

Historical and discovery narratives can help educators contextualize science within a more human context. These insights can help reduce damaging misconceptions about what science is and who gets to do it, building more inclusive science education practices. For example, realistic and accessible depictions of scientists can help readers from varied, often marginalized, social and cultural identities connect to science. Narrative materials can also be a means for educators to connect science learning activities with students' immediate (e.g., neighbourhood) and wider (e.g., current world affairs) contexts.

Narrative texts may, however, activate conflicting goals in learners. To tackle this, educators can: make explicit the connections between a text's structure, content, and function; explain that these features are flexible; and develop study strategies using narrative materials from early-on. This, in turn, will enable students to consciously adjust their expectations and adopt a more flexible approach to comprehension (Snow, 2002). Entertainment and instructing goals should be easier to reconcile in longer interventions, which provide more opportunities to alternate between the two and to integrate information. Educators may also reinforce the connections between formal and non-formal settings, taking advantage of the lower restrictiveness of the latter to use narrative-based activities (e.g., Littrel et al., 2022; Murmann & Avraamidou, 2014).

As the latter are not always viable, when planning activities, educators should at least consider the impact that certain narrative materials might have on student's cognition. This impact can be manifold and interact with education materials and activities in important ways. To name a few examples, educators can use narrative texts as scaffolding tools with less knowledgeable and/or struggling learners, but also to promote critical thinking, reflection, and autonomy. Narrative materials can also be used to capture learners' attention, interest and emotions, and to trigger mental imagery and imagination. Yet, the extent to which and the conditions in which (e.g., what text features) such processes are engaged has still to be further determined by research, which may then further inform the design and adaptation of educational practices.

Finally, when building assessment measures, educators can benefit from distinguishing between different levels of comprehension, and the C-I model can be a useful referent.

2.8. Conclusion

One way to address science learning challenges is by tightening science contents and literacy processes. Tackling such challenges and building more tailored practices can benefit from recognizing and embracing the multifaceted nature of learning. Narrative texts are a flexible educational tool that can help achieve such goals, as they connect the learners, the texts, the activities and the wider context in several and important ways. They have also been shown to improve the memory and learning outcomes of students at various educational levels, albeit not consistently. Together, these results suggest that learning from science narrative texts should be approached as a multitude of specific narrative effects that capture the complex interactions between the different elements of the learning process, instead of a single, overarching, narrative effect. Under this view, education and research should focus on what (texts), with whom (learners), when and where (activities and broader context) narrative materials can be used as effective science learning tools. A multidisciplinary theoretical framework combining complementary fields can thus be pivotal when developing practices and research based on this educational tool.

CHAPTER 3

Developing science learning materials

This chapter was based on the following manuscript: Soares, S., Simão, C., Gonçalves, M., Barata, R., Jerónimo, R., & Kolinsky, R. (2022). *How Did They Do It? The Impact of Narrating Science on Learning.* [Manuscript submitted for publication to Science Education], CIS-Iscte, Iscte–Instituto Universitário de Lisboa.

In this chapter, we included part of the main manuscript, namely the section pertaining to "Building and pretesting materials", and most of its supporting material.

Part of the results described in this Chapter were presented at the following conference: Soares, S., Jerónimo, R., & Kolinsky, R. (2019, May 16-17). *How did they do it? Incorporating discovery elements in science texts* [Oral presentation]. XV PhD Meeting in Psychology, Lisboa, Portugal.

3.1. Abstract

Narrative texts have been advocated as tools to tackle science learning challenges, and there is even the proposal of a "narrative effect" on learning, compared with the more traditional expository texts. However, there is much we still do not known about this tool, such as the conditions in which science narrative texts can prove beneficial to science learning, and the processes by which this learning occurs. To examine these questions, we built and pretested a set of narrative and expository texts, as well as corresponding learning measures, from different science topics, throughout two years. First, we ran an exploratory pretest, with the goal of testing a first attempt at material and paradigm building and collecting qualitative feedback. Performance was very low across text types, but participants provided valuable qualitative insights regarding the texts. Then, we tested new materials from chemistry and biology. The biology topic was dropped because participants were mostly drawing on related prior knowledge to answer the learning measures. A series of further analyses were performed for the chemistry responses. Finally, we conducted a think aloud protocol with the updated Chemistry materials and newly built Math materials. Learning scores were similar between these two topics and within and between text types, even if math texts were perceived as more complex and difficult than chemistry texts, but overall more familiar. The final materials were used in

the main experiment of the manuscript on which this chapter is based, which is presented in Chapter 4. As already noted by other authors, the process of building science learning materials proved challenging.

Keywords: science learning, science education materials, narrative texts, developing materials; pretests; think aloud

3.2. Introduction

We aimed at providing information on the process of development of the used learning materials (e.g., Flynn & Hardman, 2019; Klassen, 2009), as we believe it to be valuable for educators and researchers in education. The final materials were developed based on these inputs.

All materials were built in collaboration with several experts. On the one hand, it was important to guarantee that texts fell within the targeted text types, namely narrative texts (NT) and expository texts (ET), which was ascertained by one of the authors, who is a text linguist (e.g., Gonçalves, 2019). On the other hand, we needed to ensure that the texts' contents were scientifically and historically accurate. This was evaluated by science experts, which included one of the authors (e.g., History of Science; Math; Chemistry; e.g., Martins-Loução, 2020), through scales that measured a set of parameters (clarity/legibility, accuracy, completeness, complexity, density of information, coherence). We also relied on these experts' experience in science education to evaluate the learning measures (clarity/legibility, complexity, correspondence with the text's contents, correspondence with the targeted level of understanding). More qualitative and thorough feedback on the materials was provided in meetings.

We ran three pretests, all of which approved by the ethics committee of the university where they took place (agreement 25/2017). Participants did not take part in more than one pretest nor in the main study. We started by running an exploratory pretest, with the goal of testing a first attempt at material and paradigm building and collecting qualitative feedback. We then built and tested texts from two fields (Chemistry and Biology) and corresponding learning items. Finally, we conducted a think aloud protocol with the updated Chemistry materials and newly built Math materials.

3.3. Pretest 1. First Exploratory Pretest

3.3.1 Method

Participants. We examined 15 university students (93% women; M_{age} = 21.5 years, *SD =* 3.4).

Materials.

Texts. The texts' topic was the medicinal Cinchona sp. plant and mingled contents from different fields (Botany: e.g., taxonomic classification; Chemistry: isolation of chemical compounds; Physiology: symptomatology). The NT version focused on the life and actions of Bernardino António Gomes, a relatively obscure physician and scientist who isolated for the first time one of the plant's chemical compounds, the cinchonine. It provided details about his life not directly connected to the scientific contents, historical details about previous European contacts with the plant, and the scientist's thoughts and actions, which included some instances of fictionalized discourse. This passage illustrates the NT version: "As soon as Bernardino arrived in Brazil, and for the next four years, he walked through the humid forests everyday, under strong precipitations and hot temperatures. He would carefully collect several species of flora with medicinal properties (or so he hoped), which found there the ideal conditions to proliferate."

The ET version contained the same scientific contents but provided only a brief historical context and did not centre on the scientists' actions. This passage illustrates the ET version: "The Cinchona can be found in humid forests, which are characterized by strong precipitations and hot temperatures (e.g., such as the ones found in Brazil). This combination of climatic features creates the ideal conditions for the proliferation of several species of flora with medicinal properties."

Text Evaluation. Participants were required to evaluate the texts on several parameters on a scale from 1 to 7, where 1 corresponded to a low evaluation of that parameter and 7 to a high evaluation (see Table 3.1).

Learning Measures. A total of 13 questions tapped into different levels of comprehension: recall, connective inferences, predictive inferences, and application in novel situations. Answers were scored from 1 to 4 points, depending on the number of correct elements or ideas presented. Ideas with imprecisions were given 0.5 points.

Ancillary Tests. We used a set of ancillary tests that would be used in the main experiment, namely the *Teste de Idade de Leitura* (Reading Age Test), the *Author Recognition Test* and the *Scientist Recognition Test*, the reading experience and habits questionnaire, the perceived prior science knowledge questionnaire, and the sociodemographic questionnaire. We also collected some mentalizing information, by asking participants which traits they ascribed to the scientist, what could have been some of the scientist’s thoughts, and what could have been some of the scientist’s feelings. The latter task was eventually dropped.

Qualitative Feedback. We collected participants’ feedback on the materials and the paradigm (see examples of questions in Table 3.2).

Table 3.1. *Parameters of Text Evaluation Scales*

Parameter	Question
Clarity	On a scale of 1 (not clear at all) to 7 (very clear) please rate how clearly the contents were presented
Novelty of information	On a scale of 1 (not novel at all) to 7 (totally novel) please rate how novel the contents were for you
Comprehension difficulty	On a scale of 1 (not difficult at all) to 7 (very difficult) please rate how difficult it was for you to understand the text
Interest	On a scale of 1 (not interesting at all) to 7 (very interesting) please rate how interesting you found the text
Richness of vocabulary	On a scale of 1 (very poor) to 7 (very rich) please rate how rich was the text’s vocabulary
Diversity of vocabulary	On a scale of 1 (repetitive) to 7 (diversified) please rate how diverse was the text’s vocabulary
Syntax (sentence construction)	On a scale of 1 (very bad) to 7 (very good) please rate the quality of the text’s sentences
Cohesion (sentences)	On a scale of 1 (totally disperse) to 7 (totally related) please rate how well connected the sentences were among themselves
Local Coherence (ideas)	On a scale of 1 (totally disperse) to 7 (totally related) please rate how well connected the ideas were among themselves
Global Coherence (text as a whole)	On a scale of 1 (totally fragmented) to 7 (totally integrated) please rate the extent to which you felt that the text was a unified whole

General Procedure. Participants started by completing all ancillary tests (except for sociodemographic measures), after which they read the (randomly assigned) text with the purpose of evaluating its quality (e.g., Arya & Maul, 2012). They then completed the evaluation scales, the learning measures, and the mentalizing information. The session ended with the sociodemographic questionnaire and individual qualitative feedback. The session lasted about one hour.

Table 3.2. *Examples of Questions Used to Gather Qualitative Feedback in Pretest 1*

Type of question	Examples
General comprehensibility	– Did you feel uncomfortable reading about science topics you may not know much about? – What did you think about the text's length?
Text type differences	– Was the [expository] text similar to what you might find in a schoolbook? – How did you feel about the inclusion of narrative elements?
Learning measures	– Was it difficult to answer the learning measures? – Did you feel more difficulties in certain types of question? Which ones?
Reading context	– If you had to study for a test, would you use this text? Why? – If you wanted to learn more about these topics in your free time, would you read this text? Why?

3.3.2. Results

Perceived Prior Science Knowledge. As expected, participants rated as low their perceived prior knowledge of the texts' topics ($M = 1.3$, $SD = 0.4$) as other science topics ($M = 1.5$, $SD = 0.5$).

Text Evaluation. No differences were found between Pretest 1 text types ($p > .1$ for all the evaluated parameters).

Learning Measures. We were mainly interested in obtaining an overall view of the materials and paradigm, so even though statistical tests were also performed, they were not the main focus at this stage. Overall, performance was extremely low. A Kruskal-Wallis test showed that the average scores did not vary across text type (NT: $M = 0.37$; $SD = 0.65$; ET: $M = 0.31$; $SD = 0.5$; $H_{(1)} = 0.07$, $p = .792$), but they did across comprehension levels ($H_{(3)} = 10.55$, $p = .014$). Mann-Whitney tests revealed that recall questions led to significantly higher scores ($M = 0.44$; $SD = 0.65$) than inference and application questions ($M = 0.25$; $SD = 0.51$; $U = 5711$; $p = .002$), as did questions which only required information explicit in the text (recall and connective inferences, $M = 0.41$; $SD = 0.65$), as compared to questions which required extra-text information (predictive inferences and application, $M = 0.18$; $SD = 0.36$; $U = 4825$; $p = .007$).

Qualitative Feedback. All participants from Pretest 1 reported difficulties in distinguishing central from accessory information, and in retaining specific and technical information. Many also expressed lack of interest in the topics, and that they were not expecting to have to answer specific or technical questions (i.e., the learning measures).

Those who read the ET version highlighted comprehension difficulties stemming from the unfamiliarity of the concepts and from the quantity, diversity and specificity of information, troubles paying attention to technical names, and reported focusing on certain parts of the text in detriment of others. Whereas some participants found the ET similar to science school manuals, others thought it was clearer. They also reported difficulties answering the learning measures, regardless of the type of question (e.g., recall vs. interpretation).

Participants of Pretest 1 who read the NT, whose feedback we were particularly interested in, declared they had no difficulties in extracting a global idea of the text, yet struggled to retain specific scientific information. Some participants said that, because it was a narrative, they tried to understand the general outline, the order and logic of the events and the character's life. This is in line with the results of former studies reporting that young adults display greater adherence to the text's structure and order following NT reading compared to ET reading (e.g., Wolfe & Mienko, 2007), which is suggestive of the activation of different reading strategies (e.g., Hidi et al., 1982; Rastier, 2001; Zwaan, 1994). Participants of Pretest 1 also reported that some narrative details (e.g., about the scientist's life) attracted their attention further, interfering with comprehension. This agrees with the results of previous studies that have investigated the impact of seductive details (i.e., entertaining/interesting yet irrelevant information, e.g., Garner et al., 1989), which further reported that interference of these details on learning is related to factors such as overall coherence and context-dependency (Lehman et al., 2007). Scientific information should therefore be intertwined with narrative information, and be central for the story's development (e.g., Frisch, 2000; Glaser et al., 2009). However, some participants of Pretest 1 expressed that narrative and scientific information were interspersed, and that this made the text seem disperse and hard to follow.

Crucially, some participants viewed narrative information about the discovery process as enhancing both comprehension and interest, as well as a connective element between the narrative and scientific information. They also found that personalizing processes (i.e., presenting explanations based on the scientist's actions) was useful, because it provided them with applied examples. These comments are in line with claims put forward by authors who have used science NT focused on the scientist and argue that bringing learners closer to scientists' actions can promote attention and comprehension (e.g., Arya & Maul, 2021; Hong & Lin-Siegler, 2012). Although most participants found the NT easier to read, they also made clear that they would use this text if they wanted to extract a general idea of the topic (e.g., as a door opener), but would rather use a more conventional ET to study or learn more deeply about it (similar comments are reported by Prins et al., 2017).

3.3.3. Discussion of Pretest 1

This exploratory pretest had the goal of testing a first attempt at material and paradigm building and collecting qualitative feedback. The texts mingled contents from different fields (Botany, Chemistry, Physiology). Overall, the evaluation of the texts met our expectations (e.g., information was novel and somewhat challenging) and were similar for both texts. Performance was similarly low across text types, with many non-responses and errors. Participants provided valuable qualitative insights that were line with previous literature, namely regarding the quantity and diversity of technical information and the kinds of narrative information that were more useful or more distractive.

Overall, Pretest 1 was useful to get a grasp on how to build a science narrative, corresponding learning measures, and scoring guides, as well as to refine our criteria for building the materials. We also had the chance to test most of the ancillary tests, as well as the general paradigm. The quantitative data made clear that whereas texts were equivalent in a series of relevant parameters, they were not efficient learning tools, as illustrated by the poor scores observed on the learning measures. Qualitative feedback provided additional insights that helped identify both problematic and positive aspects. It became apparent that participants activated different reading strategies according to text type: the NT seemingly prompted a more generalized grasp of the texts' contents, while also invoking some resistance as an educational tool. We became aware of the need to present fewer and less diversified science contents, and that scientific information and narrative information were not intertwined, but instead interspersed. Additionally, some narrative information seemed to have worked better, whereas other may have been detrimental to comprehension of the scientific contents. Learning measures were too demanding or ambiguous, and scoring guides could benefit from further elaboration. Finally, it was also deemed useful to include a measure of interest in science in future studies, and to rethink the reading instruction

3.4. Pretest 2: Testing Science Topics

3.4.1. Method

Participants. We examined 51 university students (71% women; M_{age} = 19.18 years, *SD =* 1.27).

Materials.

Texts. Even though the contents of the text used in Pretest 1 were pretty much unknown, which was in line with our aim, it was difficult to find detailed information about the scientist's actions, so we decided to switch to other topics. We built materials on several different topics, putting into practice the idea of presenting the science contents through the actions of the scientists. This could include discovery or application of scientific ideas, practices and/or instruments. Three topics ended up not being pretested: William Herschel's discovery of the infrared radiation, Edward Jenner's development of the vaccine, and Lise Meitner's discovery of nuclear fission. Although the first two provided a detailed process of discovery that easily fit the "scientific narrative" category, it was difficult to intertwine more demanding science contents (properties of light waves and physiological mechanisms of vaccination, respectively) within this process, because this was not what these scientists discovered or dealt with directly. Nuclear fission was dropped because its process of discovery involved a strong theoretical component and because we experienced difficulties developing application questions. Fictionalized direct discourse and other more literary devices were eventually dropped by suggestion of our linguistics collaborator, according to what she learned in her research with textual corpora on science communication (e.g., Gonçalves, 2019).

Two texts were eventually used in Pretest 2. One was a Biology text about transposons or jumping genes. The NT version of the Biology text highlighted how Barbara McClintock's research eventually led her to the discovery of these genes. This passage illustrates the NT version: " However, the method Barbara used did not allow her to properly visualize and differentiate all the corn's chromosomes, so she decided to improve it. Before placing the cells under the microscope, she stained them with a red substance; this way, the structure of the chromosomes became accentuated and Barbara was able to visualize each individual chromosome under the microscope."

The ET version contained the same science contents but was not written according to her actions. This passage illustrates it: "The improvement of the method that allowed for the visualization and differentiation of chromosomes was carried out by Barbara McClintock, and it consists in staining the cells with a red substance before placing them under the microscope. This results in the accentuation of the chromosomes' structure, enabling the visualization of each individual chromosome."

The other was a Chemistry text, which dealt with the concept of molecular chirality. The text explained that crystals can have the same molecular composition but different molecular structures, providing two specific examples and describing methods that can be used to examine these parameters. The NT version was built around Louis Pasteur's discovery of molecular chirality. The NT version of the text can be illustrated in the following passage: "Louis used different methods to study these organic compounds: First, he calculated the atomic weight of the molecules that made up the crystals and wrote down the values on his notebook. Next, he dissolved the crystals in a liquid solution, as to compare their solubility, and wrote down the values as well". The ET version once again contained the same science contents, but it was not framed according the scientists' actions. This passage serves as illustration of this version of the text: "Different methods can be used to study organic compounds (Clayden et al., 2012). Chemical composition can be determined by calculating the atomic weight of the molecules that make up the crystals of the compounds, as well as by analysing the solubility level of crystals when dissolved in a liquid solution". Both texts focused on five subtopics/concepts.

Text Evaluation. We again asked participants to evaluate the presented information, this time using 5-point scale, because it was deemed easier to use than a 7-point scale. In the scale, 1 corresponded to a low evaluation of a parameter and 5 to a high evaluation.

Learning Measures. A total of 12 questions were developed for each topic. The questions tapped different levels of comprehension, this time following the strategy adopted by Arya and Maul (2012). Namely, level 1 (L1) questions tapped recall, level 2 (L2) relied on text-based inferences, level 3 (L3) evaluated the understanding of key ideas, and level 4 (L4) required application in novel situations (level 4), respectively scored from 1 to 4 points. For open-ended questions (levels 1, 3 and 4), specific scoring guides were developed to evaluate the quality of response more thoroughly (full score, 0.75, 0.5, 0.25, 0.125). In the case of L1 questions, these sub-levels reflected the accuracy and completeness of the information, whereas in higher levels (3 and 4) they also reflected degree of elaboration, for instance in establishing connections between ideas or providing explanations.

Ancillary Tests. The same set of ancillary tests as those used in Pretest 1 was applied. The perceived prior science knowledge questionnaire was adapted to include the topics of each text.

General Procedure. The procedure was similar to that of Pretest 1, except that this time we explicitly instructed participants to try to understand the texts.

3.4.2. Results

Perceived Prior Science Knowledge. Participants from both conditions rated their perceived prior knowledge of the texts' topics as low, although the average level of perceived prior knowledge on Biology was higher and presented higher dispersion (*M* = 2.2, *SD* = 0.9) than that of Chemistry (*M* = 1.3, *SD* = 0.3). Knowledge on other topics from the same field was also rated as low (Biology: *M* = 1.6, *SD* = 0.6; Chemistry: *M* = 2, *SD* = 0.6).

Text Evaluation. No differences were found between text types ($p > .085$ in all cases).

Learning Measures. We first collected data from 24 participants (aged 19 years on the average, *SD* = 0.9; 79% women) and made a global inspection of the learning results. Given the high number of systematic confusions, errors and non-responses, we decided to make modifications on the materials of Pretest 2 by reworking parts of the text and making questions more direct or clearer. We presented these modified versions on a fresh group of 27 participants the next day (aged 19 years on the average, *SD* = 1.5; 63% women). Inspection of the Biology learning measures made clear that most participants were drawing on previous general knowledge about the field, instead of using knowledge acquired from the texts. As such, these results were not analysed further. The Chemistry results improved from the previous day, so we decided to run a set of analyses on the Chemistry learning results only.

First, we analysed each participant's Chemistry learning score to get a grasp of the overall performance and identify possible outliers. As ascertained by Kruskal-Wallis tests, averaged Chemistry learning scores did not differ across text types (NT: *M* = 0.51; *SD* = 0.4; ET: *M* = 0.44; *SD* = 0.4; $H_{(1)} = 1.14$, $p = .286$), nor across comprehension levels ($H_{(3)} = 7.05$, $p = .07$). There were no outliers (i.e., participants with a score below 1.5 *SD*). We then analysed each question individually, with the purpose of identifying the number of participants who scored above and below the average value (e.g., 0.5 in a 1-point question). This gave us a sense of the difficulty of the question, and hence if it should be made easier (if many score below average) or more challenging (if many scored above average), and also helped us identify which text parts seemed to be clear and which did not. Moreover, by comparing the scores between text types we would also get a sense of possible discrepancies among them. It was important to ensure that there were no huge baseline differences, as this would suggest that one of the texts was clearly superior in terms of comprehensibility.

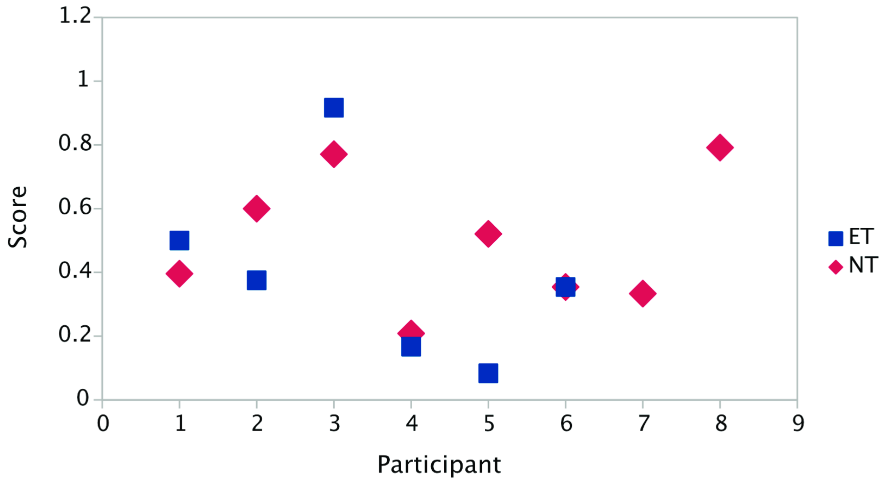


Figure 3.1*. Scatter Plot of the Scores Obtained in a L1 (Recall) Multiple-Choice Question in Pretest 1*

Note. Blue squares represent responses from participants who read the ET and pink diamond represent responses from participants who read the NT. The question was the following: "The tartaric acid and the paratartaric acid have in common: a) Chemical composition, atomic weight and structure; b) Atomic weight, structure and solubility level; **c) Chemical composition, atomic weight and solubility level**; d) None of the above." The correct answer is marked in bold.

Let us consider the scores obtained in a multiple-choice question from the recall level as an example. As illustrated in Figure 3.1, there were four options to answer the question, which referred to one of the text's crucial ideas namely that the two described acids had the same chemical composition but not the same structure. Because responses were largely distributed among the first three options, we realized that this crucial distinction had not been well understood by participants. We decided to score option marked as b) in the note of Figure 3.1 with half point (because it did not contain both chemical composition and structure), and the scatter plot shows that scores tended to be higher for participants who read the NT. We also decided to modify the question, namely, to include only two elements in each option, to rephrase it as "two of the things they have in common", and to remove the option marked as d) in the note of Figure 3.1. We further decided to make the definitions of composition and structure clearer at the beginning of the texts.

Finally, we analysed a set of central tendency and dispersion measures for each question level, to identify any major discrepancies between levels and across text types. For instance, recall level questions had higher average scores in the NT condition, as well as a higher median and maximum score, they also had higher interquartile range and standard deviation and lower minimum score (see Table 3.3). Additionally, it became clear that participants often misinterpreted details in questions and provided under-explanations in higher-level questions. To try to address these issues, we decided to write in bold key words of the questions, such as “correct”, “contrary to”, or “tartaric”, and words like “explain” and “justify”.

Table 3.3. Central Tendency and Dispersion Measures for L1 Questions (Recall) in Both Conditions in Pretest 2

	MIN	Q1	Q2	Q3	MAX	M	SD
ET_L1	0.21	0.4	0.55	0.68	0.79	0.4	0.25
NT_L1	0.09	1.27	1.33	1.38	1.43	0.5	0.57

3.4.3. Discussion

In Pretest 2, we built and tested texts from two fields (Chemistry and Biology) and corresponding learning items. For both topics, the evaluation of the two texts was again very close in all parameters. Results were similar across text types. A series of analyses were performed with the scores (participant, question, level of question), to get a better sense of the data and to modify texts and learning measures accordingly. Given the results, the Biology topic was dropped, mainly because participants were mostly drawing on prior knowledge of related topics to answer the learning measures.

Building several different materials in Pretest 2 allowed us to get a better sense of what worked and what did not, and to refine our criteria. Yet, this translated into a very laborious process, making the construction and use of these materials difficult. This is echoed by other authors, who claim that writing stories that teach scientific concepts in an appropriate, interesting and condensed fashion is a demanding and complex endeavour (e.g., Gilbert et al., 2005; Hadzigeorgiou, 2016; Kerby et al., 2018).

The results from Pretest 2 made us realize that the Biology topic was not fit for our purposes. Perhaps because participants had a slightly higher perceived prior knowledge about this topic, they felt that the contents were more familiar, and that thus they could draw on previous knowledge to answer the questions, instead of using the information presented in the text. This agrees with former observations according to which actual familiarity with the contents (e.g., Wolfe et al., 1998) or the mere subjective feeling of it (e.g., Garcia-Marques & Mackie, 2001) may trigger more heuristic and superficial processing of information, thereby failing to incorporate text information that could be potentially learned. It is possible that familiarity had been triggered by the presence of more frequent or familiar scientific terms (e.g., genetic information; cell; reproduction). The results also showed that, on the other hand, the Chemistry materials fitted our purposes. Participants were not able to draw on previous knowledge, and likely did not perceive the contents as familiar. Scores were higher than on the Pretest 1 and the set of analyses we performed proved very useful, allowing us to get a better sense of the data and to understand how we could refine the materials further. Pretest 2 also helped us refine other tasks (e.g., ART and SRT) and led us to decide to drop the mentalizing task. However, for the sake of concision and because these are less central aspects, we will not delve further into them.

3.5. Pretest 3: Examining Thought Processes with a Think Aloud Protocol

3.5.1. Method

Participants. We examined 12 university students (58% men; $M_{\text{age}} = 21.42$, $SD = 1.24$).

Materials.

Texts. Building from Pretest 2, we made further modifications to the Chemistry materials and built a new pair of Math texts using the criteria we had been refining. The contents of these texts closely matched the ones described in the main study, but the Math texts included extra information about Roman notation arithmetic (e.g., how to use a counting board).

Evaluation Scales. Participants evaluated the presented information using a 5-point scale, where 1 corresponded to a low evaluation of a parameter and 5 to a high evaluation.

Learning Measures. For each topic, two sets of 10 questions were developed, yielding a total of 20 questions per topic (so that we had one set per session for the main study, which would have two sessions).

General Procedure. Participants started by providing some sociodemographic data, after which they practised the think aloud protocol with a brief paragraph. They were encouraged to focus on the task at hands and to verbalize any thought, even if they were unsure about its relevance. Examples included whether certain parts were confusing, ambiguous or unclear or if a question was too difficult. They could take the time they needed and move backwards and forwards in the text. After ensuring that the procedure was well understood, participants read the first text and completed the corresponding evaluation scales and learning measures. This process was repeated with the second text. Participants read the narrative version of a topic and the expository version of the other. The session lasted about one and a half hours.

3.5.2. Results

Perceived Prior Science Knowledge. All participants had completed high school in social sciences/humanities or arts and were not enrolled in any natural or exact sciences' course or line of work.

Text Evaluation. Mann-Whitney tests showed that, for both topics, the two text types were deemed very similar on all parameters (all *p*s > .175). The exception was novelty of information, which for the Math topic was on average higher for the NT (4.4, *SD =* 0.9) than for the ET (*M =* 3, *SD =* 0.9; *U* = 32; *p* = .048). As determined by Wilcoxon tests, topics were evaluated as different on some relevant aspects. On average, compared to the Chemistry topic, although the Math topic was perceived as marginally less novel (4.5; *SD* = 0.7 vs. 3.6; *SD* = 1.1, respectively, *Z* = 20; *p* = .056), it was also considered as less interesting (3.9; *SD =* 0.9 vs. 3.2; *SD =* 0.9, respectively, *t*(12) = 3, *p* = .012) and less clear (4.4; *SD =* 0.8 vs. 3; *SD =* 0.5, respectively, *Z* = 66; *p* = .002). However, it was not considered as more difficult (2.8; *SD =* 1 vs. 3.3; *SD =* 1.1, respectively, *t*(12) = -1.07, *p* = .309).

Learning Measures. Wilcoxon tests revealed that learning scores were equivalent across text types (on average, NT: *M =* .59; *SD =* 0.05; ET: *M =* 0.55; *SD =* 0.14; *Z* = 51; *p* = .38) and topics (on average, Chemistry: 0.54; *SD =* 0.12; Math: 0.59; *SD =* 0.17; *Z* = 23; *p* = .233), without apparent interaction (on average, Chemistry NT: 0.54; *SD* = 0.16; Math NT: 0.66; *SD* = 0.16; Chemistry ET: 0.54; *SD* = 0.08; Math ET: 0.55; *SD =* 0.17). As two participants systematically expressed that the think aloud protocol was interfering with their reasoning while interacting with the Chemistry materials, we also reran the analysis without them, to check if this was influencing the results. The pattern of results remained the same on the 10 remaining participants.

Think Aloud Analysis. Overall, the text which was presented first received more comments than the second text, which was particularly true for negative comments (see Table 3.4). This may have been due to fatigue, as participants commented on the length and complexity of the task. ET appeared to trigger more study strategies, with people re-reading and summarizing and relating information more often with these texts. Overall, NT received much more positive comments than ET (92 vs. 65), but also somewhat more negative comments, although to a far lesser extent (97 vs. 92). Overall, Math texts gathered more negative comments than Chemistry texts, which was especially the case for the ET (see Figure 3.2), and even when participants read the Math NT and the Chemistry ET, they expressed preference for the latter. These comments were mostly directed at the perceived difficulty, lack of clarity, and even weirdness of the contents, especially the ones on the Babylonian notation. This notation was less familiar to the participants, although overall they perceived Math contents as more familiar than the Chemistry ones. These comments matched their ratings on the corresponding evaluation scales, except for difficulty, which was not rated as significantly different from the Chemistry topic.

Table 3.4. Count of Negative and Positive Comments Made on each Text from Pretest 3

	First read text	Second read text
Negative aspects	107	82
Positive aspects	82	75

Additionally, some participants expressed that they missed visual information (e.g., summarizing figure) in the Math texts, which they believed would help comprehension. This may be due to the fact that Math is more abstract than Chemistry (Aso, 2001). Yet, the fact that all the participants that made this comment were reading the ET can also suggest that this feeling was also related to structural features of the text. As for the Chemistry text, the stronger difficulties stemmed from one specific scientific content (namely, the part pertaining to the paratartaric acid crystals). However, in contrast with the Babylonian notation contents, participants seemed to grasp the contents more easily and to clear their doubts as they continued reading the text. The think aloud also provided important information on how to make modifications to specific learning questions which were perceived as ambiguous, too hard, or too easy. More globally, it also became clear that L3 (understanding key ideas) and L4 questions (applying key ideas) should focus on higher-level aspects, and not as much on detailed information.

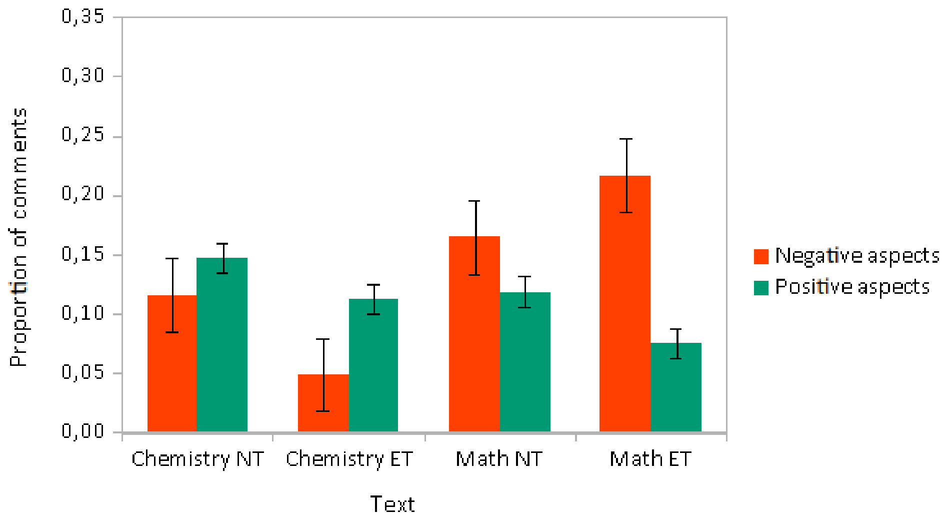


Figure 3.2. Bar Graph of the Proportion (and Standard Error) of Negative and Positive Aspects of each Text used in Pretest 3

3.5.3. Discussion

As a final pretest, we conducted a think aloud protocol with the updated Chemistry materials and newly built Math materials. Our goal was to collect rich and detailed information on participants' thoughts while they read the texts and completed the learning measures. The two versions of the texts from both topics were deemed very similar in all parameters, except for novelty of information in the Chemistry topic. Learning scores were similar between topics and within and between text types. The qualitative analysis revealed that, overall, Math texts were perceived as more familiar than Chemistry texts, but at the same time more complex and difficult. NT received more positive comments, namely when contents were understood or deemed interesting. This feedback allowed us to identify which parts of the materials were difficult or confusing and which worked well, and to get an overall sense of how the different text types and topics were perceived and processed.

The think aloud proved to be a very useful technique to test the learning materials. We were able to complement the quantitative data with rich and detailed information that we would not otherwise have access to. It provided us with a better grasp of not only which parts of the materials worked well or were unclear, confusing or ambiguous but also why participants perceived them that way. It became clear we needed to make more changes in the Math materials, namely removing some contents and making information clearer, which makes sense given that this was the first time we were testing them. The fact that the most unfamiliar Math content (Babylonian notation) was perceived as challenging made us think that it could be beneficial to maintain some familiar contents, as a way to balance things out. However, this also implied that the familiarity with the contents would play different roles in each topic, and that it could prove more difficult to create challenging yet accessible Math texts.

Even though there were no significant learning differences between text types, we were able to collect information that gave us a better sense of how the different text types were perceived and processed. Although this did not translate into statistically significant gains, the narrative versions received more positive appraisals in terms of comprehensibility and interest. Yet, the Math NT may have generated confusion in some readers' minds, who interpreted it as showing how Leonardo Pisano invented one or more notations. This highlights potentially problematic aspects that can underlie narrative processing, namely the tendency to attribute monumentality to the characters (e.g., Allchin, 2003) , and to (wrongly) infer causality between elements as a means to establish coherence and extract explanations (e.g., if this text is about a famous mathematician who used this notation, he must have invented it), also known as the "narrative fallacy" (e.g., Taleb, 2007).

Differences between think aloud and silent reading should also be considered when analysing these results. Although it provides access to participants' thoughts and thinking processes, thinking aloud may alter the process of thinking itself, because it requires cognitive processes that could otherwise be used in the primary (comprehension) task (for discussions, see Wolfe & Woodwyk, 2010; Veenman et al., 2006). The mere fact that there is someone observing can influence participant's thinking and behaviour. Indeed, some participants reported feelings of interference and distraction. However, previous research has also shown that thinking aloud may not reduce learning performance (e.g., Ariasi & Mason, 2011; Bannert & Mengelkamp 2008). In fact, it may even trigger more systematic and analytic reading strategies, such as self-monitoring and prediction generation (e.g., Chi et al., 1994; Kloser, 2013). In the present case, the fact that the study was quite long and demanding likely impacted the results as well, as participants were evaluating the full set of materials.

3.6. General Discussion of the Pretests Data

Building science narrative learning materials is often a challenging and strenuous process, as acknowledged by many educators. One way to approach these materials is to build narratives focused on the scientists' actions (i.e., that explain science contents through these actions), which has been approached in different ways (e.g., Arya & Maul, 2012; Hadzigeorgiou et al., 2012; Hong &

Lin-Siegler, 2011). The present set of three pretests allowed us to incrementally get a better sense of our materials and improve them according to our aims.

Collaborating with experts from different fields (e.g., Arya & Maul, 2012) can be extremely valuable, as they can provide feedback on complementary aspects. For instance, science experts can ascertain the accuracy of the explanations, linguists or language teachers can help determine if the text is well structured and uses appropriate discourse, as well as whether it is identifiable as a specific text type. Having the input of people experienced in educational practices can help establishing connections between the reading materials and the corresponding learning questions.

However, learners will be particularly sensible to the appropriateness of the materials, and so their feedback is hugely important. Both quantitative and qualitative data can provide valuable information at different stages of material development. Through the quantitative analysis of responses and, most importantly, global response patterns (namely, central tendency and dispersion), one can pinpoint systematic confusions and errors, the degree of difficulty of each question, and both individual and overall performance patterns. This information provides hints about the clarity and difficulty of the questions, but also about the parts of the text that may warrant further clarification or disambiguation, or that instead could be made more challenging.

Despite the usefulness of this information, interpreting quantitative feedback also requires a great amount of inferencing. Asking participants for qualitative feedback is hence a valuable way to complement and enrich quantitative outcomes. Applying a think aloud protocol can provide access to thoughts (e.g., reasoning, doubts) and feelings (e.g., frustration, fun) that could not be tapped into otherwise. This can be done either while participants are reading or completing other tasks, providing immediate or concurrent feedback, or after a section or the whole task is complete, which has a more reflective or evaluative tone to it. Both procedures can be used to frame and understand difficulties, frustrations, and errors, as well as to get a sense of participants' motivation and engagement with the materials, all which can influence either specific answers or the global pattern of performance. Additionally, it gives learners a chance to share their own ideas on how science educational materials can become clearer, more intelligible, relatable, and interesting, endowing them with an active voice and role that can be framed along the lines of citizen science projects (e.g., Herodotou et al., 2018; Pandya & Dibner, 2018).

We were also able to extract important insight from the use of different topics. A first conclusion from this approach is that it is not possible to anticipate the appropriateness and the challenging aspects of a specific topic from the head start. To some extent, it is necessary to adopt a trial-and-error approach and experiment different topics. However, this also stems from our approach to the materials, which was guided by the aim of comparing learning science from different text types. It can be extremely useful to start from a specific and pressing pedagogical question (e.g., Browning & Hohenstein, 2015, investigated whether a narrative could help younger students grasp the counter-intuitive concept of evolution). In our case, such questions could have been whether a narrative aids in the understanding of the distinction between composition and structure, or if it can help learn unfamiliar mathematical notations. This approach can make it easier to identify challenging aspects of material development, guide participants' feedback, and direct findings to specific educational contexts. Yet, it also narrows the scope of possibility and removes flexibility and adaptability. In the end, material development should be tailored to the intervention's aims, and narratives focused on the scientist's actions can fulfil different purposes. Even though many of the texts we built turned out to be inappropriate to our specific goal, they could undoubtedly be useful as door openers, to captivate interest, promote historical and sociocultural contextualization of scientific contents, or even to frame questions that are still very much pressing (e.g., vaccination; nuclear energy; women in science).

Finally, it should be noted that we are not proposing that these ideas should be applied in every learning situation. Our materials were built with specific aims, setting and population in mind. Instead, we are providing information about our process of building of these materials because we believe that it is important to share details about its challenges, compromises, as well as to discuss potential solutions to deal with specific obstacles. Whereas these aspects are frequently not shared, they can provide guidance and motivation to educators and researchers who want to develop similar materials, and who may adjust these suggestions to the specific features of their interventions. Other authors have done this to different extents (e.g., Clough, 2011; Flynn & Hardman, 2019; Klassen, 2009), although there is still a gap where young adult populations are concerned, that we hope to help call attention to. Indeed, the fact that there is a great deal of variability in the elements of the learning process (e.g., topic; population; aims; setting) fuels our claim that detailed information about the development of the materials should be shared, so that educators and researchers can benefit from a wide variety of information and choose from it what best fits their purposes. Moreover, although the learning process should always be tailored according to its specific features, as more and more varied information is shared, in time it may be possible to also draw more general conclusions about the use of narrative materials in science learning.

CHAPTER 4

How Did They Do It? The Impact of Narrating Science on Learning

This chapter was based on the following manuscript: Soares, S., Simão, C., Gonçalves, M., Barata, R., Jerónimo, R., & Kolinsky, R. (2022). *How Did They Do It? The Impact of Narrating Science on Learning.* [Manuscript submitted for publication to Science Education], CIS-Iscte, Iscte–Instituto Universitário de Lisboa.

In this chapter, we removed from the main manuscript the section pertaining to “Building and pretesting materials”, as it was presented in the previous Chapter.

Part of the results described in this Chapter were presented at the following conferences: Soares, S., Jerónimo, R., & Kolinsky, R. (2021, May 20-21). *Learning science outside the curriculum: a study with discovery narratives and expository texts* [Oral presentation]. XVI PhD Meeting in Psychology, Lisboa, Portugal; Soares, S., Simão, C., Jerónimo, R., & Kolinsky, R. (2022, June 20-22). *Learning science outside the curriculum: the impact of text type, topic, level of comprehension and delay* [Oral presentation]. XI Simpósio Nacional de Investigação em Psicologia, Trás-os-Montes, Portugal; Soares, S., Simão, C., Jerónimo, R., & Kolinsky, R. (2022, July 20-22). *Narrar ciência para a aprender: Que factores contribuem para a aprendizagem?* [Oral presentation]. 3rd Porto International Conference on Research in Education, Porto, Portugal.

4.1. Abstract

Narrative texts (NT) have been advocated as a tool to improve learning, relative to more traditional materials such as expository texts (ET). Yet, it is unclear under which the conditions science NT can prove beneficial to science learning, and this process remains fairly unexplored. In this study we examined the impact of science NT, compared to equivalent ET, on science learning, at different levels of comprehension and retention delays, as well as the impact of a set of learners’ features on this learning. 125 university students with low prior knowledge of the texts’ topics read a NT and an ET of Math or Chemistry, evaluated them, and answered corresponding tests immediately and one-week after. Learners' contact with literacy, science background, and evaluative and motivational attitudes on science and cognition were also estimated. We expected an advantage for at least one of the NT, particularly at the more elaborate comprehension levels and on delayed measures, as well as a positive impact of learners' features. Data were primarily analyzed through linear mixed models. Learning scores were higher for NT, Math, and immediate measures. Crucially, the NT advantage was only significant for the Chemistry topic, at all comprehension levels but particularly at one of the most elaborate. All learners' features positively predicted the learning score, with most main effects and interactions remaining significant. These findings highlight the importance of considering the diverse aspects that contribute to the process of learning science through narrative-based materials. Recommendations for future educational practices and research are presented.

Keywords: science learning, science narrative texts, prior knowledge, comprehension levels, learner features

4.2. Introduction

Learning about natural sciences and mathematics (referred here as *science*) takes many forms, ranging from formal education to informal initiatives. It also presents many challenges. Frequently mentioned ones include dense and technical texts (e.g., Plavén-Sigray et al., 2017; Snow, 2002), unfamiliar or counter-intuitive ideas (e.g., Browning & Hohenstein, 2015; Gilbert et al., 2005), contents decontextualized from human and cultural elements (e.g., Sánchez Tapia et al., 2018; Martin & Brouwer, 1991), as well as misconceptions on what science is and how it works (e.g., Allchin, 2003; Clough, 2011). These challenges can be to some extent traced to a gap between the everyday mode of thinking, which tends to be centred around human intentions and actions, and the science mode of thinking, rooted in abstract and formal reasoning and explanations (e.g., Bruner, 1986; see also Egan, 1997).

Such a gap is to be bridged by innovative science education practices (Klein, 2006). At the same time, there have been claims for a greater integration of literacy processes (i.e., reading and writing) in science education (e.g., Morais & Kolinsky, 2016; Norris & Phillips, 2003; Webb, 2010). One innovative science practice that fits these claims is the use of narrative-based materials (e.g., Bruner, 1991; Egan, 2005; Solomon et al., 1992). Narration is generally understood as a sequencing of temporally organized actions or events (e.g., Adam, 2011; Norris et al., 2005; Strube, 1994), and is a transcultural activity that plays a key role in cognition (Bruner, 1986; Gottschall, 2013; Telkemeyer et al., 2009).

Many empirical works comparing narrative texts (NT) with expository texts (ET) reported NT superiority in the comprehension and retention of ideas (e.g., Graesser et al., 1980; Kintsch & Young, 1984; Tun, 1989; Zabrucky & Ratner, 1992), leading to the idea of a *narrative effect* (e.g., Norris et al., 2005). Even though these results stemmed from more conventional, non-science NT, narrative materials have been advocated as powerful tools for conveying challenging information, namely science, to broader and more inclusive audiences (e.g., Bruner, 1996; ElShafie, 2018; Martinez-Conde et al., 2019). This includes the expansion of the implementation of these materials within adult and science experts contexts (e.g., Luna, 2015; Olson, 2015), thus going beyond the populations conventionally associated with narrative materials in learning contexts, children and adolescents.

Yet direct comparison between science NT and ET (more common in science instruction; Avraamidou & Osborne, 2009) has generated somewhat mixed results (Soares et al., 2022a). Some favour narrative materials (e.g., Hadzigeorgiou et al., 2012), others expository materials (e.g., Cervetti et al., 2009), and still others report differences restricted to specific conditions (e.g., immediate vs. delayed measures; Negrete & Lartigue, 2010) or no differences at all (e.g., Lamartino, 1995). Besides covering various science fields (e.g., Physics, e.g., Hong & Lin-Siegler, 2012; Biology; McQuiggan et al., 2008; Earth Sciences, e.g., Maria & Junge, 1993), science NT have incorporated diverse semantic elements, such as anthropomorphic or fantastic elements (Cervetti et al., 2009; Maria & Junge, 1993), science fiction (e.g., Negrete & Lartigue, 2010; Wolfe & Mienko, 2007), or details about scientists' life and work (e.g., Arya & Maul, 2012; Hadzigeorgiou et al., 2012; Hong & Lin-Siegler, 2012), to name a few.

NT built around scientists' actions, where science concepts and explanations are intertwined with the scientists' actions, emotions, and wider context, can be used to tackle some of the problems surrounding science learning. They provide insight on who scientists are and how they do things, retrieving the human and cultural context where science concepts and methods are discovered, developed, or applied. As such, they can arguably provide a more accurate image of science (e.g., Arya & Maul, 2021; Clough, 2011; Kubli, 2005) than traditional science educational materials, which usually present decontextualized explanations of concepts (Avraamidou & Osborne, 2009; Clough, 2011). As regards learning gains, researchers have argued that, by making the author visible (Arya & Maul, 2012), enhancing human and social presence (Hong & Lin-Siegler, 2012), and humanizing meaning (Hadzigeorgiou et al., 2012), these NT bring learners closer to science ideas, prompting attention, interest, reflection, organization of knowledge, and emulation of the scientists' practices (e.g., Arya & Maul, 2012; Hadzigeorgiou et al., 2012; Hong & Lin-Siegler, 2012).

Studies which compared this kind of science NT with more traditional, expository-based approaches, usually report better outcomes for the narrative condition. In some studies, this advantage is present both in immediate and delayed measures (e.g., Hadzigeorgiou et al., 2012), whereas in others it is more marked or limited to delayed tests (e.g., Arya & Maul, 2012; Hong & Lin-Siegler, 2012). As for young adults, one study also reported a NT advantage in delayed measures (Negrete & Lartigue, 2010), but another (Wolfe & Woodwyk, 2010) found this advantage only for measures tapping a specific level of representation, namely the *textbase level (*cf. the construction–integration model of comprehension, C-I model; e.g., Kintsch, 1998), which is a representation of the explicit meaning of the text, but not for measures tapping the *situation model*, in which that representation is integrated with prior knowledge and experiences to build a personal interpretation of the text meaning (e.g., Kintsch, 1998).

In fact, the different representation levels proposed by the C-I model (textbase and situation model levels) correspond to different outcomes (memory and learning, respectively) and are assessed with different measures (e.g., recall vs. problem-solving; Ferstl, 2001; Kintsch, 1994). Most studies that used NT built around scientists' actions have not thoroughly examined whether specific measures benefit more from these texts, because they either did not distinguish between different measures or question types (e.g., Hadzigeorgiou et al., 2012), or did not compare them (e.g., Arya & Maul, 2012; Negrete & Lartigue, 2010). Hong and Lin-Siegler (2012) provide an exception and observed both better recall and better complex problem-solving (i.e., identifying gaps and relatedness among different scientific ideas) for NT centred around scientists' struggles, compared with NT centred around the scientists' achievements and more conventional science texts. Textbook problem-solving (i.e., more straightforward problems relating to scientific concepts) did not differ significantly among the conditions.

Finally, the issue of how learning science through NT relates to different facets of the learning process remains relatively unexplored, as we will elaborate in the following section.

4.2.1. The Pedagogical Framework

The RAND Reading Study Group (Snow, 2002) endorses the view that learning through texts is a complex process involving three elements (texts, activities, learners) that interact within and with a wider sociocultural context. Regarding texts, differences in text type and content can impact learning. For example, some topics should be more easily conveyed through NT than others (Frisch, 2010; Norris et al., 2005), and the same topic may yield different results when presented through different types of text. As for learners, their learning is achieved through a series of complex features involving (prior) knowledge, affective dispositions (e.g., interest) and cognitive abilities (e.g., executive functioning). Features of the activity include its goals (e.g., studying vs. leisure) and outcomes (i.e., levels of comprehension). Learning and literacy are seen as partly cultural and historical activities. Interactions with this wider context include the execution and length of the activity (e.g., number of sessions) and the identity of participants (e.g., gender and socioeconomic status, SES).

Few studies have directly examined the impact of these elements on memory and learning outcomes when learning from science NT (e.g., text features, Wilcken, 2008; activity features, Jetton, 1994; learner features, Reuer, 2012; sociocultural features, Arya & Maul, 2012). For instance, Reuer (2012) found interest in the text to be positively correlated with science NT learning outcomes, and Wolfe and Mienko (2007) found that the advantage of NT or ET was dependent on learners' prior science knowledge (NT was more beneficial for less knowledgeable students). Arya and Maul (2012) found that students from lower SES benefited particularly from science NT (this variable was confounded with ethnicity), yet Maria and Junge (1993) did not find such benefit for less skilled readers.

4.2.1. The Current Study

We argue that the variability reported in previous studies using science NT can be explained by variability in the elements that make up the learning process. Taking them into account and gaining more knowledge about their effects can help uncover different and specific narrative effects. Results can then be more accurately interpreted and directed toward specific educational needs, providing a more complete picture of this educational issue. Our conceptual approach and research questions are therefore grounded on the RAND framework. Our aims and hypotheses mainly address features pertaining to the text (type of text, science topic), the activity (comprehension outcomes), and the learner (individual features), but interactions with the wider context, namely from the learner (sociodemographics) and the activity (length, setting), are also considered.

The main aim of the current study was to examine the impact of science NT built around scientists' actions on different levels of comprehension and the durability of these outcomes, in comparison with content equivalent ET. We are not proposing the used texts to be representative of their respective text type or topic in a way that invites generalization of results. Instead, *specific* narrative effects should arise from the specific features and interactions of the various elements of the learning process.

The participants of the current study were an understudied population in this context: young adults (for other examples, see Negrete & Lartigue, 2010; Wolfe & Mienko, 2007; Wolfe & Woodwyk, 2010). Additionally, most studies took place in schools or in articulation with curricula, and a small set took place in museums (Legare et al., 2016; C. Morais, 2015). By contrast, our study took place in a laboratory. This provided a structured and controlled environment where participants nonetheless voluntarily took part to learn outside mandatory curricula, potentially entailing new insights to this literature.

We had three main research questions. The first was whether NT focused on scientist's actions can facilitate science learning among university students with little prior knowledge on the assessed topics. We selected challenging yet accessible contents about which participants had little prior knowledge to maximize the learning experience (Wolfe et al., 1998) and better ascertain the impact of the two text types (e.g., Prins et al., 2017). Because we sought to compare the science learning outcomes provided by the NT and the ET to the same person, we adopted a within-subject design, in which participants could not read the same contents twice. Each of them read two texts on two different topics, written either as a NT or as an ET (henceforth, "Text Type" factor). Text type to topic assignment, as well as order of the texts (henceforth, "Text presentation order" factor), were counterbalanced across participants. We expected at least one of the NT to yield a learning advantage, expressed by a main effect of Text Type or by a Text Type X Topic interaction.

The second main research question asked if long-term retention would be superior for NT contents. To evaluate this, participants were presented with two equivalent sets of questions at two different moments (factor "Session"), with order of the set counterbalanced between participants (henceforth, "Test version order" factor). Session 2 was completed about a week after session 1, a common interval in the relevant literature (e.g., Hong & Lin-Siegler, 2012; Negrete & Lartigue, 2010). Based on the previously discussed studies conducted with similar NT or populations, we expected potential gains from NT to be at least maintained, and perhaps even increased, in the second session, leading to an interaction between Text Type and Session.

The third main research question focused on whether specific levels of comprehension benefit particularly from NT. To examine learning outcomes at specific levels of comprehension, we built learning items grounded on the C-I model (henceforth, "Item Level" factor). Although this model has dedicated special attention to science texts, these were usually expository, and few previous studies with science NT drew on its ideas (for exceptions, see Arya & Maul, 2012; Wolfe & Mienko, 2007; Wolfe & Woodwyk, 2010). In the present study, items were designed to tap into progressively higher degrees of elaboration of information. As such, some items required either to retrieve information that had been explicitly mentioned in the texts (Level 1) or to infer information that was not explicitly mentioned but could be extracted from information that was (Level 2), whereas others assessed the understanding of key conceptual ideas from the texts (Level 3), or required making inferences beyond the scope of the texts, such as application in a novel situation (Level 4). Drawing on the results previously presented, we expected a NT learning advantage to be observed at all levels of comprehension. Given that we aimed at creating accessible texts, based on the C-I model we reasoned this advantage might be particularly marked at more elaborate comprehension levels, and therefore predicted an interaction between Text Type and/or Item Level or between Text Type, Topic and Item Level.

In addition to these three main questions, we explored the potential impact of a set of learners' features which, as stated, remain underexplored. Our question was whether learning science is modulated by learners' contact with literacy, science background, evaluative and motivational attitudes on science and cognition, and sociodemographics, and whether there are differences according to text type. We expected at least some of these features to positively predict learning scores.

4.3. Method

4.3.1. Participants

A total of 143 university students (undergraduates, postgraduates, or Master students), participated in the study. Before being enrolled, participants completed a prescreening questionnaire to ensure they fulfilled the predefined inclusion criteria: being university students; not being enrolled in courses where Chemistry, Physics, Biology, or history of Mathematics are core subjects; being native speakers of European Portuguese; and not having a language or development disorder diagnosed. We conducted an a priori power analysis (GPower 3.1; Faul et al., 2009) for linear multiple regression model (random model) with 6 predictors (Fixed: Topic, Type, Level, Session; Random: Text presentation order; Test version order) to achieve 80% of power with a weak to moderate effect size ($\rho 2$ = 0.13). The recommended sample size was 101 participants. We used a larger sample than the recommendation to have a sample robust to attrition (dropouts and excluded participants). Results were not analyzed until all data was collected.

Data from 18 participants was removed from the analyses: Data from five participants was lost due to problems during the application or data recording; one participant was dyslexic; 10 did not show up for the second session; and two had a very low performance in Chemistry. Namely, one had no correct answers and the other's responses to open-ended questions raised doubts regarding their understanding of the contents in a fundamental level, for they mixed contents from both texts and extra-text contents in an ungraspable manner. The final sample included 125 participants (90 women, 35 men) aged from 18 to 25 years (M_{age} = 21 years; *SD* = 1.75). We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study.

4.3.1. Materials

Texts. A set of criteria for choosing the topics and building the materials was eventually established and progressively refined throughout the three pretests (see Table A1, p. 213), as were scoring guidelines for open-ended questions (e.g., Hong & Lin-Siegler, 2012). Four texts written in European Portuguese were used (see Table 4.1). The Chemistry texts dealt with the concept of molecular chirality. They explained that crystals can have the same molecular composition but different molecular structures, providing two specific examples and describing methods that can be used to examine these parameters. Math texts dealt with mathematical notation. They described three different notations in terms of representation of quantities and basic arithmetic, with a particular focus on how some challenging aspects of these notations were overcome by the Hindu-Arabic notation. Some of the contents of the Math topic were familiar to participants (Roman notation; Hindu-Arabic notation). However, the way these contents were approached and evaluated was not as familiar (e.g., position of symbols; concrete quantities vs abstract manipulation; most participants from the pretest and the main study did not realize that the Hindu-Arabic notation is the one they currently use). It is worth noting that, to the best of our knowledge, this was the first study to use topics from Math and Chemistry with young adults.

In the two ET, contents were structured according to concepts and their relations, which corresponds to how expository materials are usually organized (Meyer, 1975; Taylor, 1982). Science contents were thus explained more abstractly in the two ET compared to the NT, without being explained through scientists' actions as they were in the NT. In the latter, contents were temporally structured, which corresponds to how narrative materials are usually organized (Adam, 2011; Norris et al., 2005). Hence, in the narrative version of the Chemistry topic, the process of discovery led by Louis Pasteur guided the unfurling of the text, and so science contents were described according to his thoughts and actions. In the narrative version of the Math topic, the focus was on how Leonardo Pisano dealt with these different notations during his lifetime, with science contents also being described according to his thoughts and actions.

Table 4.1. Excerpts of the ET and NT of both the Chemistry and Math topics

Text	Excerpt
Chemistry ET	Different methods can be used to study organic compounds (Clayden et al., 2012). Chemical composition can be determined by calculating the atomic weight of the molecules that make up the crystals of the compounds, as well as by analysing the solubility level of crystals when dissolved in a liquid solution.
Chemistry NT	Louis used different methods to study these organic compounds: First, he calculated the atomic weight of the molecules that made up the crystals and wrote down the values on his notebook. Next, he dissolved the crystals in a liquid solution, as to compare their solubility, and wrote down the values as well.
Math ET	Following this logic, each column represents a value 60 times higher than the one that precedes it, rendering the position in which symbols appear, along with their value, paramount to determine the represented quantities.
Math NT	Because each column's value was 60 times higher than the one that preceded it, when Leonardo used the Babylonian notation he had to consider not only the value of the symbol but also the position in which he placed it, for the represented quantities differ drastically.

Note. The original Portuguese texts and their English translations can be found at https://osf.io/jsx5m/?view_only=9640b87c94fe4f52b01e0ba653db8e05.

To maximize the fruitfulness of NT as science learning tools, care was taken to ensure that narrative and science elements were deeply intertwined, so that the former were not accessory but instead central to the comprehension of the science contents (e.g., Fisch, 2000; Wolfe & Mienko, 2007). Because narrative and expository materials have specific organization structures, to guarantee coherence and fluidity the order in which contents appeared sometimes differed slightly between texts of the same topic.

Because the NT had additional narrative information, it was necessary to ensure that both types of text had an equivalent length. This was accomplished by adding ET-congruent information that was not relevant for learning (e.g., author names) to the ET texts. Besides length, an effort was made to ensure that other psycholinguistic parameters were equivalent between all text materials. As can be seen in Table A2 (p. 214), this equivalence was achieved for word frequency, word length, and number of words per sentence, the only exception being the number of subordinate clauses per paragraph, which were significantly more numerous in the Math NT compared to the other texts. This can be because the arithmetic procedures were presented from Leonardo Pisano's point of view, which involved inserting more conjunctions (e.g., if, that, however, because, in order to, though), that introduce subordinate clauses.

Text Evaluation. The participants from the pretests and from the main experiment were asked to rate each text they read for clarity, interest, novelty, difficulty, cohesion, local coherence, and global coherence. In the pretests, there were three additional parameters (syntax, richness of vocabulary, diversity of vocabulary) that were dropped to simplify the task for the main experiment. Parameters were chosen according to previous literature on the subject (e.g., Arya & Maul, 2012; Hong & Lin-Siegler, 2011; Limpo et al., 2014). For each one, a 5-point scale was used, where 1 corresponded to a low evaluation of that parameter and 5 to a high evaluation.

Learning Items. Following Arya and Maul's approach (2012), we established a correspondence between the learning items and the C-I model's levels of representation, hence creating items that evaluated different levels of comprehension. Such a proposal should not however be taken as a generalizable framework for science comprehension or for student's learning progression (Arya & Maul, 2012).

Items from Level 1 (L1) required retrieving information that had been explicitly mentioned in the texts, and so was related to coherence building at the level of individual propositions (defined as micropropositions in Kintsch, 1998; e.g., "in the Babylonian notation there was a symbol for the quantity 1 and a symbol for the quantity 10"). L1 items could be gap filling, short answer (enunciate or enumerate), and multiple-choice questions. Level 2 (L2) items, which were all multiple-choice, required inferring information that was not explicitly mentioned in the text but could be extracted from information that was, which requires establishing coherence between propositions (macropropositions; e.g, "their [tartaric crystals] mirror-image crystals deflect polarized light to the left", which was never directly referred to in the texts). Items from Level 3 (L3) tapped the understanding of key conceptual ideas from the texts, which requires establishing the propositional meaning of the text (textbase macrostructure; e.g., explaining how in the Babylonian notation the number of combinations between symbols was limited, and a system of columns where the position of symbols mattered was used). Finally, Level 4 (L4) items required integrating key conceptual ideas with the reader's own understanding of the text (e.g., extra-text inferences), such as applying them in a novel situation (situation model; e.g., apply the learned knowledge about molecular structure and composition to a hypothetical scenario related with glucose absorption). All L3 and L4 items were open-ended questions.

As already mentioned, we built two sets of items that targeted the same or comparable contents, to be used in two sessions (see examples in Table 4.2).

Care was taken to ensure blind scoring: Participants were attributed a new randomly generated number, and their order in the database was shuffled according to it. To match their different levels of complexity, L1 items were scored between 0 and 1 point, L2 items between 0 and 2 points, L3 items between 0 and 3 points and L4 items between 0 and 4 points. Open-ended items were scored according to four degrees of elaboration, comprehensiveness, and sophistication (see general guidelines with example of scored question in Table A3, p. 215). The first author of this study and a research assistant served as scorers, and at least 30% of the responses to the learning measures of each topic were scored by both raters. After initial training sessions, inter-rater reliability estimated by Cohen's weighed Kappa, was above 76% in all cases. Remaining differences were resolved through discussion and consensus. Scores were then normalized and averaged per participant and per level to calculate learning scores, which were the dependent variables in all analyses.

Table 4.2. Example of Learning Items of each Level of Comprehension

Level of comprehension	Example
L1	Fill in the gaps. In the Babylonian notation there was a symbol for the quantity ___ and a symbol for the quantity ___.
L2	Choose the correct option about tartaric crystals: 1 their mirror-image crystals deviate polarized light to the left 2 they do not deflect polarized light 3 their mirror-image crystals have a different chemical composition 4 they do not have the same solubility level as paratartaric crystals
L3	One of the numeric representations presented in the text is the one from Ancient Babylon. Present and explain the main representation features of this notation.
L4	Glucose is sugar's basic molecule, and it is absorbed by the human body. When analyzed with a polarimeter, its structure deflects polarized light to the right. In medical conditions such as diabetes, people should not absorb much glucose. Taking this information and the text you read into account, do you think it could be possible to create a molecular alternative that allowed these people to consume sugar? Explain how and justify your explanation.

Ancillary Tests. As we wanted to look at the influence of a set of learners' features on the learning of the science contents, participants completed a set of ancillary tests designed to examine sociodemographic variables, contact with literacy, science background, and evaluative and motivational aspects about science and cognition, and executive functions (which will not be examined here).

Sociodemographics. Sociodemographic data was collected through a written questionnaire. The collected data included age, gender, birthplace, nationality, spoken languages (native language, language spoken at home, total spoken languages), formal education (levels and fields of study), and the educational and income levels of participants' household members.

Contact with Literacy. The Teste de Idade de Leitura (TIL; Sucena & Castro, 2008; for adult norms: T. Fernandes et al., 2017) is a reading age test validated for young adult Portuguese. TIL is a sentence comprehension paper and pencil test, with 36 sentences, in which the last word of each sentence is missing. After a few training examples, participants have limited time (1 minute for adults; T. Fernandes et al., 2017) to read and choose for each sentence which of the five alternatives provides the best completion (e.g., "Durante a noite, espero que tenhas bons… sonhos, olhos, lápis, sorrisos, peixes"; meaning "during the night, I hope you have pleasant… dreams, eyes, pencil, smiles, fishes"; for illustration, the correct word is underlined).

The reading experience and habits questionnaire we used was adapted from Santos et al., (2007), a sociological survey designed to collect information on the reading habits of the Portuguese population in the context of the reading project of the Plano Nacional de Leitura (National Reading Plan; Alçada, 2016). Because it is quite long and includes questions which were not relevant for our aims, we selected 15 relevant questions, adapting them to better fit our research purposes.

To evaluate exposure to print, we adapted the Author Recognition Test (ART; Stanovich & West, 1989), in which participants are asked to tick actual literary author names among fictitious foils. We adapted the test to include Portuguese authors and to provide a balance between popularity and literariness, spanning different genres. As recommended by the author who is a text linguist, this was done by considering the reading recommendations of the Plano Nacional de Leitura (which are used in schools), the tops of famous authors from various websites, and book sales tops from big bookshops.

Science Background. In addition to assessing participants' exposure to (literature) authors, it was important for our purpose to assess their exposure to scientists. Thus, in a separate test we adapted the ART to include scientists' names (henceforth, Scientist Recognition Test; SRT). As for the authors in the ART, we included scientists with varying degrees of popularity and from diverse fields by looking into Portuguese school curricula, tops of famous scientists from diverse websites and book sales tops from big bookshops. For both the ART and SRT tests, we computed a *d′* Signal-Detection index (Macmillan & Creelman, 2005) built from the hits on authors or scientists and false alarms on foils.

To evaluate prior science knowledge, we built a brief questionnaire that presented 30 science topics, and asked participants to rate their perceived knowledge of these topics on a 5-point scale that ranged from not knowing anything at all to being able to make elaborate analysis involving the application of the concept. Five of these topics pertained to the contents of the Chemistry text, five to the contents of the Math text and the rest to general Chemistry (4), Math (4), Biology (4), Physics (4) and Astronomy topics (4). General topics were taken from science books. Mixing the texts' topics with general science topics had the double purpose of getting a sense of participants' general knowledge about science and avoiding to call attention to the texts' topics. Such a questionnaire obviously evaluates perceived prior science knowledge, so to ensure truthful answers participants were told that they could be evaluated on the mentioned topics during the session.

We also assessed participants' level of contact with exact or natural sciences during their formal education based on information provided in the sociodemographic questionnaire. Participants were grouped into three levels according to their high school and university fields of study, ranging from level A (less contact with science) to Level C (more contact with science; see Table A4, p. 216).

Evaluative and Motivational Attitudes on Science and Cognition. Besides the interest parameter included in the Evaluation Scales, evaluative and motivational attitudes were evaluated by two measures. In one, participants were asked to rate how much they enjoyed the five science fields covered by the prior science knowledge questionnaire, and how often they searched for information about these fields, using a 5-point scale (where 1 indicated low enjoyment and 5 great enjoyment). The other was the short version of the Need for Cognition questionnaire (Cacioppo et al., 1984) validated for young Portuguese adults (Gomes et al., 2013). Using again a 5-point scale, where 1 indicated "nothing" and 5 "a lot", they rated how much they identified with 18 statements related to preference for complexity, commitment of cognitive effort, or desire for understanding.

4.3.2. Procedure

Data collection took place between May 2019 and April 2020. The first session of the study took place in two universities (102 participants in one and 23 in the other). The second session took place either at the former (106 participants) or online (19 participants), following the declaration of state of emergency (COVID-19 pandemic) and subsequent closing of universities. To ensure online data collection was as equivalent as possible to in-person sessions, participants completed the session while on video call with the first author. Participants were compensated with either course credits or two 10€ vouchers. The study was approved by the same ethics committee (agreement 47/2019).

Apart from TIL, the EF tasks, and the science texts, all tasks were presented and completed on a computer via Qualtrics. All except TIL were completed at the participants' own pace. Detailed instructions were provided on the Qualtrics tabs, but the experimenter provided general instructions and encouraged participants to let them know if any doubt arose, both during in-person and online sessions.

In session 1, participants first read and signed an informed consent, being told that the study aimed at identifying and understanding the conditions that contribute to learning from texts. They were randomly assigned to one of the two main conditions (Chemistry NT and Math ET vs. Math NT and Chemistry ET) and started by completing the TIL. Then, they were presented with the prior science knowledge questionnaire, rated their interest and active search of information about those same science fields, and completed the SRT. These tasks had to be completed before reading the texts because they contained information that could be acquired through the texts (e.g., about chirality).

When participants had finished this first set of tasks, they informed the experimenter and were handed the first text. They were told that they should read it at their own pace, in order to understand it, and to re-read it if necessary; they were also asked to time each reading by using a website stopwatch. Participants were allowed to underline or write on the text. Texts were printed on paper to preserve natural reading conditions (e.g., Mangen et al., 2013) and had one and a half page each. After reading the first text, participants informed the experimenter, who collected the text, and then completed the evaluation scales and the corresponding learning measures (version 1 or version 2). They were encouraged to read and answer the questions calmly and to clarify their doubts with the experimenter. When this was completed, they were handed the second text and the procedure was repeated. Participants ended the first session by completing the ART, the reading experience and habits questionnaire, as well as the sociodemographic questionnaire. These were presented at the end of the session to minimize fatigue before reading the science text and completing the learning measures, which were challenging tasks and our main research interest.

Session 2 was completed one week after session 1. In session 2, participants began by completing the other version of the learning measures (version 2 if they completed version 1 in session 1, and vice-versa) of each topic. Before doing this, they were asked if they had any idea of the tasks they would have to complete in session 2, and if during the one week interval they had any thoughts related to the texts' topics, or tried to learn anything specific about them. The session ended with the completion of the Need for Cognition questionnaire and the EF tasks (see Figure 4.1).

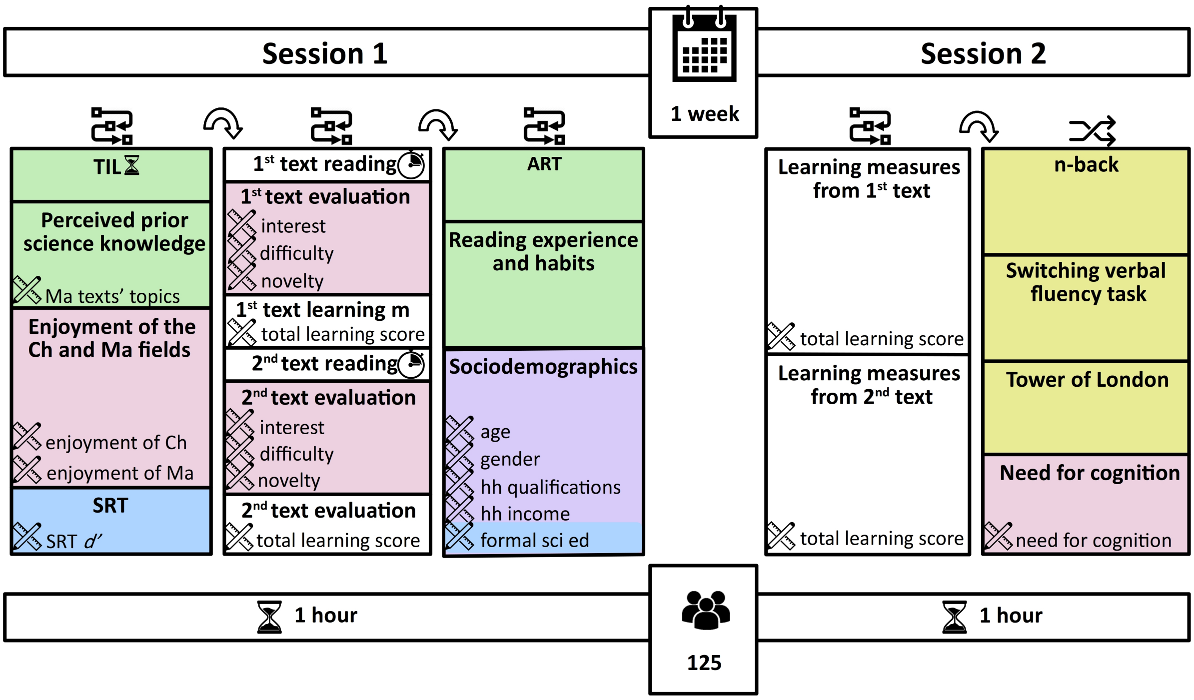


Figure 4.1. Flowchart Detailing the Study's Procedure

Note. Ma: Math, Ch: Chemistry, m: measures, hh: household, sci: science, ed: education. Order of tasks is depicted on the top of each block of tasks (apart for the last block of session 2, which was randomised, all tasks were presented in sequential order). Only the variables used in the analysis are presented inside each task block. The hourglass icon represents time limits, whereas the clock with elapsed time represents that participants timed their performance.

4.3.3. Design, Data Processing and Statistical Analyses

The design for investigating our main questions was nested, consisting of a 2 Text types (NT vs. ET) x 2 Topics (Chemistry vs. Math) x 2 Sessions (Session 1 vs. Session 2) x 4 Levels (L1 vs. L2 vs. L3 vs. L4) x 2 Text presentation orders (NT first vs. ET first) x 2 Test version orders (Test 1 first vs. Test 2 first) plan.

The first four factors (Text Type, Topic, Session, Level, Version) were manipulated within-participants. The combinations of Topic and Text type were specific to each participant, and thus were manipulated between-participants. As such, a participant who read the Chemistry NT did not read the Chemistry ET, for instance. Text presentation order (NT first or ET first) and Test version order were manipulated and randomly counterbalanced between-participants.

The dependent variable was the participants' learning scores. Variables from the ancillary tests (sociodemographics; contact with literacy, science background; evaluative and motivational aspects) were inserted as covariates in some analyses.

We mainly used linear mixed models for the analyses. These statistical models take into consideration that measurements on the same variable for the same subject are likely to be correlated and incorporate both fixed and random effects (Stroup, 2016). IBM SPSS Statistics for Linux, version 27 (IBM Corp., Armonk, N.Y., USA) was used. This study's design and its analyses were not preregistered. Materials, analyses' code (linear mixed models) and data are available on request from the corresponding author.

4.4. Results

4.4.1. Ancillary Tests

We will only present the results of the tasks that were included in the final analysis.

Sociodemographics. Participants enumerated a total of 319 people as economically active members of their households. More than half of these members (65%) had completed high school education, 35% had only finished middle school (9th grade), and 35% had attained some higher education level. More than two thirds of the participants (71%) grew up in households where all economically active members completed at least middle school, and for almost half (46%) the household's level of education was high school. Only one quarter (24%) came from a household where all economically active members completed higher education. Up to 25% of the participants came from a household whose combined value was between 500 and 1000, and up to 75% from a household whose combined value was between 1000 and 1500. As expected, the sample was thus relatively homogeneous in terms of SES. Still, for control purposes, we decided to include these variables in the analyses, together with age and gender.

Contact with Literacy and Science Background. On average, participants correctly identified 54% of the literary authors (*SD* = 0.19) of the ART, and incorrectly identified 10% of the foils as being authors (*SD* = 0.11; *d'* = 59; *SD* = 0.65). In the SRT, on average participants correctly identified 32% of the scientists (*SD* = 0.16), and incorrectly identified 9% of the foils as being scientists (*SD* = 0.09). The *d'* index average was 1.02 (*SD* = 0.53).

Overall, participants had low levels of perceived prior science knowledge. This held true both for the average of the five science topics presented in the questionnaire (*M* = 1.6, *SD* = 0.9; general Chemistry topics: *M* = 1.4, *SD* = 0.8; general Math topics: *M* = 1.7, *SD* = 0.9) and for the specific topics presented in the texts, although it was higher for the Math texts' topics than for the Chemistry texts' topics (*M* = 1.8, *SD* = 1.1 and *M* = 1.10, *SD* = 0.27 respectively, *t*(124) = -14.03, p < .001). In agreement with these data, most participants had a low (Level A: 30%) or moderate (Level B: 45%) level of contact with exact and natural sciences in their formal education, with only a minority (Level C: 25%) reporting more contact.

Text Evaluation. The two Math texts were rated as equivalent on all parameters (all *p*s > .071) except interest (*t*(123) = -2.307, p = .023), with the NT being evaluated as more interesting (*M* = 3.5, *SD* = 1.2) than the ET (*M* = 3.0, *SD* = 1.2). In contrast, the Chemistry texts were only evaluated as equivalent in terms of novelty of information (*t*(123) = -0.045, *p* = .964), with the NT receiving more favourable evaluations on all other parameters (all *p*s < .004). Of note, when the Math texts were jointly compared to the Chemistry texts, they were rated as more interesting, less difficult, and less novel (see Table 4.3).

Table 4.3. Mean and Standard Deviation of Interest, Difficulty, and Novelty per Text Topic

	Interest		Difficulty		Novelty	
	M	*SD*	*M*	*SD*	*M*	*SD*
Math texts	3.3	1.2	2.6	1	3.5	0.9
Chemistry texts	3	1.2	3	1.6	4.4	0.9
Test statistics	*t*(124) = -2.1, *p* = .038		*t*(124) = -2.91, *p* = .004		*t*(124) = -7.71, *p* < .001	

Participants also reported greater enjoyment and more search of information about the Math field (*M* = 2.1, *SD* = 2.4) than the Chemistry one (*M* = 1.1, *SD* = 1.3, *t*(124) = -2.62, *p* = .01). Both ratings were nonetheless low, as was the reported enjoyment and search of information about science fields in general (i.e., for the five science fields presented in the perceived prior science knowledge questionnaire, *M* = 2.2, *SD* = 0.7).

Finally, on average participants scored 3.6 (*SD* = 0.5) on the Need for Cognition questionnaire.

4.4.2. The Effect of Text Type, Topic, Session and Item Level on Learning Scores.

Text Type (NT, ET), Topic (Chemistry, Math), Session (Session 1, Session 2) and Item Level (from L1 to L4) were entered as fixed effects. We also entered the interaction terms of Text Type by Topic, Text Type by Item Level, Text Type by Session, as well as the Text Type by Topic by Item Level. Learning scores were the dependent variable. For random effects, we included the intercepts for participants and the Test version order, to adjust for possible variation (*estimate* = 0.03, *SE* = 0, *Wald Z* = 7, *p* < .001, 95% CI [0.24; 0.04])[1].

Overall, when the influence of the other included factors was considered, all fixed factors significantly impacted learning scores. Namely, Text Type (*F*(1, 1683.83) = 20.41, *p* < .001), Topic (*F*(1, 1683.79) = 19, *p* < .001), Session (*F*(1, 1654.21) = 28.22, *p* < .001) and Item Level (*F*(3, 771.93) = 91.23, p < .001) yielded significant main effects. There was also a significant two-way interaction between Text Type and Topic (*F*(1, 122.94) = 5.78, p = .018), as well as a significant three-way interaction between Text Type, Topic and Item Level (*F*(9, 511.34) = 9.67, *p* < .001). The Text Type by Session and Text Type by Item Level interactions were not significant (*p*s > .164) and so were removed (Stroup, 2016).

As regards main effects, pairwise comparisons revealed that Math scores (*M* = 0.47, SE = 0.18) were overall higher than Chemistry scores (*M* = 0.43, SE = 0.18; *estimate* = 0.13, *SE* = 0.05, *t*(371.91) = 2.76, *p* = .006, 95% CI [0.04; 0.22]), as were NT scores (*M* = .47, *SE* = .18) compared to ET scores (*M* = 0.42, *SE* = 0.18; *estimate* = 0.11, *SE* = 0.05, *t*(321.08) = 2.1, *p* = .036, 95% CI [0.01; 0.21]). Additionally, session 1 led to higher scores (*M* = 0.48, SE = 0.18) than session 2 (*M* = 0.42, SE = 0.18, estimate = 0.06, *SE* = 0.01, *t*(1654.21) = 5.31, *p* < .001, 95% CI [0.04; 0.08]). As for the effect of Item Level, multiple comparison tests revealed that scores significantly differed at all levels (all *p*s < .014), with performance following a downward tendency (L1: *M* = 0.55, *SE* = 0.18; L2: *M* = 0.52, SE = 0.2; L3: *M* = 0.39, *SE* = 0.19; L4: *M* = 0.34, *SE* = 0.2). Yet, pairwise comparisons showed that L4 was significantly different from both L1 (*estimate* = 0.18, *SE* = .03, *t*(384.37) = 5.28, *p* < .001, 95% CI [0.11; 0.24]) and

---

[1] Initially both Test version order and Text presentation order were included as random factors. These variables were deemed redundant by the model, a model including just one of them presented a better fit.

L2 (*estimate* = 0.26, *SE* = .04, *t*(442.65) = 7.12, *p* < .001, 95% CI [0.19; 0.34]), but not from L3 (*estimate* = -0.0, SE = 0.03, *t*(378) = -0.02, *p* = .985, 95% CI [-0.07; 0.07]).

As illustrated in Figure 4.2, decomposition of the interaction between Text Type and Topic revealed that the effect of text type was significant for Chemistry, with higher scores for the NT (*M* = 0.49, *SE* = 0.25; pink (grey) bars) than for the ET (*M* = 0.36, *SE* = 0.25, *F*(1, 153.95) = 13.45, *p* < .001; blue (black) bars). For the Math topic, however, NT (*M* = 0.46, *SE* = 0.25) and ET (*M* = 0.49, *SE* = 0.25) scores did not significantly differ from each other (*F*(1, 145.67) = 0.78, *p* = .377).

Decomposition of the three-way interaction between Text Type, Topic and Item Level showed that for the Chemistry texts, the NT always led to better comprehension than the ET, regardless of the level of comprehension, all *p*s < .036. As can be seen in Figure 4.2, this NT advantage was particularly pronounced at L3. Additionally, in contrast with the linear tendency of the general score pattern (i.e., the main effect), L2 scores were the highest, followed by L1, L3 and L4. For the Math topic there were no significant differences between text types in any level of comprehension (all *p*s > .207).

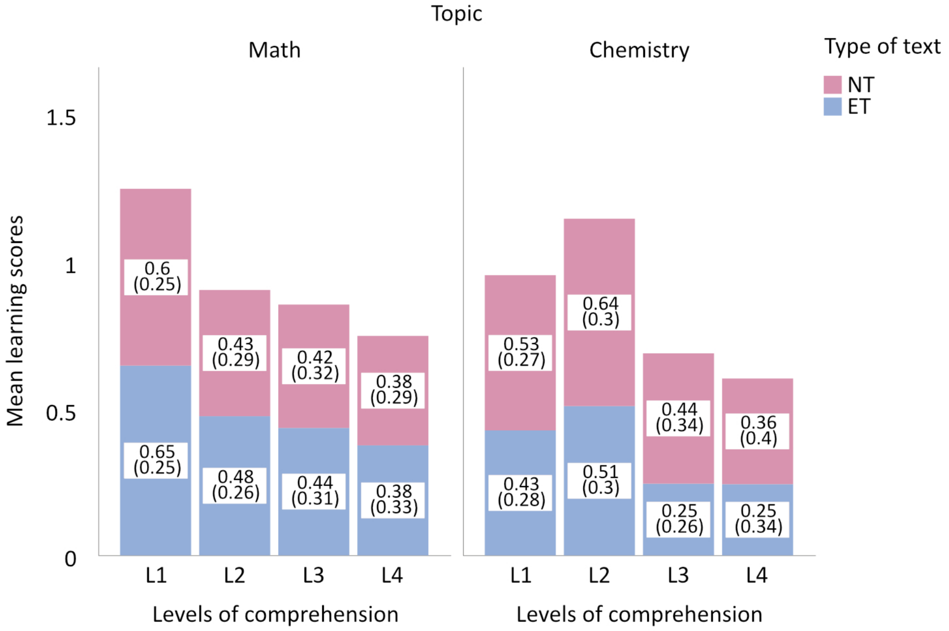


Figure 4.2. Participants' Performance in each Level of Comprehension According to Type of Text and Topic. Mean Values and Corresponding Standard Deviations Are Presented

4.4.3. Models Controlling for Sociodemographics, Contact with Literacy, Science Background and Evaluative and Motivational Attitudes.

We also aimed at gathering evidence on the potential impact of some learner features on science learning. To do so, we created four independent models similar to the previous linear mixed model, but adding relevant covariate variables, relating to sociodemographic factors, contact with literacy, science background, as well as to evaluative and motivational aspects. Other features initially included were removed due to lack of significant effects (cf. Table A5, p. 217).

Whereas the Sociodemographics model did not yield any significant results, all the other covariates positively predicted the learning score: The higher the value of the covariate, the higher the learning score (all *p*s < .03; see Table A6 p. 218). Critically, besides the significant effects of all the covariates, the main effects and the triple interaction previously described proved to be robust, remaining significant in all the models (all *p*s < .001). The Text Type by Topic interaction was significant in the evaluative and motivational attitudes model (*p* = .1), yet only reached marginal significance in the contact with literacy model (*p* = .059) and in the science background model (*p* = .063; see Table A7, p. 219).

We then re-ran each of the four models separately, according to Text type, to examine whether NT and ET generated different results in any of their variables. The results were strikingly similar, with only two exceptions. The variable SRT *d′* score, from the science background model, and the variable Enjoyment of the Chemistry and Math fields, from the evaluative and motivational attitudes model, did not predict learning science from the ET (see Annex A8 p. 220).

4.5. Discussion

In the current study, we compared the impact of two science NT and content-equivalent ET on the learning of low prior knowledge university students. Level of comprehension and retention were also investigated, as well as the impact of a set of learner features. Our main hypotheses were overall confirmed, with results showing that text type impacted learning scores, particularly under specific conditions. We also found that, as predicted, some learners' features influenced learning scores. These findings bring novel and relevant information on the process of learning science through NT, some of which have seldom been addressed before.

4.5.1. Text Type and Topic

Overall, NT scores were higher than ET scores, with at least one NT (Chemistry) generating learning gains compared with its ET counterpart. This result adds to the idea that different NT can impact learning differently, depending on the science topic (e.g., Frisch, 2010; Negrete & Lartigue, 2010), and hence that it might be more useful to study specific narrative effects, instead of aiming at a ubiquitous effect. What's more, it shows that university students can benefit from learning science from NT. This is an understudied population, partly because older students have well developed ET processing strategies, and so would not likely benefit as much from narrative learning materials, conventionally associated to younger students. Indeed, studies on young adults found better learning with the ET (Wolfe & Woodwyk, 2010), or only found a NT advantage in delayed tests (Negrete & Lartigue, 2010) or among less knowledgeable students (Wolfe & Mienko, 2007). Our results seem to be in line with the latter finding, as our participants had low levels of prior knowledge on the assessed contents, particularly on Chemistry, where the strongest NT gains were observed.

The Math NT failed to produce a learning advantage, which may be accounted by many factors. Though we attempted to ensure that the narrative elements and unfolding were central to the comprehension of the contents, some concepts may lend themselves more to be told as science NT than others (e.g., Frisch, 2010). In the Chemistry NT, the reader accompanies a succession of practical experiments that lead the scientist to a discovery, which is very fitting with the temporal organization of narratives. In the Math NT, the arithmetic procedures of different notations are personalized in the actions of the mathematician, rendering the narrative elements arguably less central to the story's development. The differences of these two writing approaches may have impacted the learning of the contents. Moreover, the Math NT had significantly more subordinate clauses per paragraph than any other text. These clauses have been found to be complex and hard to process (e.g., Lord, 2002; Wang, 1970). However, its impact varies according to the specific type of subordinate clause, as well as other text factors (e.g., M. Costa, 2005; Gayraud & Martinie, 2008), so the idea that subordination per se necessarily renders processing more difficult has been questioned (e.g., Baten & Håkansson, 2015; de Ruiter et al., 2020). In fact, learning was overall higher for Math than for Chemistry, but the possibility that the higher number of subordinate clauses contributed to the absence of a NT advantage cannot be completely excluded.

4.5.2. Session

Learning scores were significantly higher in session 1, which stands in contrast with what we predicted based on previous studies, that reported a more marked NT advantage on delayed measures. However, the lack of interaction between text type and session is indicative that the NT learning gains were kept in session 2, as expected. It is unclear how novel this result is. Hadzigeorgiou et al. (2012) reported a NT advantage on both immediate and delayed measures, but whereas the latter values were arithmetically lower they did not provide statistically comparisons. They also used the same learning items in both tests (also Negrete & Lartigue, 2010) and a greater delay (eight weeks) than the present study. The fact that we used slightly different questions between sessions may have played a role in our results. Additionally, Prins et al. (2017) reported that after one week, students retold the science NT using less factual information. NT have been claimed to improve retention compared with other text types (e.g., Mandler, 1984; Mar et al., 2021), which is one important educational goal, so the impact of science NT on this outcome warrants further examination.

4.5.3. Levels of Comprehension

When both topics are considered, scores followed a downward tendency and there was no interaction between text type and item level. The Chemistry topic revealed a different pattern of results. L2 scores (i.e., text inference items) were higher than L1 scores (i.e., recall item). It is possible that Chemistry L1 items were more challenging because they required retrieving terms which were more technical (e.g., atomic weight; polarized light; solubility) than in Math L1 items (e.g., empty space; 1 and 10; 59). The fact that all L2 questions were multiple-choice may have also provided an advantage in this topic. This difference between topics, combined with the need of creating challenging but attainable learning items, highlights the difficulties in building diverse and equivalent materials for different topics, which is a limitation of the present study.

Moreover, we expected the observed NT advantage to be more pronounced at more elaborate comprehension levels, which was partially confirmed in the interaction between text type, topic and item level. In the Chemistry topic, the NT advantage was particularly marked at L3, which was one of the most elaborate levels. The fact that the same was not observed on L4 may indicate that questions from this level were perhaps too difficult, attenuating the advantage. It can also indicate that L3 was a more sensitive measure, perhaps because it involved establishing connections and explaining challenging concepts that had just been learned. Another possibility is that either the content provided in the text, or the time participants had to elaborate on it, was not sufficient to produce effects on L4.

Moreover, most previous studies did not examine (or at least did not report) differences in levels of comprehension across text types. Importantly, Wolfe and Woodwyk (2010) reported that participants who read a science NT showed a stronger textbase representation, whereas participants who read the equivalent ET showed a marginally stronger situation model. Indeed, in a previous work with the same materials, Wolfe had already proposed that the NT seemed to promote more global processing of the text, focused on the order of events and the unfolding of the story, and that the ET induced the processing of more specific details (Wolfe & Mienko, 2007). In contrast, the pattern of the current Chemistry results seems to indicate that the NT was more efficient in promoting both levels of representation, as higher scores were attained at all levels of comprehension. These results more closely resemble the ones found by Hong and Lin-Siegler's study (2012), who reported better recall and complex problem-solving for a NT centred around scientists' struggles (albeit among high schoolers).

4.5.4. Learners' Features

Learner features pertaining to contact with literacy, science background and evaluative and motivational attitudes had an impact on science learning. These findings are significant because such connections remain relatively understudied in the literature. Participants who were more familiar with literature authors' names had higher learning scores. This result is very informing, as previous studies have not devoted much attention to the impact of contact with literacy on science learning (for an exception, see Maria & Junge, 1993, who analyzed groups based on reading skill).

Correlations with learning scores have not always been found (Wilcken, 2008), but some studies did report that prior science knowledge influenced learning from science NT, namely that less knowledgeable learners benefited more from it (e.g., Wolfe & Mienko, 2007). As stated earlier, participants from the current study had low levels of prior knowledge and, overall, those who had a stronger background in science attained better learning scores. This was true for three variables, namely the level of contact with natural and exact sciences during formal education, the perceived prior knowledge about the Math texts' topics and the familiarity with scientists' names. Yet, when text types were analyzed separately, familiarity with scientists' names only predicted learning with the NT. This may signal a difference in how NT and ET were processed, namely that a higher familiarity with scientists was helpful for processing the NT, which was built around scientists' actions.

Of note, participants' perceived prior knowledge on the Chemistry texts' topics did not predict learning, most likely because it was even lower that their perceived prior knowledge on the Math texts' topics. Indeed, the latter were rated as significantly less novel, which we were aware of, having actually used these contents to counterbalance the weirdness provoked by the Babylonian notation (as discussed in the Supporting Results from Pretest 3). Math texts were also rated as less difficult, so the higher prior knowledge on the Math texts' topic arguably accounts for the overall higher Math learning scores, and perhaps even for the absence of a NT advantage in this topic. In a previous study, differences between text types were also restricted to the most difficult text (Cervetti et al., 2009).

Another aspect that can help explain the score differences of the two topics is the interest they induced. The interest evaluations on both topics had a significant impact on learning scores: The more interesting they were perceived to be, the higher the learning score. Math texts were rated as significantly more interesting than Chemistry texts and had higher learning scores. This increased interest may help explain the absence of a NT effect in the Math topic, as one study reported a larger NT advantage in the topic deemed less interesting by participants (Arya & Maul, 2012).

More general measures of attitudes on science and cognition, namely participants' enjoyment of the Chemistry and Math fields and Need for Cognition, had the same effect on general learning. Yet, separate analysis by text type revealed that the former only predicted NT scores. This suggests that being fond of the science fields to which the texts' belonged was particularly important for NT comprehension, and may partially account for the enhanced learning scores it produced.

Importantly, it has been reported that interesting information can interfere with learning when overall coherence and context-dependency are low (Lehman et al., 2007), so the present results hint at a coherent intertwine between scientific and narrative information in the used materials. In addition, though many science NT advocates base their claims on a putative affective superiority of NT (e.g., Avraamidou & Osborne, 2009; Norris et al., 2005; Ritchie et al., 2011), few studies actually examined the impact of interest in learning scores (Reuer, 2012). Some reported more indirect evidence (e.g., Hadzigeorgiou et al., 2012; Hong & Lin-Siegler, 2012), while others failed to consistently find a NT affective superiority among participants (Cervetti et al., 2009). In the current study, interest had a positive impact on learning, arguably at least partly reflected in the main effect of topic (i.e., higher Math scores) and in the text type by topic interaction (i.e., Chemistry NT had both higher learning and interest scores than Chemistry ET). However, it did not translate completely on the main effect of text type, as both NT were rated as more interesting, but did not always produce better learning scores.

Finally, sociocultural aspects (gender, income and educational level) did not significantly impact learning scores. This was unsurprising, given the relative homogeneity of our sample.

Crucially, main effects and interactions were mostly kept in the covariate models, attesting to the robustness of the main results and pointing to a complex interaction between several factors when learning science through narrative texts, that go beyond a simple "narrative effect".

4.5.5. Recommendations for Educational Practices and Research

The present findings point to the multifaceted nature of learning and provide important clues for educators. For both text types and topics, retention of information decreased over the week following the first session, challenging the idea of a ubiquitous NT superiority. Instead, retention may have to be stimulated using other strategies (e.g., Karpicke, & Grimaldi, 2012; Roediger & Karpicke, 2006), and the factors influencing its success should be further investigated.

The fact that we found differences across topics is informative of the potentially differential effects that using NT and ET with similar contents can yield, calling attention not only to text features but also to their interaction with features from the learner, activity, and sociocultural context. Overall, the Chemistry materials were more sensitive to our manipulations, yielding important insights on the influence of text types on different levels of comprehension, particularly when informed by affective and evaluative aspects. If, as suggested by the results on interest and learning, the Chemistry topic was more challenging and/or less interesting, then the NT seemed to have been particularly useful, which is in line with claims that NT materials can be used as tools for dealing with challenging topics (e.g., Negrete & Lartigue, 2010) or as "door openers" (e.g., Gilbert et al., 2005). However, when the topic was less challenging and more interesting (judging by the learning scores and ratings on difficulty and text interest), neither text type nor level of comprehension significantly impacted learning (i.e., Math). For pedagogical reasons, it would be important to address this confound, ascertaining under which conditions NT are less useful to improve learning (e.g., increased easiness or interest). The fact that interest did not necessarily translate into discernible learning gains should not discourage its use: If NT can improve interest in science learning, they should definitely be used as engagement tools.

Results also suggest that learning through NT may engage some processes or features to a different extent, which would align with long-standing claims that NT engage readers' minds in specific ways (e.g., Britton et al., 1983; Bruner, 1986; Egan, 1997) that may improve science learning (e.g., Avraamidou & Osborne, 2009; Hadzigeorgiou, 2016). Investigating these differences and experimenting with them can help build a better understanding of who can particularly benefit from NT and how these educational tools can be integrated in classrooms and other learning contexts.

Moreover, the differences in text evaluation across text types and topics highlight that, despite concentrated attempt at building equivalent texts (see Supporting Material), it is difficult to predict how they will be perceived by participants and how they will impact learning. For these reasons, and because building materials is very demanding, we believe that it is of the utmost important to share information about this process.

From all the covariate variables, prior knowledge has been the most addressed in studies with science NT (e.g., Browning & Hohenstein, 2015; Flynn & Hardman, 2019; Wolfe & Mienko, 2007). The current results, which align with some previous data (Flynn & Hardman, 2019; Wolfe & Mienko, 2007), bring additional evidence on the impact of this variable, while also extending this concept to other science knowledge-related variables such as level of contact during formal education and familiarity with scientists' names. In contrast, literacy related aspects have been been scarcely addressed using science NT (but see e.g., Kidd & Castano, 2016; Mar et al., 2006 for investigation using conventional, non-science, NT). This issue is worth examining further, and there have been claims for a greater integration of science and literacy in science education (e.g., Morais & Kolinsky, 2016; Webb, 2010).

As for the activity, it is possible that the specificities of the learning activity (reading a text, answering written questions) affected learning from the two topics differently. The fact that the current study was not inserted in a mandatory assessment context (e.g., a course), nor in a more spontaneous and intrinsically motivated one (e.g., a museum) should be considered.

On the one hand, the current learning situation served an experimental purpose, providing greater control and internal validity, while also avoided ethical issues that could arise from unbalanced learning outcomes in a formal education setting. On the other hand, its combination of formal (e.g., structure and assessment) and informal (voluntary and non-accredited activity) elements may have created obstacles of its own. Both the texts and some learning items were challenging, and the lack of a broader educational context (and hence of ecological validity) may have interfered with people's engagement. Some participants may have even felt somewhat intimidated by this approach, with some commenting that they did not have a strong science background, or that they were not good at it. Others may have been more careless than they would be in a formal situation. Resistances to NT as a learning tool (see Chapter 3) may also be better framed in longer interventions or in more informal ones (e.g., Callanan et al., 2011; Kokkotas et al., 2010; Murmann & Avraamidou, 2014).

Although we did not delve as much in interactions with the sociocultural context, and did not find significant results related to these variables, examining this dimension can provide crucial insights (e.g., Arya & Maul, 2021; Chappell & Varelas, 2020; Lin-Siegler, 2016; Malone & Barabino, 2009). For instance, it is possible that the learning differences across topics were also influenced by contextual and sociocultural differences stemming from the Math and Chemistry disciplines.

All in all, learning science seems to engage features beyond the science contents, many of which are common to NT and ET, but some that may be specific to NT. Educators can strengthen and work on these different aspects when developing strategies to create more meaningful and effective science learning. To this end, it can be very useful to frame interventions' questions and tasks using the RAND framework, which encompasses all these facets. It can also be helpful to ground learning measures' development and evaluation on a cognitive model of text comprehension such as the C-I model, so that levels of representation and corresponding cognitive operations can be better tapped into.

4.6. Conclusion

Knowing how to best teach science is an ongoing challenge. We showed that text types can impact learning from science topics differently and that NT might be particularly useful when topics are less interesting and/or more challenging. The Chemistry NT afforded better learning than its ET counterpart at various levels of comprehension, but the learning yielded by the Math NT was equally good and this text was considered more interesting, addressing yet another challenge of science education. In fact, although the learning effects observed are robust, our results suggest that the interpretation of these outcomes benefits from considering various learner features, such as ones related to science background, contact with literacy and evaluative and motivational attitudes. All in all, the current findings underline the importance and usefulness of investigating narrative effects, made up of specific interactions between the different facets of learning, laying seeds for future research and providing diverse cues as to how educators may tackle this subject and tailor their practices using this educational tool.

CHAPTER 5

All too Human: An Eye-Gaze and Behavioural Exploratory Study on the Impact of Human Action when Learning from a Science Narrative

This chapter was based on the following manuscript: Soares, S., Frade, S., Jerónimo, R., & Kolinsky, R. (2022). *All too Human: An Eye-Gaze and Behavioural Exploratory Study on the Impact of Human Action when Learning from a Science Narrative* [Manuscript submitted for publication to Reading Research Quarterly], CIS-Iscte, Iscte–Instituto Universitário de Lisboa.

Part of the results described in this Chapter were presented at the following conferences: Soares, S., Frade, S., Jerónimo, R., & Kolinsky, R. (2022, May 19-20). *Major ToM to ground control: using Theory of Mind when learning from a science narrative* [Oral presentation]. XVII PhD Meeting in Psychology, Lisboa, Portugal.

5.1. Abstract

Science is a human endeavour, and narrative texts (NT) can be used to highlight this notion during science learning. However, the elements capturing readers' attention and the human-related thoughts triggered when reading science NT still need to be further explored, as does their impact on learning. In this study, we set out to explore how attention to and thoughts on human action occur and impact learning from a science NT at different levels of comprehension. To this aim, 44 university students with low prior science knowledge read a chemistry NT containing depictions of mental and sensorimotor actions, while their eye movements were recorded. They then answered questions about the text, assessing the extent to which they engaged in thoughts on human action and the learning of the science contents. They also completed a general measure of attribution of intentions. Ancillary measures related to learners' contact with literacy, science background, and motivational attitudes complemented the study. Linear mixed models and correlation analysis were the main methods of analysis. Results showed that attention to mental and sensorimotor actions during reading impacted learning at different levels of comprehension, as did engagement in specific human-related thoughts, some related to Theory of Mind. Crucially, these processes had joint effects on learning, which remained significant when ancillary measures were considered. These findings suggest that human action plays an important role when learning from science NT, encouraging attention to specific text elements and human-related thoughts. Recommendations for research and educational practices are presented.

Keywords: science learning, science narrative text, eye tracking, human action, Theory of Mind

5.2. Introduction

Science texts are key tools in science education (Mason et al., 2013; Sinatra & Broughton, 2011), but their trademark human and social decontextualization has been pointed out as an obstacle to science learning (e.g., Sánchez Tapia et al., 2018; Solomon, 2002). At the same time, there have been claims for a greater integration between science and literacy processes and for increasing the textual variety in science education (e.g., Morais & Kolinsky, 2016; May et al., 2020; Pappas, 2006). Narrative texts (henceforth, NT), which provide such an integration and variety, have been advocated as a tool to help overcome the human and social decontextualization of science materials, for instance by bringing learners closer to their social and cultural frames of reference (e.g., Arya & Maul, 2021; Aukerman & Schuldt, 2021; Clough, 2011).

NT are thought to capture attention because they centre on human action (e.g., Bruner, 1986; Ricoeur, 1992; Larison, 2018). Human action-related features thought to effectively capture learners' attention include being drawn to the human side of science (Arya & Maul, 2012), getting emotionally involved with characters (Banister & Ryan, 2001), and wanting to find out how the story ends (Norris et al., 2005). There have been indeed reports of students being more focused and attentive in classes based on science NT than in regular science classes (Hadzigeorgiou et al., 2012; Morais et al., 2019; Murmann & Avraamidou, 2014), which tend to be based on expository materials (e.g., Avraamidou & Osborne, 2009). Yet, further research is needed to ascertain whether human-related features actually capture readers' attention, and how dedicating attention to these features impacts science learning.

Eye tracking measures can provide valuable insights on these issues. NT and the expository texts used in science teaching usually tend to have different reading purposes and contents. Whereas science texts present abstract, and often difficult, concepts (e.g., Best et al., 2005; Graesser et al., 1991; Sinatra & Broughton, 2011), NT tend to deal with everyday events, and therefore understanding characters, intentionality, as well as temporal and spatial settings becomes crucial (e.g., Graesser et al., 1991; Zwaan & Radvansky, 1998). Consistently, eye-gaze studies examining the processes underlying the reading of expository texts and NT have found processing differences among them (e.g., Gómez-Merino et al., 2022; Kraal et al., 2019). However, to the best of our knowledge, no such study has investigated learning science from NT.

Regarding features of non-science NT, depictions of human action have been found to impact gaze behaviour (Eekhof et al., 2018; Mak & Willems, 2019). Importantly, these depictions are argued to be an important part of the NT reading experience, as readers use the described sensorimotor and mental events to enrich their representations of the text and engage in mental simulation (e.g., Mak & Willems, 2019; Zwaan, 2009). Information on science learning comes instead from studies on expository texts, which have shown that strategic re-reading is related to better learning (Hyönä et al., 2002; Ariasi & Mason 2011; Mason et al., 2013; Mikkilä-Erdmann et al. 2008).

Besides attention to human action, it is important to examine whether science NT can trigger thoughts on human action, and whether these also impact learning. As they centre around human action, NT are proposed to bring science materials closer to our everyday thinking mode (e.g., Bruner, 1986), and to scientists as well. Namely, NT can capture the human and cultural context in which science is developed, humanising scientific meaning (e.g., Egan, 1997) by encouraging the framing of science as a human and social process (e.g., Arya & Maul, 2021; Clough, 2011; Hadzigeorgiou et al., 2012; Kubli, 2005). In turn, this enhanced depiction of human and social elements is proposed to make learners vicariously experience scientists' thoughts and feelings (Arya & Maul, 2012; Hong & Lin-Siegler, 2012; Klassen, 2006), and to use perspective-taking to connect with scientists' ideas and actions (Larison, 2018; Solomon, 2002). In line with these claims, science NT have been reported to help students view science as a human process (Evangelista & Zimmermann, 2008; Leipzig, 2018), and presenting scientists' actions and struggles made students view scientists as hard-working individuals (e.g., Arya & Maul, 2021; Hadzigeorgiou et al., 2012; Lin-Siegler et al., 2016).

Feelings of connection and attempts at view things from the scientist's perspective can be connected to Theory of Mind (ToM). ToM is an umbrella term encompassing a series of abilities linked with attributing and understanding mental states in oneself and others (e.g., Baron-Cohen et al., 1985), such as being able to take other people's perspective and infer their specific thoughts and feelings (e.g., Beaudoin et al., 2020). In NT reading, engaging in such processes can be facilitated when readers feel transported into (Green & Brock, 2000), or immersed in (Ryan, 2001) or absorbed by (Kuijpers et al., 2014) the situation described and by the mental and social life of its characters (see also Dodell-Feder et al., 2013; Mar & Oatley, 2008). We will thus refer to this set of processes as ToM-related ones.

ToM-related processes have been chiefly investigated with conventional NT (i.e., non-science NT; e.g., Kim, 2015; Mar & Oatley, 2008; Oatley, 2016). NT comprehension involves a wide brain network extending beyond language areas and overlapping with regions of the ToM network (e.g., Mar 2004; Mason & Just, 2009; Xu et al., 2005). In addition, many studies present compelling evidence that NT, in particular fiction, trigger processes such as transportation into the text's world, taking the character'(s') perspective, inferring character's mental states, and feeling emotionally connected to them (e.g., Djikic et al., 2013; Mar & Oatley, 2008; Oatley, 2016). To the best of our knowledge, ToM-related processes have not been explicitly referenced in studies on science NT (Soares et al., 2022a). However, some findings suggest that students engaged in them. Studies whose narrative texts underlined scientists' struggles reported feelings of connection (e.g. Hadzigeorgiou et al., 2012; Lin-Siegler et al., 2016), and studies using narrative texts with fictional non-human characters observed attempts to adopt the character's point of view (Jetton, 1994; Murmann & Avraamidou, 2014). Immersion into the narrative learning activity has also been reported (McQuiggan et al., 2008; Murmann & Avraamidou, 2014).

Human elements are thus thought to enhance learning, as they render science materials more interesting (e.g., Arya & Maul, 2012; Hong & Lin-Siegler, 2012), relatable (e.g., Gilbert et al., 2005; Hadzigeorgiou et al., 2012), and memorable and/or understandable (e.g., Arya & Maul, 2012; Hadzigeorgiou et al., 2012; Hong & Lin-Siegler, 2012). At any rate, thinking about human action warrants more investigation in studies using science NT.

5.2.1. The Present Study

The overarching goal of the present study was to investigate how processing human-related information in science NT impacts science learning. This goal was pursued in two ways.

Firstly, we aimed at investigating, through eye gaze analysis, if dedicating attention to depictions of human action while reading a science NT impacts science learning. We analysed different kinds of human action, considering separately those that would reflect processes connected to sensorimotor simulation and those that would tap into mental simulation. Indeed, Mak and Willems (2019) found that depictions of mental states were gazed for longer than depictions of motor actions. As these depictions of human action occur at the clause/sentence level, we selected gaze measures that tap into relatively late, high-level comprehension and integration processes (e.g., Rayner et al., 2004), as in previous studies examining comprehension processes at a similar level (e.g., Ariasi & Mason, 2011; Hyöna et al., 2002). We focused on second pass reading, given that it reflects more conscious and strategic reading than first pass reading (Holmqvist & Andersson, 2017; Hyönä et al., 2003; Rayner, 2009) and on the number of regressions, given that rereading increases comprehension (Schotter et al., 2014).

Our second aim was twofold: To examine whether processes related to thinking about human action occur, and whether they impacted the learning of the science contents as well. We first examined the extent to which reading a science NT triggered thoughts on human action. One way of doing this was by examining if participants mentioned human and social elements when asked open-ended questions about the text that did not explicitly refer to such elements. Being indirect, this measure provides compelling indication that, when learning from a science NT, scientific meaning is humanised, as science is framed as a human and social activity (e.g., Arya & Maul, 2021; Clough, 2011; Hadzigeorgiou et al., 2012). Another way to examine the same question was by analysing participants' engagement in ToM-related processes. As we are not aware of tasks evaluating various ToM-related processes in connection with a specific text, we analysed participants' self-reported engagement in four processes known to be important for (non-science) NT, viz., *transportation*, *perspective-taking* and *cognitive and affective ToM* (e.g., Djikic et al., 2013; Mar & Oatley, 2008; Oatley, 2016). That is, we evaluated if and how participants felt transported into the text, tried to see things from the scientist's perspective, and inferred the scientist's thoughts and feelings, respectively. As a complement to these self-reported measures, we included an adaptation of a comic-strip task (Vistoli et al., 2015; J. Fernandes et al., 2022), which provides a general and non-verbal measure of ToM ability, unrelated to the reading of the text. This task delivers a rather pure measure of intention understanding (Eddy, 2019), as it requires participants to infer a character's intention. This measure allowed us to differentiate the contributions of ToM general abilities and text-dependent thoughts on human-action (some of which ToM-related) on science learning. Given the variety of measures tapping into thoughts on human action, to investigate their impact on learning we first examined how these different measures linked to each other.

Should both attention to and thoughts on human action significantly impact learning, we wished to explore if their impact is concurrent, or whether one of these effects has more explanatory power than the other. Besides determining their general effects on learning, we also sought to explore how attention to and thoughts on human action more specifically relate to different levels of comprehension. To this end, learning was conceptualised and operationalised in the framework of the *Construction-Integration* (C-I) model (Kintsch, 1988), widely regarded as the most comprehensive cognitive model of text comprehension (McNamara & Magliano, 2009). This model posits that comprehension is achieved at different levels of processing, namely at the *textbase* and *situation model* levels, corresponding to memory and learning, respectively. These notions have been applied in several lines of research relevant for the present work, such as studies investigating learning from science narrative texts (Arya & Maul, 2012), examining science learning using eye tracking (Mason et al., 2013), and arguing for the involvement of ToM processes in narrative text comprehension (Kim, 2015).

Complementarily, we aimed at examining if participants' contact with science, literacy, and motivational attitudes impact learning. Familiarity with literature authors and scientists are both relevant measures for the question at hands, as the text we used combines NT processing and science content learning. The impact of interest in learning is well-established (e.g., Fredricks et al., 2018; Wigfield & Guthrie, 2000), and attention to human action has been reported to be influenced by the level of appreciation of the NT (Mak & Willems, 2019). In addition, these three ancillary measures have previously been shown to impact learning from science NT (Soares et al., 2022b). Crucially, besides uncovering the potential impact of these measures, examining their effect along with the effect of attention to and thoughts on human action allowed us to further test the robustness of the latter, while also testing alternative explanations of the observed effects.

5.2.2. Hypotheses

Given how understudied all these aspects are in the context of science NT, the present study was largely exploratory. Regarding our first aim, we expected at least one kind of human action depicted in the NT to significantly impact learning, yet we were unsure whether one kind would be more impactful than the other, or which gaze measure would prove more robust. As a matter of fact, although Mak and Willems (2019) found that parts of the text depicting mental states were gazed for longer than parts depicting motor actions, their study did not involve science NT and did not connect gaze measures to learning outcomes. It was also uncertain whether attention to human action would be particularly associated to specific levels of comprehension.

As for our second aim, we expected inter-individual variability both in how much the reading of the science NT triggered thoughts on human action, and in the text-independent ability to attribute intentions, as measured by the comic-strip task. Previous studies that used this task compared these outcomes in clinical and neurotypical groups or examined exclusively male populations (e.g., J. Fernandes et al., 2022; Oker et al., 2019; Roux et al., 2016), so it was hard to predict if attribution of intentions would be reflected in both RTs and accuracy differences or in just one of them, such as RT, which may eventually be more variable. In any case, based on previous literature, we expected to find connections between these different measures of thinking about human action. To mention a few examples, feeling transported into the text facilitates adopting characters' perspectives (e.g., Mar & Oatley, 2008), perspective-taking is associated with increased discriminability of affective states (Vaccaro et al., 2022), and perspective taking is influenced by attribution of intentions (Furlanetto et al., 2013). As these measures likely engage cognition differently, we also expected them to be linked with different levels of comprehension.

5.3. Method

5.3.1. Participants

A total of 45 undergraduate students from the Psychology and Social Service courses participated in the study. To examine the specific impact of the text on science learning, participants were selected to have, in principle, little prior knowledge on exact and natural sciences. Students enrolled in courses where chemistry, physics, biology or mathematics (besides statistics) are central were thus excluded. In addition, participants had to be native speakers of European Portuguese and have no diagnosed language or developmental disorder, or uncorrected vision problems. These criteria were ascertained through a pre-screening questionnaire. Sample size was determined based on previous eye tracking studies that investigated learning from text (e.g., Ariasi et al., 2017; Catrysse et al., 2018).

Data from one participant were removed from the analyses due to poor spatial accuracy of the gaze pattern (i.e., the distance between the actual and reported gaze positions). The final sample included 44 participants (39 women, 5 men), aged from 18 to 25 years (on average, 20 years; *SD* = 2).

5.3.2. Materials, Procedures and Data Analysis

Figure 5.1. offers an overview of the main tasks and corresponding measures used in the present study, which will be presented in more detail in the following sections.

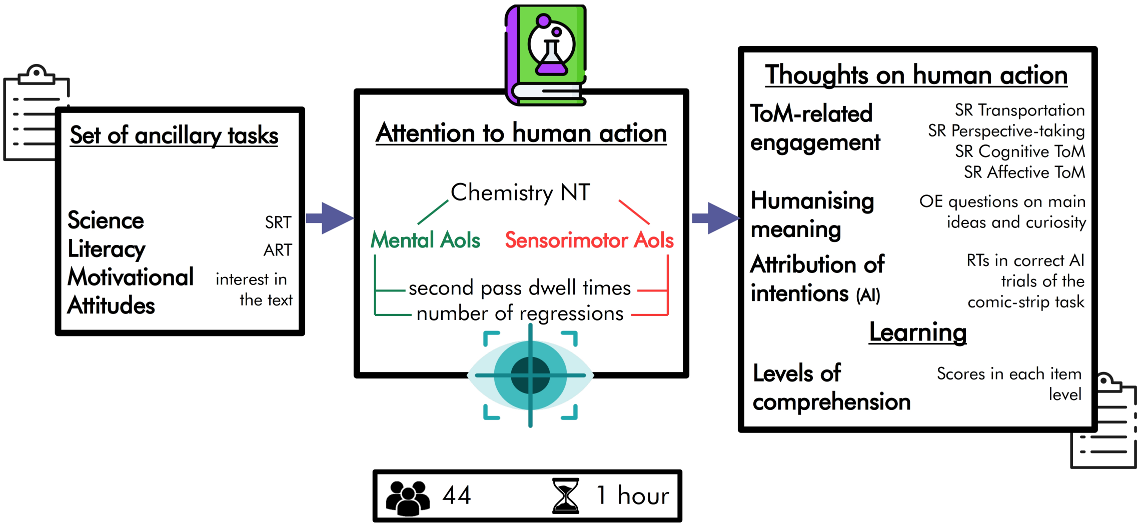


Figure 5.1. Flowchart of the General Procedure Depicting Main Stages, Tasks and Measures

Note. SRT: Science Recognition Test; ART: Author Recognition Test; SR: self-report; OE: open-ended

Attention to Human Action: Text Reading and Eye-Gaze Data Collection. We used a chemistry NT with 785 words that had been previously pretested and used in another study (Soares, 2022b), in which it promoted better learning than its expository counterpart. The text dealt with the concept of molecular chirality. It explained that crystals can have the same molecular composition but different molecular structures, providing two specific examples and describing methods that can be used to examine these parameters. The process of discovery led by Louis Pasteur guided the unfolding of the text, with science contents being described according to his thoughts and actions in a temporal succession (Adam, 2011; Norris et al., 2005). Science and narrative information were closely intertwined, and the latter was central to the comprehension of the former, thus increasing the fruitfulness of NT as science learning tools (e.g., Fisch, 2000; Glaser et al., 2009).

This passage illustrates the text: "Afterwards, **Louis decided to analyse the structure** of the crystals and to do so *he pointed a light* into the crystals and *observed the direction* to which this light was deflected as it passed through them". We identified the sections of the text that included descriptions of human actions and defined them as areas of interest (*AoIs*)[2]. We differentiated between verbs conveying mental actions (e.g., "decide"; in bold in the example provided above;

---

[2] The original Portuguese text and the English translation can be found at https://osf.io/dnwev/?view_only=362f2ec0b27241ada656d6cffd76d3a2

henceforth, *mental AoIs*) and those depicting sensorimotor actions (e.g., "pointed"; in italics in the example; henceforth, *sensorimotor AoIs*). An example is the following sentence: ". There were 17 mental AoIs (average of the number of characters without spaces: 94.61, *SD* = 28.94, from 31 to 145 characters) and 18 sensorimotor actions AoIs (average of the number of characters without spaces: 93.24, *SD* = 27.48, from 29 to 133 characters). The words immediately following the verbs were also considered if this was necessary to ensure the grammaticality of the AoI by European Portuguese standards, and hence that the AoIs could work as autonomous clauses.

The stimuli were presented using SR Research's Experiment Builder software (SR Research, Ottowa, Canada), on a 22-inch monitor with a resolution of 1600 x 900 pixels (32 bits per pixel). The text was divided into 8 sections/pages, that were presented to the participants one at a time. They were presented as black letters on a white background, in an 18-point Courier New font. Between different lines on a page, there was a 3.5-line space interval. For presentation of the sections, minimum margins of 50 pixels were used on all sides. The Experiment Builder software automatically marked each word as a segment to be analysed. There was no space between these segmentations and their boundaries were centred between horizontally and vertically adjacent words. The mean time to read each page was 40.06 second (*DP* = 15.88), the number of characters in each page ranged from 463 to 666 characters (without spaces).

Data were processed and analysed with the Eyelink Data Viewer 4.2.1 Software (SR Research, Ottowa, Canada), SPSS and R. Data quality was verified prior to data export (Orquin & Holmqvist, 2019). First, fixations that had a duration inferior to 80ms were excluded. Then, fixations that diverged too much from the lines of the text, that is, consistently falling into the interest areas of the next line, were corrected. We did a manual drift correction that shifted either all fixations or, in a few cases, the upper or bottom half of the fixations on the page, on the vertical axis (for a similar approach see Eekhof et al., 2018). Data was only considered for the first time reading each page, as there were only a few pages (1 in four participants) re-read by the participants (1.12% of the collected data).

In order to have a duration and a count measure, we extracted the second pass dwell times (summation of the duration of all fixations in AoI after first exiting the area) and the number of regressions (the number of times AoI were re-entered after moving forward in the text), respectively[3]. To control for the length of the AoIs, each one of these measures was computed as a ratio per character (Hyönä et al, 2002; Ariasi et al., 2017).

---

[3] As it is strongly correlated with dwell time, fixation count was not included in the analysis (Tullis & Albert, 2013)

Learning Evaluation. A total of 10 items assessed learning of the science content. These items belonged to four levels (henceforth, Item Level), which prompted different degrees of elaboration of information. This allowed us to tap into to four levels of comprehension within the learning scores (henceforth, Levels of Comprehension), based on the C-I model (Arya & Maul, 2012; Kintsch, 1998). Level 1 (L1) items required retrieving information that had been explicitly mentioned in the texts and so was related to coherence building at the level of individual propositions (e.g., "The light which is used to examine the structure of a crystal is called ___ light"). L1 items could be gap filling, short answer (enunciate or enumerate), and multiple-choice questions. Level 2 (L2) items, which were all multiple-choice, required inferring information that was not explicitly mentioned in the text but could be extracted from information that was, which requires establishing coherence between propositions (e.g, "their [tartaric crystals] mirror-image crystals deflect polarized light to the left", which was never directly referred to in the texts). Level 3 (L3) items tapped the understanding of key conceptual ideas from the texts, which requires establishing the propositional meaning of the text (e.g., "explain how to analyse crystals using a polarimeter, and which characteristics can be examined"). Finally, Level 4 (L4) items required integrating key conceptual ideas with the reader's own understanding of the text (e.g., extra-text inferences), such as applying them in a novel situation (e.g., "apply the learned knowledge about molecular structure and composition to a hypothetical scenario related with glucose absorption"). All L3 and L4 items were open-ended (see the full examples in Table 5.1).

Care was taken to ensure blind scoring: Participants were attributed a new randomly generated number, and their order in the database was shuffled according to it. Because items tapped different degrees of elaboration of information, they were scored in a way that matched this difference in complexity: L1 questions were scored between 0 and 1 point, L2 questions between 0 and 2 points, L3 questions between 0 and 3 points and L4 questions between 0 and 4 points. Responses to open-ended items were scored by the first author according to four degrees of elaboration, comprehensiveness, and sophistication, using the same guidelines presented in Chapter 4 (Table A4, p. 216). At least 30% of these responses were scored by a second independent rater. After initial training sessions, inter-rater reliability, estimated by weighted Cohen's kappa coefficient, was above 78% in all cases. Remaining differences were solved through discussion until a consensus was reached. Scores were then normalized and averaged per item level and per participant, so that they could be used as dependent variables.

Table 5.1. Example of Learning Items from each Item Level

Item Level	Example of Learning Item
L1	Fill in the gaps. The light which is used to examine the structure of a crystal is called ___ light.
L2	Choose the correct option about tartaric crystals: 1. their mirror-image crystals deviate polarized light to the left 2. they do not deflect polarized light 3. their mirror-image crystals have a different chemical composition 4. they do not have the same solubility level as paratartaric crystals
L3	The polarimeter is an instrument that can be used to analyse crystals. Two characteristics that can be analysed are the crystal's chemical composition and molecular structure. 1. Explain how to do analyses using a polarimeter. 2. Such analyses reveal information on the chemical composition and the molecular structure of a crystal? Explain why.
L4	Glucose is sugar's basic molecule, and it is absorbed by the human body. When analysed with a polarimeter, its structure deflects polarized light to the right. In medical conditions such as diabetes, people should not absorb much glucose. Taking this information and the text you read into account, do you think it could be possible to create a molecular alternative that allowed these people to consume sugar? Explain how and justify your explanation.

Thoughts on Human Action: Self-Reported Measures and Behavioural Tasks.

Transportation, Perspective-Taking, Cognitive ToM and Affective ToM. Immediately after reading the text, participants were asked to self-report, on a scale of 1 to 5, how much they had engaged in four ToM-related processes during reading. The items were taken or adapted from concepts and tasks from relevant literature (affective vs. cognitive ToM: Shamay-Tsoory & Aharon-Peretz, 2007; *Interpersonal Reactivity Index*: Davis, 1980; *Narrative Transportation Scale*: Green & Brock, 2000). In addition to the four items tapping into ToM-related processes, four distractor items unrelated to ToM, but instead related to the participants' lives and relationship with chemistry, were included to attenuate the social desirability bias (see Table 5.2).

When participants rated their engagement in ToM-related items as 3 or more, they were asked to further elaborate on this engagement in written responses to open-ended questions (e.g., Please try to explain how you tried to see things from Louis Pasteur's perspective. You can elaborate on your answer, and even refer back to parts of the text). Two types of measure were thus collected for each ToM-related item: Participants' ratings of their engagement (henceforth, *engagement rating*) and the degree of their engagement (henceforth, *engagement degree*). The latter could range from 0 (when engagement ratings were below 3 or no written response was provided) to 4 (written responses denoted a high level of engagement; the complete scoring guidelines can be seen in Table B1, p. 221). As each ToM-related item generated two measures (e.g., rating and degree of engagement in perspective-taking), a total of eight measures pertaining to ToM-related processes were collected.

Table 5.2. Examined ToM-Related Processes and Unrelated Aspects and Corresponding Items

Examined Process/Aspect	Item
ToM-Related Items	
Transportation/Immersion	I tried to imagine myself in the place of the events
Perspective-Taking	I tried to see things from Pasteur's perspective
Cognitive ToM	I thought about what Pasteur might have thought
Affective ToM	I thought about what Pasteur might have felt
Distractor Items	
Chemistry	I tried to recall chemistry contents
Chemistry	I thought about the chemistry classes I had
Personal Life	I thought about things that happened to me recently
Personal Life	I thought about things I need to take care of in my life

Humanising Meaning. After reporting their engagement in ToM-related processes we asked participants two more general, open-ended, questions regarding the reading of the text. The first question was: "Which main ideas do you take from this text?", which was evaluated as a measure of "humanising main ideas". The second question was: "Did the text raise any sort of question? This may include doubts, curiosities, things you would like to know or found interesting, among others", which was evaluated as a measure of "humanising curiosity". For both these questions, we scored the extent to which participants referred to human and social aspects and actions in their responses, ranging from 0 (no response) to 3 (the response made extra-text connections to human elements; see the complete scoring guidelines in Table B2, p. 222).

Attribution of Intentions. The last completed task was an adaptation (J. Fernandes et al., 2022) of the comic-strip task paradigm described by Vistoli et al. (2015). In this task, comic strips including four black and white pictures are sequentially presented. The first three pictures build up a situation that reaches its denouement in picture four, and participants must evaluate if that forth picture presents a logical or congruent ending to the preceding sequence of events (i.e., the three previous images). There are three different experimental conditions. The attribution of intentions (AI) condition depicts situations driven by the intentions of a character (e.g., fetching a chair to reach for something); the physical causality with objects condition depicts situations whose result is a strict consequence of the laws of Physics and has no human character (e.g., wind blowing out a candle); and the physical causality with character condition also depicts situations resulting from physical rules, but where a human character is present (e.g., an apple falling on someone's head). In contrast to these two latter conditions, correct responses to AI trials require participants to infer the character's intention. This setup thus discriminates the ability to decipher the character's intentions from the ability to understand physical causality, while also controlling for the mere presence of human characters.

The comic-strip task was presented using *E-Prime 3.0* (Schneider et al., 2002). The procedure was the same for each comic strip (i.e., trial): The first three pictures were shown for 2000 ms each, separated by 200-ms blank screens. The fourth picture was presented for 5000 ms and was differentiated from the previous three by a dark red border marking it as the strip's ending. Immediately after, a written question asked the participants to respond, as fast and accurately as possible, if the fourth picture presented a logical/congruent ending to the sequence of events presented in the three preceding images. The question was presented on a white background for 5000 ms, or until participants answered. Half of the fourth pictures presented congruent endings and the other half incongruent endings, which were counterbalanced between participants. Participants answered by pressing specific yes/no keys on a keyboard adapted to the dominant hand. After the participants' answer or after the answering time ran out, a black fixation-cross on a white background appeared for 1500 ms, separating trials.

Task instructions were presented on the screen and reinforced by the researcher. Each participant had training session comprising 4 trials of each of the three conditions, always different from those of the main experiment. The main experiment comprised three runs of 24 trials (8 of each condition) separated by short breaks, whose duration was controlled by participants. Strips were never repeated within the same participant. The task lasted for about 20 minutes and the collected measures were accuracy and RTs on correct responses.

Ancillary Tests. Participants completed a set of ancillary tests that assessed their contact with literacy, science background, and evaluative and motivational attitudes on science and cognition. We will only present the measures that were included in the reported analyses (based on Soares et al., 2022b); complete information on additional measures can be found in Chapter 4.

The mean rating of participants' interest in the text (on a 5-point scale, from low to high evaluation) was used as a measure of motivational attitude. This parameter was assessed along with other evaluative aspects, viz., clarity, interest, novelty, difficulty, cohesion, local coherence, and global coherence.

Measures of exposure to literacy and to scientists were provided by the *Author Recognition Test* (ART; Stanovich & West, 1989). The ART was originally designed by Stanovich and West (1989) to reflect cumulated reading practice in a way that avoids social desirability biases. The *Scientist Recognition Test* (SRT) we designed was derived from the ART. In these checklist tests, participants are asked to tick among fictitious foils either actual literary author names or actual scientist names, respectively. The ART version used in the presented study was adapted to include Portuguese authors and to provide a balance between popularity and literariness, spanning different genres. Following the recommendation of a text linguist with whom we collaborated, this was done by taking into account the reading recommendations of the Portuguese National Reading Plan (which are used in schools), the tops of famous authors from diverse websites, and book sales tops from big bookshops. It included 50 items (40 foils). Similarly, the SRT included scientist with different degrees of popularity and from diverse science fields by looking into Portuguese school curricula, tops of famous scientists from diverse websites and book sales tops from big bookshops. It included 50 items (40 foils). For both tests, we computed a *d'* Signal-Detection index (Macmillan & Creelman, 2005) built from the hits on authors or scientists and false alarms on foils.

5.3.2. General Procedure

The study was approved by the ethics committee of the university where it took place (agreement 51/20). Data collection took place in November and December 2020, and all participants received course credits for their participation. Beforehand, students completed a screening test to ensure they fit the study's participation criteria, and were informed of the conditions of data collection with eye tracking. Namely, that the recording did not involve filming the person's face nor collecting any biometric data, and that the data collection process could cause a slight discomfort. The students who passed the screening test were sent part of the ancillary tests.

The experimental session was individual and lasted about 1 hour. They took place at the laboratory of the university that participants attended, in a quiet room with constant luminosity.

Except for a reading age test (see Chapter 4), all tasks were completed at the participants' own pace on a computer. Except for the science text and the comic-strip task (see above), all computer tasks were presented through *Qualtrics*. The researcher provided general instructions and encouraged participants to speak if any doubt arose. Detailed instructions were also provided for each task on the computer screen.

Participants began by reading and signing an informed consent, being told that the study aimed at identifying and understanding the conditions that contribute to learning from text. Then, they completed the reading age test, a prior science knowledge questionnaire on five science fields, rated their interest and active search of information on the same science fields, and completed the SRT and the ART. These tasks had to be completed before reading the texts because they contained information that could be acquired through the texts (e.g., about chirality). After finishing this first set of tasks, participants read the text while their eye movements were recorded using a SR Research Eyelink Portable Duo eye-tracking system. Data were recorded with a sampling rate of 2000Hz. A standard 9-point calibration and validation procedure was used and both eyes were tracked. Participants were asked to take place on an adjusted and comfortable chair in front of the eye tracking equipment. They were told that the text would be divided in various pages and that they should read it and move through the pages at their own pace, by pressing the right arrow key. They were also told they should try to understand the text's contents and that they could go back to a previous paged if they wished, by pressing the left arrow key. They were warned that they should be as quiet as possible and avoid moving their heads. Participants then placed their head on a chin rest, which supported and stabilized their head and assured a constant distance of ≅ 70 cm between the screen and the participant's eye (cf., Catrysse et al., 2018; Korinth & Fiebach, 2018). The researcher initiated the task with a nine-point calibration and validation procedure. After successful calibration, additional drift checks were performed at the onset of each page to ensure identical start positions for the eyes.

After reading the text, participants informed the researcher, who asked them to sit in front of other computer where they would perform the remaining tasks. They began by reporting their engagement in ToM-related processes (and unrelated aspects) during the reading of the text, followed by text evaluation scales (more information in the Chapter 4). They then answered the two open-ended questions on the ideas prompted by the text, followed by the learning evaluation. The session ended with the completion of the comic-strip task (cf. Attribution of Intentions for the full procedure).

This study was not preregistered. Materials, analysis code and data are available by emailing the corresponding author.

5.4. Results

Data were analyzed using *JASP* (version 0.16.3; JASP Team, 2022) and *IBM SPSS Statistics* for Linux, version 27 (IBM Corp., Armonk, N.Y., USA). Questions were analysed through linear mixed models and exploratory correlation analysis.

In all the linear mixed model analysis, participants' learning scores served as the dependent variable, and Item Level was entered as the fixed factor. This model (henceforth, *base model*) served as a baseline to test the effects of attention to and thoughts on human action. Overall, learning scores were moderate, with an average of 51% correct responses (*SD* = 0.21), and a main effect of Item Level (*F*(3, 48.48) = 15.83, *p* < .001). Pairwise comparisons revealed that L2 scores (i.e., text inferences; *M* = 0.68, *SD* = 0.28) were significantly higher than all other ones (*p* ≤ .001 in all cases), namely, than L1 (i.e., recall; *M* = 0.48, *SD* = 0.25), L3 (i.e., understanding key ideas; *M* = 0.38, *SD* = 0.28) and L4 (i.e., applying key ideas in novel situation; *M* = 0.36, *SD* = 0.39). This result replicates the findings of Authors (2022b). In addition, L1 scores were significantly higher than both L3 (*p* = .015) and L4 (*p* = .052) scores, without significant difference between the latter (*p* = .75).

5.4.1. The Effect of Attention to Human Action on Learning

Given the exploratory nature of the current question, several models were run with the two collected measures of attention to human action (viz., second-pass dwell time and number of regressions in mental AoIs and sensorimotor AoIs) entered as covariates in the base model, separately for mental AoIs and sensorimotor AoIs. The results were the same for both types of AoI. Both measures of attention had a significant impact on learning (second pass dwell time in mental AoIs: *estimate* = 0.01, *SE* = 0.0, *t*(40.53) = 2.78, *p* = .008, 95% CI [0.0; 0.01]; second pass dwell time in sensorimotor AoIs: *estimate* = 0.0, *SE* = 0.0, *t*(40.68) = 2.62, *p* = .012, 95% CI [0.00; 0.01]; number of regressions in mental AoIs: *estimate* = 0.31, *SE* = 0.09, *t*(40.12) = 3.33, *p* = .002, 95% CI [0.12; 0.49]; number of regressions in sensorimotor AoIs: *estimate* = 0.35, *SE* = 0.14, *t*(40.54) = 2.53, *p* = .016, 95% CI [0.07; 0.63]).

As the pattern of results was mostly the same for both gaze measures, but with more interesting results in some analyses for number of regressions (e.g., in the exploratory correlation analysis; see infra), we decided to present here only the results of the analyses ran on number of regressions. The results for second pass dwell time can be found in Annex B.

Additional exploratory correlation analysis revealed that the number of regressions to mental AoIs was significantly correlated with most levels of comprehension (i.e., scores at each Item Level). The correlations with recall (L1; *r*(42) = .359, *p* = .017), text-based inferences (L2; *r*(42) = .414, *p* = .005), and understanding key ideas (L3; *r*(42) = .404, *p* = .006) were moderate and positive, but no significant correlation was found with application of key ideas in novel situations (L4; *r*(42) = .004, *p* = .978). The number of regressions to sensorimotor AoIs only had a significant moderate correlation with recall (L1; *r*(42) = .39, *p* = .009; see Table B3, p. 223 for the results with second pass dwell time).

In summary, returning to areas depicting mental and sensorimotor actions during reading, particularly to the former, significantly impacted learning scores. Participants who returned more to these text parts scored higher in most of the assessed comprehension levels, although this relation is moderate.

5.4.2. Thoughts on Human Action

Transportation, Perspective-Taking, Cognitive ToM and Affective ToM. Regarding participants' ratings of engagement in ToM-related processes during the NT reading, perspective-taking was the most often cited (80% of the participants), whereas affective ToM was the least one (55%). In addition, 77% of the participants reported feeling transported into the events of the narrative text, and 66% reported thinking about Pasteur's thoughts. Half of the participants reported engaging in up to three of the ToM processes, 39% reported engaged in all of them, and only 7% reported not engaging in any of them. Transportation and perspective-taking gathered the higher average engagement rating and engagement degree scores (see average values in Table 5.3.).

Table 5.3. Scores of Engagement Rating and Engagement Degree in the four ToM-Related Processes

ToM-Related Process	Engagement Rating (from 1 to 5)	Engagement Degree (from 0 to 4)
Transportation	3.45 (1.32)	2.07 (1.39)
Perspective-Taking	3.61 (1.17)	1.91 (1.24)
Cognitive ToM	3.32 (1.20)	1.41 (1.30)
Affective ToM	2.75 (1.35)	1.59 (1.60)

As can be seen in Table 5.3, the average degree of engagement with ToM-related processes was mostly deemed low (i.e., there is some level of engagement, but it is rather tentative or ambiguous; see meaning of score levels in Table B1, p. 221). Examples of responses are presented for one of the ToM-related processes in Table B4 (p. 224). However, despite the low average values of degree of engagement, there were rich and interesting written responses, a fact that should not go unnoticed. For instance, some participants described vivid details of the laboratory where Louis Pasteur conducted his experiments, as external observers, as a scientist, or even as Pasteur. Others tried to picture what Pasteur was seeing, particularly the crystals, whereas others imagined what he must have gone through psychologically. To a lesser extent, participants inferred his thoughts and feelings, which were particularly related to the fact that Pasteur made an important scientific discovery and went against contemporary chemists and pre-established ideas (see examples in Table B5, p. 225).

These results suggest that the science NT triggered engagement in ToM-related processes. Average scores were low, but some responses denote a deeper engagement.

Humanising Meaning. When reporting the main ideas they took from the NT, most participants stuck to the text's contents, with 45% limiting their references to science contents (i.e., science content level; see Table B2, p. 222) and 34% mentioning Louis Pasteur (i.e., human level). In average, responses were at the human level (*M* = 1.68; *SD* = 0.8). An additional 18% made extra-text connections to more global human and social aspects (i.e., social level). In contrast, only half of the participants reported some curiosity stemming from the reading of the NT (on average*,* 0.81; *SD* = 1; science content level). This curiosity mainly concerned the science contents (27%), but in some cases it was also related to Louis Pasteur (14%) or more global social aspects (19%; see examples in Table B6, p. 226).

It thus seems that reading the NT may have humanised scientific meaning to some extent, especially when it comes to the main ideas taken from the text.

Attribution of Intentions. Accuracy did not differ between AI trials, which demanded inferring intentions (on average, 0.78, *SD* = 0.13), and the two other types of trials, which only relied on physical laws (on average, physical causality: 0.79, *SD* = 0.13; physical causality with a human character: 0.76, *SD* = 0.13; *F*(2, 70.97) = 2.51, *p* = .098). Yet, participants did take longer to answer correctly to AI trials (on average, 1663 ms, *SD* = 466 ms) than to the other two types of trials (on average, physical causality: 1489.83, *SD* = 381.52; physical causality with a human character: 1472.54, *SD* = 424.66; *F*(2,7 0.97) = 2.51, *p* = .098; *F*(2, 86) = 19.71, *p* < .001, respectively).

In short, significant differences in performance were observed in the RTs on correct trials, but not in accuracy. As such, RTs on correct trials AI (henceforth, RTs on correct AI) will be used as a measure of attribution of intentions in the following analyses.

Connecting Thoughts about Human Action: Principal Component Analysis. Because we used several measures related to thoughts on human action, before examining their impact on science learning a Principal Component Analysis (PCA) was performed, to reduce the number of variables. All measures were standardized and submitted to a PCA with Oblique rotation (oblimin with the default settings; Field, 2013), as we assumed the variables to be related. The adequacy of the sample was confirmed by the Kaiser-Meyer-Olkin, Bartlett, and determinant tests, which confirmed that enough items are predicted by each factor, that the variables are sufficiently correlated, and that multicollinearity is not an issue, respectively. After an initial PCA, we removed cognitive ToM rating because it loaded on different components which were not apart by 0.2, and AI accuracy because it ended up alone in a component.

The final PCA produced three components with eigenvalues greater than 1 and explained 70.8% of variance in performance. As can be seen in Table 5.4, the first component (*Simulating Other Points of View*) was made up by four measures of self-reported ToM-related processes and one measure of attribution of intentions. The second component (*Humanising Science*) was made up by the two measures of humanisation of meaning. Finally, the third component (*Inferring Thoughts and Feelings*) was made up by three measures of self-reported ToM-related processes (cf., Tables B7-9, p. 227-29). All components had acceptable to good internal consistency (Simulating Other Points of View: α = 0.82; Humanising Science: α = 0.7; Inferring Thoughts and Feelings: α = 0.8).

Table 5.4. The Variables of the Three PCA Components, with Respective Eigenvalues and % of Explained Variance

Component	Variables	Eigenvalues	% of Variance Explained
Simulating Other Points of View	Rating of transportation rating, degree of transportation, rating of perspective-taking, degree of perspective-taking elaboration, RTs in correct AI	3.858	38.578
Humanising Science	Humanising main ideas, humanising curiosity	1.987	19.866
Inferring Thoughts	Rating of affective ToM, degree of affective	1.236	12.356

and Feelings	ToM, degree of cognitive ToM

The Effect of Thoughts on Human Action on Learning. We ran separate linear mixed models, with each resulting component as a covariate. Whereas Simulating Other Points of View significantly impacted learning scores (*estimate* = .07, $SE = 0.03$, $t(40.37) = 2.33$, $p = .025$, 95% CI [.01; .13]), Humanising Science had only a marginal impact (*estimate* = .06, $SE = 0.03$, $t(40.95) = 1.91$, $p = .064$, 95% CI [-.00; .12]). Inferring Thoughts and Feelings failed to significantly impact learning (*estimate* = -.03, $SE = 0.03$, $t(40.40) = -1.04$, $p = .305$, 95% CI [-.1; .03]), and was therefore not included in subsequent models. The effect of Item Level remained significant in each of the three models (respectively: $F(3, 48.51) = 15.40$, $p < .001$; $F(3, 47.85) = 15.98$, $p < .001$; $F(3, 48.87) = 16.61$, $p < .001$).

Additional exploratory correlation analyses were run between each component and level of comprehension. Simulating Other Points of View and Humanising Science each had a moderate positive correlation with one level of comprehension, namely with recall (i.e., L1; $r(42) = .443$, $p = .003$), and understanding of key ideas (i.e., L3; $r(42) = .331$, $p = .028$), respectively. Inferring Thoughts and Feelings, in contrast, had a moderate negative correlation with text-based inferences (i.e., L2; $r(42) = .32$, $p = .034$).

In sum, as suggested by various measures, participants engaged in thinking about human action. The different measures formed specific components, which differently impacted learning. The combination of transportation, perspective-taking and RTs on correct AI had the strongest impact, which was particularly associated with recall. Although the other components had a smaller overall impact, they too had significant associations to specific levels of comprehension.

5.4.3. The Combined Effects of Attention to and Thoughts about Human Action

We entered as covariates in the base model the number of regressions to mental AoIs and the two components of thinking about human action (entered in the separate models) that significantly impacted learning (cf. Table B10, p. 220, for results with other gaze measures).

The number of regressions to mental AoIs had a significant effect on learning when put in the same model as Simulating Other Points of View (*estimate* = .32, $SE = 0.09$, $t(39.18) = 3.77$, $p = .001$, 95% CI [.15; .5])). Simulating Other Points of View also had a significant impact on learning (*estimate* = .08, $SE = 0.03$, $t(39.18) = 2.93$, $p = .006$, 95% CI [.02; .13])), as did the Item Level ($F(3, 50.38) = 15.76$, $p < .001$).

When put in the same model as Humanising Science, the number of regressions to mental AoIs significantly impacted learning (*estimate* = .28, *SE* = 0.09, *t*(39.65) = 3.01, *p* = .005, 95% CI [.09; .47]), as did the Item Level (*F*(3, 48.5) = 16.54, *p* < .001); yet the component Humanising Science lost its previous (marginal) significance (*estimate* = .04, *SE* = 0.03, *t*(39.65) = 1.43, *p* = .16, 95% CI [-.02; .1]).

We ran additional analyses to determine whether we could tap into the specificity of attributing intentions by examining its relation with attention to human action in the text (as compared to the processing of causality with human characters). No significant results were found (see the section B11 "Additional Correlation Analysis Between Attribution of Intentions and Attention to Human Action", p. 231).

In summary, these results show that attention to and thoughts on human action concurrently impact learning.

5.4.4. Contact with Science, Literacy and Motivational Attitudes

The average rating of the interest elicited by the text, measured on a 1-to-5 scale, was moderate (3.34, *SD* = 1.12), with up to 75% of participants rating their interest as 4. Moreover, on average, participants did not correctly identify many literary authors or scientists. In the ART, participants correctly identified an average of 36% of the literary authors (*SD* = 0.14), and incorrectly judged 5% of the foils as being authors (*SD* = 0.06). In the SRT, the average of correct identification of scientists' names was 23% (*SD* = 0.13), whereas 6% of foils were incorrectly identified as being scientists (*SD* = 0.08). Hence, the *d'* average values were low: -76 (*SD* = 0.82) and -1.65 (*SD* = 0.95) for the ART and SRT, respectively.

These three ancillary measures were separately entered as covariates in the base linear mixed model, taking either as covariates the number of regressions to mental AoIs and Simulating Other Points of View, or as unique covariate Humanising Science (as it lost its explanatory power when put together with the number of regressions). With the exception of interest in the text, all variables yielded marginal to significant effects on learning on the various models (all *ps* < .054; see more details in Tables B12, p. 232, and B13, p. 232, and following discussion in section B15, p. 225).

In conclusion, literacy related factors concurrently contribute to science learning along with attention to and thoughts on human action.

5.5. Discussion

The present study investigated whether attention to and thoughts on human action impact learning science from a science NT. To the best of our knowledge, this was the first study to investigate this topic using a combination of self-report, behavioural and gaze measures. We found evidence that dedicating attention to specific kinds of depiction of human action during reading and that some processes related to thinking about human action had concurrent effects on science learning. These exploratory findings provide novel and specific insights on the kind of processes that are engaged when learning science from NT, a topic yet understudied.

5.5.1. Attention to Human Action Modulates Science Learning

In line with the idea that all knowledge is human knowledge (Egan, 1997), our results show that returning to text areas depicting human action significantly predicted learning from the science NT. This result is in line with previous eye-gaze studies on science learning, which had already established that returning to AoIs and longer reading times are related to improved comprehension outcomes (e.g., Ariasi & Mason, 2011; Catrysse et al., 2018; Hyöna et al., 2002). Although re-reading can indicate processing difficulties, it can also reflect attempts at further integration and deeper cognitive processing (e.g., Hyönä et al., 2003; Mason et al., 2013; Penttinen et al., 2013; Rayner 2009). Yet, whereas in previous eye-gaze studies on science learning AoIs were usually defined by science concepts (e.g., Ariasi & Mason, 2011; Catrysse et al., 2018), in the present study they depicted human actions associated to science concepts.

Our findings also show that both sensorimotor and mental depictions of human action were relevant for science learning. Both can presumably be useful for integrating and understanding science concepts through the scientists' actions, be it by reasoning and appraisals (mental AoIs) or by making important movements with the body (sensorimotor AoIs). Though the pattern of results of gaze behaviour was for the most part similar for these two kinds of depictions, exploratory correlations revealed some differences. Whereas returning to sensorimotor operations only related to recall performance, returning to information on mental operations was related to performance on text inferences and understanding of key ideas in addition to recall. This suggests that the impact of processing mental actions might be more pervasive than that of processing sensorimotor actions.

It may be that mental depictions were more informative or effective in rendering concepts more understandable. Alternatively, they may have been more difficult to integrate. In fact, Mak and Willems (2019) found that areas depicting mental states were gazed for longer than areas depicting motor actions. This can be especially the case in science texts, where mental state information may feel more alien than information on practical procedures. This is further suggested by the fact that there were significantly more regressions to mental areas than to sensorimotor areas (see Table B14, p. 234). At any rate, increased re-reading of mental AoIs benefited learning, and the pattern of results was for the most part similar for both kinds of AoIs, so any processing difficulties posed by mental areas were likely solved. It may also be that these narrative elements provoked more vivid imagery (e.g., Murmann & Avraamidou, 2014; Prins et al., 2017; Vrasidas et al., 2015), which has been shown to promote comprehension and retention of information of – non-science – NT (Horowitz-Kraus et al., 2013; Sadoski et al., 1990).

5.5.2. Thoughts on Human Action: Interconnections and Impact on Science Learning

Our findings suggest that the NT encouraged participants to think about human action. Many participants reported feeling transported into the events of the NT and attempting to view things from the perspective of Louis Pasteur. Although transportation reports were mostly descriptions of what their surroundings would look like (e.g., spatial locations, objects) and attempts at perspective-taking included adopting a more internal perspective of the scientist, these two reports also shared important features. Some participants described feeling transported by imagining being Pasteur or a scientist, and some described trying to adopt a more visuospatial perspective of what Pasteur was seeing, which is a more low-level kind of perspective-taking (Flavell et al., 1986). Given these features, and the fact transportation and perspective-taking were grouped with the measure of attribution of intentions, it is possible that, together, these measures reflect a set of general abilities related to simulating, either by imagining or inferring, other visuospatial and psychological points of view. The fact that transportation was grouped with these two inferential measures is congruent with the view that the ability to transport oneself into another reality may act as a catalyst for more complex inferential processes (e.g., Dodell-Feder et al., 2013). Crucially, this combination of measures was the best predictor of learning scores among measures related to thoughts on human action. It is thus likely that general imaginative and inferential processes related to human action have a part to play when learning science from NT. These processes are important for building coherence and understanding non-science NT (e.g., Graesser et al., 1994; Mason & Just, 2009), so it makes sense they would also be used to understand science NT.

Interestingly, from all the assessed levels of comprehension, it was with recall that simulation of other points of view was significantly correlated. One possible explanation is that increased mental imagery mediates this relationship. As mentioned previously, the creation of vivid and rich images has been connected to the processing of NT and proposed to improve the retention of information (e.g., Gagné, 1978; Horowitz-Kraus et al., 2013; Sadoski et al., 1990). Mental imagery is one of the components of transportation (e.g., Green, 2002), and it has also been linked to some forms of perspective-taking (e.g., empathic, Blouin-Hudon & Pychyl, 2017; visual, Ward et al., 2019), although not very robustly (e.g., Cole et al., 2022). Alternatively, people engaging in these processes of simulation during text reading may elaborate further on the information, which can improve retention (e.g., Craik & Tulving, 1975; Klein & Loftus, 1988).

On the other hand, participants were less prolific in inferring specific thoughts and feelings from Louis Pasteur (i.e., cognitive ToM and affective ToM, respectively). This is relatively unsurprising, given that the reading of the NT was part of a science learning experience, where mentions to mental states tend to be scarce, non-existing, or even discouraged (e.g., Bruner, 1986; Gilbert et al., 2005). Drawing such inferences may thus feel unfamiliar or even misplaced. It is also likely related to the broadness of the question: Asking if participants "tried to see things from Louis' perspective" may invite more diverse responses than inferring specific thoughts or feelings. Indeed, responses to the perspective-taking item were varied and included thoughts and feelings. Although inferring thoughts and feelings did not have a general significant effect on learning, it did have a moderate correlation with the text inferencing level of comprehension. Intriguingly, this correlation was negative, suggesting a possible interference between inferring the thoughts and feelings of the scientist and drawing text inferences to understand science contents.

Notably, participants' written responses on self-reported engagement in cognitive ToM and in affective ToM were mostly related to the fact that Pasteur made an important scientific discovery going against contemporary chemists and pre-established ideas. Previous studies with science NT have reported that participants felt connected to the scientist when the NT talked about the scientists' struggles (e.g., Lin-Siegler et al., 2016) or their defiance of conventions (Hadzigeorgiou et al., 2012). Importantly, this finding suggests that these human and social elements became salient in learners' minds.

The salience of such aspects is further suggested by the results obtained in the two open-ended questions on the ideas prompted by the text. In their responses, participants connected human, and ever wider social elements, to science contents, despite not having been explicitly prompted to establish these connections. Together, these findings strengthen the idea that NT promote the framing of science as a human activity, humanising scientific meaning (e.g., Egan, 1997; Hadzigeorgiou et al., 2012). Many authors have championed the idea that exposing such connections may enhance learning (e.g., Arya & Maul, 2021; Clough, 2011). In the present study, the humanising science component only had a marginal effect on general learning scores. However, the fact that this component moderately correlated with the understanding of key ideas can hint at its involvement in science knowledge construction, even though it is not possible to draw such a conclusion with the present data.

Finally, additional analyses including ancillary measures suggest that the effects of both simulating other points of view and humanising science were to some extent dependent on participants' interest on the text, as these effects overlapped/cancelled each other.

5.5.3. Room for Both: Attention to and Thoughts about Human Science

Crucially, we found evidence that, when considered together, both attention to and thoughts on human action (namely simulating other points of view) made significant contributions to science learning. The fact that they made distinct contributions confirms that in the present study attention to and thoughts on human action were, to at least some extent, different processes. Moreover, it shows that several processes had an impact on the knowledge that was built from the science NT, namely: Dedicating attention to depictions of human action while reading the text, engaging in ToM-related processes pertaining to these actions, and general ToM abilities. Moreover, the effects of attention to and thoughts on human action remained significant when other factors were jointly considered, namely general (science and literature) literacy and text-related motivational aspects, providing further evidence of the robustness of these effects.

5.5.4. Limitations and Future Research

Despite our best efforts, it is unclear whether response desirability was reliably attenuated on the self-reported ToM-related measures. Therefore, we cannot be sure that participants actually engaged in these processes while reading the text (as instructed), or if reading the items during the task prompted them to do so. Although conclusions should be taken with caution, these self-reports provided a window into participants' thoughts on human-related aspects of a science NT, showing that detailed and rich responses can be obtained via self-report.

Additionally, it should be noted that the PCA analysis was run under acceptable, but not ideal, conditions (e.g., few variables overall), rendering results preliminary and exploratory. Despite this limitation, it provided interesting hints for future research. For example, the exploratory correlations suggest that processes reflected by the three components (Simulating Other Points of View, Humanising Science, and Inferring Thoughts and Feelings) may establish specific relations with distinct levels of comprehension, that would be interesting to examine more directly. One way of doing so would be by varying the reading instructions given to participants and examine the impact on different levels of comprehension. Participants could be explicitly asked to imagine themselves in the place of the events or to actively attempt to take the perspective of the scientists (both to examine the simulation of other points of view), or to try to infer the scientists' thoughts or feelings during reading (to trigger inferences of thoughts and feelings), or to think about human and social elements (to humanise scientific meaning). Possible negative or hindering effects of ToM related processes to learning science could be explored as well.

In addition, developing and assessing learning items according to a cognitive model of text comprehension, as we have done presently and in a prior study (the C-I model, Soares et al., 2022b), allows to distinguish between different levels of comprehension and cognitive operations. Future studies may benefit from adopting a similar approach to learning evaluation.

It would also be very important to better investigate different depictions of human action, that is, to examine what kind of processes they trigger, and how these connect to the engagement in ToM-related processes and other human-related thoughts. As we determined that the two sets of processes (attention and thoughts on human action) jointly contributed to learning, it seems unlikely that depictions of human action are merely triggering processes connected to mental state inferences. It also seems that attention to depictions of human action is not merely reflecting participants' interest in the text, as indicated by supplementary analysis with ancillary measures. Uncovering more about the workings of these depictions of human action can, in turn, facilitate the task of determining their role on science knowledge construction, and on distinct levels of comprehension. Yet, a limitation of the present study should be acknowledged, namely that the depictions of mental and sensorimotor actions were identified by the authors of the paper, and not through pretesting (e.g., Mak & Willems, 2019) or by using already validated materials (albeit the latter was not an option for European Portuguese).

Finally, it should be kept in mind that the fact that the NT was used with the goal of learning from the activity, instead of entertaining (as is social and culturally expected), sets the present study apart from most other science learning or NT processing studies. In addition, some contextual features may have influenced participants' learning and engagement. Namely, the fact that the study was not inserted in a continuous and mandatory learning context (e.g., a course), nor in a more spontaneous and intrinsically motivated one (a museum), and the eye tracking data collection procedure (which may have felt weird or uncomfortable).

5.5.5. Applications in Education

Even though the present study is quite exploratory, its findings bring promising insights to applied educational contexts, especially for educators interested in approaching science learning through a more integrative lens. Namely, our results should encourage science educators to integrate depictions of human action when building texts, or to select text materials that do so, as dedicating attention to such depictions did not hamper learning, much on the contrary. From an educational perspective, this is very important, because there are concerns that narrative elements may act as seductive details, providing interesting yet irrelevant information that can compromise concept-related comprehension (e.g., Garner et al., 1989; Hidi et al., 1982). Our findings show that coherent and well-intertwined narrative information does not have this negative effect (Arya & Maul, 2012; Glaser et al., 2009; Lehman et al., 2007), and can be used to introduce variability in textual science educational materials (e.g., May et al., 2020). Importantly, despite the fact that NT are usually used and investigated in populations of younger learners, our set of findings also shows that it is possible to use these materials in young adult learning. In line with previous studies with teenagers, focusing on scientist's struggles (e.g., Lin-Siegler et al., 2016) and their defiance of conventions (Hadzigeorgiou et al., 2012), our results suggest that the use of science NT can be a good way of capturing young adult's attention and trigger thoughts on these human actions. Moreover, framing science as a human activity can help bring learners closer to relevant social and cultural contexts (Arya & Maul, 2021; Aukerman & Schuldt, 2021) and fight misconceptions about scientists and science workings, which are quite ample and can damage learning (e.g., Clough, 2011; Hong & Lin-Siegler, 2012).

Additionally, the less palpable results in cognitive and affective ToM measures should not discourage educators to promote these processes. We believe that they can be part of the science learning experience; it is just that they can be more difficult to achieve and/or may seem weird or even intimidating at first, given how separated from this kind of elements science education materials and modes of thinking have usually been. Another good reason for promoting these ToM abilities is their connection to empathetic and prosocial behaviours (e.g., Imuta et al., 2016; Mar & Oatley, 2008; transportation has also been connected to it, e.g., Johnson et al., 2012). While such connections remain to be further ascertained with science NT, using science texts depicting human emotions and struggles will likely engage these abilities to some extent, especially if overtly addressed and discussed. Indeed, although our study shows that the reading of the science NT can encourage unprompted connections to human and social elements, it also shows that when prompted by more direct questions, participants can provide elaborate responses on specific human-related processes. Thus, scaffolding learners, be it through guided discussions, visual prompts, or inquiry methods, to name a few, may help further integrating these elements with the science contents.

The multidimensionality of the science learning process is not limited to a place for depictions of human action. Learners' literacy, be it related to science or to literature, also has a role to play. Whereas the first may come as expected, the second has received less attention, and both relate to claims for a greater integration between science education and language and literacy processes (e.g., Morais & Kolinsky, 2016; May et al., 2020). These aspects should thus be integrated and fostered during science learning. Furthermore, motivational aspects such as interest in the text may be tightly connected to engagement in thoughts on human action during science learning, and may even determine the extent to which learners engage in such processes.

5.6. Conclusion

Learning science has a human facet, and the results of the present study help to determine the importance of highlighting it. Our results showed that readers that dedicated more attention to depictions of mental and sensorimotor action in a science NT obtained higher science learning scores. Results also suggest a role for thinking about human action when learning science from a science NT, with different processes yielding impact on learning. The various depictions of human action and thoughts on human action may also be related to distinct levels of comprehension. Importantly, the impact of attention to and thoughts on human action on learning proved to be a concurrent advantage, and so both sets of processes should be jointly considered. Together, the findings suggest that learning science is more than understanding complex concepts about the workings of the physical world: It is human, *all too human.*

CHAPTER 6

Science learning… What comes to mind?: People's perceptions and the role of narrative texts and museums

This chapter was based on the following manuscript: Soares, S., Elvas, M. C., Batel, S., Barata, R., Jerónimo, R. (2022). *Science learning… What comes to mind?: People's perceptions and the role of narrative texts and museums* [Manuscript submitted for publication to Public Understanding of Science], CIS-Iscte, Iscte–Instituto Universitário de Lisboa

6.1. Abstract

Learning natural and formal sciences poses various challenges relating to different components of the learning process, as conceived by pedagogical frameworks such as the Reading for Understanding (RAND) framework. As scientific knowledge fills an important social role, and most people acquire this knowledge non-formally, creating lifelong learning opportunities and understanding people's perceptions on science learning become crucial. In this study, we set out to investigate general perceptions on how science learning and literacy develop. We were particularly interested in the roles of narrative texts and museums. Three groups of people with different levels of contact with science participated in an online non-formal learning experience that included science texts and a learning activity facilitated by a museum science educator. The activity was followed by focus group discussions on the topic of science learning and literacy. Three themes were identified: smoothing the way, which depicted key facilitators in initial learning phases; weaving webs, which centred around the importance of establishing connections between learners, science knowledge, and science institutions; and deep dives, which concerned the development of deeper science knowledge. The themes and corresponding subthemes captured different elements and stages of the process of science learning, establishing different connections to our tools of interest. Findings can inform the development of science educational materials and activities that can help create bridges between science (e.g., institutions, schools, academia) and people.

Keywords: perceptions on science learning, scientific learning, science and people, non-formal education, narrative texts, museums, RAND framework

6.2. Introduction

As attested by widespread reports on student engagement and performance (e.g., National Center for Education Statistics, 2022; Organisation for Economic Cooperation and Development, 2019), learning natural and formal sciences (henceforth, *science*) is prone to difficulties. These difficulties start in school and continue throughout adult life (e.g., Field and Powell, 2001), and, although most people do not become proficient in the language and methods of science (e.g., Bruner, 1990), they still benefit from acquiring a derived or functional literacy to address personal needs or inform participation and decisions in discussions taking place in the wider sociocultural milieu (e.g., Norris and Phillips, 2003; Sadler, 2009). In fact, as knowledge production increases, so does the need of assessing scientific evidence in daily life (Albuquerque and El-Hani, 2021), making crucial to understand the public understanding and perceptions of science and devise means for lifelong science learning (e.g., Falk et al., 2007; Field and Powell, 2001). In the present paper we will be particularly interested in exploring perceptions on the role of specific types of texts and museums.

The role of non-formal educational settings in lifelong learning has been increasingly focused (e.g., Callanan et al., 2011; Falk and Dierking, 2010; Murmann and Avraamidou, 2014a), and museums are prime examples of learning settings that aim at reaching various and diverse people (e.g., Barata et al., 2017; Feinstein and Meshoulam, 2014). The fact that non-formal science learning takes place within a wider community than its formal counterpart, combined with the eminently social purpose of developing non-specialized scientific literacy, underline the importance of seeing science learning as a social and cultural process, besides an individual one (e.g., Callanan et al., 2011; Rennie et al., 2003).

Moreover, many people learn science through texts materials: They have been a cornerstone of formal science education (Kloser, 2013; Van den Broek, 2010), but also have an important role in more widespread science communication, which may include popular science books, posts in social media, and museum plates (Negretti, 2022; Ravelli, 2007). This role is further substantiated by claims for a greater integration of language and literacy processes in science education (e.g., Klein, 2006; Morais and Kolinsky, 2016; Snow, 2010; Soares et al., 2022a). Texts thus offer an important widow to understand science learning difficulties and develop solutions.

However, science texts can pose difficulties to science learning. The language of science texts is often seen as dense and technical (e.g., Snow, 2010), and readers find many of its ideas unfamiliar or even counter-intuitive (e.g., Gilbert et al., 2005). Learning materials and activities tend to be framed in authoritative and dogmatic discourses (e.g., Kloser, 2013), and are usually decontextualized from human, social, and cultural aspects (e.g., Sánchez Tapia et al., 2018). These problems apply to museums as well, not only because they also use texts, but because similar criticism has been more generally directed to them (e.g., Archer et al., 2016; Dawson, 2014; Kelsey and Dillon, 2016). Importantly, these learning challenges further reflect the multidimensionality of the process of learning science through reading (e.g., Soares et al., 2022a), captured in pedagogical frameworks such as the Reading for Understanding (RAND) framework (Snow, 2002). RAND endorses the view of text comprehension as a specific combination between features from the text, the reader, and the activity, elements which interact both within and with a wider sociocultural context (see Figure 6.1). This framework provides a very useful backbone of the conditions that should be considered when planning science learning interventions through texts, or analysing its outcomes.

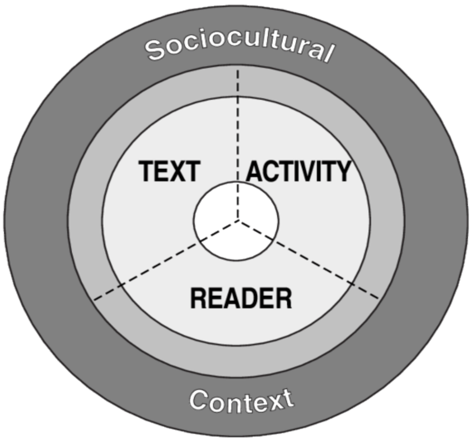


Figure 6.1. The RAND framework (taken from Snow, 2002)

The outlined challenges can also be traced to a gap between the scientific and the everyday thinking modes (Bruner, 1986; Klein, 2006). The scientific thinking mode is based on general and formal causes, proofs, and theories, and is usually conveyed using expository texts. These texts tend to be abstract and challenging, as contents are structured according to concepts and their relations (e.g., Graesser et al., 1991). By contrast, more familiar and pervasive in everyday life is the mode of thought based on human intentions and actions. This mode is better captured by the temporal structure of narration (e.g., Graesser et al., 1991), which is integral to every culture and has a privileged status in cognition (e.g., Bruner, 1986; Gottschall, 2013).

Fittingly, narrative-based tools are proposed to tackle this set of science learning challenges. Authors underline narrative's accessible language (e.g., Avraamidou and Osborne, 2009), ability to arouse interest (e.g., Arya and Maul, 2021) and enhanced contextualization (e.g., Kubli, 2005) between science contents and learners' personal experiences and daily situations (Murmann and Avraamidou, 2014b; Mutonyi, 2016), and between science contents and narrative contents such as human, social and cultural elements of scientific practices (e.g., Arya and Maul, 2021; Hong and Lin-Siegler, 2012). These pedagogical and engagement features have attracted the attention of science museums, that have successfully applied in exhibitions narrative-based methodologies not limited to texts (e.g., Frykman, 2009; Kerby et al., 2018; Murmann and Avraamidou, 2014b).

Importantly, a gap is also perceived in the social relationship between the general public and science: there is a gulf between the diversity of the general public and the public who visits museums (e.g., Feinstein and Meshoulam, 2014); surveys consistently reveal lack of understanding and also of trust in science within some parts of the population (e.g., Achterberg et al., 2017; Bensaude-Vincent, 2001); many public discussions perpetuate misconceptions and anti-scientific attitudes (e.g., climate change; A. Costa, 2021).

In short, given that learning science is a social and cultural activity, and that its importance goes beyond specialization ambitions, it becomes crucial to reach out to people to better understand their perceptions of the lifelong process of science learning and literacy. People seem to have an affinity for narrative texts, and non-formal settings use texts and have an affinity for innovative tools, so these are two tools relevant to explore. In turn, these insights may be used to adapt how science is communicated and taught, in ways that resonate with different people.

6.2.1. Aims

We set out to investigate general perceptions on how science learning and literacy occur, and more specifically on how people related to texts (narrative and expository) and museums. Our main goal was to understand how these tools were perceived in their own right, but we were also interested in finding out if people's perceptions were suggestive of combinations between them.

Our first aim was to connect participants to these tools. Namely, they were shown a video filmed at a science museum, took part in a non-formal learning activity facilitated by a science educator from that museum, and read (narrative and expository) science texts. We expected that these instances of our tools of interest helped activate perceptions on those tools, but also of the topic of science learning and literacy more generally. Our second aim was to tap into these perceptions in focus group discussions.

6.3. Methods

We ran a total of three focus group discussions, each with 6-7 participants (e.g., Gaskell, 2000; Krueger and Casey, 2009), which were preceded by a learning activity developed by the authors. The study was to take place in the science museum where two of authors work in the education department, but had to be adapted and run online due to the onset of Covid 19 and ensuing security measures. Despite this change, the learning activity was still facilitated by a science educator from the museum (second author). The sessions took place in Zoom in December 2020, and lasted around 2h (the full session protocol can be seen in Table C1, p. 237). The study was approved by the ethics committee of the institution from which three of the authors were from (agreement 48/20).

Table 6.1. Sociodemographic information about the participants of each focus group

Focus group	N	Gender	Age range	Field of study (level of contact with science)	Field of work (level of contact with science)	Experience in science education	Number of read science books
1	7	4 men	29–39	Total (2 people) Some (3 people)	Total (1 person) Some (4 people)	1 person (non-formal)	Most 1 per month at most
2	6	3 men	22–55	Total (3 people) Some (1 person)	Total (3 people) Some (1 person)	2 people (both non-formal and one formal as well)	Most 1 per month
3	6	4 women	19–48	Total (2 people) Some (1 person)	Total (2 people) Some (2 people)	1 person (non-formal)	All less than 1 per month

Note. Areas such as Physics or Informatics were considered as providing "total" contact with natural or formal sciences. Areas such as Psychology or Architecture were considered has providing "some" contact.

Participants were recruited through social media to participate in a study about non-formal science learning, and received two 10€ vouchers for their participation. Table 6.1 presents a global characterization of the participants from each focus group (more specific information can be found in Table C2, p. 238). Participants were not intended to be representative of the general population, but we encouraged people from different backgrounds to participate, aiming at including people from at least varying gender and age and different levels of contact with science. A total of 22 participants (10 women, 9 men, age range 19-55) took part in the study. Participants were from diverse specialization backgrounds (e.g., Design, Physics, Psychology, Tourism, Translation) and had varying science reading habits. Interestingly, albeit non intentionally, each focus group had at least one participant with experience as a formal and/or non-formal science educator.

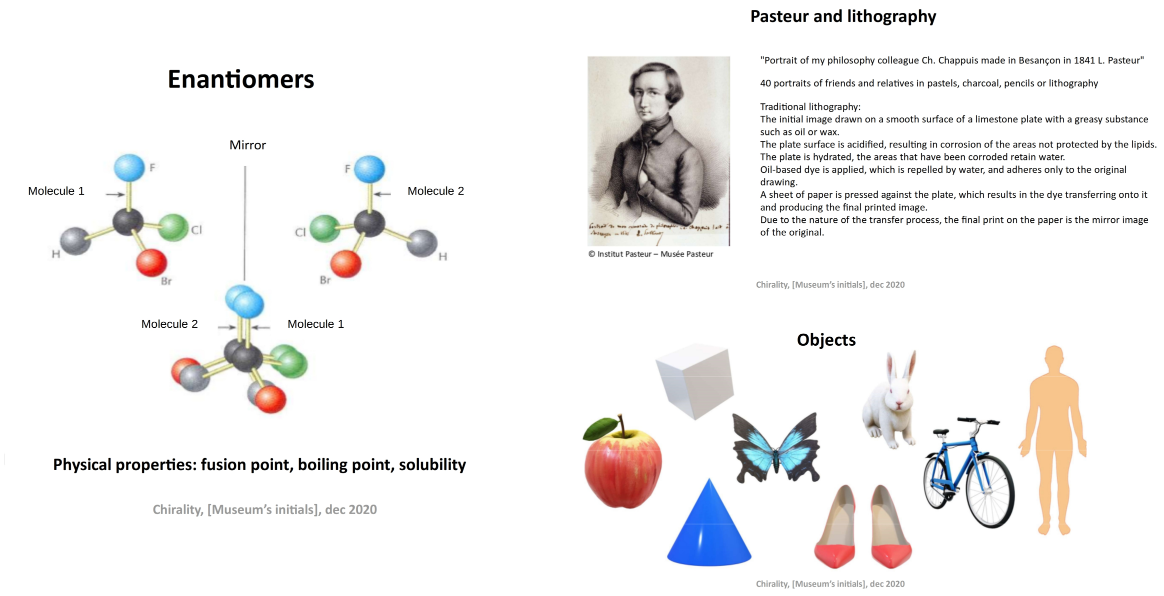
To illustrate different types of science texts, participants started by reading about the Chemistry concept of chirality through an expository text or a narrative text. The expository text only presented the concepts and its relations, whereas the narrative text detailed the process of discovery carried out by Louis Pasteur, describing science contents according to his thoughts and actions, in a temporal succession. Participants then answered six questions about the contents presented in the text. These questions required participants to think about the contents of the text in different ways (e.g., recalling information, making inferences among text contents or beyond), also in an attempt to more fully engage participants in the topic of science learning.

one point was given to each correct answer to this end, merely to ascertain whether participants paid attention to the texts contents and were able to understand its contents.

6.3.1. Learning activity

The texts had been previously developed to be used in studies that addressed factors and processes involved in science learning (Soares et al., 2022b; Soares et al., 2022c), and the topic was purposely selected to depict a science discovery that would be practically unknown to most people. The questions were based in the questions used in these previous studies, but were developed in collaboration with the second author to be more aligned with the learning activity of the current study. Both texts and questions can be found in sections C3 and C4 (p. 238-39).

After reading the text and answering the questions, a small video filmed for the study was shown. The video took participants on a virtual tour of the location where the study was to take place, namely the Chemistry laboratory of the museum and its historical equipment, accompanied by a voice-over explanation given by the second author, who works as a science educator there. The video aimed at further connecting participants with this non-formal setting and immersing them in the learning activity. Right after, the second author facilitated a non-formal learning activity, previously developed based on the texts' contents. The activity was presented in PowerPoint, and in it the second author provided further details and explanations on the concept of chirality. As can be seen by the examples in Figure 6.2, the activity this explanation combined different elements: exposition and explanation of science contents (2A), narrative details about Pasteur's personal and professional life (2B), and a more practical and interactive moment of interpreting examples (i.e., applying the concept of chirality; 2C). Participants were also told they could pose questions at any moment of the activity.



Pasteur and lithography

"Portrait of my philosophy colleague Ch. Chappuis made in Besançon in 1841 L. Pasteur"

40 portraits of friends and relatives in pastels, charcoal, pencils or lithography

Traditional lithography:
The initial image drawn on a smooth surface of a limestone plate with a greasy substance such as oil or wax.
The plate surface is acidified, resulting in corrosion of the areas not protected by the lipids.
The plate is hydrated, the areas that have been corroded retain water.
Oil-based dye is applied, which is repelled by water, and adheres only to the original drawing.
A sheet of paper is pressed against the plate, which results in the dye transferring onto it and producing the final printed image.
Due to the nature of the transfer process, the final print on the paper is the mirror image of the original.

© Institut Pasteur – Musée Pasteur

Chirality, [Museum's initials], dec 2020

Figure 6.2. Examples of the three different elements incorporated in the learning activity.

Note. 2A presents representations of enantiomers; 2B describes how Pasteur's drawing skills, namely in lithography, may have contributed to his discovery of chirality due to the mirror features both share; 2C shows the examples in which participants were encouraged to apply their knowledge on chirality.

6.3.2. Focus groups and thematic analysis

Focus group discussions, lasting around 1h, ensued, with the purpose of tapping into participants' perceptions on science learning and literacy, and the specific roles of different text types and museums. As can be seen in Table 6.2, our questions were organised around two axes.

The central topic of the discussion was how science learning and scientific literacy occur. The moderator (the first author) began by asking participants what they understood by these concepts, as to encourage them to apply and discuss their own definitions. The discussion progressively moved on to more specific aspects of science learning and literacy, such as barriers, facilitators, and the role of different tools, including text types and museums. The elements from the preceding learning activity, which also had a "prompting" role, were used as reference by the moderator to help introduce or illustrate the questions. Participants were also given the opportunity to read the other text (i.e., the expository or narrative, depending on the one read at the beginning of the activity). The moderator provided or asked for clarifications and examples whenever needed, but mostly let the discussion flow naturally, as long as it remained on topic. Although the goal was for the dialogue to flow organically, being an online discussion (e.g., Halliday et al., 2022) the availability of visual and bodily cues was reduced, and so participants were asked to "sign-up" on the chat before speaking. If a participant would consecutively sign-up, while others would do it less often, the moderator would intervene and ask whether the latter would like to share something before the more loquacious participant intervened again.

Recordings were transcribed and the transcripts were coded and analysed using thematic analysis (Braun and Clarke, 2006). The purpose of thematic analysis is to identify recurrent themes, which are sometimes nuanced and should be judged on the basis of salience/relevance, instead of frequency.

Following familiarisation with the data, codes were identified by the first author through flexible and iterative engagement. These codes were phrases or extracts describing ideas, experiences and feelings of the participants in connection to the research questions (e.g., "what should be there from the beginning", "the importance of seeing connections with the real world"). Attention was also given to the use of personal examples, emotional responses, and implicit beliefs. Although codes largely incorporated semantic aspects (i.e., surface meanings), latent aspects of the data (i.e., underlying concepts and ideas) were also included. Codes were then grouped into relevant themes that expressed clear patterns and discussed with one of the other authors. Subthemes captured distinct aspects of a theme, while sharing the same central organising concept. For example, "show some connections" and "see things happening" capture specific aspects of establishing connections, the organising concept of the theme weaving webs. Before being defined, however, themes and subthemes were reviewed and refined, to ensure they were related to the central concept of the theme and had enough support, relevance, and distinctiveness in the data. Despite their distinctiveness, care was taken to ensure that the different themes related to each other, as well as with the research questions (Braun and Clarke, 2006). A researcher who was not an author of the study reviewed the themes for consistency, both with the central organising concept of the theme and the research questions. Disagreements between researchers were always resolved through discussion. Short definitions for each theme and subtheme were then written, and illustrative extracts selected.

Although we were particularly interested on the role of narrative texts and museums, which was patent in the focus group script (see Table 6.2), we were also interested in the science learning and literacy processes more generally, and it was not possible, nor desirable, to limit the free flowing discussion to specific tools. As such, if references to other materials or contexts were more salient and better captured the central concept of a theme, they were selected as illustrative examples. Finally, data was weaved together with analysis and connections to the adopted pedagogical framework and previous relevant literature on the topic of science learning.

Table 6.2. Script used in the focus group discussions

Axe	Questions
Science learning and literacy (beliefs, practices, preferences, previous experiences)	What is science learning/scientific literacy? How do you image science learning/scientific literacy unfolding? Is it important to learn science/develop science literacy? And why/in which life dimensions? What kind of barriers to science learning do you perceive? And what could help improving these processes?
The role of specific tools (with a focus on different types of texts and museums)	What are the best ways to learn science? What is the importance you attribute to texts? How are the science learning texts you tend to find? Do you think there is place for different types of text (built around the contrast of expository and narrative texts)? What kind of role do science museums play?

6.4. Results

Regardless of the text they read, participants answered most learning questions, with no marked advantage for either text type. Moreover, the fact that some participants had a high level of experience with science, whereas others “ran away” from it, did not prevent the latter from sharing their thoughts and feelings during the focus group, nor did it create noticeable tension or discomfort. Instead, the focus group discussions were engaging and diversified, with people organically taking part in the discussion and interacting with each another. Participants also used the texts they read or the ensuing learning activity to illustrative specific ideas or opinions.

Three themes were identified through thematic analysis. These were smooth the way, weaving webs (both with three subthemes; see Table 6.3) and deeper dives, which will be presented in the following subsections.

Table 6.3. Themes and subthemes from the thematic analysis

Themes	Subthemes
Smooth the way	“it should arouse my curiosity” “give smaller doses” “I needed something more visual”
Weaving webs	“show some connections” “see things happening” “there is, in fact, a huge gap”
Deeper dives	

6.4.1. Smooth the way

The first theme concerned conditions and features that can act as facilitators of the learning process in initial and less specialized stages. They seemed to be regarded as key if any further learning is to occur.

“it should arouse my curiosity”. The importance of hooking learners’ curiosity and interest was a recurring theme among participants when discussing the meaning and process of learning science. Similarly, the lack of such factors was regarded as a barrier.

> Often, the difficulty of the process of teaching/learning [science] is to arouse interest.
> (P8, science school teacher and informal facilitator)

Narrative texts were consistently pointed out as very efficient in capturing readers’ interest in initial learning stages. However, a contrast between reading narrative texts and deeply processing or absorbing information would also sometimes surface.

> Up to adulthood, the narrative context is ideal, because what’s important is not to absorb the greatest amount of information possible, but to attract people to that

> knowledge [science] and that they want to explore further from there, and that it is not a barrier to begin with (P1, does research in Cognitive Science)

The texts used were often referred to, and the superiority of the narrative text over the expository text, in terms of interest and engagement, was highlighted both by people who are not comfortable with science and people who are:

> The first text I got was super hard. [laughs]. (…) This second text, the narration, I mean I think it… immediately transports (P16, works in tourism)

> Although the texts have the same contents, I feel that the first, this is at a more personal level, for a first approach I would become interested (P18, works as a facilitator in informal science learning contexts)

"give smaller doses". The need of rendering science contents intelligible was also systematically stressed. It was characterized by the use of simple language and few concepts when first approaching, and getting a general view of, science topics.

> For instance, Einstein's theory for 10-year olds. That can totally happen, but what is the language being used to (…) write the relativity equation to them? [laughs] (P12, visual artist and former geography teacher).

> Much of the science communication is made for other scientists (…) and that is elitist and a type of language that is complicated to break in smaller pieces if you never head it before (P5, does research in Cognitive Science)

The advantages of using narrative texts at this stage, as compared to expository texts, were again highlighted, even when participants were quite familiar with a scientific field (albeit not necessarily with the specific contents of the texts).

> The narrative text (…) was like a soft introduction, it was like more hand in hand. (…) the first [expository] text (…) is more technical, I would be a bit blocked because... it's too direct, too cold, too many concepts. (P15, designer and animator)

> I received the more formal (…) more expository, and it was in fact quite dense (…) and so more difficult to understand, yes. (P9, does research in Physics)

"I needed something more visual". The importance of using visual tools and promoting visual strategies was also recurrently shared. The narrative text was considered to help visualize science contents.

> I got the more expository version (…) and was having some trouble understanding without drawing a schema (…) Maybe a text with a story would be… easier to understand, as it helps creating images (P6, works in Science Managing)
>
> If I had to study I think I would start by drawing a schema next to it [expository text], you know? To visualize and get the idea, as it is more difficult that wit the narrative text (P18, works as a facilitator in informal science learning contexts)

Importantly, museums were seen as places which provide that visual component, and overall as being able to make knowledge more accessible to people in general.

> I museums play that part of introducing people in an easier way, and so it does that translation (…) you have the experience, the visual, a person explaining in a very didactic manner (P4, works as an architect)

6.4.2. Weaving webs

The second theme was built around the idea that establishing connections is crucial for science learning to be more fully grasped and put to use. Different kinds of connections between learners, science knowledge, and science institutions, were captured by the different subthemes. It was when participants were establishing this sort of connections that personal experiences and memories more often emerged, accompanied by displays of emotion.

"show some connections". The importance of connecting scientific knowledge to other kinds of knowledge was underlined, notably of contextualizing and situating science learning within broader, familiar or important, frames. Contextualizations taking place in the classroom setting were the more pervasive examples.

> My Math teacher from the 10th grade, C., in the first class he taught us, the first thing he did was to explain Einstein's relativity theory (…) but he did it to introduce the fact that

> Math is the language through which we understand the universe. (…) it was like contextualizing it in an important...context (P14, does research in Psychology).

In contrast, not being able to establish this context was framed as a frustrating experience.

> I thought of myself as a rather interested person (...) I would always try to (…) connect them [sciences] to my daily life (…) and for me that was a barrier, a reason for running away from the sciences. (P16, works in tourism)

“see things happening”. Besides seeing science contents in context (i.e., seeing what is not there) participants mentioned the importance of seeing their practical and applied side (i.e., get a better grasp of what is there). On the one hand, this included learning by doing, that is, of seeing things happening and doing them first-hand.

> Examples like Chemistry I. I need to see things, I’m very visual, so I have to (…) see things reacting with one another, and then go back and test it again myself. (P4, works as an architect)

Additionally, seeing applications and the practical side of science knowledge was seen as helping uncover its usefulness. Importantly, the contents of the texts participants read, and the contents of museums, were evoked as examples.

> I’m thinking back to the exercise we just did about (….) Chirality. And I thought it was interesting (…) but at this moment it is not useful to me at all. But with the example of masks, on the other hand… (P10, works as a translator)

> One of the great obstacles (…) is the vision that people in general increasingly have of museums. That they only have ancient things exposed, that are not relevant to anyone, that are not useful, etc. (P7, works in tourism).

“there is, in fact, a huge gap”. Yet another relevant form of connection is established at a more societal level, which is perfectly crystallised in the following comment:

> Truth is there is, in fact, a huge gap, between school, and not only academia, but society at large. (P1, does research in Cognitive Science)

The gaps were diversified, but consistently pointed towards communication and connection problems between science and people. Given this diversity and the importance of this subtheme, as it connects different social levels (i.e., museums, schools, groups of people), we will include a few more examples. For instance, the lack of adequate museum communication and outreach campaigns were pointed out, both by people who work in that setting and people who do not.

It's great that we go to museums in school fieldtrips but then it seems like museums do not invest in attracting adult public. (P18, works as a facilitator in informal science learning contexts)

And so museums. I think that maybe they will have to surrender to (…) more digital approaches… I don't know if everyone has the same ability to visit a museum in person, even outside the context of the pandemic, even as an institution in general (P14, studied and does research in Psychology)

Moreover, the COVID-19 pandemic was a recurring example of polarization: Of how these gaps become apparent, but also how a social crisis can bridge them:

> A problem with politics, namely populism, is (…) deconstructing simple arguments (…) to me that has a scientific method component. And it should be way more taught and education is not fulfilling that role. (P1, does research in Cognitive Science)

> We see people turning away from… science (…) what is also curious is seeing how during the pandemic people seemed to somehow reconnect to science (P9, studies and does research in Physics)

6.4.3. Deeper dives

The third and final theme concerns the development of deeper scientific knowledge. The adverb *afterwards* (or after, or then) is pervasive, usually following an initial understanding or building of interest. It was related to the reading of more specialized expository texts, believed by most participants to be better for conveying specific and rigorous information than narrative texts.

> If we aim at (…) something more technical, more specialized, I think that it would make way more sense [to read] more concise texts (…) so that afterwards you are able to (…) speak a language, right? (P3, does research in Modern History)

Although most participants saw narrative texts as less appropriate to this end, it was not completely discarded as an effective learning tool.

> I think that a narrative can give the impression that concepts are not as well learned, and yet they are way better learned that with other [texts]. Because… these is less confusion. That is, our memory is narrative, isn't it? (P1, does research in Cognitive Science)

An additional idea was that of the importance of gathering knowledge from recognised knowledge authorities. Museums were regarded having a role to play as part of the latter.

> Everybody is certain, and has a theory as to why there are fires (…) And it is worrying that they don't even (...) ask for the opinion of someone who actually knows about that matter (P11, works in the field of ecology)

> Maybe the role that museums could have, in that sense, is not so much of creation and being one more person... in being one more organism in the cake, but maybe, more of a role of digital curatorship. (P9, does research in Physics)

6.5. Discussion

The current study had the goal of examining perceptions on how science learning and literacy occur, shared in the context of a non-formal learning activity. In general terms, the process of science learning was depicted as being made up of different elements and stages, which established specific relations to different text types and museums. In addition, the RAND model offered a useful pedagogical framework in which to frame our interpretations and connect the ideas expressed by the participants with the different components of the learning process.

Findings can inform the development of science educational materials and activities related to our tools of interest that consider learner's needs, while also framing these needs within wider contextual and social aspects. We will discuss the results in connection with such insights.

The present findings bring further evidence on the sort of difficulties that can arise during science learning, and provide information of the kind of elements and approaches that educators and educational settings can use to provide more meaningful learning experiences. Participants made clear that interest and intelligibility of the learning materials should be prioritized over exhaustiveness and complexity of content, and that visual tools and strategies should be further used and stimulated. Importantly, regardless of participants' level of contact with science, narrative texts were regarded as a useful tool to tackle these issues of interest, readability, and visualization of science contents. There is indeed evidence that science texts can arouse interest (e.g., Arya and Maul, 2021; Hadzigeorgiou et al., 2012) and be particularly effective among low prior knowledge or struggling learners (e.g., Hong and Lin-Siegler, 2012), and they are also claimed to promote visualization and mental imagery further (e.g., Akarsu et al., 2015; Prins et al., 2017). What is more, these perceptions echo previous ones from researchers, teachers and students (e.g., Murmann and Avraamidou, 2014b; Prins et al., 2017; Ritchie et al., 2011); the fact that they were shared in the context of a non-formal activity strengthens these claims.

Participants' perceptions of narrative texts are consistent with the common characterization of narratives as "door openers" in initial learning stages (e.g., Kubli, 2005; Gilbert et al., 2005). These perceptions were contrasted with striving for more than a global overview of contents, contended as usually occurring *after* interest and/or a global idea of the topic have been captured, which emphasises the importance of meeting these conditions beforehand. Consistent with its official status in specialized science communication (e.g., Avraamidou and Osborne, 2009), expository texts were regarded by most participants as appropriate tools for deeper learning. Still, narrative texts were not completely discarded, echoing claims that they should be used with broader audiences (e.g., Olson, 2015). These results also reflect the fact that narrative texts and expository texts have different purposes (e.g., entertain vs instruct), which are associated to different social contexts (e.g., Adam, 1997; Soares et al., 2022a) and may therefore activate different reading strategies (e.g., Hidi et al., 1982). Despite their contrasting nature, as regards the RAND framework, all these ideas mainly concerned interactions between learners (interest, prior knowledge), texts (linguistic and conceptual density; text structure), and learning activities (learning goals, namely general vs deep learning). As such, suggestions concerning narrative and expository texts were more strongly associated to individual learning strategies, and thus to the bridging of individual gaps between people and science.

Additional suggestions for making science learning more meaningful included situating science contents into a wider context and connecting it with other kinds of knowledge, as well as showing the applied side of science contents and promoting hands-on-activities. These proposals match well-established frameworks such as situated (Sadler, 2009) and inquiry-based (Martins-Loução et al., 2020) science learning, and hint at the notion that disengagement with science learning is related to its wider decontextualization (e.g., Gilbert et al., 2005; Sánchez Tapia et al., 2018). Additionally, the specific claim that the usefulness of science contents should be made clear nods at the notion that scientific truth should fulfil a social function. This idea has a long history (Bernal, 1939), and has been at the centre of recent debates on science and society, such as science denialism and vaccination (e.g., Allchin, 2022; A. Costa, 2021). Fittingly, many ideas on the importance of the practical and useful side of science were imbued with references to the pandemic, highlighting the connections between the lack of communication between science and society and major social upheavals. The underlying perception of a gap between everyday and scientific thinking and matters (Bruner, 1986; Klein, 2006) thus extended to a wider social level (e.g., Bensaude-Vincent, 2001; Feinstein and Meshoulam, 2014; A. Costa, 2021). Although they relate to different social levels, they can impact one another: Perceptions of a gap at the personal level can contribute to enlarge the gap at the wider societal level, creating a self-feeding cycle of disengagement and lack of a common basis of communication and understanding.

Yet, there was also the perception that some people "returned to science" during the pandemic, which is further strengthened by the fact that participants established these connections during the discussion. This polarization in people's response to science during the pandemic has been described in the literature (Krause et al., 2019; Reif and Guenther, 2021). In terms of the RAND framework, these ideas were less connected to texts, instead emphasising interactions between learners, learning activities and a wider social context, such as schools, museums, research and academia). Of note, these ideas were overall more filled with personal examples, memories, and emotional displays than the other two themes, suggesting that establishing connections with daily life and social issues resonated with participants at a more personal level. Science learning seemed to become more meaningful when put in a wider, familiar and relevant, context (e.g., Kubli, 2005; Mutonyi, 2016).

Importantly, it was when the need of connections between science and people was being shared that museums were more consistently mentioned. The fact that museum's outreach and inclusiveness were put into question (echoing long-standing concerns, e.g., Feinstein and Meshoulam, 2014; Postolache et al., 2022), combined with references to museums when the need for recognised science knowledge authorities was discussed, strongly suggest that museums were seen as important communication interfaces. Nonetheless, museums should further focus on expanding their reach and becoming more inclusive. This includes improving and modernising their communication and exhibition displays, both of which involve digitalisation to some extent. Thus, by strengthening the described connections in their exhibits and communication, museums can play a pivotal role in rendering science learning more meaningful for broader and more diverse audiences, and thus establish larger scale connections between science and people. In addition, by making more use of narrative texts (or other mediums, given the expressed need for increased visual tools and modernisation), museums have the potential to simultaneously bridge gaps at individual and social levels. In fact, ideas concerning these two levels (e.g., the need for accessible language and of seeing the applied side of science, respectively) align with the proposal that that learning a new skill becomes easier when language and daily life are connected (Freire, 1970; Sadler, 2009), and depict challenges to the pursue of deeper science learning that can feed each other off (e.g., using daily examples can be a way of capturing initial interest, and the lack of an accessible language prevents people from contextualizing science knowledge).

Although this may be easier to accomplish in non-formal settings such as museums (Callanan et al., 2011), school education can also benefit form a broader use of these tools, especially when first presenting students to specific topics. As they may prove of inspiration for educators from varied settings, we added to section C5 (p. 240) a set of more specific ideas and suggestions on how to improve science communication and learning shared by participants.

This set of findings further highlights the benefits of examining groups of people with different points of view (e.g., Knudsen et al., 2015). We were able to tap into the experiences of people who have always struggled with science learning and people who have built a path on a science field, even as educators themselves. Unfortunately, the discussions did not include people who see science as wholly uninteresting or irrelevant. Hearing the concerns and perceptions of these groups of people is crucial to bridge the outlined gaps, but can also prove more difficult to achieve. People will be less likely to participate in this sort of studies, as they focus precisely on what does not interest them (science), or may be very resistant and reactive if they do participate. At any rate, such opportunities for sharing points of view and collective discussion can be very useful both in research and educational contexts (e.g., Jerónimo and Reis, 2016; King et al., 2021; Solomon, 1992). In turn, collecting rich and diversified input can be a very useful complement to empirical studies on learning and, most importantly, to strengthen the connections between theory, practice and policy.

6.6. Conclusion

Even outside specialization ambitions, scientific information permeates several aspects of our personal and social life, making it beneficial to develop some level of scientific literacy. By examining the perceptions of people with different levels of contact with science, our findings underline the kind of difficulties that can be felt, but also the ways in which these difficulties can be counteracted, so that more effective materials and practices can be developed. Narrative materials can help reduce the gap between learner's everyday thinking mode and the scientific thinking mode, by providing an interesting, accessible, and more visible bridge. Moreover, contextualising science and revealing its pervasiveness in daily life makes science more relevant and meaningful for learners throughout their lives. Museums can provide this crucial bridge between science and people, while also taking advantage of the more individual narrative bridge. Yet, they must insure their communication reaches a wide and inclusive audience. Connecting people's perceptions with theoretical ideas and empirical results provides a much needed bridge between science (e.g., institutions, schools, academia) and people, and is a step towards reducing gaps and making science learning more inclusive and pervasive.

CHAPTER 7

General discussion

The proposal of using narrative texts as a science education tool consists in using a highly pervasive tool to make sense of the also pervasive, yet often alien, presence of scientific concepts and developments in daily life. However important, science learning faces a variety of challenges (e.g., unfamiliarity and counterintuitiveness of science ideas; human and cultural decontextualization of science materials), and narrative texts can address many of these challenges through a literacy-based approach. Yet, there is still much that we do not know about the outcomes they produce, and about the learning process leading to such outcomes.

Although previous research has reported encouraging results for the use of narrative texts as science education tools (e.g., Arya & Maul, 2012; Hadzigeorgiou et al., 2012), it has also produced mixed results (e.g., Jetton, 1994; McQuiggan et al. 2008), which have not been satisfactorily interpreted and explained. In addition, despite narrative's conceptual association to a set of processes, little is known about the actual processes that are engaged when learning science from narrative texts, and thus on how the processes triggered by narrative texts impact science learning. Specifically, little is known about the features of the learner, text, activity and wider context impacting, modulating, or even restricting the occurrence of a narrative learning effect. Moreover, a theoretical dialogue with literacy aspects at its centre seems to be missing, even though narrative texts are a literacy-mediated science learning tool. Such a theoretical backbone could prove key for making sense of previous results, as well as designing new research and educational interventions.

In the present work, we set out to investigate the potentially symbiotic relationship between science learning and narratives, returning to a long-standing idea (i.e., the use of narrative texts to teach science) with a fresh look and new ideas. Namely, our research was rooted on the premise that the multidisciplinarity inherent to this topic should be embraced and had the overarching aim of building a more complete picture by focusing on three dimensions: theoretical backbone, learning outcomes, and learning process. Our first research question (RQ 1) tapped into the first of these dimensions (theoretical backbone), asking which theoretical frameworks would prove more useful to investigate the issue of learning science from narrative texts. In Chapter 2, we conducted a theoretical review to answer it and used the resulting theoretically-grounded mapping when addressing the following research questions. Our second research question (RQ 2) concerned learning outcomes (the second of the outlined dimensions), and touched on various aspects. We started by analysing, in previous studies, the consistency of these learning outcomes at different educational levels, by means of a theoretical review that drew on the proposed theoretical backbone (Chapter 2). We then conducted empirical behavioural studies with young adults with low levels of prior relevant science knowledge that assessed learning science from narrative texts at different levels of comprehension (Chapters 4 and 5). In Chapter 4, we compared this learning with the one obtained by reading expository texts with equivalent scientific contents. The used learning materials (texts and corresponding learning items) were developed by us beforehand (Chapter 3).

Finally, our third research question (RQ 3) asked how the process of learning science through narrative texts occurs. This final question concerned the last outlined dimension, the learning process, and was firstly addressed in the theoretical review conducted in Chapter 2, in which we gathered and linked ideas and evidence regarding the conditions and underlying mechanisms involved in learning from science narrative texts, based on our proposed theoretical backbone. Furthermore, the third question addressed three facets of the learning process. In Chapter 5 we investigated the influence of human-related elements in science learning at different levels of comprehension (RQ 3a). More specifically, we examined the attention dedicated to human depictions in science narrative texts by means of eye tracking, which, to our knowledge, is unprecedented. We also analysed the human-related thoughts triggered by the reading of the science narrative text, using tasks we developed for this purpose, as well as more general human-related thoughts, assessed through a task that had not been previously used to in this body of literature (i.e., learning from science narrative texts). In addition, we analysed the impact on science learning of set of features from the learners (RQ 3b), through independent tasks (i.e., not pertaining to the reading of the texts; Chapters 4 and 5). In Chapter 6, we investigated people's perceptions on science learning and literacy, and particularly on the role played by narrative texts and museums (RQ 3c). A qualitative study, combining focus-group and thematic analysis, was conducted.

Overall, the findings reported in the present dissertation successfully addressed the research questions and objectives, shedding light on theoretical aspects, as well as on learning outcomes and the process leading to such outcomes. Regarding theoretical aspects, in Chapter 2 we proposed a literacy-focused dialogue between ideas from text linguistics, cognitive psychology and pedagogy, which can help make sense of previous results while also contributing for future conceptualization, research and educational interventions. As for learning outcomes, the findings from Chapter 2 help determine the extent of a generalized "narrative effect" in previous studies. The results from Chapters 4 and 5 provide evidence on whether young adults can benefit from learning science from narrative texts; Chapter 4, in specific, delivers further evidence on whether this benefit extends to different science topics, levels of comprehension, and retention intervals. As regards the learning process, these two Chapters also report evidence on the kind of more general learner's features contributing to learning science from narrative texts (and from expository texts as well, in Chapter 4). Moreover, the findings from Chapter 5 highlight the kind of human-related elements that influence learning from science narrative texts. These include elements depicted or triggered by the science narrative text itself, or pertaining to more general, text-unrelated, processes. In Chapter 6, we revealed the kind of difficulties people feel during lifelong science learning, and how tools such as narrative texts and museums can help ameliorate these challenges. Still regarding the learning process, Chapter 2 delivered a mapping of theoretically-grounded conditions and processes that are involved when learning from science narrative texts. Lastly, a supporting contribution to these major contributions is found in Chapter 3. The information we shared information, on the process of developing science texts and corresponding learning items, may be useful for educators and researchers interested in using narrative texts as science learning tools.

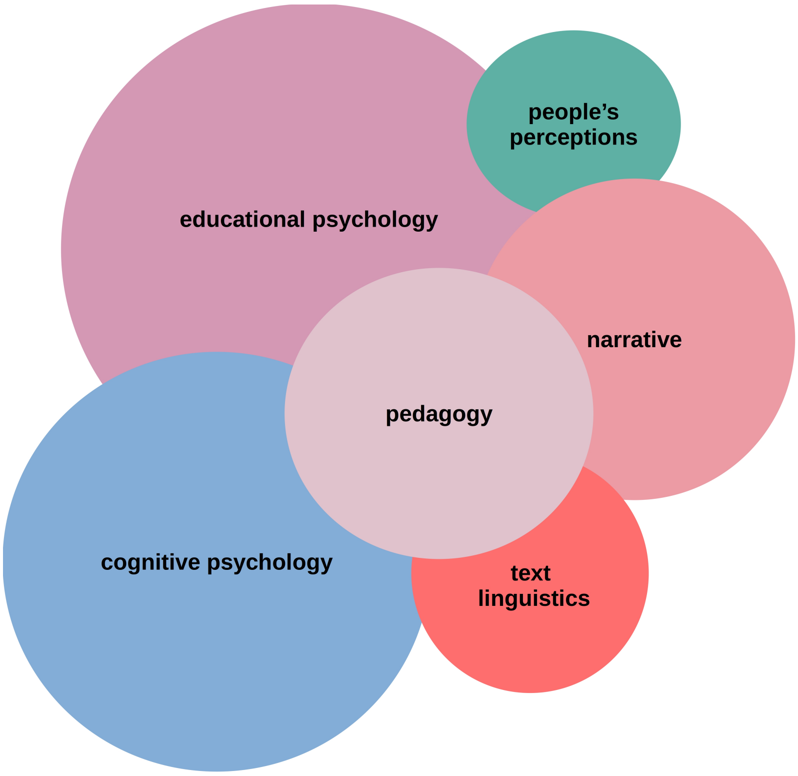


Figure 7.1. The areas the present dissertation was built on

In this seventh, and last, chapter, we will summarize the main findings and implications of our studies, discuss the major theoretical and applied contributions of this work, and point out its overall limitations. We will also identify aspects that warrant further research while discussing these three aspects. As a final remark, it is important to point out that this dissertation was built on a multidisciplinary effort, as illustrated in Figure 7.1.

7.1. Summary of main findings and implications

RQ 1: Theoretical backbone – *Better understanding the issue of learning science from narrative texts: Which theoretical frameworks would be useful?*

We addressed the question of which theoretical frameworks would be particularly useful to draw from by means of a theoretical review (Chapter 2). In this review, we adopted a literacy-focused approach, highlighting the importance of tackling challenging aspects of science learning by improving language and literacy processes. Given that texts, particularly narrative texts, were our tools of interest, we based our proposal on the relevance of considering texts' features (e.g., Adam, 1997), the cognitive representations and corresponding processes by which texts are processed by readers during learning activities (e.g., Kintsch, 1998), and the interactions of these elements with each other within and with a wider context during the learning process (e.g., Adam, 1997; Snow, 2002). The conceptualization of text features came from text linguistics, which conceives texts as social objects made up of textual and contextual features (Gonçalves, 2019). Knowledge on cognitive representation and the processes underlying comprehension came from the Construction-Integration model (e.g., Kintsch, 1988, 1998), which posits that readers construct and integrate information into representations through an interactive interplay of processes, and has been the predominant paradigm when conceptualizing basic processes and pedagogical practices for comprehension (e.g., Pearson & Cervetti, 2015). Finally, the RAND framework gave information on the elements and interactions that make up the process of learning from text. Importantly, this framework recognises the multifaceted nature of learning and aims at providing guidelines for education research and development.

Our proposal is the first to offer a multidisciplinary dialogue to tackle the topic of learning science through narrative texts. In general, a framework that goes beyond ideas on the concept of narrative was missing. Founding ideas on narrative are very important and should be present, but they are not enough for more directed and theoretically-guided research and educational practices to be developed. Although some theoretical works and empirical studies have drew on features from narrative, these were usually mostly focused on textual features (Avraamidou & Osborne, 2009, Norris et al., 2005, Prins et al., 2017), failing to more deeply considering how pragmatic and textual features interact within a wider sociocultural context, and the consequences this holds for using science narrative texts. Furthermore, although many works adopted constructivist pedagogical frameworks (Hadzigeorgiou et al., 2012; Klassen, 2007; Kokkotas et al., 2010), and others drew on cognitive frameworks of text processing such as the Construction-Integration model (Arya & Maul, 2012; Wolfe & Mienko, 2007; Wolfe & Woodwyk, 2010), none had combined input from all the different fronts we propose.

First and foremost, the combined insights from these different frameworks had important implications for the development of the dissertation project. The proposal outlined by the Reading for Understanding (RAND) framework provided the pedagogical backbone from which we framed the two questions addressed in the theoretical review and in the following studies. This framework helped us recognise the multifaceted and complex nature of learning through text, made up of features from the text, the reader, the activity, and the wider sociocultural context, and use it to gain a deeper understanding of the results from our own studies, and the results from previous literature. One characteristic that stands out is the fact that most former studies focused on just one or two of these elements. While there is nothing wrong with paying particular attention to one element, the fact that the other elements might impact the results should not be forgotten, even if those elements are not under direct examination. Importantly, this understanding was crucial for developing our idea that it is more sensible to envision several specific narrative effects, instead of a generalized and singular narrative effect.

The view of texts put forward by text linguistics offered us a better grasp of the features that make up science narrative texts. The notion of texts as highly heterogeneous objects was particularly important, as was the distinction between sequences (focused on textual features) and genres (encompassing wider sociodiscursive practices). These concepts are often used interchangeably in the literature on using narrative texts in science education, which tends to obscure the role that social practices and expectations play in text processing (i.e., genre), such as expected contents and purposes, and focus solely on recognisable textual features (i.e., sequences). This knowledge influenced the development of the texts (Chapter 3) that we used in following studies (Chapters 4-6), but especially our expectations and interpretation of the results, while making us aware of the limits of our approach. Namely, that the narrative and expository texts we developed in compliance with highly controlled parameters, are fundamentally different from the ones usually found outside the laboratory, which are more heterogeneous, and are therefore more aptly described as sequences (i.e., highly homogeneous pieces of text).

Finally, the Construction-Integration model gave us a solid and comprehensive cognitive framework from which we developed and evaluated the more cognitive features of this project. Namely, it was crucial for developing learning items that tapped into different levels of comprehension, providing us with a more complete and robust view of the underlying representations and processes of the overarching educational outcomes "memory" and "learning". In addition, the notion that several processes influence the building of cognitive representations (and thus, of learning) equipped us with a framework for investigating the processes that underlie the learning of science through narrative texts, instead of examining them only as outcomes. That is, it helped us retrieve the "why" of the learning outcomes afforded by science narrative texts, an issue that has seldomly been addressed in the literature.

This dialogue between theoretical frameworks was then used to guide our examination of previous literature, namely through the questions *If* (i.e., focused on the learning outcomes) and *How* (i.e., focused on the learning process), as well as the subsequent studies we conducted to empirically address these questions. Although we described the implications of our findings in relation with the present dissertation, similar implications can, of course, be extended to research and education in general.

In sum, we believe that this multidisciplinary lens, which places literacy at its core, is highly relevant, as texts are a cornerstone of science learning and of many of the challenges to science learning.

RQ 2: Learning Outcomes – *Can people, in particular young adults with little prior science knowledge, learn science from narrative texts at different levels of comprehension, and even better than from expository texts?*

The question pertaining to the learning outcomes afforded by science narrative texts was addressed in Chapters 2, 4 and 5. More specifically, each Chapter used different approaches and touched on different aspects included in the question (e.g., whether people in general can benefit from narrative texts as a science learning tool; whether narrative texts benefit different levels of comprehension; comparing the outcomes from narrative texts and expository texts).

In Chapter 2, by means of a theoretical review, we examined previous evidence on memory and learning outcomes stemming from science narrative texts at different levels of education. This analysis revealed encouraging results for the use of science narrative texts, both when narrative texts were used as a stand-alone tool, and when they were being compared with expository texts. The results further showed that science narrative texts can be useful learning tools at different educational levels, including young adults, although there were not many studies with this population. This is an important finding, given the lingering pragmatic association between narrative texts and younger students. Yet, results were suggestive that students with lower levels of prior knowledge may reap more benefits from this learning tool, which preserves the association between narrative texts and early states of learning, regardless of age. Results further indicated that memory and learning benefits of science narrative texts may more noticeable in the long term.

In Chapter 4 and 5, we conducted empirical studies with young adults with low levels of prior knowledge of the science topics presented in the texts, using linear mixed model analysis. We observed that young adults in initial stages of learning did indeed learn science from the narrative texts they read. However, in Chapter 4, where the durability of the memory and learning results was tested, participants' performance actually declined after one week, in contrast to what was mostly observed in the analysis conducted in Chapter 2. Additionally, and importantly, learning was observed at different levels of comprehension: Namely, in the recall of information mentioned in the text, inferences between different parts of the text, comprehension of key ideas, and application of key ideas in a novel situation. This finding is informative because, even though this question is of great relevance for educators and researchers alike, few previous studies have reported learning results at different levels of comprehension, and so it remains unclear which, if any, level benefits particularly from narrative texts. By drawing on a cognitive model of text comprehension, we were able to distinguish the impact of science narrative texts on different levels of comprehension, hence shedding light on the cognitive operations underlying the observed outcomes. This information can then be used to help understand why specific learning results can be obtained, or why they fall short of expectation.

Moreover, the average results were medium in the learning items tapping into levels of comprehension requiring less elaboration, and lower in the learning items requiring more elaboration. It is possible that the materials were somewhat challenging, and it is important to keep in mind that it was the first time participants read the described contents, most of them reading the text only once before answering the questions. When the results from both math and chemistry were considered, in Chapter 4, scores followed a downward tendency; yet, in both Chapters 4 and 5, chemistry's learning results were higher in questions requiring inference between different parts of the text, signalling that different science topics can yield different learning results.

Crucially, in Chapter 4 this difference between science topics extended to the effect of text type. Indeed, the analysis from Chapter 2 had already revealed that narrative texts did not always provide an advantage to expository texts, being sometimes indifferentiable, and other times lower. In Chapter 4, the results produced by the narrative and the expository maths texts did not differ significantly, which may have owned itself to higher familiarity with the contents of the maths texts, as compared with the chemistry texts. On the other hand, a narrative advantage was observed in the chemistry topic, extending to all assessed levels of comprehension, but particularly in one of the more elaborate comprehension levels. Besides being rated as less familiar, these texts were further considered to be more difficult and less interesting, highlighting the importance of examining other features and processes that intervene in the learning processes. Most importantly, this finding adds to the idea, already put forward in Chapter 2 and by other authors, that learning some science topics might benefit particularly from the use of science narrative texts. Thus, to our understanding, based on the present findings and the theoretical background we proposed, the notion of a generalized narrative effect does not seem to hold, even when just two topics are being compared. Despite this, it is worth nothing that the memory and learning outcomes observed in Chapters 4 and 5, whether favouring narrative texts or not, proved to be robust, as they remained significant even in the presence of other variables that also contributed significantly to the learning results.

One of the main implications of these findings is that the potentialities of narrative texts as learning tools are not restricted to younger children and to school contexts: Instead, they can be used to teach science to older learners, in the context of lifelong learning, at least in initial learning stages. Crucially, narrative texts do not necessarily entail memory and learning superiority in comparison with the more established expository, and so narrative effects should be studied at their own, specific, right, instead of being assumed from the get-go and treated like a homogeneous effect. Importantly, examining the learning process has the potential of bringing relevant information that enables a better comprehension of the observed results.

In sum, these findings shed light into the issue of learning science from narrative texts in several ways, from gathering and analysing a comprehensive set of previous results and ideas that can be very useful in future research, to the investigation of empirical effects of narrative texts as stand alone tools and in comparison with expository texts with equivalent contents. A nuanced view of the narrative effect is proposed, and the importance of understanding the process leading to the observed outcomes becomes evident. Moreover, further studies are needed to better ascertain the memory and learning benefits reaped by older learners, and whether these benefits are limited to initial learning stages, as well as the durability and pervasiveness (in terms of level of comprehension) of the learning effects. Lastly, we gathered and analysed the learning outcomes from a representative set of previous studies which used science narrative texts, which may prove useful for researchers and educators interest in this topic.

RQ 3: Learning Process – *How does the process of learning science through narrative texts occur?*

The final research question, pertaining to the process of learning through science narrative texts, was also approached in different Chapters and in different ways. The fact that this question entailed three facets of the learning process is of significance, as they corresponded to different pieces, and levels, of the same puzzle, helping build a more complete picture of this issue. These facets were: processes triggered specifically by human-related elements from narrative texts (RQ 3a); processes stemming more generally from learners' features (RQ 3b); and more global perceptions that people have on science learning and the use of narrative texts in science education (RQ 3c).

In Chapter 2, drawing on the proposed dialogue between theoretical frameworks, we gave an outline for the investigation of the process of learning science through narrative texts, translated in a set of conditions and underlying processes deemed relevant for the matter at hands. We established connections with a broader literature, in search of hints on what these conditions and processes might be, complementing them with evidence whenever available. More specifically we identified specific features of the used science narrative texts (temporal structure; contents, namely the inclusion or not of fictional elements), of the learning activities using these texts (aims; duration of intervention; variety of extra-text activities), and of the learners that learned from these texts (social and cultural identities; socioeconomic status). We also identified processes proposed to be engaged by science narrative texts, namely integration with prior knowledge, affective dispositions, and cognitive abilities. We sought to address some of these issues using different approaches in the following empirical studies (Chapters 4, 5 and 6).

Overall, we found that many of these conditions and processes are either understudied or have yielded varying, and at times conflicting, results. As regards conditions, we found that a major distinction in science narrative texts pertains to the incorporation of fictional elements with factual science contents, or restriction to factual contents (i.e., scientific and narrative). The latter seem to more consistently have yielded good learning results, although both seem to produce other positive outcomes, such as engagement and connection with the (fictional or factual) characters. However, higher interest in narrative texts is not ubiquitous (e.g., Cervetti et al., 2009). This result is in line with the results of our following studies (Chapter 4 and 5). In addition, science narrative texts seem to elicit some confusion in terms of reading purpose, as they combine entertainment and instructional goals. On the other hand, it seems clear that learning science from narrative texts benefits from longer interventions, and teachers experienced in using such materials tend to prefer to use them in longer units (e.g., Kokkotas et al., 2010). Thus, even though we observed encouraging learning results from the use of science narrative texts in Chapters 4 and 5, it is possible that the benefits would be even more pronounced and extensive with a longer intervention. Furthermore, there is some evidence that students belonging to marginalized groups, such as less dominant social and cultural identities and low-income backgrounds, can benefit especially from science narrative texts. As regards the empirical studies from Chapters 4 and 5, participants were from a relatively homogeneous background, for the most part Portuguese, middle class, cisgender, white students. Our results thus indicate that students with these features are likely to benefit, at least to some extent, from narrative texts when learning science. All these issues provide interesting hints that warrant further and more direct investigation.

Regarding processes, integration with prior knowledge has been the more consistently examined aspect (e.g., Browning & Hohenstein, 2015; Flynn & Hardman, 2019; Wolfe & Mienko, 2007), and students with lower levels of previous knowledge seem to particularly benefit from learning science through narrative texts. This finding is again in line with the results of our empirical studies described in Chapter 4 and 5, and with the perceptions gathered in our study presented in Chapter 6. Additionally, there is evidence that science narrative texts can positively impact affective processes, yet the specific ways in which these affective processes in turn impact learning remains to be further ascertained. Exceptions include Reuer's results (2012), which report a significant correlation between the level of interest elicited by science narrative texts and the learning outcomes they yielded, and our own results. As a matter of fact, in Chapter 4 we ascertained that the level of interest in the texts predicted both the learning outcomes produced by the science narrative texts, and by their expository counterparts. This result becomes particularly significant when we consider that, although not all the texts used in Chapter 4 yielded equivalent learning outcomes, the interest the texts elicited in participants impacted their learning. Moreover, although the math narrative text did not yield better results than its expository counterpart, it was considered to be significantly more interesting. This result nods at the aforementioned finding that, even though not all narrative texts consistently generate better learning outcomes than expository texts (e.g., narrative texts with fictional elements), they nonetheless tend to generate positive affective outcomes. Moreover, by increasing interest, narrative texts can lead to educational benefits other than learning (e.g., engagement, proactivity), which may eventually contribute to improve learning in the long term. Of note, Reuer's study (2012), in which a significant correlation between interest and learning from science narrative texts was observed, was a semester-long intervention.

Finally, although a set of cognitive abilities were identified in Chapter 2, these were investigated at their own right to an even lesser extent in previous studies, and much less in connection to observed learning outcomes stemming from science narrative texts. It can thus be concluded from the analysis made in Chapter 2 that, even though there are many interesting leads, the conditions and processes contributing to learning science from narrative texts are still quite unknown, warranting further and more direct investigation.

Besides the results described in the previous paragraphs, we also investigated the impact on science learning of other general learner features and features and processes related to human elements depicted in a narrative text. In Chapter 4, we further determined that a greater science and literacy background, as well as more higher scores on measures of attitudes on science and cognition, had a positive impact on learning science from both the narrative and the expository texts. This constituted a novelty, as this set of measures had not been used in previous studies of the relevant literature. Literacy-related measures, in specific, have seldomly been investigated (for an exception, see Maria & Junge, 1993), and although prior knowledge has been more thoroughly investigated, other aspects pertaining to prior science background, which we have examined, have not. Part of these measures were also again found to predict learning from the chemistry narrative text in Chapter 5, crucially even in the presence of measures related to attention to and thoughts on human action. Furthermore, we also determined that some of these measures did not predict learning from the expository texts, whereas all predicted learning from the narrative texts, indicating processing differences when learning from these two text types.

In addition, in Chapter 5 we delved deeper into the processing of the chemistry narrative text. Namely, we directly examined the impact on science learning of depictions of human action in narrative texts, as well as the impact of human-related thoughts related to the that text, and to more general socio-cognitive abilities. Human-related features and processes are a defining feature of narrative, and have been closely examined in the context of more conventional (i.e., non-science), narrative texts (e.g., Kidd & Castano, 2016; Mar et al., 2006). Yet, even though these features are often evoked to justify enhanced engagement and learning from science narrative texts, previous evidence using these texts is usually more indirect (e.g., Jetton, 1994; Murmann & Avraamidou, 2014; Lin-Siegler et al., 2016). For these reason, even though the study conducted in Chapter 5 was very exploratory, it gives key insights on this important issue.

Specifically, using eye tracking and linear mixed model analysis, we found that dedicating attention to mental and sensorimotor depictions of human action during narrative text reading significantly and positively predicted science learning scores. Importantly, our study is the first to establish such a connection, as it is the first study to use eye tracking to investigate such issues using science narrative texts. Moreover, we also found that some thoughts on human action also predicted science learning scores and that attention to and thoughts on human action made joint contributions to learning. Additionally, exploratory correlation analysis indicated that some of these attention and thought measures correlated with specific levels of comprehension of the science contents presented in the text.

It is worth noting that participants engaged in a set of thoughts on human action related to the reading of the science narrative texts. Some of them were related to theory of mind measures, and thus more directly connected to the scientist's actions (e.g., Larison, 2018; Solomon, 2002); others were related to more general humanising of scientific meaning (e.g., Egan, 1997; Hadzigeorgiou et al., 2012), some of it pertaining to higher level connections to social aspects of the scientific process. We also included a more general socio-cognitive measure, in this case of attribution of intentions, which was yet another novel aspect of Chapter 5, given that previous studies on science narrative texts have not directly assessed theory of mind abilities. Of note, the measure we used (J. Fernandes et al., 2022) delivers a rather pure measure of intention understanding, is non-verbal, and was not related to the reading of the narrative text or its science contents. As such, it allowed us to differentiate the contributions on science learning of a theory of mind ability from those of participants' literacy skills and science-related knowledge. Furthermore, using a principal component analysis, we determined how these different measures were linked to each other (i.e., through components), and whether they contributed to science learning. Specifically, we found that the independent measure of attribution of intentions was grouped with measures of transportation and perspective-taking. We considered that this combination reflected a set of general abilities related to simulating, either by imagining or inferring, other visuospatial and psychological points of view. From all the measures related to thoughts on human action, this component was the best predictor of science learning, being particularly correlated to the level of recall. Another component, formed by our two measures of humanisation of scientific meaning, was particularly correlated to the level of comprehension of key ideas, despite only having a marginal general impact on science learning. Finally, the component formed by measures of cognitive theory of mind and affective theory of mind did not have a general impact on science learning, but was negatively correlated to the level of text-based inferences.

Moreover, we found that the interest elicited by the texts did not predict learning scores when put together with measures of attention to and thoughts on human action, but also pointed out that this was probably due to the overlap between interest in the text and the latter measures (i.e., thoughts on human action). It thus seems that the impact on learning of attention to depictions of human action is not merely reflecting participants' interest in the text, but the impact of measures of thoughts on human action seem to be partly dependent on it.

Finally, besides collecting and analysing a series of quantifiable measures, we gathered as well more qualitative perceptions on the role that narrative texts and other potential lifelong learning tools may play on science learning. This was done in Chapter 6, and in a way, represented an extension of the process of collecting participants' perceptions on the science learning materials we built, that was made in two of the pretests from Chapter 3. These more qualitative perceptions were analysed through thematic analysis, in contrast with, and as a complement to, the quantitative analyses from previous Chapters. Also, unlike the previous Chapters, in Chapter 6 these perceptions were gathered from people with varying levels of contact with science, including people who worked on science-related fields and had experience as science educators. This allowed us to build a rich mosaic of ideas, perceptions and feelings on the topic, and is a relevant contribution, as previous studies analysing perceptions on the use of narrative texts in science learning mostly examined school students and teachers (e.g., Prins et al., 2017; Ritchie et al., 2011). Despite this divergence from previous literature and the different backgrounds of the participants, their perceptions converged with ideas gathered from analysis of the previous studies (particularly in Chapter 2), and with empirical results from Chapters 4 and 5. In specific, narrative texts were regarded as a particularly effective tool for initial learning stages, given their perceived potential to capture learners' interest, offer a more familiar and easier language, and promote visual strategies. Yet, though a few people also saw their potential beyond "door openers", most people deemed expository texts to be more appropriate for subsequent stages of learning, after learners are equipped with the basic conditions to pursue deeper processing of information, and motivated to do so.

Another novel contribution of the study reported in Chapter 6 was concomitantly analysing more general perceptions on how science learning and literacy occur. Specifically, people highlighted the importance of integrating and connecting other kinds of knowledge when learning science, an idea that aligns with the results from previous Chapters, namely that other, non-science related, aspects contribute to the learning of the science contents. Additionally, participants underlined the importance of seeing the applied, and also useful, side of science, a result that nods at the outlined social and pervasive nature of science contents and developments. This connection was further cemented by the role that was attributed to museums, seen as important lifelong science communication hubs that connect different levels of society, namely science and people. Yet, participants also stated the need of improving that communication, both by modernising the way in which it is done and by reaching a broader, more inclusive audience. These perceptions, which had already been pointed out in the relevant literature (e.g., Achterberg et al., 2017; Feinstein and Meshoulam, 2014), add to the perceived challenges and difficulties outlined in the Introduction and in previous Chapters. Fittingly enough, the COVID-19 pandemic was a recurrent example not only of the failures of communication and connection between science and society, but also of how people can reconnect to science, namely to scientific knowledge authorities, during major social upheavals. These results further highlight the human facet of science explored in Chapter 5, albeit in a more qualitative and global, manner. Of note, Chapter 6 resulted from a collaboration with a science museum, Museu de História Natural e da Ciência de Lisboa, connecting scientific research, non-formal learning, and people, and thus extending the reach of the present dissertation.

The main implication of this set of findings is the enhanced knowledge of the process of learning science from narrative texts, in specific, but also of more global aspects pertaining to the process of learning science. Specifically, we found that there is some knowledge on the conditions and processes involved in this process, but also that many of them have not been more directly and thoroughly examined in connection to learning outcomes. Still, we gathered these meaningful ideas and organised them according to our proposed multidisciplinary theoretical mapping, which may inspire hypotheses in future research, as well as educational practices, using science narrative texts. Additionally, we were able to provide evidence on the impact of some aspects and processes on science learning, some which had already been addressed in previous literature, but others which were examined for the first time, or were at least for the first time more directly connected to science learning outcomes. We found that some of these processes were common to outcomes stemming from both narrative and expository texts, and that some were specific to the former. Moreover, and importantly, we made a deep exploration of the role of several different human-related elements and processes when learning science from a narrative text. Finally, we collected more qualitative and global insights on the role of narrative texts as lifelong science learning tools, as well as on more global aspects of this socially important issue. These findings, which pertain to different facets of the learning process, combine different kinds and levels of information: cognitive processes specifically related to human elements, cognitive processes more generally stemming from learners' features, and discursive cues resulting from more general perceptions from people.

In sum, these findings point to the centrality of human and social factors for science learning, further confirming the important role that narrative texts can thus play on this learning. These inputs highlight the importance of building a more complete picture of this learning process, by connecting learners to science texts and activities, but science to people more globally.

7.2. Major Contributions for Theory and Practice

In this section, we focus on what we believe to be the major contributions to the topic of learning science through narrative texts of the three main research questions, and corresponding findings. We identified four higher-order theoretical and applied contributions, which we will now highlight: 1) Theoretical dialogue with literacy at its core is key; 2) The importance of understanding the learning process to explain many narrative effects; 3) Science education materials with a human facet; and 4) Science learning is a fundamentally social endeavour. The main contributions pertaining to each chapter are summarized in Figure 7.2.

7.2.1. Theoretical dialogue with literacy at its core is key

First, the present work highlights that literacy-related aspects are key to tackle many of the challenges to science learning, and to understand the relationship between narrative texts as a learning tool and science education. Texts have been a central instrument in science education, in both formal and non-formal contexts, so it should come as no surprise that literacy-related aspects should be considered when tackling the challenging aspects of science learning.

To the best of our knowledge, we elaborated the first proposal of dialogue between theoretical frameworks from distinct fields that, nonetheless, focus on different and complementary aspects of the process of learning through text. These were text linguistics (text features), cognitive psychology (text comprehension) and pedagogy (the elements and interactions involved in the global process of learning through text); together, these frameworks deliver complementary insights that form a solid theoretical backbone. This theoretical backbone can then be used to interpret previous, often mixed, evidence, as they offer not only a more complete picture of the pieces of the overall puzzle, but also a nuanced and multidisciplinary-lens view of the topic at hands. Research and educational interventions that place science texts as a centrepiece may therefore benefit from the proposed theoretically-grounded mapping.

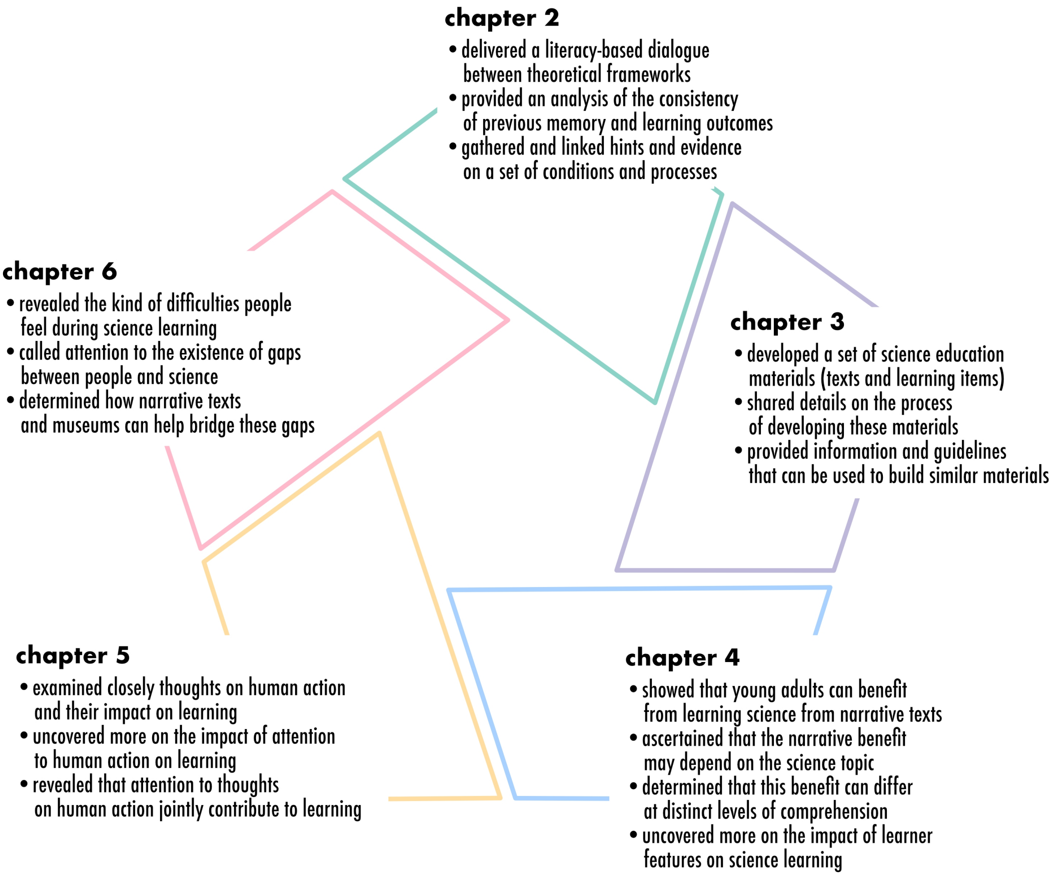


Figure 7.2. Main contributions of the dissertation by Chapter

As an illustration, educators and researchers may use this theoretical backbone to conceptualize the problematic of learning science through narrative texts, using it to shape other ideas and concepts relevant for their work (be it on narrative, be it on other relevant concepts and methods). Educators and researchers may also use it to develop text materials while keeping in mind the textual and pragmatic features of texts, and the wider mediation of the sociocultural context. A useful example is considering that science narrative texts can tap into learners' previous knowledge and expectations and activate ideas and processing strategies that can be detrimental to science learning. For instance, that narrative texts only serve to entertain and that, because they tend to include non-factual information, they are unreliable for science learning. Knowing this in advance, researchers can predict potentially conflicting or mixed results, and educators can take steps to better accommodate these learning tools, for instance by discussing these issues with learners explicitly. Moreover, drawing on insights from a text comprehension framework, such as the Construction-Integration model, may also help researchers and educators conceptualizing learning in a way that considers different levels of comprehension, and building assessment items that tap into these different levels of comprehension. In addition, educators can draw on the insights afforded by the proposed theoretical backbone to plan interventions that consider the several elements that make up the complex process of learning from text, even if they decide to focus their attention further in some of these elements. Educators can further strengthen and work on these different aspects when developing pedagogical practices, to create more meaningful and effective science learning. To this end, it can be very useful to frame interventions' questions and tasks using the RAND framework, which encompasses all these facets.

Additionally, by drawing on a multidisciplinary view of the topic at hands, the proposed theoretical dialogue may eventually encourage other kinds of dialogue and collaboration, namely between different people and institutions who take an interest in topics of science learning and literacy, and in the use of innovative tools such as narrative texts. Examples include, but are not limited to, school teachers, museum educators and other science communicators, linguists, psychologists (e.g., educational, cognitive, social), cognitive scientists, writers, and policy-makers.

However, it should also be noted that we do not mean to suggest that these are the only useful theoretical frameworks to be considered when tackling the issue of the use of narratives in science learning. Rather, we propose that adopting a multidisciplinary proposal that draws on different aspects of literacy-based learning can prove valuable to address it.

Overall, the proposed theoretical backbone can be used in different and complementary ways to grapple with the challenges of science learning, aiding in the designing and planning of future research or educational interventions. This multidimensionality further underlines the importance of grasping the different aspects of the learning process, which give rise to multiple narrative effects.

7.2.2. The importance of understanding the learning process to explain many narrative effects

Second, and in articulation with the first point highlighted, one of the major takeaways of the present work is that it is crucial to build a better and more complete understanding of the learning process, rather than merely focusing on its outcomes. Intimately connected to this idea is the proposal that specific and multiple narrative *effects* should be investigated, rather than conceiving the benefits afforded by narrative texts as a generalized or homogeneous effect.

Truth be told, we too started off this project enthusiastic with the idea of a relatively generalized narrative effect, so we can recognise the appeal of this conception. However, the incorporation of different theoretical frameworks, combined with the results of our studies, unequivocally point towards a more nuanced view of this effect. As a result, any observed narrative effect should therefore be made up of specific interactions between the different facets of learning. It becomes clear from the analysis conducted in Chapter 2 that previous studies had already exemplified different conditions in which the effect appeared, or fell short of being observed; yet, ours was the first work to more directly try to explain why this was the case, and offer a more specific proposal to explain previous results and plan future interventions. As a research and educational framework, we thus propose that it is necessary to specify with what (texts), with who (learners), when and where (activities and wider context) these narrative effects occur. We believe this proposal can offer diverse cues as to how educators may deal with this matter and tailor their practices using science narrative texts as an educational tool. It also lays seeds for future research that is more aware of the complexity of this issue. In fact, as we underlined when summarizing our main theoretical findings, even if researchers end up choosing to focus on specific aspects in detriment of others, they may nonetheless actively recognise the aspects that are not being addressed, yet likely contribute to the observed results. More comprehensive suggestions for future research can then be extracted.

As regards texts, there are many ways a narrative can be written, especially in terms of the different, and at times conflicting, elements it may contain (e.g., fantasy vs. historical accuracy). The problem is not the variety of narratives *per se*, which testifies to the creative nature of narrative (e.g., Bruner & Lucariello, 1989), but rather the fact that different narrative texts may interact in fundamentally different ways with learners, activities and sociocultural contents. For example, processes like imagination, emotional involvement, or perspective-taking, to name a few, will likely be elicited to different extents depending on the elements included in the narrative text. The same applies to features of the learners themselves (e.g., while one learner may loathe science-fiction, another may find it most stimulating), and also of the wider context in which the activity is taking place (e.g., some elements may not resonate within specific cultures, or even be offensive). Importantly, as regards learners, an important implication of the present dissertation is that narrative texts can be used to teach science to young adults, which aligns with previous claims of a more widespread use of these materials (e.g., Olson, 2015; Luna, 2015). It remains to be ascertained whether this use is limited to initial learning stages, as our findings mostly suggest, or can be used in further stages of learning.

Future research may also investigate more thoroughly which particular narrative elements have the potential to reach specific learners, and why, so that educators can draw on this information to tailor their practices according to specific needs and proclivities. Complementarily, through the use and continued exploration of narrative texts as science learning tools, educators may find some narrative elements to be particularly suited to stimulate interest, others to improve memory, and still others to approach a particularly difficult (or boring) topic. In a related vein, based on previous studies and our own empirical results, it is important to consider that narrative texts can yield more positive results than expository texts in outcomes such as interest, without these translating, at least straight away, into enhanced learning. This is not to say that educators should refrain from using narrative texts in short interventions, but merely that memory and learning benefits may be more noticeable in the long term. However, and importantly, the fact that learning may not improve in the short-term (e.g., in one session), does not imply that it will not be improved in the long-term. As a matter of fact, though our empirical results and previous studies show that narrative texts can afford learning benefits in short-term interventions, our theoretical review indicated that they may prove particularly beneficial in longer interventions (e.g., spanning several weeks). One of the reasons for this may be that increased interest exerts a more substantial effect in the long term: Interested learners will arguably engage more actively in the learning process, which may eventually improve their learning. In addition, longer interventions may also offer an opportunity to attenuate potential processing conflicts elicited by narrative texts (e.g., to entertain vs to learn), as there is more time to accommodate and integrate different reading purposes, as well as to deconstruct associated learning resistances (e.g., that narrative texts are not a "serious" learning tool). Notably, if the use of narrative texts becomes more widespread and common, they will become more familiar and cemented in science learning settings, further contributing to attenuating processing conflicts and resistances. Nonetheless, educators may still find that some narrative elements generate trade-offs between effects (e.g., interest vs learning accuracy), which again highlights the importance of adopting a nuanced approach to the use of this tool, tailored to specific educational needs and aims.

Furthermore, our findings on the processes contributing to the learning outcomes afforded by narrative texts are supportive of the notion that learning through science narrative texts may engage some processes or features to a different extent than expository texts. This echoes long-standing claims that narrative texts engage readers' minds in specific ways (e.g., Bruner, 1986; Egan, 1997), which may then improve learning. Of relevance, within our theoretical approach, was the finding that general literacy knowledge (e.g., familiarity with literary authors) contributes to the observed learning outcomes. Combined with the finding that various aspects can make similar contributions, and with participants' shared perception of the importance of further linking science knowledge with other kinds of knowledge, daily life and contemporary matters, our findings encourage a diversified, multidisciplinary-leaning, take on science learning. Educators may integrate different aspects to create more meaningful science learning.

Yet, there is still much to find out about the way science narrative texts are processed. As we pointed out in the theoretical review from Chapter 2, there are many processes which have been investigated using non-science narrative text, but not using their science-conveying counterparts. Future research may take note of these hints, and even repeat some of these studies using science narrative texts. Our findings hence highlight the relevance of combining educational practices with cognitive investigation, as the latter has been considerably less examined in the literature of learning science from narrative texts. In our understanding, science educators and cognitive scientists represent two distinct epistemological stances, with different aims and priorities. Even though both are interested in learning, many science educators cannot afford to focus on designing studies to investigate how narrative texts engage specific processes. Instead, their aims will likely be born of practical needs and focus on whether science narrative texts can help students change their thinking and ideas regarding a science subject, within the context of wider educational practices. Narrative texts are thus another tool meant to address general learning challenges by maximizing engagement and learning, making difficult to draw conclusions on its more specific effects and the role they play on generating learning outcomes. However, education and research can greatly benefit from going hand in hand. For instance, translating practical needs into specific research questions can generate new empirical studies, whose results can both help understand how science narrative texts work and inform educators on how to adjust their practices. In this sense, drawing on the same theoretical frameworks can be a way of building common ground and solid collaborations.

In sum, building on the idea of multiple narrative effects reveals the multidimensionality of the process of learning science through narrative texts, and highlights the importance of understanding and addressing different dimensions of this process. The human facet is of particular importance and will be examined more closely in the next section.

7.2.3. Science education materials with a human facet

The humanness inherent to science narrative texts is an aspect that pervades all the Chapters of this dissertation, and there is an overall convergence in terms of the salience of this feature in science narrative texts. The mere fact that the science narrative texts yielded learning outcomes is arguably linked to its inclusion of human elements. The narrative texts were built around scientists' thoughts and actions, and so the science contents were described, and consequently learned, in tight connection to these thoughts and actions. This aspect was given special attention in Chapter 5, enabling us to extract more comprehensive implications. These implications are not only particularly relevant for educators and researchers interested in building science narrative texts, but can also guide selection and adaptation of these educational materials.

First, the fact that dedicating attention to descriptions of mental operations and movements in the narrative text contributed to improve learning at different levels of comprehension indicates that this information was useful for integrating and understanding the science contents. This is an important finding, as there are concerns that narrative elements might be distractive and end up hampering learning (e.g., Garner et al., 1989; Hidi et al., 1982). Educators interested in using such materials can thus take note of the importance of ensuring that narrative and scientific information are tightly intertwined, something which had already been pointed out by other researchers (e.g., Arya & Maul, 2012; Lehman et al., 2007; Wolfe & Mienko, 2007).

Second, the fact that reading the narrative text triggered a series of thoughts on human action, some of which also made joint contributions to learning, has important theoretical and applied implications. On the one hand, it establishes connections to a body of literature which has developed important research on these processes using non-science narrative texts (e.g., Djikic et al., 2013; Mar & Oatley, 2008; Oatley, 2016), but that is not mentioned in studies with science narrative texts, despite their common focus on human-related aspects of narrative texts. Of note, while there are a few mentions to "perspective taking" in works on the use of narrative texts in science education, to the extent of our knowledge, concepts such as "attribution of intentions" and "theory of mind" are seemingly absent. As such, this was the first time a measure of socio-cognitive ability was included in a study on science narrative texts. In addition, given the shortage of theory of mind tasks targeting normative adult populations, the inclusion of this task and its corresponding findings also offer a more general contribution for studies interested in expanding their "theory of mind toolbox".

On the other hand, it calls attention to the fact that thinking about human action can have a positive impact on science learning. In the present case, we found thoughts related to taking the scientists' perspective and of immersion or transportation into the story, as well as the more general ability of attributing intentions, to be particularly good predictors of learning score. These results bring evidence to claims that science narrative texts prompt attention, interest, reflection, as well as organisation and recall of knowledge (e.g., Arya & Maul, 2012; Hadzigeorgiou et al., 2012; Hong & Lin-Siegler, 2012), specifically connecting depictions of human action and related thoughts on human action to these various educational outcomes. Moreover, even though not all the thoughts on human action participants engaged in had palpable effects on learning, the fact that they engaged in them is in itself important and relevant for educators interested in using narrative texts in science education. Of note, most of participants' written responses on thoughts on human action were related to the scientist's struggles and his defiance of social conventions. This result is in line with previous theoretical ideas and empirical evidence that such elements can engage readers (e.g., Arya & Maul, 2021; Hadzigeorgiou et al., 2012; Lin-Siegler et al., 2016), and thus indicate that narrative texts including those elements can be particularly effective to capture readers' attention. In addition, there was seemingly some degree of overlap between thoughts on human action and participants' interest in the text. This suggests both that these processes likely feed off each other, but also that interest in the text, thoughts on human action, and science learning, are not merely independent outcomes, but are instead connected and, to some extent, interdependent. In fact, it indicates that narrative texts encouraged participants to engage in a variety of thoughts related to human action, which may then lead to other positive educational outcomes, and improve learning in the long run. Engaging in such human-related processes may even yield more generalized positive social outcomes. For example, previous studies have found that engaging in processes such as theory of mind can promote empathetic and prosocial behaviours (e.g., Imuta et al., 2016; Mar & Oatley, 2008). Future research should thus examine this issue using science narrative texts.

These findings are of great relevance for research, namely on the processes underlying learning science from narrative texts, and have important implications for applied educational contexts and practices. Specifically, findings imply that this kind of human-related processes should be encouraged, and perhaps even explicitly discussed with learners, when using science narratives to teach science. A corollary of these results is that affective and socio-cognitive processes should not be estranged from science learning; in fact, these different and apparently opposed modes of thought (everyday/narrative and scientific) can be connected through the reading of science narrative texts, a connection that can result in positive learning outcomes. They should therefore be addressed and promoted together in rich and multidimensional learning environments.

In sum, by retrieving the humanness of science, science narrative texts have the potential of bridging the gap between everyday thinking and scientific thinking, and between affective and socio-cognitive aspects and science concepts. As science narrative texts may be used to further connect learners to the wider social and cultural contexts, their impact can be further extended to more encompassing societal levels and issues.

7.2.4. Science learning is a fundamentally social endeavour

The final major contribution of this work pertains to the relationship between science learning and the wider social and cultural context. The human and cultural decontextualization of science contents has been regarded as one of its most challenging aspects (e.g., Arya & Maul, 2021; Solomon, 2002), entailing problems for science learning and for the perception people have on science more generally that can feed off each other. The present findings contribute to highlight the importance of situating science in the wider social and cultural context in which it takes places.

Our findings (particularly from Chapter 5) suggest that human and social elements became salient in learners' minds when reading a science narrative text, even in the absence of explicit prompts. This suggests that the science narrative text encouraged readers to frame science as a human and social activity, also known as humanising of scientific meaning (e.g., Egan, 1997). The salience of wider social aspects is further cemented by the fact that participants tended to build their human-related thoughts around the information that Pasteur made an important scientific discovery going against contemporary chemists and pre-established ideas. Besides encouraging learners to feel more connected and closer to scientists, these thoughts and framing can help build a more accurate, socially and culturally-grounded, image of the nature and workings of science, challenging misconceptions that can be damaging to learning (e.g., that scientists are genius or that science is based on unquestionable truths; Allchin, 2003; Clough, 2011; Hong & Lin-Siegler, 2012). Historical and discovery-based narratives, such as the ones we presented, may be particularly relevant in this regard, helping educators contextualize science within a wider, and more human, context. Importantly, we found that the humanisation of meaning contributed to the learning of the science contents and was particularly related to the comprehension of key ideas. This contribution to learning also presented overlap with the interest elicited by the text, once again highlighting how learning outcomes and affective aspects of the learning activity are connected.

As such, the literacy-mediated tool of science narrative texts has the potential of retrieving social and cultural dimensions of learning through reading (C. Morais, 2015; Morais & Kolinsky, 2021). This potential assumes important epistemological implications, as many authors have pointed out how cognitive psychology and educational psychology, core areas of the present work, have tighten its connections to areas such as computation and artificial intelligence, overlooking its deep social and cultural roots and determinants (e.g., Bruner, 2004; Morais & Kolinsky, 2021).

One way of retrieving this dimension is by investigating and incorporating people's perceptions of these processes. Besides addressing individual gaps between everyday thinking and scientific thinking, this effort has the potential of bridging yet another, wider, gap between science and people. As outlined in the introduction, lifelong science learning has a pivotal socially rooted role that goes well beyond the limits of formal instruction, where science narrative texts have been chiefly applied and investigated. Concomitantly, texts maintain an important role in more widespread science communication in an increasingly digitalized world, such as through posts in social media, popular science books, and museum plates (Negretti et al., 2022; Ravelli, 2007). It is therefore very relevant that we determined that science narrative texts have a multilayered impact on lifelong science learning. These range from measurable learning outcomes, to encouraging a human and social framing of science, to more generalized perceptions that their features make them particularly useful in specific learning stages.

Importantly, our findings suggest that narrative texts and other non-formal tools, such as museums, can have complementary roles in connecting people to science in the context of lifelong science learning. This idea is particularly interesting for museum educators. On the one hand, narrative texts were perceived by participants as an efficient tool to bridge more individual gaps in science learning (e.g., to capture interest); on the other hand, museums were regarded as instrumental for connecting science and people (e.g., wider science communication). This entails a potentially powerful combination: Museums can take advantage of narrative's features to create more inclusive science learning spaces for several different people and, in doing so, contribute to bridge wider societal gaps between everyday and science thinking. A further implication of these findings is that science learning should dialogue with people and consider their ideas and ways of thinking. Most people will not get to think effectively in the scientific mode (Bruner, 1986; 1990) and it does not help that science is predominantly portrayed as a decontextualized body of knowledge that people are required to learn, instead of something relevant to their cultures, knowledge, or interests (Calabrese et al., 2003; Gilbert et al., 2005). Initiatives such as developing communities of practice around specific topics, which have strong connections to narrative thinking and storytelling, can be a way of connecting people to science knowledge in a way that is relevant to them (e.g., Jerónimo & Reis, 2016; Lave & Wenger, 1991; King et al., 2021). Thus, narrative texts can be a way of bringing together and empowering groups of people, for instance by encouraging them to develop and share narrative-based science materials. This implication is somewhat akin to citizen science methodologies, whose potentialities have been increasingly focused (e.g., Van Haeften et al., 2020).

Crucially, applying this set of measures to connect people to science can in turn have very palpable and generalized effects in society at large. We currently deal with a polarization in people's response to science, and on one side of that polarization is a deep lack of trust in it (e.g., Achterberg et al., 2017; Li & Qian, 2022). This lack of trust is often fuelled by lack of understanding of science and of the scientific process (e.g., Kreps & Kriner, 2020; Scheufele & Krause, 2019), and entails a myriad of societal dangers. Two very familiar examples are the effects on individual and social behaviours regarding the ongoing COVID-19 pandemic, and vaccination more generally (e.g., the idea that there is causal link with autism), and climate change. However, there are other examples. One comes in the form of policies directed at reducing transgendered people's rights (e.g., bathroom bills, medical discrimination), therefore endangering their lives. The groups behind these policies often claim to base their arguments in "biology", when, in reality, they are not up to date with the non-binary quality of human biology (e.g., de Vries & Södersten, 2009; Joel et al., 2015; Wilhelm et al., 2007). Another example is believing that the earth is flat. While they might seem harder to fully grasp at first sight, the wider psychological and social consequences of holding this set of beliefs were eloquently described in a article from the Atlantic magazine: "Conspiracy theorists can fall into vicious cycle of alienation and acceptance, pulling them away from society at large and further into the circle of believers" (Weill, 2022). It is imperative to underline that distrust in science and reluctance in acquiring scientific knowledge is often deeply intertwined with social and political ideology, and should therefore be tackled and understood in this wider context. In this sense, dealing with the problem of distrust in science is a bilateral process: People and groups have the responsibility of questioning the limitations and dangers of their beliefs, but science also needs to become more approachable and consider people's needs (Douglas et al., 2017) and different epistemological stances, or narratives (A. Costa, 2021). As such, narrative texts are the perfect candidate to increase connections between different personal and collective narratives and science.

Our findings thus underline that, even when topics pertain to chemical components or arithmetic operations, science is a fundamentally social and cultural practice. While it is important to address individual gaps, such as connecting people to science materials and improving their learning, it is also important to further connect science to people and increase its relevance and reach. Being a familiar and pervasive instrument, narrative texts can thus help retrieve the social and cultural component of science and, when coupled with other tools such as museums, help people connect with science topics on a wider social scale, generating more widespread social impact.

7.3. General limitations of this work

Even though our findings offered novel and important insights into the topic of learning science using narrative texts, it is also important to point out some of the limitations of the present work.

One potential limitation of the present work actually resides in one of its more appealing aspects: Its "kaleidoscopic" approach. The fact that we used several different lens (both in terms of theoretical ideas and methodological approaches) to tackle our topic of interest, endows the work with a wide scope and with several different kinds of evidence and reflections. However, it also necessarily entails that it was not possible, in due course, to delve deeper and focus on certain aspects, or on “each lens”. The fact that we first needed to develop the science learning materials, through a series of pretests detailed in Chapter 3, further contributed to this limitation, as this process was quite time-consuming. This limitation is echoed by other authors, who claim that developing narrative texts that teach scientific concepts in an appropriate, interesting, and condensed fashion is a demanding and laborious endeavour (e.g., Hadzigeorgiou, 2016; Kerby et al., 2018).

Another potential limitation regards the kind of participants involved in the studies, namely the fact that they were not completely estranged from science. In Chapter 4 it was difficult to ensure that all participants had as little contact with science as possible. Many participants had studied science in high-school, and some had subjects involving mathematics/statistics or biology in their present undergraduate courses. On the one hand, this ensured some level of variability in the results, which is important for statistical reasons. On the other hand, it would have been interesting to either include a sample of students with virtually no contact with science, or to include several levels of contact. However, this would have added yet another layer of complexity to an already laborious and complex study. Similarly, in Chapter 6, it would have been interesting to include people which did not enjoy science in the discussion, but as pointed out in that Chapter, it would have been more difficult to ensure, as these people will not likely wish to spend their free time talking about science. At any rate, these limitations reflect a more general limitation of scientific studies, which often are unable to ensure a wide scope of inclusion.

Additionally, although we had initially planned to conduct studies with school children and/or adolescents, we ended up deciding to focus on young adults, a relatively understudied population in the topic at hands. While this is not a limitation it itself, it is something worth examining in the future. Indeed, even though these younger populations have been quite studied in the context of learning science through science narrative texts, this is not yet the case in Portugal.

Another limitation, that relates to this more global potential limitation of scientific practices, is the fact that we did not include a wider discussion with students examined in the studies of Chapters 4 and 5, beyond taking part in a series of experimental tasks and having their learning assessed. Narrative texts are claimed to facilitate collective discussions on science topics (e.g., Dinsmore et al., 2017; Solomon, 2002), and this active engagement and discourse can be an important foundation of scientific literacy, one of our main concerns and interests in the present work. In this sense, the topics we selected were not very familiar or dealt with pressing matters (e.g., current social issues), which would have made more difficult to generate such discussions and reflections among students. The fact that the topics were unknown filled an experimental purpose aligned with our aims, and, in turn, our aims were not particularly directed at generating discussions, but it is still relevant to point this out, as it relates to our more general aim of tackling challenges in the development of scientific literacy and can inspire future studies to address these issues.

As for the learning materials themselves, they were perhaps a bit more challenging than we would have initially hoped. In addition, we did not include a more diverse range of learning items, such as tasks assessing more physical applications (e.g., draw or build/mount something), which were something discussed in the beginning of the project. One way of tackling this kind of difficulties in future studies is by involving students in the conception and development of learning items. For instance, an independent group of students (i.e., not taking part in the main studies) could have taken part in these tasks. This would have a double effect: The students from the main studies would complete learning items more akin to their ways of thinking and thus potentially more engaging and relevant for them, whereas the students developing these materials would still be part of science learning experience, akin to citizen science projects or communities of practice.

A further limitation of the present work pertains to the limited predictive power of our proposed theoretical framework. Whereas social science strives for the development of models which enable behavioural predictions, and it would undoubtedly be useful, for instance for designing and tailoring educational practices according to specific needs, our proposed theoretical dialogue is mostly descriptive. It provided a mapping from which to interpret evidence and plan research and educational interventions, but not to predict specific effects based on the combination of its described features. On the one hand, this lack of predictive power stems from the lack of current knowledge on the processes that underlie learning from science narrative texts, as this prevents us from understanding how these tools work. On the other hand, however, we consider that the lack of predictive power also derives from our multidisciplinary approach: Acknowledging the multi-layered complexity of these learning phenomena should create a trade-off with the ability to predict its effects, at least to some extent.

Lastly, the lack of wider learning context likely influenced the learning process and its outcomes, perhaps hindering them to some extent. This is true both of the laboratory studies from Chapters 4 and 5 and the focus-group study from Chapter 6, the latter which had to take place online instead of in the museum. Even though the former filled experimental purposes and the latter had the silver-lining of bringing a learning activity to people's homes during confinement, this limitation still needs to be acknowledged. Learning is a slow-burn process, requiring time, continuity, and integration with other information and processes taking place in our lives. As such, we likely merely accessed a tiny fraction of the potential outcomes of using narrative texts as science learning tools. However, instead of focusing on the limitations this absence poses, we may choose to see the unforeseen opportunities and possibilities it leaves room for, and use it as an inspiration to continue exploring this tool.

7.4. Concluding remarks

Throughout this project, I was often haunted by a relentless idea. I was afraid that participants, having low prior knowledge in science, would feel that their level of knowledge or intelligence was not "good enough", and that they would feel frustrated, or even anguished, as a result of taking part in my studies. This was, to some extent, confirmed by the comments of some participants, although many did not feel this way, or at least not to extent of feeling compelled to share it (and what a relief that was!). Yet, this idea and the uneasiness that accompanies it, was not new to me; in fact, it is deeply entrenched in my path. I had felt it while working as an explainer in a science museum, and as a cognitive tutor. Most importantly, I had felt it long before that, throughout my life, with my own self and my learning processes. Formal and natural sciences have this impact on many people, making us doubt our abilities, and putting scientists on a pedestal. And, sometimes, narratives reinforce these ideas, with common tropes and stereotypes that we see time and time again in films and TV shows. And then life imitates art, and art imitates life.

Some weeks ago, I finished reading a tale by Ted Chiang, "Stories of your life". This science fiction tale draws on concepts from linguistics, with which I feel comfortable with, given by a linguist (a woman), but, crucially, it also draws on concepts from physics, given by a physicist (a man). At a certain point, I was struggling to understand Fermat's principle, namely because of its use of teleological language, which to me felt very weird and counter-intuitive, even counter-scientific. And this had not been the first time I had a similar reaction to explanations from formal and natural sciences. In fact, as you may recall, I did a whole PhD project on it. And yet, my first reaction was again to assume I was the one "not getting it", most likely due to my own intellectual shortcomings. And this is what I feared my participants would feel as they read the texts I wrote and completed the learning measures I developed. I did not want to evoke in them the same uncomfortable feelings, I did not want to be yet another snobby academic making other people question their intelligence and, by extent, their worth. We should not be afraid of having doubts, of questioning science's ideas and language. This is what rigidity of thinking looks like, it is what authoritarianism feeds on. And I truly believe these fears and cognitive dissonances help fuel many of the major social and political upheavals we are now witnessing in all its splendor. It generates a sense of asymmetry, jeopardizes the motivation to find common ground, and, thus, of making a concerted effort to create effective change.

And then, some pages later, something wonderful happened: The linguist, who was obviously a very bright woman, nonetheless told the physicist that she was struggling with the same diffulty as I. I immediately felt my doubt was valid, that I should not question my intelligence based on it, the same way I was not questioning the linguist's intelligence. And this is the power of stories, and why I candidly believe stories are powerful instruments of change. Like Bruner, I endorse the view that "'world making' is the principal function of mind, whether in the sciences or in the arts" (1991, p. 691). Stories are not mere instruments for representing reality, it is through them that we imagine new realities and begin transforming the present reality to create them. And so, it is not enough to use stories to talk about the present world. We need to consider how we tell those stories, who we include in them, the purpose of telling them. And this is why science can never be neutral, because its stories are deeply intertwined to all the other stories we tell, to the broader act of storytelling itself. Stories about science cannot be extricated from the wider world: They are made by human hands and seen by human eyes, the world is a web of stories, a matrix of events with a common narrative thread.

And If we can change the way we tell stories, the agents we include in them, the reasons why we tell them, then maybe there is hope. Maybe we can change the world.

References

Abd-El-Khalick, F. (1999). Teaching science with history. *The Science Teacher, 66*(9), 18–22.

Abdool Karim, S. S. (2022). Public understanding of science: Communicating in the midst of a pandemic. *Public Understanding of Science, 31*(3), 282-287. https://doi.org/10.1177/09636625221089391

Achterberg, P., Koster, W. d., & van der Waal, J. (2017). A science confidence gap: Education, trust in scientific methods, and trust in scientific institutions in the United States, 2014. *Public Understanding of Science, 26*(6), 704–720. https://doi.org/10.1177/0963662515617367

Adam, J. M. (1997). Genres, textes, discours: pour une reconception linguistique du concept de genre. *Revue belge de philologie et d'histoire, 75*, 665-681

Adam, J. M. (2011, May 20-21). *The narrative sequence : history of a concept and a research area* [Paper presentation]. 1er Colloque International organisé par le Réseau Romand de Narratologie (RRN), Fribourg, Switzerland

Akarsu, B., Kariper, A., & Coşkun, H. (2015). The effect of using scientific stories on teaching science and on the academic achievement of the students. *Mersin Üniversitesi Eğitim Fakültesi Dergisi, 11*(2)

Albert, B., & Tullis, T. (2013). *Measuring the user experience: collecting, analyzing, and presenting usability metrics*. Newnes

Albuquerque, U., & El-Hani, C. N. (2021). In a world in shadows and flames scientists and laypeople need better understanding of how science works. *Ethnobiology and Conservation, 10*:13.

Alçada, I. (2016). *Plano Nacional de Leitura: Fundamentos e Resultados* [National Reading Plan: Fundaments and Results]. Leya

Allchin, D. (2003). Scientific myth-conceptions. *Science Education, 87*, 329–351

Allchin, D. (2022). Who speaks for science?. Science & Education, 31(6), 1475-1492. https://doi.org/10.1007/s11191-021-00257-4

Alvermann, D. E., Hynd, C. E., & Qian, G. (1995). Effects of interactive discussion and text type on learning counterintuitive science concepts. *The Journal of Educational Research, 88*, 146-154. https://doi.org/10.1080/00220671.1995.9941293

Archer, L., Dawson, E., Seakins, A., & Wong, B. (2016). Disorientating, fun or meaningful? Disadvantaged families' experiences of a science museum visit. *Cultural Studies of Science Education, 11*(4), 917-939. https://doi.org/10.1007/s11422-015-9667-7

Ariasi, N., & Mason, L. (2011). Uncovering the effect of text structure in learning from a science text: An eye-tracking study. *Instructional science, 39*(5), 581-601. https://doi.org/10.1007/s11251-010-9142-5

Ariasi, N., Hyönä, J., Kaakinen, J. K., & Mason, L. (2017). An eye-movement analysis of the refutation effect in reading science text. *Journal of Computer Assisted Learning*, *33*(3), 202-221. https://doi.org/10.1111/jcal.12151

Arya, D. J., & Maul, A. (2012). The role of the scientific discovery narrative in middle school science education: an experimental study. *Journal of Educational Psychology, 104*, 1022–1032. https://doi.org/10.1037/a0028108

Arya, D. J., & Maul, A. (2021). Why Sociocultural Context Matters in the Science of Reading and the Reading of Science: Revisiting the Science Discovery Narrative. *Reading Research Quarterly, 56,* S273-S286. https://doi.org/10.1002/rrq.393

Aso, K. (2001). Visual images as educational materials in mathematics. *Community College Journal of Research and Practice, 25*. 355-360. https://10.1080/106689201750192184

Aukerman, M., & Chambers Schuldt, L. (2021). What matters most? Toward a robust and socially just science of reading. *Reading Research Quarterly, 56*, S85-S103. https://doi.org/10.1002/rrq.40

Avraamidou, L., & Osborne, J. (2009). The role of narrative in communicating science. *International Journal of Science Education, 31,* 1683–1707. https://doi.org/10.1080/09500690802380695

Baddeley, A. D., & Hitch, G. (1974). Working memory. *Psychology of Learning and Motivation, 8*, 47–89. doi:10.1016/S0079-7421(08)60452-1

Bakhtin, M. M. (1984). *Esthétique de la création verbale*. Gallimard

Ball, P. (2017). It’s not just you: science papers are getting harder to read. *Nature*, *30*. https://doi.org/www.nature.com/news/it-s-not-just-you-science-papers-are-getting-harder-to-read-1.21751.

Ball, P. (2021). What the COVID-19 pandemic reveals about science, policy and society. *Interface Focus*, *11*(6), 20210022. https://doi.org/10.1098/rsfs.2021.0022

Banister, F., & Ryan, C. (2001). Developing science concepts through story-telling. *School Science Review, 82*, 75-84

Bannert, M., & Mengelkamp, C. (2008). Assessment of metacognitive skills by means of instruction to think aloud and reflect when prompted. Does the verbalisation method affect learning?. *Metacognition and Learning, 3*(1), 39-58. https://doi.org/10.1007/s11409-007-9009-6

Baram-Tsabari, A., & Yarden, A. (2005). Text genre as a factor in the formation of scientific literacy. *Journal of Research in Science Teaching, 42*(4), 403-428. https://doi.org/10.1002/tea.20063

Barata, R., Castro, P. & Martins-Loução, M.A. (2017) How to promote conservation behaviours: the combined role of environmental education and commitment. *Environmental Education Research, 23*(9), 1322-1334. DOI:10.1080/13504622.2016.1219317

Baron-Cohen, S. (2001). Theory of mind in normal development and autism. *Prisme*, *34*(1), 74-183.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a “theory of mind”?. *Cognition, 21*(1), 37-46. https://doi.org/10.1016/0010-0277(85)90022-8

Baten, K., & Håkansson, G. (2015). The development of subordinate clauses in German and Swedish as L2s: A theoretical and methodological comparison. *Studies in Second Language Acquisition, 37*(3), 517-547. https://doi.org/10.1017/S0272263114000552

Beaudoin, C., Leblanc, É., Gagner, C., & Beauchamp, M. H. (2020). Systematic review and inventory of theory of mind measures for young children. *Frontiers in psychology*, *10*, 2905. https://doi.org/10.3389/fpsyg.2019.02905

Bell, P., Lewenstein, B., Shouse, A. W., & Feder, M. A. (2009). *Learning science in informal environments: People, places, and pursuits* (Vol. 140). Washington, DC: National Academies Press.

Bellet, P. S., & Maloney, M. J. (1991). The importance of empathy as an interviewing skill in medicine. *Jama, 266*(13), 1831-1832. https://doi:10.1001/jama.1991.03470130111039

Bensaude-Vincent, B. (2001). A genealogy of the increasing gap between science and the public. *Public Understanding of Science, 10*(1), 99. https://10.1088/0963-6625/10/1/307

Bernal, J. D. (1939). *The social function of science.* Routledge

Best, R. M., Floyd, R. G., & McNamara, D. S. (2008). Differential competencies contributing to children's comprehension of narrative and expository texts. *Reading psychology, 29*(2), 137-164. https://doi.org/10.1080/02702710801963951

Best, R. M., Rowe, M., Ozuru, Y., & McNamara, D. S. (2005). Deep-level comprehension of science texts: The role of the reader and the text. *Topics in language disorders*, *25*(1), 65-83. https://doi.org/10.1097/00011363-200501000-00007

Bevan, B. (2016). The museum of pink: Retheorizing the science museum. In *Intersections of Formal and Informal Science* (pp. 67-77). Routledge.

Blotnicky, K. A., Franz-Odendaal, T., French, F., & Joy, P. (2018). A study of the correlation between STEM career knowledge, mathematics self-efficacy, career interests, and career activities on the likelihood of pursuing a STEM career among middle school students. *International journal of STEM education, 5*, 22. https://doi.org/10.1186/s40594-018-0118-3.

Blouin-Hudon, E. M. C., & Pychyl, T. A. (2017). A mental imagery intervention to increase future self-continuity and reduce procrastination. *Applied Psychology*, *66*(2), 326-352. https://doi.org/10.1111/apps.12088

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology, 3*(2), 77-101

Britner, S. L., & Pajares, F. (2006). Sources of science self-efficacy beliefs of middle school students. *Journal of Research in Science Teaching, 43*(5), 485-499. https://doi.org/10.1002/tea.20131

Britton, B. K., Graesser, A. C., Glynn, S. M., Hamilton, T., & Penland, M. (1983). Use of cognitive capacity in reading: Effects of some content features of text. *Discourse Processes, 6*(1), 39-57. https://doi.org/10.1080/01638538309544553

Broemmel, A. D., & Rearden, K. T. (2006). Should teachers use the teachers' choices books in science classes?. *The Reading Teacher, 60*, 254-265. https://doi.org/10.1598/RT.60.3.5.

Bronckart, J-P. (1997). *Activité langagière, textes et discours.* Lausanne, Delachaux et Niestlé

Brown, A. (2003). In the beginning was the worm: finding the secrets of life in a tiny hermaphrodite. Columbia University Press

Browning, E., & Hohenstein, J. (2015). The use of narrative to promote primary school children's understanding of evolution. *Education 3-13, 43*, 530-547. https://doi.org/ 10.1080/03004279.2013.837943

Bruner, J., & Lucariello, J. (1989). Monologue as Narrative Recreation. *Narratives from the crib*, *73*.

Bruner, J. (2004). Life as narrative. *Social research: An international quarterly*, *71*(3), 691-710. https://10.1353/sor.2004.0045

Bruner, J. (1986). *Actual minds, possible worlds.* Harvard University Press

Bruner, J. (1988). *Research currents: Life as narrative*. Language Arts, 65(6), 574-583.

Bruner, J. (1990). *Acts of meaning*. Harvard University Press

Bruner, J. (1991). *The narrative construction of reality*. Critical Inquiry, 18(1), 1–21

Bruner, J. (1996). *The culture of education*. Harvard University Press

Cacioppo, J. T., Petty, R. E., & Kao, C. E. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment, 48*, 306–307. https://doi.org/10.1207/s15327752jpa4803_13

Callanan, M., Cervantes, C., & Loomis, M. (2011). Informal learning. *Wiley Interdisciplinary Reviews: Cognitive Science, 2*(6), 646-655. https://doi.org/10.1002/wcs.143

Catrysse, L., Gijbels, D., & Donche, V. (2018). It is not only about the depth of processing: What if eye am not interested in the text?. *Learning and Instruction*, *58*, 284-294. https://doi.org/10.1016/j.learninstruc.2018.07.009

Cervetti, G. N., Bravo, M. A., Hiebert, E. H., Pearson, P. D., & Jaynes, C. A. (2009). Text genre and science content: Ease of reading, comprehension, and reader preference. *Reading Psychology, 30*, 487-511. https://doi.org/10.1080/02702710902733550

Chappell, M. J., & Varelas, M. (2020). Ethnodance and identity: Black students representing science identities in the making. *Science Education, 104*(2), 193-221. https://doi.org/10.1002/sce.21558

Cheng, T. L., & Goodman, E. (2015). Race, ethnicity, and socioeconomic status in research on child health. *Pediatrics, 135*(1), e225-e237. https://doi.org/10.1542/peds.2014-3109

Chi, M. T. H., de Leeuw, N., Chiu, M., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*(3), 439-477. https://doi.org/10.1016/0364-0213(94)90016-7

Claus, B., & Kelter, S. (2006). Comprehending Narratives Containing Flashbacks: Evidence for Temporally Organized Representations. *Journal of Experimental Psychology: Learning, Memory and Cognition, 32,* 1031-1044. https://doi.org/10.1037/0278-7393.32.5.1031

Clough, M. P. (2011). The story behind the science: bringing science and scientists to life in post-secondary science education. *Science & Education, 20*, 701–717. https://doi.org/10.1007/s11191-010-9310-7

Cole, G. G., Samuel, S., & Eacott, M. J. (2022). A return of mental imagery: The pictorial theory of visual perspective-taking. *Consciousness and Cognition*, *102*, 103352. https://doi.org/10.1016/j.concog.2022.103352

Corni F., Giliberti E., and Mariani C. (2010, August 22-27). *A Story as Innovative Medium for Science Education in Primary School* [Paper presentation]. GIREP-ICPE-MPTL International Conference, Reims, France

Corni, F., Giliberti, E., & Fuchs, H. U. (2014). Student teachers writing science stories: A case study. In C. P. Constantinou, N. Papadouris, & A. Hadjigeorgiou (Eds.), *E-Book Proceedings of the ESERA 2013 Conference: Science Education Research For Evidence-based Teaching and Coherence in Learning*. (pp. 2494-2505). European Science Education Research Association

Costa, A. (2021). Da verdade inconveniente à suficiente: cosmopolíticas do antropoceno. C*ognitio-Estudos: revista eletrônica de filosofia, 18*(1), 37-49.

Costa, M. A. (2005). *Processamento de frases em Português Europeu. Aspectos cognitivos e linguísticos implicados na compreensão da língua escrita* [Sentence processing in European Portuguese. Cognitive and linguistic aspects involved in written language comprehension]. Fundação Calouste Gulbenkian

Craik, F. I., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of experimental Psychology: general*, *104*(3), 268. https://doi.org/10.1037/0096-3445.104.3.268

Cromley, J. G., Snyder-Hogan, L. E., & Luciw-Dubas, U. A. (2010). Reading comprehension of scientific text: A domain-specific test of the direct and inferential mediation model of reading comprehension. *Journal of Educational Psychology*, *102*(3), 687. https://doi.org/10.1037/a0019452

Curie, M. (1904). *Radioactive substances*. Chemical News Office.

Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology, 10*, 85. https://doi.org/10.1037/0022-3514.44.1.113

Dawson, E. (2014). “Not designed for us”: How science museums and science centers socially exclude low-income, minority ethnic groups. *Science education, 98*(6), 981-1008. https://doi.org/10.1002/sce.21133

De Pryck, K., & Gemenne, F. (2017). The denier-in-chief: Climate change, science and the election of Donald J. Trump. *Law and Critique*, *28*(2), 119-126. https://doi.org/10.1007/s10978-017-9207-6

de Ruiter, L. E., Lieven, E. V., Brandt, S., & Theakston, A. L. (2020). Interactions between givenness and clause order in children’s processing of complex sentences. *Cognition, 198*, 104130. https://doi.org/10.1016/j.cognition.2019.104130

de Vries, G. J., & Södersten, P. (2009). Sex differences in the brain: the relation between structure and function. *Hormones and behavior, 55(*5), 589-596. https://doi.org/10.1016/j.yhbeh.2009.03.012

Deci, E. L., & Ryan, R. M. (1982). Curiosity and self-directed learning: The role of motivation in education. In L. Katz (Ed.), *Current topics in early childhood education: Volume IV* (pp. 71–85). Albex

Diamond, A. (2013). Executive functions. *Annual review of psychology*, *64*, 135. https://10.1146/annurev-psych-113011-143750

Dib, C. Z. (1988, October). Formal, non-formal and informal education: concepts/applicability. In A*IP conference proceedings* (Vol. 173, No. 1, pp. 300-315). American Institute of Physics. https://doi.org/10.1063/1.37526

Dinsmore, D. L., Zoellner, B. P., Parkinson, M. M., Rossi, A. M., Monk, M. J., & Vinnachi, J. (2017). The effects of different types of text and individual differences on view complexity about genetically modified organisms. International *Journal of Science Education, 39*(7), 791-813. https://doi.org/10.1080/09500693.2017.1298871

Djikic, M., Oatley, K., & Moldoveanu, M. C. (2013). Reading other minds: Effects of literature on empathy. *Scientific study of literature*, *3*(1), 28-47. https://doi.org/10.1075/ssol.3.1.06dji

Dnieprov, A. (1969). The crabs take over the island. In R. Magidoff (Ed.). *Russian science fiction.* University of London Press

Dodell-Feder, D., Lincoln, S. H., Coulson, J.P., & Hooker, C. I. (2013). Using Fiction to Assess Mental State Understanding: A New Task for Assessing Theory of Mind in Adults. *PLoS ONE 8*(11): e81279. https://doi.org/10.1371/journal.pone.0081279

Douglas, K. M., Sutton, R. M., & Cichocka, A. (2017). The psychology of conspiracy theories. *Current directions in psychological science, 26*(6), 538-542. https://10.1177/0963721417718261

Eason, S. H., Goldberg, L. F., Young, K. M., Geist, M. C., & Cutting, L. E. (2012). Reader–text interactions: How differential text and question types influence cognitive skills needed for reading comprehension. *Journal of educational psychology, 104*, 515-528. https://doi.org/10.1037/a0027182

Eddy, C. M. (2019). What do you have in mind? Measures to assess mental state reasoning in neuropsychiatric populations. *Frontiers in psychiatry, 1*0, 425. https://doi.org/10.3389/fpsyt.2019.00425

Eekhof, L. S., Eerland, A., & Willems, R. M. (2018). Readers' insensitivity to tense revealed: No differences in mental simulation during reading of present and past tense stories. *Collabra: Psychology, 4*(1), 16. https://doi.org/10.1525/collabra.121

Egan, K. (1997). *The educated mind: how cognitive tools shape our understanding*. The University of Chicago Press

Egan, K. (2005). *An imaginative approach to teachin*g. Jossey-Bass

ElShafie, S. J. (2018). Making science meaningful for broad audiences through stories. *Integrative and comparative biology, 58*(6), 1213-1223. https://doi.org/10.1093/icb/icy103

Erten, S., Kiray, A., & Sen-Gumus, B. (2013). Influence of scientific stories on students ideas about science and scientists. *International Journal of Education in Mathematics, Science and Technology, 1*, 122–137

Evangelista, P. C. Q. (2008). *Produção de histórias infantis científicas no curso de Pedagogia* [Master's thesis, Universidade de Brasília]. Open Access Theses and Dissertations

Evangelista, P., & Zimmermann, E. (2008, October 21). *Produção de histórias científicas infantis na formação de pedagogos para o ensino de física* [Paper presentation]. XI Encontro de Pesquisa em Ensino de Física, Curitiba, Brasil

Falk, J. H., & Dierking, L. D. (2010). School is not where most Americans learn most of their science. *American Scientist, 98*(6), 486. https://doi.org/10.1511/2010.87.486

Falk, J. H., Storksdieck, M., & Dierking, L. D. (2007). Investigating public science interest and understanding: Evidence for the importance of free-choice learning. *Public understanding of Science, 16*(4), 455-469. https://doi.org/10.1177/0963662506064240

Farah, M. J. (2017). The neuroscience of socioeconomic status: Correlates, causes, and consequences. *Neuron, 96*(1), 56-71. https://doi.org/10.1016/j.neuron.2017.08.034.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior research methods*, *41*(4), 1149-1160. https://doi.org/10.3758/BRM.41.4.1149

Feinstein, N. W., & Meshoulam, D. (2014). Science for what public? Addressing equity in American science museums and science centers. *Journal of Research in Science Teaching, 51*(3), 368-394. https://doi.org/10.1002/tea.21130

Fernandes, J. M., Soares, S., Lopes, R., Jerónimo, R., & Barahona-Corrêa, J. B. (2022). Attribution of intentions in autism spectrum disorder: A study of event-related potentials. *Autism Research, 15*(5) 847-860. https://doi.org/10.1002/aur.2702

Fernandes, T., Araújo, S., Sucena, A., Reis, A., & Castro, S. L. (2017). The 1-min Screening Test for reading problems in college students: Psychometric properties of the 1-min TIL. *Dyslexia, 23*(1), 66-87. https://doi.org/10.1002/dys.1548

Ferstl, E. C. (2001). Learning from text. In N. J. Smelser, & P. B. Baltes (Eds.), *International encyclopedia of the social and behavioral sciences* (pp. 8605-8609). Elsevier. https://doi.org/10.1016/B0-08-043076-7/01478-9

Ferstl, E. C. (2001). Learning from text. In N. J. Smelser, & P. B. Baltes (Eds.), *International encyclopedia of the social and behavioral sciences* (pp. 8605-8609). Elsevier. https://doi.org/10.1016/B0-08-043076-7/01478-9

Field, A. (2013). *Discovering statistics using SPSS.* Sage

Field, H., & Powell, P. (2001). Public understanding of science versus public understanding of research. *Public Understanding of Science, 10*(4), 421–426. https://doi.org/10.3109/a036879

Fisch, S. M. (2000). A capacity model of children's comprehension of educational content on television. *Media psychology*, *2*(1), 63-91. https://doi.org/10.1207/S1532785XMEP0201_4

Fisher, W. (1987). *Human communication as narration: towards a philosophy of reason, value and action*. University of South Carolina Press

Flavell, J. H., Green, F. L., Flavell, E. R., Watson, M. W., & Campione, J. C. (1986). Development of knowledge about the appearance-reality distinction. *Monographs of the society for research in child development*, i-87. https://doi.org/10.2307/1165866

Flynn, S., & Hardman, M. (2019). The Use of Interactive Fiction to Promote Conceptual Change in Science. *Science & Education, 28*, 127-152. https://doi.org/10.1007/s11191-019-00032-6

Fredricks, J. A., Hofkens, T., Wang, M. T., Mortenson, E., & Scott, P. (2018). Supporting girls' and boys' engagement in math and science learning: A mixed methods study. *Journal of Research in Science Teaching*, *55*(2), 271-298. https://doi.org/10.1002/tea.21419

Freire, P. (1970). *Pedagogy of the oppressed.* Continuum

Frisch, J. K. (2010). The stories they'd tell: pre-service elementary teachers writing stories to demonstrate physical science concepts. *Journal of Science Teacher Education, 21,* 703-722. https://doi.org/10.1007/s10972-010-9210-3

Frykman, S. G. (2009). Stories to tell? Narrative tools in museum education texts. *Educational Research, 51*(3), 299-319. https://doi.org/10.1080/00131880903156898

Fuchs, H. (2015). From stories to scientific models and back: Narrative framing in modern macroscopic physics. *International Journal of Science Education, 37*, 934-957. https://doi.org/10.1080/09500693.2015.1025311.

Furlanetto, T., Cavallo, A., Manera, V., Tversky, B., & Becchio, C. (2013). Through your eyes: incongruence of gaze and action increases spontaneous perspective taking. *Frontiers in human neuroscience*, *7*, 455. https://doi.org/10.3389/fnhum.2013.00455

Gagné, R. M., & White, R. T. (1978). Memory structures and learning outcomes. *Review of Educational Research*, *48*(2), 187-222. https://doi.org/10.3102/0034654304800218

Garcia-Marques, T., & Mackie, D. M. (2001). The feeling of familiarity as a regulator of persuasive processing. *Social Cognition, 19*(1), 9-34. https://doi.org/10.1521/soco.19.1.9.18959

Garner, R., Gillingham, M. G., & White, C. S. (1989). Effects of "seductive details" on macroprocessing and microprocessing in adults and children. *Cognition and Instruction, 6*, 41–57. https://doi.org/10.1207/s1532690xci0601_2

Gaskell, G. (2000). Individual and group interviewing. *Qualitative researching with text, image and sound, 38-56*.

Gayraud, F., & Martinie, B. (2008). Does structural complexity necessarily imply processing difficulty?. *Journal of Psycholinguistic Research*, *37*(1), 21-31. https://doi.org/10.1007/s10936-007-9057-8

Gilbert, J., Hipkins, R., & Cooper, G. (2005, May 30 – June 1). *Faction or fiction: Using narrative pedagogy in school science education* [Paper presentation]. Redesigning Pedagogy: Research, Policy, Practice conference, Nanyang, Singapore

Glaser, M., Garsoffky, B., & Schwan, S. (2009). Narrative-based learning: Possible benefits and problems. *Communications, 34,* 429-447. https://doi.org/10.1515/COMM.2009.026

Gomes, A., Santos, J. V. D., Gonçalves, G., Orgambidez-Ramos, A., & Giger, J. C. (2013). Validation Study of the Need for Cognition Scale with a Portuguese sample. *Avaliação Psicológica*, *12*(2), 179-192

Gómez-Merino, N., Fajardo Bravo, I., Ferrer Manchón, A. M., & Joseph, H. (2022). Eye movements of deaf students in expository versus narrative texts. *American Annals of the Deaf*. https://doi.org/10.1353/aad.2022.0030

Gomez-Zweip, S., & Straits, W. (2006). Analyzing anthropomorphisms. *Science and Children, 44*(3), 26–29

Gonçalves, M. (2019). Quelques réflexions autour des activités de langage, des genres textuels et de l ´ implicite. *L'implicite dans les langues romanes*, 66-71

Gottschall, J. (2013). *The Storytelling Animal. How Stories Make Us Human*. Mariner Books

Graesser, A. C., Hauft-Smith, K., Cohen, A. D., & Pyles, L. D. (1980). Advanced Outlines, Familiarity, and Text Genre on Retention of Prose. *The Journal of Experimental Education, 48*, 281-290. https://doi.org/10.1080/00220973.1980.11011745

Graesser, A. C., Hauft-Smith, K., Cohen, A. D., & Pyles, L. D. (1980). Advanced Outlines, Familiarity, and Text Genre on Retention of Prose. *The Journal of Experimental Education, 48*, 281-290. https://doi.org/10.1080/00220973.1980.11011745

Graesser, A. C., Golding, J. M., & Long, D. L. (1991). Narrative representation and comprehension. In R. Barr, M. L. Kamil, P. B. Mosenthal, & P. D. Pearson (eds.), *Handbook of Reading Research,* Vol. 2 (pp. 171–205). Lawrence Erlbaum Associates, Inc

Graesser, A. C., Olde, B., Klettke, B. (2002). How does the mind construct and represent stories? In M. Green, J. Strange, & T. Brock (Eds.), *Narrative impact* (pp. 229-262). Lawrence Erlbaum Associates

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review, 101*, 371–395. https://doi.org/10.1037/0033-295X.101.3.371

Green, M. C., & Brock, T. C. (2000). The role of transportation in the persuasiveness of public narratives. *Journal of Personality and Social Psychology, 79*, 701–721. https://doi.org/10.1037//0022-3514.79.5.701

Hadzigeorgiou, Y. (2016). *Imaginative Science Education. The Central Role of Imagination in Science Education*. Springer International

Hadzigeorgiou, Y., & Schulz, R. M. (2019). Engaging Students in Science: The Potential Role of "Narrative Thinking" and "Romantic Understanding". In *Frontiers in education* (Vol. 4, p. 38). https://doi.org/10.3389/feduc.2019.00038

Hadzigeorgiou, Y., Klassen, S., & Klassen, C. F. (2012). Encouraging a 'romantic understanding' of science: the effect of the Nikola Tesla story. *Science & Education, 21*, 1111–1138. https://doi.org/10.1007/s11191-011-9417-5

Halliday, M., Mill, D., Johnson, J., & Lee, K. (2022). Online focus group methodology: Recruitment, facilitation, and reimbursement. In Contemporary Research Methods in Pharmacy and Health Services (pp. 433-445). Academic Press.

Hazelkorn, E., Ryan, C., Beernaert, Y., Constantinou, C. P., Deca, L., Grangeat, M., & Welzel-Breuer, M. (2015). *Science education for responsible citizenship. Report to the European Commission of the Expert Group on Science Education*. Publications Office of the European Union. https://doi.org/10.2777/12626.

Hellstrand, A., & Ott, A. (1995). The utilization of fiction when teaching the theory of relativity. *Physics Education, 30*(5), 284

Herodotou, C., Aristeidou, M., Sharples, M., & Scanlon, E. (2018). Designing citizen science tools for learning: lessons learnt from the iterative development of nQuire. *Research and Practice in Technology Enhanced Learning*, *13*(1), 1-23. https://doi.org/10.1186/s41039-018-0072-1

Heyvaert, M., Maes, B., & Onghena, P. (2013). Mixed methods research synthesis: definition, framework, and potential. *Quality & Quantity, 47*(2), 659-676. https://doi.org/10.1007/s11135-011-9538-6

Hidi, S., Baird, W., & Hildyard, A. (1982). That's important but is it interesting? Two factors in text processing. In A. Flammer, & W. Kintsch (Eds.), *Advances in psychology: Discourse processing* (Vol. 8, pp. 63-75). North-Holland. https://doi:10.1016/S0166-4115(08)62681-3

Hoffman, R. (2014). The tensions of scientific storytelling. *American Scientist, 102*, 250-253.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & van de Weijer, J. (2017). *Eye tracking: A comprehensive guide to methods*. Oxford University Press

Hong, H.-Y., & Lin-Siegler, X. (2012). How learning about scientists' struggles influences students' interest and learning in physics. *Journal of educational psychology, 104*, 469-484. https://doi.org/10.1037/a0026224Hadzigeorgiou, 2016

Horowitz-Kraus, T., Vannest, J. J., & Holland, S. K. (2013). Overlapping neural circuitry for narrative comprehension and proficient reading in children and adolescents. *Neuropsychologia*, *51*(13), 2651-2662

Hyönä, J., Lorch, R. F., & Kaakinen, J. K. (2002). Individual differences in reading to summarize expository text: Evidence from eye fixation patterns. *Journal of Educational Psychology, 94*(1), 44–55. https://doi.org/10.1037//0022-0663.94.1.44

Hyönä, J., Lorch, R. F., & Rinck, M. (2003). Eye movement measures to study global text processing. In J. Hyönä, R. Radach, & H. Deubel (Eds.). *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 313–334). Amsterdam: Elsevier Science

Imuta, K., Henry, J. D., Slaughter, V., Selcuk, B., & Ruffman, T. (2016). Theory of mind and prosocial behavior in childhood: A meta-analytic review. *Developmental Psychology, 52*(8), 1192–1205. https://doi.org/10.1037/dev0000140

Jackson, M. C., Galvez, G., Landa, I., Buonora, P., & Thoman, D. B. (2016). Science that matters: The importance of a cultural connection in underrepresented students' science pursuit. *CBE—Life Sciences Education, 15*(3), ar42. https://doi.org/10.1187/cbe.16-01-0067

JASP Team (2022). *JASP (Version 0.16.3)* [Computer software]

Jerónimo, R. & Reis, F. (2016). Uma comunidade de prática de integração da investigação no ensino. *Jornadas da Escola de Ciências Sociais e Humanas do ISCTE-IUL*.

Jetton, T. L. (1994). Information-driven versus story-driven: what children remember when they are reading informational stories. *Reading Psychology, 15*, 109–130. https://doi.org/10.1080/0270271940150203

Joel, D., Berman, Z., Tavor, I., Wexler, N., Gaber, O., Stein, Y., Shefi, N., Pool, J., Urchs, S., Margulies, D. S., Liem, F., Hänggi, J., Jäncke, L., & Assaf, Y. (2015). Sex beyond the genitalia: The human brain mosaic. *Proceedings of the National Academy of Sciences, 112*(50), 15468-15473. https://doi.org/10.1073/pnas.1509654112

Johnson, D. R. (2011). Women of color in science, technology, engineering, and mathematics (STEM). *New Directions for Institutional Research*, *2011*(152), 75-85. https://doi.org/10.1002/ir.410

Johnson, D. R. (2012). Transportation into a story increases empathy, prosocial behavior, and perceptual bias toward fearful expressions. *Personality and individual differences*, *52*(2), 150-155. https://doi.org/10.1016/j.paid.2011.10.005

Kalogiannakis, M., & Violintzi, A. (2012). Intervention strategies for changing pre-school children's understandings about volcanoes. *Journal of Emergent Science, 4*, 12-19

Kane, L. (2004). Educators, learners and active learning methodologies. *International Journal of Lifelong Learning 23*, 275–86. https://doi.org/10.1080/0260/37042000229237

Karimi, S., Pohl, S., Scholer, F., Cavedon, L., & Zobel, J. (2010). Boolean versus ranked querying for biomedical systematic reviews. *BMC medical informatics and decision making, 10*, 58. https://doi.org/10.1186/1472-6947-10-58

Karpicke, J. D., & Grimaldi, P. J. (2012). Retrieval-based learning: A perspective for enhancing meaningful learning. *Educational Psychology Review, 24*(3), 401-418. https://doi.org/10.1007/s10648-012-9202-2

Kelsey, E., & Dillon, J. (2016). 'If the Public Knew Better, They Would Act Better': The Pervasive Power of the Myth of the Ignorant Public. In *Towards a Convergence Between Science and Environmental Education* (pp. 323-336). Routledge.

Kerby, H. W., DeKorver, B. K., & Cantor, J. (2018). Fusion Story Form: a novel, hybrid form of story that promotes and assesses concept learning. *International Journal of Science Education, 40*(14), 1774-1794. https://doi.org/10.1080/09500693.2018.1512172

Kidd, D., & Castano, E. (2016). Reading literary fiction and theory of mind: Three preregistered replications and extensions of Kidd and Castano (2013). *Social Psychological and Personality Science, 10*(4), 522-531. https://doi.org/10.1177/194855061877541

Kim, Y. S. G. (2015). Developmental, component-based model of reading fluency: An investigation of predictors of word-reading fluency, text-reading fluency, and reading comprehension. *Reading research quarterly, 50*(4), 459-481. https://doi.org/10.1002/rrq.107

King, B., Borland, R., Morphett, K., Gartner, C., Fielding, K., O'Connor, R. J., Romijnders, K., & Talhout, R. (2021). 'It's all the other stuff!'How smokers understand (and misunderstand) chemicals in cigarettes and cigarette smoke. *Public Understanding of Science*, *30*(6), 777-796. https://doi.org/10.1177/0963662521991351

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review, 95*(2), 163–182. https://doi.org/10.1037/0033-295X.95.2.163

Kintsch, W. (1994). Text comprehension, memory, and learning. *American Psychologist, 49*, 294-303. https://doi.org/ 10.1037//0003-066x.49.4.294

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge University Press

Kintsch, W. (2012). Psychological models of reading comprehension and their implications for assessment. In J. P. Sabatini, E. R. Albro, & T. O'Reilly (Eds.), *Measuring up: Advances in how we assess reading ability* (pp. 21–38). Rowman & Littlefield Education

Kintsch, W., & Rawson, K. A. (2005). Comprehension. In M. J. Snowling & C. Hulme (Eds.), Blackwell handbooks of developmental psychology. *The science of reading: A handbook* (p. 209–226). Blackwell Publishing. https://doi.org/10.1002/ 9780470757642.ch12

Kintsch, W., & Van Dijk, T. A. (1975). *Comment on se rappelle et on résume des histoires*. Langages, (40), 98-116

Kintsch, W., & Young, S. R. (1984). Selective recall of decision-relevant information from texts. *Memory & cognition, 12*, 112-117. https://doi.org/10.3758/BF03198424

Klassen, S. (2006). A theoretical framework for contextual science teaching. *Interchange, 37*(1), 31-62. https://doi.org/10.1007/s10780-006-8399-8

Klassen, S. (2007). The application of historical narrative in science learning: The Atlantic cable story. *Science & Education, 16*, 335–352. https://doi.org/10.1007/s11191-006-9026-x

Klassen, S. (2009). The construction and analysis of a science story: A proposed methodology. *Science & Education, 18,* 401-423. https://doi.org/10.1007/s11191-008-9141-y

Klein, P. D. (2006). The challenges of scientific literacy: From the viewpoint of second-generation cognitive science. *International Journal of Science Education, 28,* 143-178. https://doi.org/10.1080/09500690500336627

Klein, S. B., & Loftus, J. (1988). The nature of self-referent encoding: The contributions of elaborative and organizational processes. *Journal of personality and social psychology*, *55*(1), 5. https://doi.org/10.1037/0022-3514.55.1.5

Kloser, M. (2013). Exploring high school biology students' engagement with more and less epistemologically considerate texts. *Journal of Research in Science Teaching, 50*, 1232-1257. https://doi.org/10.1002/tea.21109

Knudsen, J. K., Wold, L. C., Aas, Ø., Haug, J. J. K., Batel, S., Devine-Wright, P., Qvenild, M., & Jacobsen, G. B. (2015). Local perceptions of opportunities for engagement and procedural justice in electricity transmission grid projects in Norway and the UK. *Land use policy, 48,* 299-308. https://doi.org/10.1016/J.LANDUSEPOL.2015.04.031

Kokkotas, P., Rizaki, A., & Malamitsa, K. (2010). Storytelling as a strategy for understanding concepts of electricity and electromagnetism. Interchange, 41, 379–405. https://doi.org/10.1007/s10780-010-9137-9

Kolinsky, R. (2015). How learning to read influences language and cognition. In A. Pollatsek & R. Treiman (Eds.), *The Oxford Handbook of Reading*. New York, NY: Oxford University Press (pp. 377–393). https://doi.org/10.1093/oxfordhb/9780199324576.013.29.

Korinth, S. P., & Fiebach, C. J. (2018). Improving silent reading performance through feedback on eye movements: A feasibility study. *Scientific Studies of Reading*, *22*(4), 289-307. https://doi.org/10.1080/10888438.2018.1439036

Kraal, A., van den Broek, P. W., Koornneef, A. W., Ganushchak, L. Y., & Saab, N. (2019). Differences in text processing by low-and high-comprehending beginning readers of expository and narrative texts: Evidence from eye movements. *Learning and Individual Differences*, *74*, 101752. https://doi.org/10.1016/j.lindif.2019.101752

Krause, N. M., Brossard, D., Scheufele, D. A., Xenos, M. A., & Franke, K. (2019). Trends—Americans' trust in science and scientists. *Public Opinion Quarterly, 83*(4), 817-836. https://doi.org/10.1093/poq/nfz041

Krueger, R., & Casey, M. (2009). *Focus groups: A practical guide for applied research.* Sage Publications

Kubli, F. (2001). Can the theory of narrative help science teachers become better storytellers? *Science & Education, 10*, 595–599. https://doi.org/10.1007/978-94-010-0730-6_12

Kubli, F. (2005). Science teaching as a dialogue–Bakhtin, Vygotsky and some applications in the classroom. *Science & Education, 14*(6), 501-534. https://doi.org/10.1007/s11191-004-8046-7

Kuijpers, M. M., Hakemulder, F., Tan, E. S., & Doicaru, M. M. (2014). Exploring absorbing reading experiences: Developing and validating a self-report scale to measure story world absorption. *Scientific Study of Literature, 4*(1), 89–122. https://doi.org/10.1075/ssol.4.1.05kui

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in psychology, 4,* 863. https://doi.org/10.3389/fpsyg.2013.00863

Lamartino, A. (1995). *Science and reading* [Unpublished master's thesis]. Kean University

Larison, K. D. (2018). Taking the Scientist's Perspective. *Science & Education*, *27*(1-2), 133-157. https://doi.org/10.1007/s11191-018-9957-z

Lee, O., & Luykx, A. (2007). Science education and student diversity: Race/ethnicity, language, culture, and socioeconomic status. *Handbook of research on science education, 1*, 171-197

Legare, C. H., Lane, J. D., & Evans, E. M. (2013). Anthropomorphizing science: How does it affect the development of evolutionary concepts?. *Merrill-Palmer Quarterly, 59*(2), 168-197. https://doi.org/10.13110/merrpalmquar1982.59.2.0168

Lehman, S., Schraw, G., McCrudden, M. T., & Hartley, K. (2007). Processing and recall of seductive details in scientific text. *Contemporary Educational Psychology, 32,* 569–587. https://doi.org/10.1016/j.cedpsych.2006.07.002Lamartino, 1995

Leipzig, P. (2018). *What's the story?: the effects of narratives in science classrooms* [Doctoral dissertation, Colorado State University]. Open Access Theses and Dissertations

Lemke, J. (1998). Multiplying meaning: Visual and verbal semiotics in scientific text. In J. R. Martin & R. Veel (Eds.), ience (pp. 87–113). Routledge.

Levi, P. (1985). *The Periodic Table.* Penguin Books

Li, N., & Qian, Y. (2022). Polarization of public trust in scientists between 1978 and 2018: insights from a cross-decade comparison using interpretable machine learning. *Politics and the Life Sciences, 41*(1), 45-54. https://doi.org/10.1017/pls.2021.18

Limpo, T., Alves, R. A., & Fidalgo, R. (2014). Children's high-level writing skills: development of planning and revising and their contribution to writing quality. *British Journal of Educational Psychology, 84*, 177-193. https://doi.org/10.1111/bjep.12020

Lin-Siegler, X., Ahn, J. N., Chen, J., Fang, F. F. A., & Luna-Lucero, M. (2016). Even Einstein struggled: Effects of learning about great scientists' struggles on high school students' motivation to learn science. *Journal of Educational Psychology, 108*(3), 314. https://doi.org/10.1037/edu0000092

Linderholm, T., & van den Broek, P. (2002). The effects of reading purpose and working memory capacity on the processing of expository text. *Journal of educational psychology, 94*(4), 778. https://doi.org/1778.0.1037//0022-0663.94.4.778

Logan, M., & Skamp, K. (2008). Engaging students in science across the primary secondary interface: Listening to the students' voice. *Research in science education*, *38*, 501-527.

Lord, C. (2002). Are subordinate clauses more difficult. *Complex sentences in grammar and discourse*, 224-233.

Luna, R. E. (2015). The storytelling scientist. *Science, 350*(6259), 391–391. https://doi.org/10.1126/science.aad2179

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide*. Lawrence Erlbaum.

Mak, M., & Willems, R. M. (2019). Mental simulation during literary reading: Individual differences revealed with eye-tracking. *Language, Cognition and Neuroscience*, *34*(4), 511-535. https://doi.org/10.1080/23273798.2018.1552007

Malone, K. R., & Barabino, G. (2009). Narrations of race in STEM research settings: Identity formation and its discontents. *Science Education, 93*(3), 485-510. https://doi.org/10.1002/sce.20307

Mandler, J. M., & Johnson, N. S. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive psychology, 9*(1), 111-151.

Mandler, J. M. (1984). *Stories, scripts, and scenes: Aspects of schema theory*. Erlbaum.

Mangen, A., Walgermo, B. R., & Brønnick, K. (2013). Reading linear texts on paper versus computer screen: Effects on reading comprehension. *International journal of educational research, 58*, 61-68. https://doi.org/10.1016/j.ijer.2012.12.002

Mar, R. A. (2004). The neuropsychology of narrative: Story comprehension, story production and their interrelation. *Neuropsychologia, 42*, 1414-1434. https://doi.org/10.1016/ j.neuropsychologia.2003.12.016

Mar, R. A., & Oatley, K. (2008). The function of fiction is the abstraction and simulation of social experience. *Perspectives on psychological science*, *3*(3), 173-192. https://doi.org/10.1111/j.1745-6924.2008.00073.x

Mar, R. A., Oatley, K., Hirsh, J., Dela Paz, J., & Peterson, J. B. (2006). Bookworms versus nerds: Exposure to fiction versus non-fiction, divergent associations with social ability, and the simulation of fictional social worlds. *Journal of research in personality, 40*(5), 694-712. https://doi.org/10.1016/j.jrp.2005.08.002

Maria, K., & Johnson, J. M. (1989, November 28 – December 2). *Correcting misconceptions: Effects of Type of Text*. [Paper presentation]. Annual meeting of the National Reading Conference, Austin, TX, United States of America

Maria, K., & Junge, K. (1993, December). *A comparison of fifth graders' comprehension and retention of scientific information using a science textbook and an informational storybook* [Paper presentation]. Annual Meeting of the National Reading Conference, Charleston, SC, USA

Marleau, J. N., & Girling, K. D. (2017). Keeping science's seat at the decision-making table: Mechanisms to motivate policy-makers to keep using scientific information in the age of disinformation. *Facets, 2*(1), 1045-1064. https://doi.org/10.1139/facets-2017-0087.

Martin, B. E., & Brouwer, W. (1991). The sharing of personal science and the narrative element in science education. *Science Education, 75*(6), 707-22. https://doi.org/10.1002/sce.3730750610

Martinez-Conde, S., Alexander, R. G., Blum, D., Britton, N., Lipska, B. K., Quirk, G. J., Macknik, S. L. (2019). The Storytelling Brain: How Neuroscience Stories Help Bridge the Gap between Research and Society. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 39*(42), 8285–8290. https://doi.org/10.1523/JNEUROSCI.1180-19.2019

Martins-Loução, M. A., Gaio-Oliveira, G., Barata, R., & Carvalho, N. (2020). Inquiry-based science learning in the context of a continuing professional development programme for biology teachers. *Journal of Biological Education, 54*(5), 497-513. https://doi.org/10.1080/00219266.2019.1609566

Mason, L., Pluchino, P., & Tornatora, M. (2013). Effects of Picture Labeling on Science Text Processing and Learning: Evidence From Eye Movements. *Reading Research Quarterly, 48*(2), 199–214. https://doi:10.1002/rrq.41

Mason, R. A., & Just, M. A. (2009). The role of the theory-of-mind cortical network in the comprehension of narratives. *Language and Linguistics Compass, 3*(1), 157-174. https://doi.org/10.1111/j.1749-818X.2008.00122.x

Matthews, M. (1994). *Science teaching: The role of history and philosophy of science*. Routledge.

May, L., Crisp, T., Bingham, G. E., Schwartz, R. S., Pickens, M. T., & Woodbridge, K. (2020). The durable, dynamic nature of genre and science: A purpose-driven typology of science trade books. *Reading Research Quarterly, 55*(3), 399-418. https://doi.org/10.1002/rrq.274

McNamara, D. S., & Magliano, J. (2009). Toward a comprehensive model of comprehension. *Psychology of learning and motivation, 51*, 297-384. https://doi.org/10.1016/S0079-7421(09)51009-2

McNamara, D. S., Ozuru, Y., & Floyd, R. G. (2011). Comprehension challenges in the fourth grade: The roles of text cohesion, text genre, and readers' prior knowledge. *International electronic journal of elementary education, 4*(1), 229-257.

McQuiggan, S. W., Rowe, J. P., Lee, S., & Lester, J. C. (2008). Story-Based Learning: The Impact of Narrative on Learning Experiences and Outcomes. In: B.P. Woolf, E. Aïmeur, R. Nkambou, & S. Lajoie (Eds.) *Intelligent Tutoring Systems. ITS 2008. Lecture Notes in Computer Science* (vol 5091, pp. 530-539). Springer. https://doi.org/10.1007/978-3-540-69132-7_56

McQuiggan, S. W., Rowe, J. P., Lee, S., & Lester, J. C. (2008). Story-Based Learning: The Impact of Narrative on Learning Experiences and Outcomes. In: B.P. Woolf, E. Aïmeur, R. Nkambou, & S. Lajoie (Eds.) *Intelligent Tutoring Systems. ITS 2008. Lecture Notes in Computer Science* (vol 5091, pp. 530-539). Springer. https://doi.org/10.1007/978-3-540-69132-7_56

Meyer, B. J. (1975). *The organization of prose and its effects on memory* (Vol 1). North-Holland Publishing Company

Mikkilä-Erdmann, M., Penttinen, M., Anto, E., & Olkinuora, E. (2008). Constructing mental models during learning from science text. In *Understanding models for learning and instruction* (pp. 63-79). Springer, Boston, MA.

Montgomery, S. L. (1996). *The scientific voice.* Guilford Press.

Morais, C. (2015). Storytelling with chemistry and related hands-on activities: Informal learning experiences to prevent ‘chemophobia’ and promote young children’s scientific literacy. *Journal of Chemical Education, 92*, 58–65. https://doi.org/10.1021/ed5002416

Morais, C., Araújo, J. L., & Saúde, I (2019). Awakening to chemistry through storytelling and practical activities: middle school students interacting with pre-school children. *Chemistry Education Research and Practice,* 20(1), 302-315. https://doi.org/10.1039/C8RP00096D

Morais, J. (2015). *Lire, écrire et être libre: de l'alphabétisation à la démocratie*. Odile Jacob.

Morais, J., & Kolinsky, R. (2016). Literacia científica: leitura e produção de textos científicos. *Educar em Revista*, (62), 143-162

Morais, J., & Kolinsky, R. (2021). Seeing thought: a cultural cognitive tool. *Journal of Cultural Cognitive Science*, *5*(2), 181-228. https://doi.org/10.1007/s41809-020-00059-0

Murmann, M., & Avraamidou, L. (2014a). Narrative as a learning tool in science centers: Potentials, possibilities and merits. *Journal of Science Communication, 13*(2), A02. https://doi.org/10.22323/2.13020202

Murmann, M., & Avraamidou, L. (2014b). Animals, Emperors, Senses: Exploring a story-based learning design in a Museum setting. *International Journal of Science Education, Part B, 4*, 66-91. https://doi.org/10.1080/21548455.2012.760857

Mutonyi, H. (2016). Stories, proverbs, and anecdotes as scaffolds for learning science concepts. *Journal of Research in Science Teaching, 53,* 943–971. https://doi.org/10.1002/tea.21255

National Center for Education Statistics. (2022). Science Performance. *Condition of Education.* U.S. Department of Education, Institute of Education Sciences. https://nces.ed.gov/programs/coe/indicator/cne

Negrete A., (2005). Facts via Fiction Stories that Communicate Science. In N. Sannit (Ed.), *Motivating Science. Science communication from a philosophical, educational and cultural perspective* (pp. 95-102). Luthon

Negrete, A., & Lartigue, C. (2004). Learning from education to communicate science as a good story. *Endeavour, 28*(3), 120-124. https://doi.org/10.1016/j.endeavour.2004.07.003

Negrete, A., & Lartigue, C. (2010). The science of telling stories: Evaluating science communication via narratives (RIRC method). *Journal Media and Communication Studies, 2*, 98-110. https://doi.org/10.5897/JMCS.9000080

Negretti, R., Persson, M., & Sjöberg-Hawke, C. (2022). Science stories: researchers' experiences of writing science communication and the implications for training future scientists. *International Journal of Science Education*, Part B, 1-18. https://doi.org/10.1080/21548455.2022.206053

Norris, S. P., & Phillips, L. M. (2003). How literacy in its fundamental sense is central to scientific literacy. *Science Education, 87*(2), 224-240

Norris, S. P., Guilbert, S. M., Smith, M. L., Hakimelahi, S., & Phillips, L. M. (2005). A theoretical framework for narrative explanation in science. *Science Education, 89*(4), 535–563

Oatley, K. (2016). Fiction: Simulation of social worlds. *Trends in cognitive sciences*, *20*(8), 618-628. https://doi.org/10.1016/j.tics.2016.06.002

Organisation for Economic Cooperation and Development (OECD). (2019). *PISA 2018 results: What students know and can do* (Vol. I)

Oker, A., Del Goleto, S., Vignes, A., Passerieux, C., Roux, P., & Gouet, E. B. (2019). Schizophrenia patients are impaired in recognition task but more for intentionality than physical causality. *Consciousness and Cognition*, *67*, 98-107. https://doi.org/10.1016/j.concog.2018.11.007

Olson, R. (2015). *Houston, we have a narrative: Why science needs story*. University of Chicago Press

Orquin, J. L., & Holmqvist, K. (2019). A primer on eye-tracking methodology for behavioral science. In *A handbook of process tracing methods* (pp. 53-64). Routledge

Pandya, R. E., and Dibner, K. A. (2018). *Learning Through Citizen Science: Enhancing Opportunities by Design*. National Academies Press

Pappas, C. C. (2006). The information book genre: Its role in integrated science literacy research and practice. *Reading Research Quarterly, 41*(2), 226-250. https://doi.org/10.1598/RRQ.41.2.4

Pearson, P. D., & Cervetti, G. N. (2015). Fifty years of reading comprehension theory and practice. Research-based practices for teaching Common Core literacy, 1-24.

Penttinen, M., Anto, E., & Mikkilä-Erdmann, M. (2013). Conceptual change, text comprehension and eye movements during reading. *Research in Science Education*, *43*(4), 1407-1434. https://doi.org/10.1007/s11165-012-9313-2

Phillips Galloway, E., McClain, J.B., & Uccelli, P. (2020). Broadening the lens on the science of reading: A multifaceted perspective on the role of academic language in text understanding. *Reading Research Quarterly, 55*(S1), S331–S345. https://doi.org/10.1002/rrq.359

Phillips, L. M., & Norris, S. P. (2009). Bridging the gap between the language of science and the language of school science through the use of adapted primary literature. *Research in Science Education, 39*(3), 313-319. https://doi.org/10.1007/s11165-008-9111-z

Plavén-Sigray, P., Matheson, G. J., Schiffler, B. C., & Thompson, W. H. (2017). The readability of scientific texts is decreasing over time. *Elife, 6*, e27725.

Popper, K. (1972). *The Logic of Scientific Discovery*. Hutchinson.

Postolache, S., Torres, R., Afonso, A. P., Carmo, M. B., Cláudio, A.P., Domingos, D., Ferreira, A., Barata, R., Carvalho, P., Coelho, A.G., Duarte, M.C., Garcia, C., Leal, A.I. & Redweik, P. (2022). Contributions to the design of mobile applications for visitors of Botanical Gardens. *Procedia Computer Science, 196*, 389-399. https://doi.org/10.1016/j.procs.2021.12.028

Prins, R., Avraamidou, L., & Goedhart, M. (2017). Tell me a story: the use of narrative as a learning tool for natural selection. *Educational Media International, 54*(1), 20-33. https://doi.org/ 10.1080/09523987.2017.1324361

Rastier, F. (2001). *Arts et sciences du texte*. Presses universitaires de France

Ravelli, L. (2007). *Museum Texts: Comunication Frameworks.* Routledge

Rayner, K. (2009). Eye movements in reading: Models and data. *Journal of eye movement research*, *2*(5), 1

Rayner, K., Warren, T., Juhasz, B. J., & Liversedge, S. P. (2004). The Effect of Plausibility on Eye Movements in Reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(6), 1290–1301. https://doi.org/10.1037/0278-7393.30.6.1290

Reif, A., & Guenther, L. (2021). How representative surveys measure public (dis) trust in science: A systematisation and analysis of survey items and open-ended questions. *Journal of Trust Research*, 1-25. https://doi.org/10.1080/21515581.2022.2075373

Rennie, L. J., Feher, E., Dierking, L. D., & Falk, J. H. (2003). Toward an agenda for advancing research on science learning in out-of-school settings. *Journal of research in science teaching, 40*(2), 112-120.

Reuer, M. D. (2012). *Backroads to learning: the use of narratives in high school biology* [Master's Thesis]. https://scholarworks.montana.edu/xmlui/handle/1/2123

Ricoeur, Paul, 1992. *Oneself as Another.* The University of Chicago Press

Ritchie, S. M., Tomas, L., & Tones, M. (2011). Writing stories to enhance scientific literacy. *International Journal of Science Education, 33*, 685-707. https://doi.org/ 10.1080/09500691003728039

Ritchie, S., Rigano, D., & Duane, A. (2008). Writing an ecological mystery in class: Merging genres and learning science. *International Journal of Science Education, 30*, 143-166. https://doi.org/ 10.1080/09500690601161783

Roberts, D. A. (2007). Scientific literacy/science literacy. In S. K. Abell, & N. G. Lederman (Eds.), *Handbook of Research on Science Education* (pp. 729-780). Lawrence Erlbaum Associates

Roediger, H.L., & Karpicke, J.D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*(3), 181-210. https://doi.org/10.1111/j.1745-6916.2006.00012.x

Rosa, V. (2010). *Ciência em forma de História para a compreensão dos conceitos científicos* [Doctoral dissertation, Universidade Regional de Blumenau]. Open Access Theses and Dissertations

Roux, P., Brunet-Gouet, E., Passerieux, C., & Ramus, F. (2016). Eye-tracking reveals a slowdown of social context processing during intention attribution in patients with schizophrenia. *Journal of Psychiatry and Neuroscience*, *41*(2), E13-E21. https://doi.org/10.1503/jpn.150045

Rowe, M. P., Gillespie, B. M., Harris, K. R., Koether, S. D., Shannon, L. J. Y., & Rose, L. A. (2015). Redesigning a general education science course to promote critical thinking. *CBE—Life Sciences Education*, *14*(3), ar30. https://doi.org/10.1187/cbe.15-02-0032

Rukeyser, M. (1968). *The Speed of Darkness*. Random House

Ryan, M. L. (2001). *Narrative as Virtual Reality: Immersion and Interactivity in literature and Electronic Media*. Johns Hopkins University Press.

Sadler, T. D. (2009). Situated learning in science education: socio-scientific issues as contexts for practice. *Studies in science Education, 45*(1), 1-42. https://doi.org/10.1080/03057260802681839

Sadoski, M., Goetz, E. T., Olivarez Jr, A., Lee, S., & Roberts, N. M. (1990). Imagination in story reading: The role of imagery, verbal recall, story analysis, and processing levels. *Journal of Reading Behavior*, *22*(1), 55-70. https://doi.org/10.1080/10862969009547694

Sanacore, J. (1991). Expository and narrative text: Balancing young children's reading experiences. *Childhood Education*, *67*(4), 211-214.

Sánchez Tapia, I., Krajcik, J., & Reiser, B. (2018). “We do not know what is the real story anymore” : Curricular contextualization principles that support indigenous students in understanding natural selection. *Journal of Research in Science Teaching, 55*(3), 348–376 https://doi.org/10.1002/tea.21422

Santos, M. D. L. L. D., Neves, J. S., Lima, M. J. S. A. P. D., & Carvalho, M. D. D. (2007). *A leitura em Portugal* [Reading in Portugal]. GEPE-Gabinete de Estatística e Planeamento da Educação

Scells, H., & Zuccon, G. (2018, July 8-12). Generating better queries for systematic reviews. In *The 41st international ACM SIGIR conference on research & development in information retrieval* (pp. 475-484). https://doi.org/10.1145/3209978.3210020

Schiefele, U. (1996). Topic interest, text representation, and quality of experience. *Contemporary Educational Psychology, 21*, 3–18. https://doi.org/0361-476X/96

Schiefele, U., & Krapp, A. (1996). Topic interest and free recall of expository text. *Learning and Individual Differences, 8*(2), 141–160. https://doi.org/10.1016/S1041-6080(96)90030-8

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime reference guide.* Psychology Sofware Tools Inc

Schotter, E. R., Tran, R., & Rayner, K. (2014). Don’t believe what you read (only once) comprehension is supported by regressions during reading. *Psychological science, 25*(6), 1218-1226. https://doi.org/10.1177/0956797614531148

Shamay-Tsoory, S. G., & Aharon-Peretz, J. (2007). Dissociable prefrontal networks for cognitive and affective theory of mind: a lesion study. *Neuropsychologia*, *45*(13), 3054-3067 https://doi.org/10.1016/j.neuropsychologia.2007.05.021

Shanahan, M. C. (2012). Reading for evidence in hybrid adapted primary literature. In S. P. Norris (Ed.), *Reading for evidence and interpreting visualizations in mathematics and science education* (pp. 41–63). Sense Publishers.

Silva, H. M., & Prado, I. G. O. (2010). Creationism and intelligent design: Presence in the Brazilian educational policy. *Procedia-Social and Behavioral Sciences*, *2*(2), 5260-5264. https://doi.org/10.1016/j.sbspro.2010.03.856.

Sinatra, G. M., & Broughton, S. H. (2011). Bridging reading comprehension and conceptual change in science education: The promise of refutation text. *Reading Research Quarterly, 46*(4), 374-393. https://doi.org/10.1002/RRQ.005

Sinatra, G. M., Broughton, S. H., & Lombardi, D. O. U. G. (2014). Emotions in science education. In R. Pekrun, & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 415–436). Routlege

Skydsgaard, M. A., Møller Andersen, H., & King, H. (2016). Designing museum exhibits that facilitate visitor reflection and discussion. *Museum Management and Curatorship, 31,* 48-68. https://doi.org/10.1080/09647775.2015.1117237

Snow, C. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. Rand Corporation

Snow, C. (2010). Academic Language and the Challenge of Reading for Learning About Science. *Science, 328,* 450-452. https://doi.org/10.1126/science.1182597

Soares, S., Gonçalves, M., Jerónimo, R., & Kolinsky, R. (2022a). *Narrating Science: Can It Benefit Science Learning, and How? A Theoretical Review* [Manuscript accepted for publication with pending minor revisions], Journal of Research in Science Teaching

Soares, S., Simão, C., Gonçalves, M., Barata, R., Jerónimo, R., & Kolinsky, R. (2022b). *How Did They Do It? The Impact of Narrating Science on Learning* [Manuscript submitted for publication to Science and Education], CIS-Iscte, Iscte–Instituto Universitário de Lisboa

Soares, S., Frade, S., Jerónimo, R., & Kolinsky, R. (2022c). *All too Human: An Eye-Gaze and Behavioural Exploratory Study on the Impact of Human Action when Learning from a Science Narrative* [Manuscript submitted for publication to Reading Research Quarterly], CIS-Iscte, Iscte–Instituto Universitário de Lisboa

Solomon, J. (2002). Science Stories and Science Texts: What Can They Do for Our Students?, *Studies in Science Education, 37*, 85–105. https://doi.org/10.1080/03057260208560178

Solomon, J., Duveen, J., Scott, L. and McCarthy, S. (1992). Teaching about the nature of science through history: action research in the classroom. *Journal of Research in Science Teaching*, *29,* 409–421. https://doi.org/10.1002/tea.3660290408

Squire, L. R., Genzel, L., Wixted, J. T., & Morris, R. G. (2015). Memory consolidation. *Cold Spring Harbor perspectives in biology, 7*(8), a021766. https://doi.org/10.1101/cshperspect.a021766

Stanovich, K. E., & West, R. F. (1989). Exposure to print and orthographic processing. *Reading research quarterly*, 402-433. https://doi.org/10.2307/747605

Stroup, W. W. (2016). *Generalized Linear Mixed Models: Modern Concepts, Methods and Applications*. CRC Press

Strube, P. (1994). Narrative in the science curriculum. *Research in Science Education, 24* (1), 313-321. https://doi.org/10.1007/BF02356358

Sucena, A., & Castro, S. L. (2008). *Aprender a ler e Avaliar a Leitura* [Learning how to read and the assessment of reading]. Almedina

Sullivan Jr, W. J., Strathdee, S. A., & Hotez, P. J. (2021). Making it personal: science communication for the masses. *Trends in Parasitology, 37*(8), 684-686. https://doi.org/10.1016/j.pt.2021.05.010

Swami, V., Stieger, S., Pietschnig, J., Nader, I. W., & Voracek, M. (2012). Using more than 10% of our brains: Examining belief in science-related myths from an individual differences perspective. *Learning and Individual Differences*, *22*, 404-408. https://doi.org/10.1016/j.lindif.2011.12.005.

Swami, V., Thorn, L., Husbands, D., Tran, U. S., Nader, I. W., von Nordheim, L., Pietschnig, J., Stieger, S., & Voracek, M. (2015). Are the Scope and Nature of Psychology Properly Understood? An Examination of Belief in Myths of Popular Psychology among University Students. In A. Columbus (Ed.), *Advances in Psychology Research* (Vol. 101, pp. 9-32). Nova Science Publishers.

Sweet, A. P., & Snow, C. E. (Eds.). (2003). *Rethinking reading comprehension.* Guilford Press.

Taleb, N. N. (2007). *The black swan: The impact of the highly improbable* (Vol. 2). Random house

Taylor, B. M. (1982). Text structure and children's comprehension and memory for expository material. *Journal of educational psychology, 74*(3), 323. https://doi.org/10.1037/0022-0663.74.3.323

Telkemeyer, S., Rossi, S., Koch, S. P., Nierhaus, T., Steinbrink, J., Poeppel, D., Hellmuth , O., & Wartenburger, I. (2009). Sensitivity of newborn auditory cortex to the temporal structure of sounds. *Journal of Neuroscience, 29*(47), 14726-14733

Tobias, S. (1990). *They're not dumb, they're different: Stalking the second tier.* Research Corporation.

Tomas, L., Ritchie, S. M., & Tones, M. (2011). Attitudinal impact of hybridized writing about a socioscientific issue. *Journal of Research in Science Teaching, 48*, 878-900. https://doi.org/10.1002/tea.20431

Tun, P. A. (1989). Age differences in processing expository and narrative text. *Journal of Gerontology, 44*(1), 9-15. https://doi.org/10.1093/geronj/44.1.P9

Vaccaro, A. G., Heydari, P., Christov-Moore, L., Damasio, A., & Kaplan, J. T. (2022). Perspective-Taking is Associated with Increased Discriminability of Affective States in Ventromedial Prefrontal Cortex. *Social Cognitive and Affective Neuroscience*. https://doi.org/10.1093/scan/nsac035

Van den Broek, P. (2010). Using texts in science education: Cognitive processes and knowledge representation. *Science, 328*(5977), 453–456. https://doi.org/10.1126/science.1182594

Van den Broek, P., Young, M., Tzeng, Y., & Linderholm, T. (1999). The Landscape Model of reading: Inferences and the online construction of memory representation. In H. van Oostendorp & S. R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 71–98). Lawrence Erlbaum Associates

Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. Academic Press

van Prooijen, J. W. (2017). Why education predicts decreased belief in conspiracy theories. *Applied cognitive psychology*, *31*, 50-58. https://doi.org/10.1002/acp.3301.

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*, 3–14. https://doi.org/10.1007/s11409-006-6893-0

Visintainer, T. (2020). "I think at first glance people would not expect me to be interested in science": Exploring the racialized science experiences of high school students of color. *Journal of Research in Science Teaching, 57*, 393– 422 https://doi.org/10.1002/tea.21597

Vistoli, D., Passerieux, C., El Zein, M., Clumeck, C., Braun, S., & Brunet-Gouet, E. (2015). Characterizing an ERP correlate of intentions understanding using a sequential comic strips paradigm. *Social neuroscience, 10*(4), 391-407.

Vrasidas, C., Avraamidou, L., Theodoridou, K., Themistokleous, S., & Panaou, P. (2015). Science Fiction in Education: case studies from classroom implementations. *Educational Media International, 52*, 201-215. https://doi.org/10.1080/09523987.2015.107510

Wang, M. D. (1970). The role of syntactic complexity as a determiner of comprehensibility. J*ournal of Verbal Learning and Verbal Behavior, 9*(4), 398-404. https://doi.org/10.1016/S0022-5371(70)80079-2

Wang, S., Scells, H., Koopman, B., & Zuccon, G. (2022). Automated MeSH Term Suggestion for Effective Query Formulation in Systematic Reviews Literature Search. *Intelligent Systems with Applications*, 200141. https://doi.org/10.1016/j.iswa.2022.200141

Ward, E., Ganis, G., & Bach, P. (2019). Spontaneous vicarious perception of the content of another's visual perspective. *Current Biology*, *29*(5), 874-880. https://doi.org/10.1016/j.cub.2019.01.046

Webb, P. (2010). Science education and literacy: Imperatives for the developed and developing world. *Science, 328*(5977), 448-450. https://doi: 10.1126/science.1182596. PMID: 20413487

Weill, K. (2022, April 9). When Your Friends Fall off the Edge of the Earth. *The Atlantic.* https://www.theatlantic.com/family/archive/2022/02/flat-earther-social-isolation/622908/

Wells, G. (1986). The meaning makers: Children learning language and using language to learn. *Heinemann.*

Wharton, C., & Kintsch, W. (1991). An overview of construction-integration model: a theory of comprehension as a foundation for a new cognitive architecture. *ACM Sigart Bulletin, 2*(4), 169-173. https://doi.org/10.1145/122344.122379

Wigfield, A., & Guthrie, J. T. (2000). Engagement and motivation in reading. *Handbook of reading research*, *3*, 403-422.

Wilcken, W. M. (2008). *Comprehension and retention: The effect of concrete details and causal structure in scientific narrative* [Master's thesis]. https://scholarsarchive.byu.edu/cgi/viewcontent.cgi?article=2649&context=etd

Wilhelm, D., Palmer, S., & Koopman, P. (2007). Sex determination and gonadal development in mammals. *Physiological reviews*. https://doi.org/10.1152/physrev.00009.2006

Wolfe, M. B., & Mienko, J. A. (2007). Learning and memory of factual context from narrative and expository text. *British Journal of Educational Psychology, 77*, 541-564. https://doi.org/10.1348/000709906X143902

Wolfe, M. B., & Woodwyk, J. M. (2010). Processing and memory of information presented in narrative or expository texts. *British Journal of Educational Psychology, 80,* 341-362. https://doi.org/10.1348/000709910X485700

Wolfe, M. B., Schreiner, M. E., Rehder, B., Laham, D., Foltz, P. W., Kintsch, W., & Landauer, T. K. (1998). Learning from text: Matching readers and texts by latent semantic analysis. *Discourse Processes, 25,* 309–336. https://doi.org/10.1080/01638539809545030

Wu, Y., Barquero, L. A., Pickren, S. E., Barber, A. T., & Cutting, L. E. (2020). The relationship between cognitive skills and reading comprehension of narrative and expository texts: A longitudinal study from Grade 1 to Grade 4. *Learning and individual differences, 80,* 101848. https://doi.org/101848. 10.1016/j.lindif.2020.101848

Xu, J., Kemeny, S., Park, G., Frattali, C., & Braun, A. (2005). Language in context: emergent features of word, sentence, and narrative comprehension. *Neuroimage, 25,* 1002-1015. https://doi.org/10.1016/j.neuroimage.2004.12.013

Yang, X., Chen, L., & Ho, S. S. (2020). Does media exposure relate to the illusion of knowing in the public understanding of climate change?. *Public Understanding of Science, 29*(1), 94-111. https://doi.org/10.1177/0963662519877743

Yang, Y. (2003). Dimensions of socio-economic status and their relationship to mathematics and science achievement at individual and collective levels. *Scandinavian journal of educational research, 47*(1), 21-41. https://doi.org/10.1080/00313830308609

Young, K., & Saver, J. L. (2001). The neurology of narrative. *SubStance, 30*(1), 72-84. https://doi.org/ 10.1353/sub.2001.0020.

Zabrucky, K. M., & Moore, D. (1999). Influence of text genre on adults' monitoring of understanding and recall. *Educational Gerontology, 25*, 691–710. https://doi.org/10.1080/036012799267440

Zabrucky, K., & Ratner, H. H. (1992). Effects of passage type on comprehension monitoring and recall in good and poor readers. *Journal of Reading Behavior, 24*(3), 373-391. https://doi.org/10.1080/10862969209547782

Zwaan, R. A. (2016). Situation models, mental simulations, and abstract concepts in discourse comprehension. *Psychonomic bulletin & review*, *23*(4), 1028-1034. https://doi.org/10.3758/s13423-015-0864-x

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123*(2), 162–185. https://doi.org/10.1037/0033-2909.123.2.162

Annex A: Chapter 4

Table A1. Final Criteria for Developing the Materials (Texts and Corresponding Learning Measures)

Materials	Criteria
Texts	1) Each text should focus on one topic, and a few corresponding subtopics/concepts, from one science field;
	2) Balance quantity and complexity of contents: they should be challenging but accessible;
	3) It should be clear which ideas are central, and they should be well connected to each other;
	4) Avoid accessory details and unnecessary complexity that may obscure the main ideas and interfere with learning; critical – to be learned – information should be novel (i.e., low prior knowledge);
	5) The narrative versions will focus on relevant scientists' actions, so sufficiently detailed information about these scientific actions must be available in external sources;
	6) Refrain from fictionalizing contents and using discursive strategies such as direct discourse and exclamation marks;
	7) The scientists' actions must involve practical or experimental procedures (i.e., not only theoretical/reasoning-based) and be intertwined with the story's development as much as possible (e.g., Fisch, 2000; Glaser et al., 2009).
Learning measures	8) It must be possible to build questions at all the contemplated comprehension levels;
	9) The questions should be sensible to the fact that participants are reading and thinking about most of the text's contents for the first time;
	10) Questions should be formulated in a simple and direct way, with the possibility of writing key words for the comprehension of the question in bold or underlined.

Table A2. Psycholinguistic Parameters Controlled for in the Main Study’s Texts

	N WRD	N SNT	N PAR	N WORD/ SNT		WRD LNT		WRD FREQ		N SUB PAR	
				M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Math NT	785	25	6	31.96	9.81	5.31	3.22	2.70	1.42	15.83	5.42
Math ET	780	26	6	30.31	9.88	5.21	3.37	2.78	1.39	7.33	4.93
Chemistry NT	785	27	7	29.04	12.13	5.34	3.12	2.62	1.50	11.0	3.79
Chemistry ET	783	24	7	32.92	12.27	5.36	3.28	2.67	1.45	8.71	3.50
				F(3, 2840) = 1.35, *p* = .257		*F*(3, 3113) = 0.33, *p* = .807		*F*(3, 98) = 1.62, *p* = .605		*F*(3, 22) = 4.36, *p* = .015	

Note. N WRD: number of words, N SNT: number of sentences, N PAR: number of paragraphs; N WORD/ SNT: number of words per sentence; WRD LNT: word length; WRD FREQ: word frequency; N SUB PAR: number of subordinate clauses per paragraph. The respective count or mean and standard deviation values are presented. There were no differences in word frequency, word length, and number of words per sentence (all *p*’s > .257). The Math NT had significantly more subordinate clauses (*M* = 15.83, *SD* = 4.42) than both the Math ET (*M* = 7.33, *SD* = 4.93, *p* = 0.02) and the Chemistry ET (*M* = 8.71, *SD* = 3.5, *p* = 0.05).

Table A3. General Scoring Guidelines Used in the Main Study

Score	Evaluation Criteria	Example
0	No response or response is incorrect	(a) I do not remember (b) I think I would not be able to analyse those characteristics using the polarimeter, seeing that there are other methods/procedures to do so.
25%	There is a correct idea	a) One just has to direct the polarimeter, in a place with light, towards a crystal because the latter reflects polarized light. b) I do not know
50%	It is more incomplete or there are more/bigger imprecisions	a) It is a device that can emit a directed light beam. b) No, using this device one could just observe the direction that light takes after hitting the compounds
75%	There is an important element missing, a huge imprecision, or the explanation is not clear	One must project the polarized light into the crystal and verify the direction in which the light that comes through the crystal is deflected. It shows the side to which and the way in which the molecules are connected among themselves, therefore we can understand more about their chemical composition
100%	The response has all required elements and they are completely correct or with minor imprecision	(a) One must project the polarized light into the crystal to observe the deviation that it causes (b) It only provides information on the molecular structure, different light deflections, or no deflection at all, indicate different molecular structures

Note. This general logic was adapted to the features and requirements of each level of comprehension and of each specific question (except to multiple-choice questions, which were only scores as completely correct or completely incorrect. Full scores could be 1, 3 or 4 (not 2 because all L2 questions were multiple-choice). An example of scoring translated from the Portuguese original is provided, using a L3 question from Chemistry. The question stated: “The polarimeter is an instrument that can be used to analyze crystals. Chemical composition and molecular structure are two characteristics from crystals that can be analyzed. a) Explain how to perform analyses using a polarimeter b) These analyses provide information about the chemical composition and the molecular structure of crystals? **Explain why**.

Table A4. Contact with Natural and Exact Sciences in Formal Education

	High School	University
Level A	Social Science /Humanities/Arts	Social Science /Humanities/Arts
Level B	Social Science/Humanities/Arts	Psychology Business/ Sports/Informatics
	Socioeconomics Science & Technology	Social Science /Humanities/Arts
Level C	Socioeconomics Science & Technology	Psychology Business/ Sports/Informatics

Note. The areas from which participants graduated high school and university from were combined to create three different levels of gradually higher contact

Table A5. Variables that Were Removed from the Independent Models due to Lack of Significant Effect

	Removed Variable(s)	Test Statistics
Contact with Literacy	TIL (proportion of correct responses)	*estimate* = 0.21, *SE* = 0.2, *t*(114.58) = 1.02, *p* = .311, 95% CI [-0.2; 0.61]
	Quantity of books read in the last year	*F*(4, 114.58) = 1.25, *p* = .295
Science Background	Prior knowledge of the Chemistry texts' topics	*estimate* = -0.05, *SE* = 0.9, *t*(115.43) = -0.62, *p* = 0.539, 95% CI [-0.23; 0.12]
	Prior knowledge of the five science topics	*estimate* = 0.03, *SE* = 0.06, *t*(115.43) = 0.53, *p* = .596, 95% CI [-0.09; 0.15]
Evaluative and Motivational Attitudes	Search of information about the Math and Chemistry fields	*estimate* = -0.01, *SE* = 0.02, *t*(115.36) = -0.53, *p* = .596, 95% CI [-0.06; 0.04]

Note. When variables are continuous or have two categories, *t* test values are presented. When variables have more than two categories, *F* test values are presented.

Table A6. Summary of Results of each Covariate Variable in the Four Independent Models

Model	Covariates		Values	Test Statistics
Sociodemographics	Age		Mean values	*estimate* = 0.01, *SE* = 0.1, *t*(118.55) = 0.6, *p* = .55, 95% CI [-0.01; 0.03]
	Gender		Two categories	*estimate* = -0.04, *SE* = 0.04, *t*(118.59) = -1.01, *p* = .285, 95% CI [-0.12; 0.04]
	Household educational level	Up to 9th grade	Two categories	*estimate* = -0.02, *SE* = 0.01, *t*(1638.84) = -1.32, *p* = .186, 95% CI [-0.05; 0.01]
		Up to high school	Two categories	*estimate* = -0.00, *SE* = 0.01, *t*(1641.56) = -0.24, *p* = .812, 95% CI [-0.03; 0.02]
	Household Income level		Mean values	*estimate* = -0.01, *SE* = 0.01, *t*(1608.31) = -1.21, *p* = .223, 95% CI [-0.02; 0.0]
Contact with Literacy	ART *d′* score		Mean values	*estimate* = 0.08, *SE* = 0.03, *t*(119.59) = 3.01, *p* = .003, 95% CI [0.03; 0.13]
Science Background	Level of formal science education		Three categories	*F*(2,116.28) = 5.67, *p* = .004
	Prior knowledge of the Math texts' topics		Mean values	*estimate* = 0.08, *SE* = 0.02, *t*(117.39) = 3.06, *p* = .003, 95% CI [0.03; 0.13]
	SRT *d′* score		Mean values	*estimate* = 0.07, *SE* = 0.03, *t*(117.39) = 2.21, *p* = .029, 95% CI [0.01; 0.13]
Evaluative and Motivational Attitudes	Enjoyment of the Chemistry and Math fields		Mean values	*estimate* = 0.04, *SE* = 0.02, *t*(116.27) = 2.20, *p* = .030, 95% CI [0.00; 0.08]
	Interest in the Math text		Mean values	*estimate* = 0.03, *SE* = 0.01, *t*(116.27) = 2.89, *p* = .005, 95% CI [0.01; 0.06]
	Interest in the Chemistry text		Mean values	*estimate* = 0.05, *SE* = 0.01, *t*(116.27) = 4.66, *p* < .001, 95% CI [0.03; 0.08]
	Need for Cognition questionnaire		Mean values	*estimate* = 0.08, *SE* = 0.03, *t*(116.27) = 2.97, *p* = .004, 95% CI [0.03; 0.13]

Note. When variables are continuous or have two categories, *t* test values are presented. When variables have more than two categories, *F* test values are presented.

Table A7. Main Effects and Interaction Results in the Independent Models whose Covariate Variables Predicted Learning

Covariate	Main Effect of Text Type	Main Effect of Topic	Main Effect of Session	Main Effect of Level of Comprehension	Text Type by Topic Interaction	Text Type by Topic by Level of Comprehension Interaction
Contact with Literacy	*F*(1, 1682.66) = 20.38, *p* < .001	*F*(1, 1682.55) = 19.07, *p* < .001	*F*(1, 1653.83) = 28.19, *p* < .001	*F*(3, 771.92) = 91.23, *p* < .001	*F*(1, 121.99) = 3.62, *p* = .059	*F*(9, 524.15) = 9.66, *p* < .001
Science Background	*F*(1, 1681.44) = 20.42, *p* < .001	*F*(1, 1681.72) = 18.96, *p* < .001	*F*(1, 1655.23) = 28.14, *p* < .001	*F*(3, 771.89) = 90.97, *p* < .001	*F*(1, 119.1) = 3.52, *p* = .063	*F*(9, 528.22) = 9.66, *p* < .001
Evaluative and Motivational Attitudes	*F*(1, 1676.16) = 20.36, *p* < .001	*F*(1, 1676.30) = 18.86, *p* < .001	*F*(1, 1658.59) = 28.38, *p* < .001	*F*(3, 776.33) = 91.56, *p* < .001	*F*(1, 119.7) = 6.9, *p* = .010	*F*(9, 506.79) = 9.67, *p* < .001

Note. F tests are presented for each main effect and interaction effect.

Table A8. Differences in the Results of Covariate Variables Between Text Types

Covariate	NT	ET
SRT *d'* score	*estimate* = 0.07, *SE* = 0.03, *t*(113.79) = 2.13, *p* = .035, 95% CI [0.01; 0.14]	*estimate* = 0.06, *SE* = 0.03, *t*(114.86) = 1.8, *p* = .073, 95% CI [-0.01; 0.12]
Enjoyment of the Chemistry and Math fields	*estimate* = 0.05, *SE* = 0.23, *t*(113.49) = 2.21 *p* = .029, 95% CI [0.01; 0.1]	*estimate* = 0.04, *SE* = 0.02, *t*(115.13) = 1.51, *p* = .133, 95% CI [-0.01; 0.08]

Note. t test values are presented for each covariate variable effect.

Annex B: Chapter 5

Table B1. Scoring Guidelines for Evaluating Engagement in the Four ToM-Related Processes

Score Level	Meaning
0	No reported engagement: Participants rated their engagement in the process as below 3 or provided no written answer
1	Ambiguous engagement: It is unclear whether there is an attempt to engage in the process; responses are very tied to the formulation of the item (i.e., recapitulation) or to science contents
2	Tentative/Low engagement: There is some level of engagement in the process
3	Clear engagement: There is a clear engagement in the process, with at least one accompanying instance or example
4	High engagement: Engagement is clear and more detailed (e.g., more than one example; linking different ideas; embedded in a deeper explanations)

Table B2. Scoring Guidelines for Evaluating the Level of Humanisation of Meaning in Participants Responses to Open-Ended Questions

Score Level	Meaning
0	No response: No response is provided
1	Science content level: The response only refers to scientific contents
2	Human level: The response mentions Louis Pasteur but keeps to the contents of the text
3	Social level: The response makes extra-text connections, such as linking the science contents or Pasteur to society at large, or extracts higher-level appraisals based on Pasteur's actions

Table B3. Results of Exploratory Correlation Analysis Between Second Pass Dwell Time in Mental AoIs and Sensorimotor AoIs and Levels of Comprehension

Gaze Measure	Level of Comprehension	Correlation Result
Second Pass Dwell Time on Mental AoIs	L1	*r*(42) = .425, *p* = .004
	L3	*r*(42) = .328, *p* = .030
Second Pass Dwell Time on Sensorimotor AoIs	L1	*r*(42) = .416, *p* = .005
	L3	*r*(42) = .340, *p* = .024

Note. Only significant results are reported.

Table B4. Example of Responses to one ToM-Related Item (Perspective-Taking), for each Level of Score

Participants’ Response	Score
“Louis Pasteur wanted to assert his perspective.”	1
“I tried to imagine what Pasteur was seeing, when he was doing the microscope experiment, despite not knowing the shape of the crystals.”	2
“I did not try very hard to see things from Pasteur's perspective, but, in the part of the text where he decides to redo the experiment, I thought that if it was me, I would have persisted too, to really understand why it happens “	3
“I tried to imagine that Louis was discovering what had not yet been discovered around me to understand the difficulty of reaching such a conclusion, and then I tried to imagine what it would be like if I had been able to reach it, as Louis did.”	4

Table B5. Examples Responses to Each of the Four ToM-Related Items that Denote High Engagement

ToM-Related Item	Participants' Response	Score
Transportation	"I imagined a small, dimly lit room, almost like a basement with only one window. A small night-light, many shelves with lab material, test tubes, flasks, etc."	1
Perspective-Taking	"I imagined myself as a person from that era, without current knowledge, curious about how the things that were being studied at that time worked, intrigued about how different compounds act in different ways, happy/ proud in a way because I could have found something new to contribute to society."	2
Cognitive ToM	"Louis Pastor questioned himself about a new (in his time) parameter for the evaluation of chemical elements, since he was faced with a situation that intrigued him about the possible molecular organization/structure that would differentiate chemical elements with the same degree of solubility and atomic weight, going against the belief, at the time, that if these aspects of a chemical element were the same, the structure would also have to be the same. Louis intended, therefore, to innovate within chemistry and obtain new answers, to expand humanity's knowledge in this field. "	3
Affective ToM	"I think he felt joy, fun, a little satisfaction that he had gotten to where no one else could at that time, and he may have felt a little sadness as well, because the mystery had come to an end."	4

Table B6. Example of Responses to the Two Open-Ended Questions, for each Level of Scoring

Question	Participants' Response	Score
Humanising Main Ideas	"different compounds react in different ways to the incidence of light"	1
	"A chemist discovers that a given crystal did not have the composition that others believed."	2
	"That the scientist did not let himself be taken by what was already known and was ambitious enough to disagree with what was known to the scientific community and searched for an answer that would satisfy him. I also got, that, were it not for this ambition that motivated him and many other scientists in their pursue for answers, many theories that exist today would not exist. Finally, I conclude that science is not something exact, but rather something that is constantly changing."	3
Humanising Curiosity	"I found it very curious and I think I'd like to see a video, or a visual representation, of the experiments he did"	1
	"How did the question come to be?"	2
	"How were the other researchers/scientists react after they heard that Louis was able to successfully complete his theory/research?"	3

Table B7. Total Variance Explained by the Components

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	3.858	38.578	38.578	3.858	38.578	38.578	3.11
2	1.987	19.866	58.445	1.987	19.866	58.445	1.85
3	1.236	12.356	70.801	1.236	12.356	70.801	2.794
4	.956	9.559	80.359				
5	.597	5.971	86.331				
6	.478	4.778	91.109				
7	.4	4.001	95.11				
8	.249	2.485	97.595				
9	.146	1.464	99.059				
10	.094	.941	100				

Table B8. Pattern Matrix

	Component		
	1	2	3
Degree of transportation	.801		
AI correct RT	.738		-.333
Rating of transportation	.699		
Rating of perspective-taking	.676		.3
Degree of perspective-taking	.638		.332
Humanising curiosity		.87	
Humanising main ideas		.824	
Rating of affective ToM			.911
Degree of affective ToM			.903
Degree of cognitive ToM			.593

Note. Rotation converged in 17 iterations.

Table B9. Structure Matrix

	Component		
	1	2	3
Degree of transportation	.828		
Rating of transportation	.78	.333	.406
Rating of perspective-taking	.761		.473
Degree of perspective-taking	.751	.368	.505
RTs in correct AI	.623		
Humanising curiosity		.867	
Humanising main ideas		.828	
Rating of affective ToM			.908
Degree of affective ToM			.891
Degree of cognitive ToM	.414		.654

Table B10. Combined Effects of Attention to and Thoughts on Human Action When Other Gaze Measures Are Entered (Instead of Number of Regressions to Mental AoIs)

Measure of Thoughts on Human Action	Gaze Measure	Effect of Level of Item	Effect of Measure of Thinking About Human Action	Effect of Gaze Measure
Simulating Other Points of View	Number of Regressions to Sensorimotor AoIs	*F*(3, 50.12) = 14.8, *p* < .001	*estimate* = 0.08, *SE* = 0.03, *t*(38.97) = 2.78, *p* = .008, 95% CI [0.02; 0.13]	*estimate* = 0.38, *SE* = 0.13, *t*(38.97) = 2.93, *p* = .006, 95% CI [0.12; 0.63]
	Second Pass Dwell Time in Mental AoIs	*F*(3, 50.53) = 14.76, *p* < .001	*estimate* = 0.08, *SE* = 0.03, *t*(38.87) = 3, *p* = .005, 95% CI [0.03; 0.14]	*estimate* = 0.01, *SE* = 0.0, *t*(38.87) = 3.36, *p* = .002, 95% CI [0.0; 0.01
	Second Pass Dwell Time in Sensorimotor AoIs	*F*(3, 50.28) = 14.46, *p* < .001	*estimate* = 0.08, *SE* = 0.03, *t*(38.83) = 2.91, *p* = .006, 95% CI [0.02; 0.14]	*estimate* = 0.01, *SE* = 0.0, *t*(38.83) = 3.15, *p* = .003, 95% CI [0.0; 0.01]
	Number of Regressions to Sensorimotor AoIs	*F*(3, 48.65) = 15.73, *p* < .001	*estimate* = 0.05, *SE* = 0.03, *t*(39.88) = 1.52, *p* = .137, 95% CI [-0.02; 0.12]	*estimate* = 0.31, *SE* = 0.03, *t*(39.88) = 2.21, *p* = .033, 95% CI [0.03; 0.59]
Humanising Science	Second Pass Dwell Time in Mental AoIs	*F*(3, 48.78) = 15.78, *p* < .001	*estimate* = 0.05, *SE* = 0.03, *t*(39.88) = 1.52, *p* = .138, 95% CI [-0.02; 0.11]	*estimate* = 0.01, *SE* = 0.0, *t*(39.88) = 2.48, *p* = .017, 95% CI [-0.0; 0.01]
	Second Pass Dwell Time in Sensorimotor AoIs	*F*(3, 48.72) = 15.52, *p* < .001	*estimate* = 0.04, *SE* = 0.03, *t*(39.95) = 1.44, *p* = .157, 95% CI [-0.02; 0.11]	*estimate* = 0.0, *SE* = 0.0, *t*(39.95) = 2.26, *p* = .029, 95% CI [0.0; 0.01]

B11. Additional Correlation Analysis Between Attribution of Intentions and Attention to Human Action

As an additional test, we examined whether we could tap into the specificity of attributing intentions, as measured by RTs correct AI. To do this, we compared whether different measures from the comic-strip task that involved human characters established different correlations with measures of attention to human action in the text. This comparison would provide us with a sense of the strength of the specificity of the process of attributing intentions, as compared to the mere presence of characters.

The used measures from the comic-strip task were RTs on correct AI, the aggregated correct RTs on trials with character (i.e., average of AI and physical causality with character trials; henceforth, *aggregated character RTs*), and the difference between RTs on correct AI and RTs on correct physical causality with character trials (henceforth, *difference character RTs*). The measures of attention to human action used were the number of regressions to mental AoIs and the number of regressions to sensorimotor AoIs. The number of regressions to mental AoIs did not significantly correlate with any of the comic-strip measures (RTs in correct AI: $r(42) = .163$, $p = .29$; aggregated character RTs: $r(42) = .18$, $p = .243$; difference character RTs: $r(42) = -.008$, $p = .96$). A similar pattern of results was observed with the number of regressions to sensorimotor AoIs (RTs in correct AI: $r(42) = .122$, $p = .43$; aggregated character RTs: $r(42) = .118$, $p = .447$; difference character RTs: $r(42) = .054$, $p = .73$).

Table B12. Summary of Effects When each Ancillary Measure is Separately Entered as Covariate in the Model of Combined Effects of Attention to and Thoughts on Human Action

	Effect of Level of Item	Effect of Simulating Other Points of View	Effect of Number of Regressions to Mental AoIs	Effect of Ancillary Measure
Science Background (SRT)	*F*(3, 51.790) = 15.42, *p* < .001	*estimate* = .07, *SE* = 0.02, *t*(38.59) = 3.01, *p* = .005, 95% CI [.02; .12]	*estimate* = .31, *SE* = 0.08, *t*(38.59) = 3.98, *p* < .001, 95% CI [.15; .46]	*estimate* = .08, *SE* = 0.03, *t*(38.59) = 3.23, *p* = .003, 95% CI [.03; .13]
Contact with Literacy (ART)	*F*(3, 51.98) = 15.42, *p* < .001	*estimate* = .07, *SE* = 0.02, *t*(38.37) = 2.78, *p* = .008, 95% CI [.02; .12]	*estimate* = .3, *SE* = 0.08, *t*(38.37) = 3.8, *p* = .001, 95% CI [.14; .45]	*estimate* = .1, *SE* = 0.03, *t*(38.37) = 3.18, *p* = .003, 95% CI [.03; .16]
Motivational Attitudes (Interest in the Text)	*F*(3, 51.24) = 15.49, *p* < .001	*estimate* = .06, *SE* = 0.03, *t*(37.9) = 2.21, *p* = .033, 95% CI [.01; .12]	*estimate* = .33, *SE* = 0.08, *t*(37.9) = 4, *p* < .001, 95% CI [.16; .5]	*estimate* = .05, *SE* = 0.02, *t*(37.9) = 1.9, *p* = .065, 95% CI [-.00; .1]

Note. When variables are continuous or have two categories, *t* test values are presented. When variables have more than two categories, *F* test values are presented.

Table B13. *Summary of Effects When each Ancillary Measure is Separately Entered as Covariate in the Model of the Effects of Humanising Science*

	Effect of Level of Item	Effect of Humanising Science	Effect of Ancillary Measure
Science Background (SRT)	*F*(3, 49.04) = 15.66, *p* < .001	*estimate* = .06, *SE* = 0.03, *t*(40.32) = 1.95, *p* = .059, 95% CI [-0.0; .11]	*estimate* = .09, *SE* = 0.03, *t*(40.32) = 3, *p* = .004, 95% CI [0.03; .15]
Contact with Literacy (ART)	*F*(3, 49.48) = 15.56, *p* < .001	*estimate* = .06, *SE* = 0.03, *t*(40.21) = 2.1, *p* = .042, 95% CI [0.0; .11]	*estimate* = .12, *SE* = 0.03, *t*(40.21) = 3.41, *p* = .001, 95% CI [0.05; .19]
Motivational Attitudes (Interest in the Text)	*F*(3, 48.36) = 15.66, *p* < .001	*estimate* = .04, *SE* = 0.03, *t*(39.71) = 1.39, *p* = .174, 95% CI [-0.02; .11]	*estimate* = .05, *SE* = 0.03, *t*(39.71) = 1.6, *p* = .116, 95% CI [-0.01; .1]

Note. When variables are continuous or have two categories, *t* test values are presented. When variables have more than two categories, *F* test values are presented.

Table B14. Comparison of the Number of Regressions to Mental AoIs and to Sensorimotor AoIs

	M	*SD*	Test Statistic
Mental AoIs	0.52	0.34	*t*(49) = 2.83, *p* = .007
Sensorimotor AoIs	0.44	0.23	

B15. Discussion of Supplementary Results

As can be seen in Table B12 (p. 232), most ancillary measures yielded significant results on learning, despite concurrent significant effects of variables more closely related to the processing of the learning materials (number of regressions to mental AoIs, Item Level). On the one hand, the fact that Item Level, the number of regressions to mental AoIs, and Adoption of Other Points of View were significant attests to the robustness of these effects.

On the other hand, these findings show that more general contact with science and literature had a robust impact on learning, explaining something about the scores that neither of the variables related to the processing of the materials and ToM-related processes did. Namely, the greater the science and literature literacy, the greater the learning scores. On the other hand, a more motivational aspect pertaining to participants' appreciation of the text, yielded only a marginal effect at most. This indicates that either our measure of attention to human action, our measure of thoughts on human action, or both, make similar contributions to learning as interest in the text.

To determine whether the overlap in variance was common to both variables or restricted to one of them, we entered as covariates to the base model the interest in the text along with either the number of regression to mental AoIs, or Simulating Other Points of View. Whereas when included with the measure of thoughts on human action, none of the variables had a significant impact on learning (Simulating Other Points of View: *estimate* = 0.06, *SE* = 0.03, *t*(39.18) = 1.73, *p* = .092, 95% CI [-0.01; 0.12] interest in the text: *estimate* = 0.04, *SE* = 0.03, *t*(39.18) = 1.43, *p* = .160, 95% CI [-0.02; 0.01]), when included with the measure of attention to human action, both variables significantly impacted learning (number of regression to mental AoIs: *estimate* = 0.32, *SE* = 0.09, *t*(38.99) = 3.74, *p* = .001, 95% CI [0.15; 0.5] interest in the text: *estimate* = 0.06, *SE* = 0.02, *t*(38.99) = 2.67, *p* = .011, 95% CI [0.02; 0.11]). These results confirm that interest in the text and simulation of other points of view make similar contributions to learning scores, whereas it offers a distinct contribution to learning scores than attention to human action (number of regressions to mental AoIs). It thus seems that engaging in simulation of other points of view and the level of interest in the text may be to some extent related, or even dependent on one another.

A similar pattern of results can be seen for the Humanising Science variable, in Table B13 (p. 233), with the exception that the number of regressions to mental AoIs was not included in this model (as it had already been determined that the marginal effect of Humanising Science was lost in the presence of this variable). In contrast, the effect of Humanising Science slightly improved in the presence of the (also significant) effects of the ART and the SRT *d′* scores. This shows that both

Humanising Science and general literacy measures (related to either science or literature) made specific and concurrent contributions to science learning. When put together with interest in the text, however, neither variable has a significant impact on learning scores, likely because they make similar contributions to learning scores. As with simulating other points of view, it is possible that humanising science is to some extent related to, or dependent on, the interest elicited by the text.

Nonetheless, as measures tapping into different processes and factors jointly influenced learning scores, these results support a multidimensional conception of science learning (i.e., that varied aspects and processes are involved; e.g., Snow, 2002), and can provide interesting cues for future research and applications.

Annex C: Chapter 6

Table C1. Study protocol

Instrument	Brief description
Informed consent	Read and signed online before session
Sociodemographic questionnaire	Completed online before session
Welcome	Participants were greeted by the first and second authors. The first author presented the study and provided general instructions
Reading the texts	In the texts, it was explained that crystals can have the same molecular composition despite different molecular structures. Two specific examples were described, as well as the methods that can be used to examine these parameters. There was no time limit, but participants usually read it under 5/7 minutes. The texts were presented using the Netlify platform: Narrative text: https://pasteur-museu.netlify.app/nd Expository text: https://pasteur-museu.netlify.app/te
Answering the questions	Presented using the SurveyHero platform. There was no time limit
Video	Lasted 4:45 minutes and presented the historical Chemistry laboratory of the museum. The second author guided the "virtual tour", explaining the uses of different rooms and instruments and providing fun facts and trivia
Learning activity	Was presented in PowerPoint and lasted between 30-40 minutes. Participants could pose questions during the activity
Focus group discussion	Took place in Zoom and was recorded for subsequent analysis
Wrapping-up and farewell	The moderator thanked participants for their participation. Participants made comments on the activity, with some asking to be informed of a future publication

Table C2. More specific information of each participant of the study

Focus group	No.	Gender	Age	Field of study	Current field of work	Experience as science educator	Science reading habits
1	P1	Man	32	Physics; Cognitive Science	Research in Cognitive Science		Less than 1 per month
	P2	Man	32	Informatics	Informatics		Less than 1 per month
	P3	Woman	29	Modern History	Research in Modern History		2-5 per month
	P4	Woman	34	Architecture	Urbanism		2-5 per month
	P5	Man	39	Rehabilitation and social integration; Cognitive Science	Cognitive Science	Non-formal	2-5 per month
	P6	Woman	30	Psychology	Science managing		1 per month
2	P7	Man	32	Artistic Promotion and Heritage; Tourism	Museum guard		Less than 1 per month
	P8	Woman	32	Biology	Science school teacher	Formal and non-formal	1 per month
	P9	Man	25	Physics	Physics PhD candidate	Non-formal	1 per month
	P10	Man	41	Humanities	Translation		Less than 1 per month
	P11	Woman	55	Ecology	Ecology		1 per month
	P12	Man	31	Humanities and Arts	Visual artist		1 per month
	P13	Woman	22	Psychology	Master student in Psychology		1 per month
3	P14	Man	27	Psychology	Research in Psychology		Less than 1 per month
	P15	Woman	28	Design	Motion Design and Animation		Less than 1 per month
	P16	Woman	48	Tourism	Tourism technician		Less than 1 per month
	P17	Woman	19	Languages and humanities	Psychology undergrad		Less than 1 per month
	P18	Woman	31	Biology	Biologist and non-formal educator	Non-formal	Less than 1 per month
	P19	Man	33	Informatics	Data Engineering		Less than 1 per month

C3. Texts used in the learning activity

Chemistry narrative text

Louis Pasteur wanted to know the elements that made up living things: the field of organic Chemistry fascinated him. Like other chemists of the 19th century, Louis gathered organic compounds and extracted from them small crystals, which were solid materials with specific geometric shapes, that he analyzed in his laboratory. He knew that the shapes of the crystals were due to the molecules that composed them and the way in which they were organized. His notebooks were filled with notes on the chemical composition and molecular structure of different compounds.

For Louis, two organic compounds were a cause of great bewilderment: the tartaric and the paratartatic compounds. His analyses showed that they had the same chemical composition, for they had the same atomic weight and the same solubility level; yet his analyses also showed that they did not have the same molecular structure, which was very odd. Louis decided to extensively study the crystals of these two compounds and, amidst this process, he discovered in 1848 a structural feature of these molecules yet unknown, that he named chirality.

Louis used different methods to study these compounds: first, he calculated the atomic weight of the molecules that made up the crystals and wrote down the values on his notebook. Next, he dissolved the crystals in a liquid solution, as to compare their solubility, and noted their level of solubility. By verified that the values he had written down were the same, and concluded that the tartaric and the paratartaric crystals had the same chemical composition. Afterwards, Louis decided to analyze the structure of the crystals and to do so he pointed a light into the crystals and observed the direction to which this light was deflected as it passed through them. This way, Louis could understand how the molecules that made up the crystals were organized, in other words, what was the molecular structure of the crystals.

However, Louis could not simply project the light from a lamp, for it traveled in all directions, preventing him from observing its deflection: he needed to use polarized light, that is, a light that traveled only in a specific direction. To accomplish this, he fetched an instrument called a polarimeter and placed a polarizing filter in front of the lamp before projecting the light into the crystals, causing the light to travel in only one direction; only then did he point this polarized light at the jars where the crystals were dissolved. When Louis pointed the polarized light at the solution that had the tartaric crystals he observed it being deflected to the right; but when he directed polarized light into the solution with the paratartaric crystals, it was not deflected in any direction.

To Louis, different interactions with polarized light indicated that the molecular structure of the two compounds was different, as their molecules were organized differently. However, this defied the views of many chemists of the time, who argued that the compounds must have had the same molecular structure, seeing they had the same chemical composition. Refusing to accept this conclusion, Louis decided to conduct a set of experiments in his laboratory.

He began by preparing samples of tartaric crystals and paratartaric crystals, which he observed under the microscope. Examining the tartaric crystals, Louis ascertained, ashe expected, that they were all turned to the right. However, upon closer examination of the paratartaric crystals, he noticed that some were turned to the right, like the tartaric crystals, but some were turned to the left. The two types of crystal formed mirror images of each other, which were mixed in the compound. To him, this proved that molecules with the same composition could be organized differently.

Puzzled, Louis decided to do more experiments, this time using only the crystals from the paratartaric compound. He manually separated the right-oriented crystals from the left-oriented crystals, placing them into two separate piles; he then dissolved the crystals from each pile and placed them in the polarimeter. When pointing the polarized light at the right-facing crystals, Louis found the light to be deflected to the right; similarly, by pointing the polarized light at left-facing crystals, the light deflected to the left. For Louis, the conclusion was clear: when separated, the two crystals of the paratartaric compound deflected polarized light; however, when mixed in the compound, they deflected the polarized light in opposite directions, neutralizing each other's deflection. Louis could not help but notice that this symmetric mirroring of the paratartaric crystals was similar to that of his own hands. Remembering the Greek word for hand, *kheir,* translated as chiral, he decided to name the structural feature he had discovered as chirality.

Chemistry expository text

The scientific field of Chemistry includes, as defined in specialized textbooks, the study of the elements that make up organic matter (i.e., living things) and how these elements are organized (see, for instance, Clayden, Greeves, & Warren, 2012). Studies are conducted in a laboratory setting by extracting small crystals of organic compounds. Crystals are solid materials with specific geometric shapes and owe these shapes to their specific atoms and molecules, and the way these elements are organized, making it is possible to extract information regarding the chemical composition and molecular structure of different organic materials (e.g., Glusker, Lewis, Rossi, 1994).

Different methods can be used to study organic compounds (Clayden et al., 2012). Chemical composition can be determined by calculating the atomic weight of the molecules that make up the crystals of the compounds, as well as by analyzing the solubility level of crystals when dissolved in a liquid solution. Additionally, it is also possible to observe the interaction of the dissolved crystals with light, namely by observing the direction in which the light is deflected after being directed onto the crystals. The direction in which light is deflected is dependent on the specific organization of the molecules that make up the crystal are, and therefore gives information about the molecular structure of the crystals.

However, this analysis cannot be done with natural light, and the light must be polarized. When light is projected from a source (e.g., a lamp) it travels in all directions (e.g., Shipman, Wilson, & Higgins, 2015), making it impossible to observe its deflection. Using an instrument called a polarimeter, it is possible to place a polarizing filter in front of the light source and have the light travel in a specific direction (e.g., De Martino, Kim, Garcia-Caurel, Laude, & Drévillon, 2003).

Compounds may present the same chemical composition, as determined by the analysis of their atomic weight and solubility level, yet present different results when it comes to the analysis of their structure. Crystal structure analyses relate to the concept of chirality (Flack, 2009; Kauffman & Myers,1998), a molecular structural feature discovered through systematic studies conducted on crystals of the tartaric and the paratartaric acids by the chemist Louis Pasteur (Pasteur, 1848). Conducting chemical analyses reveals that the two acids have identical chemical composition and solubility level but interact differently with polarized light.

By dissolving crystals and observing the resulting solutions in the polarimeter, it can be seen that a solution with tartaric acid crystals causes the polarized light beam to be deflected to the right as it passes through; however, a solution with paratartaric acid doesn't affect the light's direction in any way. The fact that the tartaric and the paratartaric acids have the same chemical composition meant that until the 19th century it was argued by many chemists that their molecular structure was the same, but one deflected polarized light, the other did not. Today it is known that differences in the way two compounds interact with light indicate differences in their molecular structure (see, on this topic, Kauffman & Myers, 1998).

Observation of the tartaric and the paratartaric acid crystals reveals the specific orientation that characterizes each crystal. The crystals of the tartaric acid are all facing the right, whilst in the paratartaric acid some crystals are right-oriented (like the tartaric crystals) but others are left-oriented, thus forming mirror images of one another that are mixed in the acid (e.g., Gal, 2008; Nagendrappa, 2007). This attests to the fact that molecules that have the same composition can nonetheless be organized differently.

The two types of paratartaric acid (i.e., right-oriented and left-oriented) can be further analyzed with the polarimeter using polarized light. By manually separating each type of crystal and preparing liquid solutions with it, it is possible to observe that the solution with right-oriented crystals deflect the polarized light to the right and the solution with left-oriented crystals deflects it to the left. This observation allows the conclusion that when separated, the two types of crystal of the paratartaric compound deflect the polarized light; when mixed in together in the same compound, the crystals deflect the polarized light in opposite directions, neutralizing each other's deflection (e.g., Flack, 2003; Fox & Whitesell, 2004). This particular type of symmetric mirroring is similar to that of human hands, which is the reason why this structural feature was named chirality (i.e., *kheir*; Greek word for hand; chiral).

C4. Questions about the text used in the learning activity

1- Light is a type of electromagnetic radiation:

Which spreads in a straight line in all directions

Spreads in a straight line in only one direction

Does not spread in a straight line

2- The property of polarized light that interests us:

It is not affected by the structure of molecules

It can be deflected to the left or right when passing through chemical compounds

It changes the structure of molecules

3- The polarimeter is a piece of equipment that:

Produces polarized light

Uses polarized light

Does not work with polarized light

4- An object is said to be chiral when:

Its image doesn’t overlap its mirror image

Its image overlaps its mirror image

It is symmetrical object

5- Symmetrical objects:

Can be chiral or non-chiral

Are chiral

Are non-chiral

6- The concept of chirality is a property:

Unique to Chemistry

Unique to Biology

Present in all areas of our daily lives

C5. More specific ideas and suggestions from participants

P14: Portraying museums as (…) fascinating places, places of learning, of things that are, that (...) bring something to people's lives.

P1: It could be companies, it could be universities... exhibitions on certain topics, those fairs for "hunting" students for certain faculties, etc. this could be done in a much more constant way, without that voracious interest in capturing students...

P17: maybe if there was more marketing like that encouraging going to museums... it would be much better, because I used to go during school trips and then after a while I no-noticed that I hadn't gone since

P4: There is this lack of seeing the museum as an extension of the classroom. It's a little bit like expanding the idea of museums and, perhaps, also centering the museum itself as a classroom and producing a curriculum.

P1: Because in reality either young people go with their parents, or start going with their parents to museums as kids, or, otherwise, it's hard to go later on if they lacked contact with them before, right? It's actually a bit like reading, and I think it has to stop being so dependent on the parents and come from the school. Because not only does that completely skew the availability of thinking outside the box, but it doesn't get past the social barriers, that only a certain social class goes to museums, etc., right? And I think that if museums were... (...) more present in schools, that could help a lot

P16: Maybe museums must do more... I don't know if it's advertising, but maybe more... more invitations. Maybe the ticket prices aren't very “attractive”, are they? There must be some mechanism or some support from the-the state to make it easier to go to museums.

P1: The museums, for example, I think the connection with being presented in schools, which might be a bit of a megalomaniac job, but there should be a presentation tour in schools or school clusters. I think it can be a great advantage. I feel that there aren’t many presentations of things outside the schools in the schools. There are very few visits, mainly from academia, academia lacks interest in going to schools to talk to their their potential...to their future...apprentices, students, etc. (...) Because I think schools are actually very interested in having these kinds of activities

P9: Maybe the, the schools, or even the museums... should have kits or prepare... or even find kits to give, for example, like giving it to a father and saying "Look... You can even assemble with stuff you have at home, and you'll discover this concept with your son" And have this creation of... of knowledge in an informal environment... provided perhaps by schools and by... for example, by museums.

P9: One thing that museums can, at least in Portugal, and-and should do more of-is to take… what they do well to the digital. So that a person doesn't just have to go to the museum, but can, even in the context of the current pandemic, go to the museum whilst staying home. And read stories, and have those diagrams, and watch videos that take that knowledge that people who work in museums, and who do communication for museums, already do so well, and put it on a, shall we say, on a on demand basis. Where people can search for that specific kind of information on that topic or they can even browse around the museum. Something that teachers can also send to the students, and say "Look, go to the museum's website, they have a set of presentations, they have a set of videos". And then sometimes you can even have a-a little quiz, and I think that this would be a very good thing that museums and any place that does informal learning can do, to go further. Because the hard part they already do well. Now they just have to change their... their range, and change the people they reach.

P9:...there's a lot of content nowadays. Maybe the role that museums could have, in that sense, is not so much of creation and being one more person in the... in the... in being one more organism in the cake, but maybe, more of a role that maybe a little bit of digital curatorship. Instead of museums having "Look, here is this little thing that we did" and having "Look, here are two things that we talked about in our museum, you can see the exhibition", maybe something and so on, but then saying "Okay, here are several videos that we..." whether it be videos, books, texts "...that we selected (...)

P15: I think it is to have a modernization in the way museums communicate to the public, introducing more in the digital area and even more in the interactive part, being that our generation and the younger ones live with the... constant inter-interactivity. Looking at, for example, art or design museums, for example, many of them have interactive exhibits and... that appeals to an audience that I still think museums related more to... history and with science and not with such artistic branches could take as an example... those more artistic museums in the way they communicate, how they do marketing, how they do invitation videos

P15: (...) and also this idea of museums doing street activities, this is also an idea... quite an interesting idea because... there it is, we are outside and we see a group of people, it captures our interested.

P16: The last museum that I visited had screens to place us in history, in the age when those drawings were made in the stones, the kind of people who created those drawings, what those drawings meant, and all that with a game and using giant screens, and it was very... it was a lot of fun, it was a lot of fun.

P12: As far as informal spaces go, I already had the experience of being in charge of one and there was a methodology of the before, during, and after. Because... before, the person gets there only minimally prepared, and then you must find a methodology that makes them aware towards arriving to the space, to a moment of a set of experiences in that formal or informal educational space

P19: Remembering some museum visits and my frustration when, for example, I would go to an interactive exhibit and it was interactive because there was a lot of digitalization present in the exhibit but there was no way for me to gather more information about what was in front of me. And the fact that I couldn't, at that very moment, travel in history and go to that moment in history to understand the cultural-historical, social context... that makes it more difficult for me to try to understand. Instead of being able to at that very moment interact with the item in the sense of "let me read more, let me know more, let me understand its importance" The audio guides don't always because they seem to be going very quickly over the topic, I feel like, that it's very-that it provides a good introduction but it's very limited... having a guide actually talking about the topic and bringing a better explanation helps, or the possibility that I could pay for a guide for that. It's not always clear whether it's something I can do or not. Or even just simply being able to use my phone with a QR code or whatever to get more information on the topic.

P8: For example, now I'm teaching in [a specific city], I tell them a lot about things from [that city]. I use examples that are close to their homes. And I think that goes halfway to arousing their interest because these are things that they see in their daily lives.

P8: The dinosaur museum of [a specific city] that's on the beach where they found the dinosaurs. I mean, it's when you have a museum that talks about something from the area where you are. And then, even if you have no interest in dinosaurs, you find yourself in [that specific city] and you start seeing dinosaurs everywhere and that can arouse your interest.

P7: A few days ago I saw a documentary that started by mentioning the relationship between bees and chocolate. What the hell do bees have to do with chocolate? And this will make people a little more interested: "What do you mean? If there are no bees I won’t have chocolate, but I like chocolate so much." So, this could be something that click with me, right? And museums can do similar things

P19: I like it a lot, maybe because I come from computer science, and I like understanding how one arrives to something, or what made them think of something, right? And why was it so different from others who were also doing research at that time.

P5: I think it's good, and it's useful for people to know the process and get enthusiastic about this discovery process and... and also to know that there's a path, that it's not... that ideas and discoveries are not "Ah! It's here, that's it!" No. It's an idea path, which is a continuum, and that there can be mistakes.

P18: Sometimes it's important not only to call things by their names but also know their practical applications. If we can translate that into more practical models or apply to everyday things, people get interested. I can give a very simple example right now. I think we are all now feeling a bit of winter. I have the fireplace on all day and the heat just escapes... very quickly, isn’t it? If I do my own house plan one day and I know, through the laws of physics, that the heat is going to rise, right? How can I adapt and plan my house to better take advantage of that heat? Just a possible example. Or if we take a vitamin that our body isn’t absorbing but maybe if we know that there are vitamins that associate with each other...

P18: Hands-on activities make sense to me. When we are confronted with something we don't understand, then we have the curiosity to research it and propose hypotheses. Through questioning, experimenting, observing, testing

P9: I think it's very important that we also give people that honesty, that science is made of 20000 failures and one success. To remove this idea, especially among kids, that "Oh, a scientist was banging their head on the walls all night..." or like thinking and then it was a linear process. But that it's not it. That it’s a construction and, it’s increasingly becoming, a collective construction.

P8: For me, what has been more effective is to-to take everyday things. In knowing what you go through every day (...)I mean, it's both, isn't it? Taking the day-to-day to know science, and knowing science towards applying it to the day-to-day.

P10: One idea, I think, would be to present the information on several levels, right? For example, you’d enter a room in the museum, but also, for example, in a book or any pedagogical material, it will already have a-a title "Ah! This is the room about theme such and such" This title already gives some background information. Then there's a... I don't know, a paragraph that explains a bit more, and then a paragraph with a more complete explanation, and so the person chooses-starts reading... because it already has a little information, and if they want to know more, they can progress into... a more advanced level... and... can choose how deep they want to dive into the question.

P19: All the teachers I had known up to that point couldn't really make me get into the subjects and this one teacher did. This was due to two things... firstly, he presented an extremely complex subject, something quite difficult to understand. It was like "No way! This is only for... for super smart people" and out of the blue he would start to deconstruct it... into simple pieces. That made all of us, even the worst students in the class, gain interest in the subject. They felt like they were learning and didn't feel judged or as being treated like they were dumb. I think that was the core of it, that when faced with a complex topic you aren’t made to feel judged, nor like you have to be the smartest person, right? That it makes you feel that your intelligence level is perfectly capable and valid for-for you to understand that theme, and to be able to enjoy it, right?

P8_F_C+E: for the kids what I have been doing the most is practical activities, always with the theoretical part, but I think that practical activities when they have to look for... and when I say practical, it doesn't have to be in a lab doing an activity or an experiment, it can be just them searching for the information by themselves

P13: It also goes through the schools, for example, field trips, for example in science, they teach a certain subject and take the class to... to science labs for example. (...) it also involves making the classes a little more dynamic to... to captivate the students... Or even, I don't know if I can address the part about work here, for example, if-ten students want to practice medicine, then, in the 10th grade, maybe we could have uh, a field trip or a day to go see the medical profession, to see what it's like. And even turn the students into teachers. For example, the teach... teaching a class and once a week, or once every two weeks, a student has to read about something and present it by themselves.

P18: And in the same way, connect it to our daily lives. Maybe if you're a teacher who gets their news from the TV, for example, and brings to the classroom something that everyone has heard about and then, maybe from an article, you can get to certain scientific concepts. It will be that way, yes. Taking examples from the day-to-day and showing that science is not outside of it, that it is all around us.

P18: In school when we learn science it seems like a separate subject... and in fact maybe we can talk about science and mix it with history when we talk about maritime exploration, we talk about the ships, we can talk about their physical properties, chemical properties, we can talk about what floats and what doesn't float. Maybe when we talk about spices we can introduce concepts of, I don't know, taste, so I don't think we need to put science in a separate box, we can connect it with a lot of what we learn in

P8: Show them a real situation, a daily situation, and from there they will also get information. That is, it's no longer just this is this because so and so, but rather they have to interpret the situation

P1: I think that the components that are supposedly more focused on informal education today should be integrated into formal education. But they should be formalized, in the sense of being standardized.