



ELSEVIER

Contents lists available at ScienceDirect

Learning and Instruction

journal homepage: www.elsevier.com/locate/learninstruc



Tell me how: Teaching science with narration vs exposition

Sara Soares^{a,b,*}, João O. Santos^{c,2}, Cláudia Simão^{d,3}, Matilde Gonçalves^{e,4},
Raquel Barata^{f,g,5}, Rita Jerónimo^{a,6}, Régine Kolinsky^{b,h,7}

^a Iscte-Instituto Universitário de Lisboa, CIS-Iscte, Lisboa, Portugal

^b Unité de Recherche en Neurosciences Cognitives (Unesco), Center for Research in Cognition and Neurosciences (CRCN), Université Libre de Bruxelles (ULB), Belgium

^c ISPA - Instituto Universitário de Ciências Psicológicas, Sociais e da Vida, Portugal

^d Católica-Lisbon School of Business and Economics, Universidade Católica Portuguesa, Lisboa, Portugal

^e Centro de Linguística da Universidade NOVA de Lisboa, Faculdade de Ciências Sociais e Humanas, Universidade NOVA de Lisboa, Lisboa, Portugal

^f Jardim Botânico, Museu Nacional de História Natural e da Ciência, Universidade de Lisboa, Lisboa, Portugal

^g Centre for Ecology, Evolution and Environmental Changes, Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal

^h Fonds de la Recherche Scientifique-FNRS, Belgium

ARTICLE INFO

Keywords:

Narrative texts
Science education
Text comprehension processes
RAND characteristics
Construction-Integration model

ABSTRACT

Background: Narrative texts can support science comprehension. Yet, it remains unclear under which conditions they are beneficial.

Aims: We compared the comprehension benefits of science narrative versus expository texts across comprehension levels and retention intervals and identified reader and activity characteristics predicting science comprehension.

Sample: A sample of 125 university students with low prior knowledge of the content of the texts volunteered to participate.

Methods: Participants read both a narrative and an expository text and then completed comprehension questions targeting different levels of comprehension. The contents were related to math or chemistry. Participants completed the comprehension items on that session, and one week after reading the texts. We also assessed participants' pre-reading perceived prior knowledge of the science topics and their post-reading perceptions of the texts' difficulty and interest.

Results: Comprehension scores were higher for the narrative than for the expository text only with the chemistry topic, at all comprehension levels. Scores were higher in immediate comprehension assessments than in delayed assessments. Exploratory analyses suggest that narrative texts support comprehension in part by increasing perceived interest and lowering perceived difficulty.

Conclusion: The narrative advantage was specific to the less familiar content, boosting comprehension across all levels of representation, partly by reducing perceived difficulty and increasing interest — underscoring the importance of aligning text type with learners' prior knowledge and task demands.

* Corresponding author. Centro de Investigação e de Intervenção Social, Iscte – Instituto Universitário de Lisboa, Av. das Forças Armadas, 1649-026, Lisboa, Portugal.

E-mail addresses: sapss@iscte-iul.pt (S. Soares), joao.santos@ispa.pt (J.O. Santos), csimao@ucp.pt (C. Simão), matilde.goncalves@fcsh.unl.pt (M. Gonçalves), arbarata@museus.ulisboa.pt (R. Barata), rita.jeronimo@iscte-iul.pt (R. Jerónimo), kolinsky.regine@ulb.be (R. Kolinsky).

¹ Sara Soares: cognitive and social processes, especially in learning and education.

² João O. Santos: social cognition and connecting philosophy of science with research methods.

³ Cláudia Simão: prosocial behavior, decision-making, public policy design, and positive emotions.

⁴ Matilde Gonçalves: the relationship between social activities, genres and texts and literacy.

⁵ Raquel Barata: evaluation of museum education, including participatory methodologies.

⁶ Rita Jerónimo: interaction with the social and non-social world via cognitive processes.

⁷ Régine Kolinsky: the brain, cognitive, and sociocognitive consequences of literacy.

<https://doi.org/10.1016/j.learninstruc.2026.102436>

Received 1 November 2025; Received in revised form 1 July 2026; Accepted 9 July 2026

Available online 23 July 2026

0959-4752/© 2026 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

1. Introduction

Science education has a multifold nature which encompasses the understanding of science concepts and processes, coping with challenges in everyday life, preparing students for careers in science, and fostering informed participation in society (Stuckey et al., 2013).

Narrative materials combine elements from everyday life with scientific concepts and processes. As such, they have been proposed as a useful text type to make science meaningful and understandable for learners (e.g., Bruner, 1996; Golke et al., 2019; Zabel & Averdunk, 2024). Teaching science with narrative texts can yield better results than relying on conventional expository texts, but evidence is mixed and seemingly dependent on the specific conditions in which it is applied (e.g., Dinsmore et al., 2017; Golke & Wittwer, 2024; Hong & Lin-Siegler, 2012; Maley & Seyedi, 2022; Tobler et al., 2025; Wolfe & Mienko, 2007; for a review, see Soares et al., 2023). To better understand these conditions, it is crucial to examine the psychological processes and the reader, text, and activity characteristics underlying comprehension. These can reveal *how* and *why* narrative texts help (or fail to help) science comprehension under specific conditions, thereby enabling the design of science materials that deliberately leverage the respective advantages of narrative and expository texts.

Existing research has prioritized comprehension outcomes over the processes and characteristics that drive them (Soares et al., 2023). The present study addresses this asymmetry by examining science comprehension at different representational levels and retention intervals, using a theory-driven framework that integrates reader, text, and activity variables (Dinsmore et al., 2017; Soares et al., 2023). We also examined whether reader and activity characteristics predicted comprehension.

1.1. The case for narrative in science education

Expository materials and discourse have been the backbone of science education (Avraamidou & Osborne, 2009). Scientific ideas are often unfamiliar and counterintuitive for novices (e.g., Browning & Hohenstein, 2015). In expository texts, they are structured according to concepts or themes and their relations (e.g., Meyer, 1975). However, this removes them from their human and cultural context (e.g., Chappell & Varelas, 2020), often resulting in technical and dense texts (e.g., Plavén-Sigray et al., 2017; Snow, 2002).

In contrast, narration is a cross-cultural activity (Bruner, 1986) in which information is structured in sequences of temporally and causally linked actions and events (e.g., Adam, 2011). Bruner (1986) distinguished between this narrative mode and the paradigmatic mode, built on general laws based on logic and abstract reasoning.

Although these modes are complementary (Bruner, 1986), narration plays a key role in making sense of the world and remembering information (Bruner, 1996; Newkirk, 2014). Narrative features, such as chronological or causal framing, aid comprehension (e.g., Bolkan et al., 2020); simultaneously, the presence of human agents promotes engagement (e.g., Kubli, 2005) and makes narrative texts memorable (Hong & Lin-Siegler, 2012; Schank & Berman, 2002).

Following Brunauer's proposal of a narrative and a paradigmatic mode (Bruner, 1986), Dahlstrom (2014) argued that narratives can be used with non-expert adults for informational purposes. Bruner himself (1996) argued that the narrative mode, being more intrinsic to the human mind, can be useful at any age. This is an enticing idea, as narrative texts tend to be more common among younger students, who, by the time they reach university, have been exposed mainly to expository texts and have developed corresponding reading strategies (Kloser, 2013). This issue also connects back to calls for greater flexibility in science textual materials (e.g., Webb, 2010).

These claims supported our focus on university students. However, we also draw on studies with younger students whenever relevant.

1.2. Beyond outcomes, A multi-feature approach

In the present study, science comprehension refers to readers' ability to construct a coherent mental representation of the ideas expressed in science texts (Kintsch & Mangalath, 2010; Snow, 2002). This goes beyond decoding to encompass connecting concepts, inferring causal and mechanistic relations, and integrating text with existing scientific knowledge.

In the context of science comprehension, some authors have argued that science education should deliberately integrate literacy practices into science instruction and materials. These literacy practices include reading, writing, discussion, and argumentation (e.g., (Morais & Kolinsky, 2016)). In addition, this integration should involve working with a wider range of science text types rather than relying solely on traditional expository textbooks (e.g., May et al., 2020). Such an integration is essential, as science texts are conceptually and linguistically complex and require specialized vocabulary, causal reasoning, evidence evaluation and active knowledge construction (e.g., Kintsch & Mangalath, 2010; McNamara et al., 1996; Morais & Kolinsky, 2016). LiterFollowing Brunauer's proposal of a narrative and a paradigmatic mode (Bruner, 1986), Dahlstrom (2014) argued that narratives can be used with non-expert adults for informational purposes. Brunauer himself (1996) argued that the narrative mode, being more intrinsic to the human mind, can be useful at any age. This is an enticing idea, as narrative texts tend to be more common among younger students, who, by the time they reach university, have been exposed mainly to expository texts and have developed corresponding reading strategies (Kloser, 2013). This issue also connects back to calls for greater flexibility in science textual materials (e.g., (Webb, 2010)).acy practices systematically develop these abilities by fostering decoding, vocabulary growth, inferencing, and strategic processing (e.g., Barth et al., 2014; Spencer et al., 2019). Moreover, textual variety trains flexibility and genre-specific reading strategies. An example is handling different cohesion patterns and discourse structures, which, in turn, enables more active and deeper construction of knowledge (e.g., Kulesz et al., 2016; McNamara et al., 1996).

The construction-integration model (Kintsch, 1998) is widely recognized and extensively applied in text and discourse comprehension research across cognitive psychology, education, and computational modeling (Graesser et al., 2011; McNamara & Magliano, 2009; Goldman et al., 2016). It has been used to investigate comprehension of both narrative and expository texts (e.g., Neri et al., 2023; Zhang et al., 2023).

According to this model, comprehension involves two processes, assumed to be dynamic and iterative: constructing (i.e., representing meanings) and integrating (creating a coherent representation) meaning from text. In the construction of a rough draft, *everything* related to each sentence is automatically activated and retrieved, such as text meaning, prior knowledge, and context. Integration acts like a selection mechanism: strongly connected ideas are more activated, reinforce each other, and are maintained, while less connected ideas (peripheral, weak or irrelevant) fade away, leaving a clear mental representation.

The model also assumes three levels of representation, each supported by distinct cognitive processes. The *surface level* encodes literal wording and syntax (via lexical access and syntactic parsing) and is assumed to minimally contribute to understanding (McNamara & Magliano, 2009). The *textbase level* constructs an explicit representation of the text meaning supported by processes such as local coherence inferencing and explicit meaning integration. This is done via a network of propositions – complete ideas that are interrelated and based on words, syntactic relationships, and cohesion-related inference generation (Kintsch, 1998). These idea units are organized into larger meaning units according to global topics and their interrelationships, that link larger portions of the text. At the level of the *situation model*, propositions are further elaborated and integrated with readers' long-term memory – such as prior science knowledge – to form a coherent mental representation of the phenomena described. This is supported by

knowledge retrieval and global coherence inferencing.

The textbase representation is usually assessed by literal recall, paraphrasing, and low-level inferencing, whereas the situation model assessment relies on higher level inferencing and transfer (Ferstl, 2001; Kintsch, 1994). Although Kintsch's model (1988) does not make text type-based specific predictions, some studies have investigated differences between levels of representation using science narrative texts and expository texts.

Zwaan and collaborators (e.g., Zwaan & Radvansky, 1998) deeply investigated situational models in narrative texts and discuss how they enable rich, multidimensional models (time, space, causation, motivation, protagonist). Wolfe and Woodwyk (2010) found that, among university students, narrative texts promoted a stronger textbase representation by directing the focus to the chronology of events ($d = 0.4$). However, expository texts provided a better situation model than narrative texts by facilitating connections between new information and prior knowledge ($d = 0.4$). Yet, Wannagat et al. (2022) found null results regarding the relative effectiveness of narrative and expository texts for textbase and situation models in elementary school students (both $r = 0.03$). In an effort to highlight the underlying processes, Arya and Maul (2012) proposed different levels of comprehension based on the construction-integration model but did not include them as a factor in their analysis of middle school students.

Considering different levels of comprehension can yield important insights for understanding comprehension outcomes, as they are supported by different cognitive processes. Findings in this area have been mixed (Soares et al., 2023). In general, however, there is evidence that narrative texts may benefit comprehension across different educational and comprehension levels (Arya & Maul, 2012; Browning & Hohenstein, 2015; Dinsmore et al., 2017; Hadzigeorgiou et al., 2012; Hong & Lin-Siegler, 2012; Tobler et al., 2025; see Soares et al., 2023; Tobler et al., 2024 for reviews). Among university students, it has been found that biology students benefited from narrative texts in items measuring transfer of knowledge (i.e., open questions measuring students' ability to transfer the learned knowledge to new contexts; $d = 0.46$) on enzyme energetics and molecular stochasticity (Tobler et al., 2025). Narrative texts also promoted knowledge complexity (i.e., conscious reflection on science evidence; $\eta^2 = 0.14$) on genetically modified organisms (Dinsmore et al., 2017). Both transfer and knowledge complexity pertain to the situational model level of representation.

However, other studies have failed to find a narrative comprehension advantage among university students, observing no differences or even worse performance with narrative texts on biology topics (Golke et al., 2019; Golke & Wittwer, 2024; Maley & Seyedi, 2022; Wolfe & Mienko, 2007) and technical and natural processes (Maley & Seyedi, 2022). For instance, Golke and Wittwer (2024) found that comprehension was significantly higher with the expository text when reading about hormones and blood circulation ($\eta^2 = 0.14$).

Studies have also investigated retention measures, which give information on both initial encoding and the durability of memory traces (American Psychological Association, 2018; Kintsch & Mangalath, 2010). Textbase representations decay more quickly than situation models, and differences in how readers construct meaning in working memory also shape what is retained over the long term (Kintsch & Mangalath, 2010).

One study using recall tasks found a narrative advantage primarily at the textbase level: narrative introductions on biological, technical, or natural processes were better recalled than expository ones among university students (Mensink et al., 2021; $\eta^2 = 0.06$). In Negrete and Lartigue (2010), the expository group outperformed the narrative group on several immediate memory tasks for both the Crabs (biology) and Nitrogen (chemistry) texts, but the gap narrowed at the delayed test.

For Crabs, the second-session differences were small to moderate across retell, identify, recall, and context, with the factual group still ahead on identify and recall. In Nitrogen, the narrative group did better on identify but the factual group retained an edge on recall and context.⁸

Studies that used broader comprehension measures give a more mixed picture, but they do not always separate recall from deeper comprehension, so they cannot be cleanly classified by representational level. Even so, these studies found narrative advantages in both immediate and delayed assessment (Arya & Maul, 2012, immediate assessment of Galilean telescope: $d = 0.17$; delayed assessment of both Galilean telescope and radioactivity: $d = 0.95$; Hadzigeorgiou et al., 2012, immediate assessment of Nikola Tesla's story: $d = 1.31$, delayed assessment of Nikola Tesla's story: $d = 1.72$). Another study found narrative benefits for recall at the delayed assessment – students who read about scientists' struggles showed better recall and complex problem solving of physics' topics than those who read the expository version (Hong & Lin-Siegler, 2012; $d = 0.9$). Others found no differences between text types in immediate assessments of recall: Arya and Maul (2012; radioactivity topic: $d = 0.17$) and Wannagat et al., 2022 (various science topics; 2022; $r = 0.03$) for younger students and Wolfe and Mienko (2007; circulation system: $d = 0.27$) for undergraduates.

Whereas the construction-integration model specifies the cognitive architecture of comprehension, the RAND Reading Study Group framework (Snow, 2002), contextualizes it – mapping the reader, text, and activity variables that jointly determine comprehension outcomes. RAND is also an extensively used model (Kendeou et al., 2014; McNamara, 2010). Akin to the construction-integration model, reading comprehension is conceptualized as “the process of simultaneously extracting and constructing meaning through interaction and involvement with written language” (Snow, 2002, p. xiii). Achieving this comprehension is understood as a long-term developmental process, which culminates in proficient readers able to read varied materials for varied purposes with ease and interest, even when materials are not considered particularly easy or interesting.

RAND frames comprehension as a dynamic interaction between three factors: the reader, the text, and the activity. The reader brings to the table their cognitive abilities (e.g., memory, inferencing), motivation (e.g., interest, need for cognition), knowledge (linguistic, topic, comprehension strategies), and experiences. These can present inter-individual differences as well as intra-individual differences (e.g., with different texts or activities). Research has shown that these characteristics can influence comprehension, yet how to enhance them during instruction is less well understood (Snow, 2002). Every text has specific features which may impact comprehension processes: differing structures (e.g., temporal, conceptual), text types (e.g., narrative, expository) and contents (e.g., imaginary realms, scientific). Texts also vary in cohesion and coherence levels (McNamara et al., 2010).

It is important for educators to acknowledge these features and know how to apply (text features) or engage (reader features) them in specific activities. The activity concerns the purposes or tasks, the operations that sustain text processing, and the outcomes of that activity. Purposes can be multifold (e.g., entertainment, learning, passing time) and can change throughout reading. The cognitive operations refer to the processes that, according to Kintsch and Mangalath (2010), support the three levels of representation (e.g., decoding the text, inferencing, integration, and retention). The extent to which these processes are involved is highly dependent on reader and text features, which can make the text content more or less difficult or more or less novel. Outcomes may include satisfying curiosity, solving problems, increasing knowledge, and fostering emotional engagement, and they may not always match the initial purpose. While learning outcomes guide instructional design, the comprehension processes that drive them

⁸ Because the paper reports only summary percentages and standard deviations, any effect-size estimate is approximate.

remain underexplored. One example is precisely how text type affects textbase and situation-model construction.

Reader, text and activity are further embedded in a wider sociocultural context – shaped by values, socioeconomic background, and historical experience – which, while beyond the scope of the present study, forms part of the full RAND model (Snow, 2002).

Several studies have reviewed reader variables and text variables in general narrative texts (vs. expository texts). The reviewed variables include, respectively, decoding, working memory (Hannon, 2022) and prior knowledge (Mar et al., 2021), and text length and difficulty (Hannon, 2022; Mar et al., 2021). However, evidence concerning science narrative texts remains sparse.

Some research looked at the impact of text, reader, and activity features when learning science from narrative and expository texts. Particularly of interest to the present study are studies that included personal, historical and/or a discovery element to the narratives; these seem to be particularly helpful for science comprehension among both middle school students (Arya & Maul, 2012; Hadzigeorgiou et al., 2012) and high school students (Hong & Lin-Siegler, 2012). As aforementioned, narrative texts featuring scientists' struggles during their scientific work led to better performance not only compared with expository texts ($d = 0.9$; Hong & Lin-Siegler, 2012) but also compared with narrative texts describing only the scientists' achievements ($d = 0.81$). Adding features such as a sense of wonder and the questioning of conventions to the description of a scientist's achievements also led to better results than simply presenting the scientific content ($d = 1.72$; Hadzigeorgiou et al., 2012). Finally, focusing on the discovery aspect of science topics also proved useful for understanding scientific content (Arya & Maul, 2012; $d = 0.43$ and $d = 0.17$).

Still concerning the text factor, different scientific information may yield different comprehension outcomes, even within the same study (Arya & Maul, 2012; Cervetti et al., 2009; Hong & Lin-Siegler, 2012). For instance, for the topic of radioactivity, Arya and Maul (2012) found significant differences between the narrative and expository texts only at the one-week delayed assessment, whereas for the topic Galilean telescope differences were found both in immediate and delayed assessments.

Regarding reader features, studies have mostly examined the role that prior science knowledge plays in the comprehension of narrative and expository texts (e.g., Tobler et al., 2022; Tobler et al., 2025; Wolfe & Mienko, 2007; Wolfe & Woodwyk, 2010). Prior science knowledge is linked to the level of novelty and difficulty students feel while reading a text. Studies with university students tend to conclude that students with lower prior knowledge benefit more from the narrative version (Tobler et al., 2025; Wolfe & Mienko, 2007; Wolfe & Woodwyk, 2010). Wolfe and Woodwyk (2010) found a positive correlation between prior knowledge and learning from the expository text ($r = .44$; see Browning & Hohenstein, 2015 for younger students and Tobler et al., 2022 for the opposite result with university students). In contrast, students' own perceptions of their prior knowledge – how knowledgeable they feel before reading – have not been systematically examined in the narrative-expository science literature. And yet, they are theoretically important as metacognitive indicators of knowledge availability and task readiness (Hattan, Alexander, & Lupo, 2023; Thiede et al., 2003).

Affective aspects, such as interest, are crucial in science education (e.g., Fredricks et al., 2018). Authors have consistently suggested that narratives have a positive effect on students' affective dispositions (e.g., Avraamidou & Osborne, 2009; Bruner, 1986; Norris et al., 2005). Yet, as regards science narrative texts, this effect is sometimes observed (e.g., Golke & Wittwer, 2024; Hong & Lin-Siegler, 2012), but is inconsistent (e.g., Cervetti et al., 2009; see the review by Mar et al., 2021; for narratives in general).

Moreover, affective variables tend to be examined as an outcome rather than as a characteristic contributing to comprehension outcomes. Still, there is some evidence for such a relationship as regards science narrative texts. A positive correlation between interest in science

materials and their comprehension has been observed ($R^2 = 0.61$; Reuer, 2012). A mediation of situational interest (including aspects like need for cognition, need for affect, and individual interest) in the effect of text type on science understanding was also reported ($\beta = 0.10$, $b = 1.96$, $SE = 0.88$, 95% CI [0.41, 3.87]; Golke & Wittwer, 2024). These examples tapped into interest elicited by the text which, rather than reflecting a stable reader characteristic, operates here as part of the activity feature. This is because it emerges from the *processing* of the specific text during the task. Ease of reading, a proxy for subjective text difficulty, is another activity feature examined in one study on science narrative texts (Cervetti et al., 2009).

Drawing on the RAND framework and the construction-integration model, we propose that a multi-factor lens can reveal the conditions under which specific narrative texts benefit science comprehension. In the present study, this lens encompasses representational level, retention interval, perceived prior knowledge, text difficulty, and interest in the text. Such an approach may reveal *specific narrative effects* that emerge under particular configurations of these characteristics.

1.3. The current study

We set out to answer three main research questions and a fourth exploratory one: 1) To what extent do narrative texts – specifically those featuring human agents doing science in a historical context – benefit science comprehension compared to expository texts?; 2) Is the difference between narrative and expository texts more pronounced at more complex levels of representation?; 3) To what extent are narrative vs. expository texts effective at promoting delayed retention?; and 4) Do reader and activity characteristics (such as perceived pre-reading prior knowledge, and post-reading text difficulty and interest in the text) predict science comprehension from different text types?

Following Soares and collaborators (2023), we framed the variables according to the features of the comprehension process proposed by RAND, namely characteristics of the text (expository vs. narrative with scientists doing science), of the reader (perceived prior knowledge) and of the activity (levels of comprehension, retention interval, difficulty, interest). We based our understanding of different levels of comprehension and their retention over time on the construction-integration model, which is shared by the RAND framework.

To investigate our research questions, we compared how university students understood science contents from a narrative text and a more conventional expository text. To ensure that each participant would read a text of each type, we created texts on two different science topics. Crucially, the different text types (narrative vs. expository) on a given topic presented the same contents. Beyond this methodological requirement, this diversity is also theoretically relevant, as the literature has shown that different topics may yield different results (Arya & Maul, 2012; Cervetti et al., 2009).

Whereas most previous studies took place in schools, our context was the laboratory. Conducting the study in a controlled laboratory setting enabled precise isolation of text-type effects, independent of curriculum demands, teacher influence, or peer dynamics that characterize naturalistic classroom research. As participation is voluntary, students may also have been more motivated. Such an approach can reveal baseline conclusions for adult science comprehension, establishing causal mechanisms before designing practices and applying them in educational settings.

As regards our theoretical background, it is not clear whether young adults would benefit more from the narrative or the expository text. On the one hand, they have learned to read expository texts and developed corresponding strategies, so expository texts may facilitate integration with prior knowledge (Kintsch, 1988; Snow, 2002; Wolfe & Woodwyk, 2010). On the other hand, we selected our sample to ensure that prior knowledge of the presented topics was very low. It is therefore possible that narrative texts provide an advantage at both the textbase (Wolfe & Woodwyk, 2010) and situation-model levels, as they have the potential

to generate richer situation models (Zwaan & Radvansky, 1998).

The reviewed literature on comprehension points in the same direction, be it with university students (Dinsmore et al., 2017; Tobler et al., 2025; Wolfe & Woodwyk, 2010) or younger students, but using similar paradigms (Arya & Maul, 2012; Cervetti et al., 2009) or text features (Arya & Maul, 2012; Hong & Lin-Siegler, 2012). Thus, we expected the narrative text to yield better comprehension than the expository text for at least one of the science topics, especially in questions targeting the situation level and its intersection with the textbase level.

Narrative texts' causal/temporal structure can both aid proposition linking at the textbase level (Wolfe & Mienko, 2007) as well as building richer and more coherent situation models (Zwaan & Radvansky, 1998). Drawing on this rationale and similar studies, we expected information to be more integrated and consolidated with narrative texts than expository texts after a week (Arya & Maul, 2012; Hong & Lin-Siegler, 2012).

Finally, we formulated some exploratory hypotheses concerning selected reader and activity characteristics. First, we expected pre-reading perceptions of knowledge (reader feature) to predict comprehension. This relationship might vary across science topics, with participants who perceive themselves as having lower prior knowledge benefiting more from the narrative version (Negrete & Lartigue, 2010; Tobler et al., 2025; Wolfe & Mienko, 2007). Second, we expected post-reading evaluations of difficulty of the text and interest (activity features) to help explain a potential narrative advantage and, possibly, a topic-related advantage. The RAND framework proposes that comprehension emerges from dynamic interactions between the reader, the text, and the activity. Accordingly, perceived cognitive demands of the task (e.g., difficulty) and motivational responses (e.g., interest) to texts should relate differently to actual understanding depending on the reader's prior knowledge and the characteristics of the text.

We expected narrative texts to be considered more interesting and less difficult, although the literature on this issue is mixed (e.g., Cervetti et al., 2009; Hong & Lin-Siegler, 2012; Reuer, 2012). By treating these three variables as subjective perceptions that may shape or reflect the unfolding of comprehension, we aim to help fill a gap in the literature concerning the processes underlying the comprehension of narrative science texts.

2. Method

2.1. Design and participants

The experiment employed a 2 (text type: narrative texts vs. expository texts) $\times$ 2 (topic: chemistry vs. math content) $\times$ 2 (session: 1 vs. 2) $\times$ 4 (comprehension level: L1–L4) factorial design. The dependent variable was the *comprehension score*. Text type, session, and comprehension level were manipulated within participants. The expository text condition served as the control condition. The different text types presented the same content.

Sample size was informed by a power analysis using G*Power (Faul et al., 2009). We conducted an a priori power analysis for linear multiple regression model (random model) with six predictors – text type, topic, item level, session, text order; item set order – targeting 80% of power with a small to medium effect size ($p2 = 0.13$). Although the recommended sample size was 101 participants, we recruited a larger sample, to make it more robust to attrition (dropouts and excluded participants). Results were not analyzed until all data was collected.

A total of 143 university students from different universities and fields took part in the study. Participants were recruited through two universities' study participation reward system and by flyers. Inclusion criteria were the following: being undergraduates, postgraduates, or Master students; not being enrolled in courses where chemistry, physics, biology, or history of math were core topics; being native speakers of European Portuguese; not having a diagnosed language or

developmental disorder; completing both sessions.

For data analyses, we removed one participant who had a development disorder, five for problems in data collection or recording, two because of fundamental concerns regarding their comprehension of the texts, and 10 who did not complete the second session. The final sample included 125 participants (90 women, 35 men) aged from 18 to 25 years ($M_{age} = 21$ years; $SD = 1.75$; see Table S1 for socioeconomic information). Most participants (75%) had a low to moderate level of formal contact with exact and natural sciences, as determined by their chosen academic tracks in secondary school and university (see Table S2). This proportion targeted low-knowledgeable students on the evaluated topics (e.g., Prins et al., 2017) while ensuring some level of variability to avoid a floor effect.

2.2. Materials

2.2.1. Designing Principles

Across three pilot studies, material design was refined together with experts in text linguistics and science, using both quantitative (see FS1 and Table S5) and qualitative (Table S4 and Fig. S2) analyses (for a detailed account, see subsection S1.2, pp. 3–22, of the Supplementary Materials).

In line with text linguistic models (e.g., Adam, 2011), expository texts were structured around conceptual relations, conveying contents in an abstract manner, whereas narrative texts were temporally structured, embedding scientific reasoning within the sequence of actions (i. e., Pasteur's experiments; Leonardo Pisano's notation work). Narrative and science components were tightly integrated to ensure comprehension was mediated by the narrative frame, with as few fictional elements as possible. Other criteria used in the development of the texts included: avoiding non-essential details, balancing the quantity and the complexity of contents, and making clear which ideas were central and how they were connected (see Table S3).

Despite the necessary differences between text types, we ensured textual equivalence across them. Namely, additional narrative information was matched by expository-congruent information, such as in-text citations, to ensure equal length, and texts were balanced in a set of psycholinguistic parameters (e.g., number of words, sentence length, frequency; see Table 1). Pilot ratings based on relevant literature (Arya & Maul, 2012; Limpo et al., 2014) confirmed that, in terms of text type, narrative texts were generally judged clearer, more globally coherent and interesting, and less difficult than expository texts (see Table 2; see Appendix A and Table A1 for more criteria). In terms of science topic, math texts were rated more familiar, easier and more interesting than chemistry texts; see Table 3).

Regarding the comprehension items that measure our main dependent variable, they were designed to cover a set of comprehension levels using simple and direct sentences. This was important, as participants would give their responses after a first quick approach to these novel science contents. Four levels of comprehension items (L1 – L4) were developed by aligning item demands with levels of representation from the construction-integration model (Kintsch, 1998; see Arya & Maul, 2012 for a related approach). Item format (L1 – L4) varied to match the cognitive demands of each representation type to isolate specific processes. L1 items used direct, verbatim prompts to isolate surface-textbase access, avoiding inference demands. L2 items assessed a higher textbase level via multiple-choice to promote basic connections of low-level inferencing. Both L3 and L4 were open-ended. L3 items assessed the interface between levels of representation, as they targeted understanding and explaining key conceptual ideas (both meaning and integration based). L4 items asked participants to transfer and apply these key ideas to novel contexts, targeting the situation model level (Ferstl, 2001; Kintsch, 1994). By administering them both immediately and after a delay, we were able to assess retention over time.

Texts and comprehension items underwent independent evaluation

Table 1

Values and Averages and Standard Deviations (When Possible) of the Psycholinguistic Parameters Controlled for.

Text	N WRD	N SNT	N PAR	N WRD/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 WRD/SNT: number of words per sentence; WRD LNT: word length; WRD FREQ: word frequency; N SUB PAR: number of subordinate clauses per paragraph; NT: narrative text; ET: expository text. 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 > 0.257). The math narrative text had significantly more subordinate clauses ($M = 15.83, SD = 5.42$) than both the math expository text ($M = 7.33, SD = 4.93, p = .02$) and the chemistry expository text ($M = 8.71, SD = 3.5, p = .05$); this is likely because many conjunctions that introduce subordinate clauses (e.g., if, that, however, because, in order to, though) were inserted to present arithmetic procedures from Leonardo Pisano's point of view.

Table 2

Averages (and Standard Deviations) of Part of the Post-Reading Evaluations Made by Participants of each Text Type Read.

Parameter	Question	NT		ET		<i>t</i>
		M	SD	M	SD	
Clarity	On a scale of 1 (not clear at all) to 5 (very clear) please rate how clearly the contents were presented	3.88	0.93	3.44	1	3.6 ***
Comprehension difficulty	On a scale of 1 (not difficult at all) to 5 (very difficult) please rate how difficult it was for you to understand the text	2.61	1.05	3.06	1.12	−3.12 **
Interest	On a scale of 1 (not interesting at all) to 5 (very interesting) please rate how interesting you found the text	3.39	1.21	2.84	1.18	2.9 ***
Global Coherence (text as a whole)	On a scale of 1 (totally fragmented) to 5 (totally integrated) please rate the extent to which you felt that the text was a unified whole	4.13	0.90	3.73	0.99	3.29 **

Note. The df were 123. ET = expository text, NT = narrative text.

by science and linguistics experts based on the final criteria for developing the materials (see Table S3 of the Supplementary Materials).

2.2.2. Texts

The following concerns the “text” component of the RAND framework. Two pairs of texts (chemistry and math, each in an expository and narrative version) were developed. The chemistry texts dealt with the concept of molecular chirality, explaining that some crystals have the same molecular composition yet different molecular structures, and describing the methods used to examine these parameters. Math texts dealt with mathematical notation, describing three different notations in terms of representation of quantities and basic arithmetic, with a particular focus on their challenging and advantageous aspects. The science topics selected for the texts should not be taken as representative

Table 3

Participants' Ratings of the Post-Reading Perceived Difficulty and Novelty of each Text they Read, as Well as of Their Pre-Reading Perceived Prior Science Knowledge.

Measure	Chemistry texts		Math texts		<i>t</i>
	M	SD	M	SD	
Level of difficulty of the text (evaluation scale)	3.04	1.16	2.62	1.01	−2.91 **
Novelty of the text (evaluation scale)	4.45	0.88	3.54	0.92	−7.71 ***
Perceived prior knowledge on the text's contents	1.10	0.39	1.82	1.13	−14.3 ***
Interest in the text	2.96	1.24	3.27	1.19	2.10 *

Note. The df were 124. These measures were evaluated on 5-point increasing scales.

of their broader fields. Full texts and English translations are available at https://osf.io/jr89m/overview?view_only=04bd14106be74371afcecd9d9a024287; excerpts are provided in Table 4.

2.2.3. Comprehension items

The comprehension items (item levels and retention interval of application) pertain to the cognitive operations underlying comprehension, and hence are part of the “activity” element of the RAND framework. For each topic, two sets of different but equivalent items (one for each session) were created following the construction-

Table 4

Excerpts of each of the used texts.

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 analyzing 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, 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. ET = expository text, NT = narrative text. Each text had one and a half pages. The original texts and their English translations can be found at https://osf.io/jr89m/overview?view_only=04bd14106be74371afcecd9d9a024287.

Table 5
Example of Items of each Level of Comprehension.

Item Level	Example	Level of representation (C-I model)
L1	Fill in the gaps. In the Babylonian notation there was a symbol for the quantity $\frac{1}{2}$ and a symbol for the quantity $\frac{1}{3}$.	building coherence at the level of individual propositions (micro propositions)
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	establishing coherence between propositions (macro propositions)
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.	establishing the propositional meaning of the text (textbase macrostructure)
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.	integrating key text ideas with participants' own knowledge and understanding (situation model)

integration model (see Table 5 for examples). These were applied right after the reading of the texts and one week later.

2.2.4. Tasks measuring Reader's characteristics

Participants completed a set of tasks directed at reader and activity features (see the complete set of tasks in https://osf.io/jr89m/overview?view_only=04bd14106be74371afcecd9d9a024287). The focal tasks pertained to knowledge (perceived prior knowledge) and activity (difficulty, interest).

Participants' perceived prior science knowledge was assessed with a questionnaire in which participants rated 30 science topics (including the ones from the texts) on a 5-point scale. The scale ranged from not knowing anything at all to being able to make elaborate applied analysis. Five of these topics pertained to the contents of the chemistry text, five to the contents of the math text, and the remaining to one of the following topics: general chemistry (4), math (4), biology (4), physics (4), and astronomy (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 calling attention to the texts' topics. Such a questionnaire naturally 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.

2.2.5. Tasks measuring activity characteristics

We included the items from the difficulty and interest evaluation scales. As explained in the Designing Principles (see 2.2.1) and can be seen in Table 2, these items were measured on a 5-point scale (1: no interest or no difficulty; 5: a lot of interest or difficulty), right after participants read each text.

2.3. Procedure

The study was approved by the institutional review board (agreement 47/2019). Participants were treated in accordance with APA ethical guidelines. All signed an informed consent form prior to completing the study activities and were free to skip any part of the study or drop out without penalty. They received either course credits or two 10€ vouchers for their participation.

The first session took place in two universities. For most participants, the second session took place in one of these universities. However, due to the declaration of a state of emergency during the COVID-19 pandemic and the subsequent closing of universities, 19 participants completed the second session online. To ensure equivalence with in-person data collection, all online sessions were conducted via video call, following the same procedure, timing, and instructions. This protocol maintained direct experimenter oversight and allowed for real-time clarification of instructions, minimizing procedural differences between collection conditions.

All sessions were conducted by the first author and adhered to the planned duration and procedure. Participants were individually and randomly assigned to one of the two main conditions, which served as the unit of analysis. Participants started session 1 by completing the science background items, as they contained information that was to be acquired through the texts. They then read the texts. Given that they read both text types, topic and text type combinations were counter-balanced between participants to control for topic effects (see Arya & Maul, 2012; Cervetti et al., 2009). Participants thus read one combination, either chemistry narrative text with math expository text, or chemistry expository text with math narrative text.

Participants were instructed to read each text at their own pace. They were allowed to underline or re-read portions of the text if they felt insecure. Permitting re-reading aimed at reducing test anxiety and approximate naturalistic reading conditions (e.g., Peterson, 1991), thereby improving ecological validity within the controlled setting.

Once they had read the text, the experimenter collected it, and participants completed the evaluation scale for that text (which included difficulty and interest in the text). They then answered the comprehension items from the first text. The procedure was then repeated for the second text. Presentation order of texts and item sets was counter-balanced across participants to minimize order and practice effects. At the end of the session, participants filled out individual measures pertaining to literacy and sociodemographic variables (see Materials section above for details).

The second session took place one week after the first, a common interval in the relevant literature (e.g., Arya & Maul, 2012; Hong & Lin-Siegler, 2012; Negrete & Lartigue, 2010). Participants completed a second set of different but matched comprehension items, with presentation of item sets also counterbalanced across participants. The session ended with individual measures pertaining to motivational dispositions.

Most tasks were presented and completed on a computer via Qualtrics at the participants' own pace. Detailed instructions were provided on the Qualtrics tabs, but the experimenter provided general instructions and encouraged participants to clear any doubts.

For an illustration of the flow of the procedure, see Appendix B (Table B1).

3. Results

To test the focal hypotheses, we fitted a linear mixed model on the comprehension scores, with text type, topic, item level, and session as fixed effects,⁸ and with random intercepts and slopes of topic and item level per participant, using afexmixed() (Singmann & Kellen, 2019) with Kenward-Roger denominator degrees of freedom. For exploratory purposes, we fitted two mixed models. First, we examined how text type and topic predicted participants' ratings of perceived prior knowledge,

difficulty, interest, and novelty, with measure, text type, and topic as fixed effects, and random intercepts and slopes for topic by participant. Second, we fitted a similar model, replacing text type with standardized comprehension scores as a continuous fixed predictor. Estimated marginal means, trends, and multiple comparisons were computed using *emmeans* (Lenth, 2024). No outliers were detected in any model (all Cook's distances <0.08). The datasets and data analysis scripts are available at https://osf.io/jr89m/overview?view_only=04bd14106be74371afcecd9d9a024287.

3.1. Preliminary analysis

Participants were attributed a new randomly generated number, and their order in the database was shuffled according to it. Missing responses were rare and were treated as missing values (10.8% for chemistry and 6.16% for math). Although some individual item responses were missing, no participant omitted all items. Mean scores were calculated based on the available valid items. We addressed reliability at two levels: inter-rater reliability for the subjectively scored items, and internal consistency for the full comprehension battery.

For the open-ended items (L1, L3 and L4), which required rater judgment, one independent rater scored a random subset of 697 item responses. Inter-rater reliability between the independent rater and the first author was excellent: ICC(1) = 0.97, 95% CI [0.966, 0.974], with 91.8% exact agreement [p. 29]. The L2 (multiple-choice inference) items were objectively scored and required no rater judgment. Internal consistency of the comprehension instrument was estimated separately for each version of each topic (chemistry/math x v1/v2), as each version x topic combination constitutes a distinct instrument. Cronbach's α was acceptable for both chemistry versions (v1: α = .72; v2: α = .76) but lower for the math versions (V1: α = .52; V2: α = .63). These lower values are partly attributable to the structure of the battery: the 10 items deliberately span four distinct comprehension levels (L1–L4), each tapping different cognitive processes.

Participants' perceived familiarity with the overall science contents was low (M = 1.64, SD = 0.93), but higher for math than for chemistry (see Table 3). Participants' interest and perceived difficulty was moderate for both the chemistry and math texts (see Table 3). Narrative texts were perceived as less difficult and more interesting than expository texts (see Table 2), and interest was higher for the math than for the chemistry texts (see Table 3). We averaged, separately, the four items of each of the scales measuring participant's pre-reading perceived prior knowledge on the math and chemistry texts' topics. For the math perceived knowledge items, responses were somewhat variable, though still clustered toward the lower end of the scale. This matched our aim of low (perceived) prior knowledge (M = 1.82, SD = 1.13, Min = 1.00, Max = 4, $Range$ = 3). Internal consistency was modest (Cronbach's α = .63).

For the chemistry perceived knowledge items, responses showed a likely floor effect, with very low ratings and little dispersion (M = 1.10, SD = 0.22, Min = 1.00, Max = 2.50, $Range$ = 1.50). Moreover, internal consistency was low (Cronbach's α = .36). This variable was excluded as it was expected to cause range restriction issues in the analysis. For each science topic, we combined participants' ratings of the narrative and expository versions into single composite scores for perceived difficulty and interest.

Responses to the comprehension items were blindly scored using a rubric summarizing five levels of completeness and accuracy (0%, 25%, 50%, 75% and 100%; see Table 6). We used five levels to allow for finer differentiation between levels of partial understanding, as open-ended responses often show gradations in conceptual accuracy. At least 30% of items from each science topic were also scored by a research assistant, with inter-rater agreement at 91.8% (and ICC = 0.97). Discrepancies were resolved through discussion. Scores were normalized and averaged separately per participant and per item level to yield comprehension scores—the dependent variable in the analyses.

Table 6
General scoring guidelines.

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 analyze 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 incomplete or there are more/ bigger imprecision	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. To match their different levels of complexity, L1 items were initially 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. 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 of 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."

3.2. Main analysis

3.2.1. Overview

We first report the focal mixed-model results organized by research question (3.2.1–3.2.5), followed by exploratory analyses of reader and activity characteristics (3.2.5).

The model revealed no significant evidence of higher-order interactions, topic x item level x session interaction: $F(3, 1427.9) = 0.86$, $p = .461$, $\eta_p^2 = 0.005$; all other F s < 1). Similarly, the interactions between text type and session, text type and item level, as well as item level and session, were all non-significant, respectively, $F < 1$; $F(3, 1427.9) = 1.63$, $p = .181$, $\eta_p^2 = 0.003$; $F < 1$.⁹ Fig. 1 depicts the results from this model, showing the estimated marginal means, and 95% CI for comprehension scores across topic, text type and item level.

3.2.2. Comprehension differences between narrative and expository texts

Narrative texts outperformed expository texts when aggregating across topic, item level, and session. Overall, comprehension was better for math than chemistry.

Importantly, there was a significant interaction between text type and topic, $F(1, 123) = 6.10$, $p = .015$, $\eta_p^2 = 0.047$. Multiple comparisons

⁹ The patterns are similar across data collection modes, suggesting no online-specific deviation in performance (see pp. 22-24 of the Supplementary Materials).

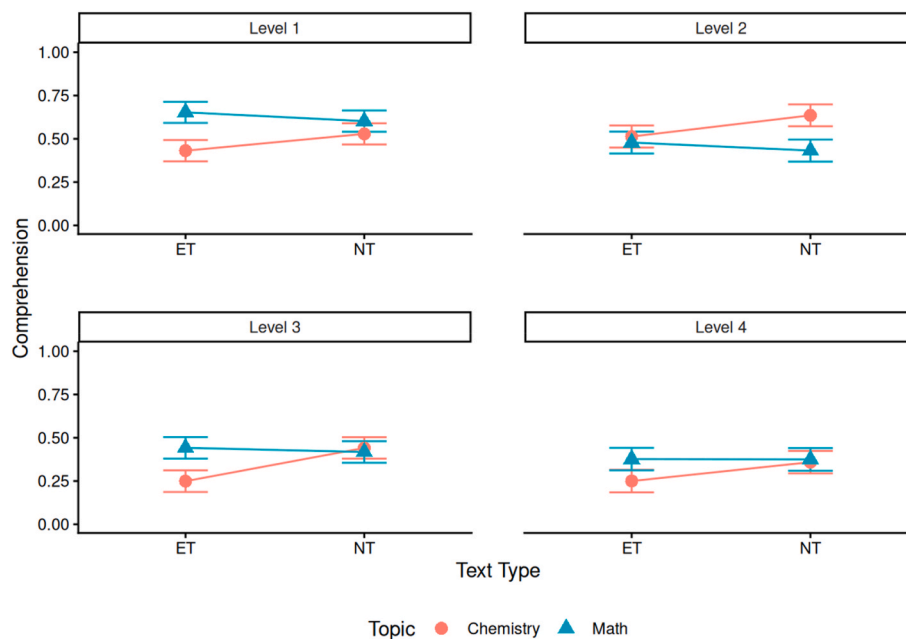


Fig. 1. Comprehension by text type and topic across item level (Normalized proportion of correct answers).

Note. NT = narrative text, ET = expository text. Dots represent the marginal means estimated from the linear mixed model, and the error bars depict the 95% CI for those means. The narrative advantage is visible only for the chemistry topic across all comprehension levels; no reliable difference emerges between text types for the math topic.

revealed that narrative texts led to better comprehension than expository texts for chemistry, but not math, $t(176) = 3.61$, $SE = 0.03$, $p < .001$; $t < 1$, respectively.

3.2.3. Narrative–expository differences across levels of representation

When we look at comprehension for each item level (aggregating across topic, text type, and session), we see that it was progressively lower, $t_{L2-L1}(560) = -2.45$, $SE = 0.02$, $p = .015$; $t_{L3-L2}(553) = -7.78$, $SE = 0.02$, $p < .001$; $t_{L4-L3}(131) = -2.58$, $SE = 0.02$, $p = .015$. Moreover, the model revealed a significant interaction between topic and item level, $F(3, 1427.9) = 31.89$, $p < .001$, $\eta_p^2 = 0.063$. Chemistry and math differed significantly at all levels.

Math outperformed chemistry at three of the four levels (L1, L3, L4); the exception was L2, where chemistry comprehension was higher, $t_{\text{chemistry-math } L1}(565) = -6.28$, $SE = 0.02$, $p < .001$; $t_{\text{chemistry-math } L2}(565) = 5.07$, $SE = 0.02$, $p < .001$; $t_{\text{chemistry-math } L3}(565) = -3.60$, $SE = 0.02$, $p < .001$; $t_{\text{chemistry-math } L4}(565) = -3.04$, $SE = 0.02$, $p = .002$. Multiple consecutive comparisons (with adjusted p -values, Benjamini & Hochberg, 1995) showed that, for math, comprehension was always better for the previous level than the next, except between the second and third levels, $t_{L2-L1}(1178.95) = -8.03$, $SE = 0.02$, $p < .001$; $t_{L3-L2}(1149.87) = -1.17$, $SE = 0.02$, $p = .242$; $t_{L4-L3}(336.19) = -2.31$, $SE = 0.02$, $p = .033$. However, for chemistry, comprehension was significantly better for the second than the first level, $t_{L2-L1}(1179) = 4.4$, $SE = 0.02$, $p < .001$. It then declined significantly from the second to the third level, $t_{L3-L2}(1150) = -10.51$, $SE = 0.02$, $p < .001$, although the decline from the third to the fourth level is not significant at the conventional alpha of 0.050, $t_{L4-L3}(336) = -1.75$, $SE = 0.02$, $p = .082$.

3.2.4. Narrative–expository differences in delayed retention

Comprehension was better in the first session than in the second, $F(1, 1427.9) = 33.56$, $p < .001$, $\eta_p^2 = 0.02$; $M_{\text{First}} = 0.48$, 95% CI [0.44, 0.52]; $M_{\text{Second}} = 0.42$, 95% CI [0.38, 0.46]. This decline is qualified by a topic $\times$ session interaction, $F(1, 1427.9) = 7.37$, $p = .007$, $\eta_p^2 = 0.005$: the decline is significant for both chemistry and math, $t(1428) = 6.02$, $p < .001$; $t(1428) = 2.18$, $p = .029$, respectively, but larger in chemistry than math, $t(1428) = 2.71$, $p = .007$. Consequently, topic domains are

comparable in Session 1, $t < 1$, but math outperforms chemistry in Session 2, $t(247.9) = -3.97$, $p < .001$.

3.2.5. Reader and activity predictors of comprehension

We began by conducting two analyses. First, we explored how text type (narrative vs. expository) and topic (chemistry vs. math) affected participants' ratings of difficulty, interest, knowledge, using mixed-effects models (Analysis 1; see Appendix C). Narratives were rated as less difficult and more interesting than expository texts across both topic domains; these contrasts were larger for chemistry, where prior knowledge was near floor, consistent with the idea that narrative framing has a stronger experiential impact when content is unfamiliar (see Tables C1–C3).

Following these results, we tested whether comprehension performance predicted participants' ratings of difficulty, interest, knowledge, and novelty as a function of content (chemistry vs. math) using a linear mixed-effects model (Analysis 2; see Appendix D). Higher comprehension performance is associated with lower perceived difficulty ($b = -0.285$, 95% CI [−0.362, −0.208]) and higher interest ($b = 0.427$, 95% CI [0.350, 0.504]) in chemistry, and with higher interest in math ($b = 0.246$, 95% CI [0.160, 0.332]). Higher comprehension is also associated with lower perceived novelty in both chemistry ($b = -0.110$, 95% CI [−0.186, −0.033]) and math ($b = -0.178$, 95% CI [−0.263, −0.092]). There is no significant evidence that comprehension performance is related to prior knowledge in chemistry ($b = 0.003$, 95% CI [−0.074, 0.079], $p = .948$), whereas there is significant evidence of a positive relationship in math ($b = 0.117$, 95% CI [0.031, 0.203], $p = .007$; see Table D1). Predicted ratings at three comprehension levels are presented in Table D2.

Analysis 2 identifies which subjective perceptions most reliably signal comprehension success, setting up a mediation analysis to test whether those perceptions help account for the narrative advantage observed with chemistry.

To examine whether subjective text experiences help explain the text-type effect on comprehension, we conducted mediation analyses with difficulty and interest as mediators using the *mediation* R package (Tingley et al., 2014; Analysis 3; see Appendix E). This analysis was

restricted to the first-session data, as ratings were collected immediately after reading, and to chemistry comprehension, as the text-type effect (the one that might be mediated) appeared only for that topic. Comprehension scores were averaged across item levels within each participant $\times$ text type combination, and the resulting values were standardized. Both the mediator model (rating $\sim$ text type) and the outcome model (comprehension $\sim$ text type + rating) were estimated using ordinary least squares regression. Text type (narrative text vs. expository text) was specified as the treatment, the rating (difficulty or interest, also standardized) as the mediator, and comprehension as the outcome. All indirect effects were estimated via 5000 bootstrapped samples. There was clear evidence that both difficulty and interest mediated the narrative effect.

For difficulty, the average causal mediation effect (ACME) was 0.150, 95% CI [0.04, 0.30], $p = .003$, indicating that narrative texts supported comprehension in part by reducing perceived difficulty relative to expository texts. The corresponding direct effect (ADE) remained significant after adjusting for difficulty (0.43, 95% CI [0.07, 0.76], $p = .019$), and the total effect of text type on comprehension was 0.58, 95% CI [0.23, 0.90], $p < .001$, with approximately 26% of this effect mediated by difficulty. For interest, the mediation pattern was even stronger: ACME = 0.25, 95% CI [0.08, 0.44], $p = .005$; ADE = 0.32, 95% CI [0.015, 0.625], $p = .040$; total effect = 0.58, 95% CI [0.245, 0.911], $p = .001$; approximately 44% of the chemistry text-type effect on comprehension was mediated by interest. Both results indicate partial mediation, consistent with the interpretation that narrative texts in chemistry boosted comprehension partly by making the text feel less difficult and more interesting, while additional mechanisms may also contribute (see also [Tables E1 and E2](#)).

4. Discussion

The present study aimed at investigating how specific text characteristics, namely science narrative texts featuring scientists doing science, impact different levels of comprehension and delayed retention in non-expert young adults. We also sought to provide exploratory clues on reader and activity features contributing to comprehension.

The experimental manipulation of text type produced a content-dependent pattern of comprehension: narrative texts yielded clearer benefits in the chemistry texts than in the math texts. While interpreting our findings, we will also include possible alternative explanations that are derived from our design's characteristics and possible limitations.

Regarding the characteristics of the narrative texts, temporality and causality were built through the scientists' actions according to the specificities of each science topic. The chemistry narrative text is mostly focused on temporal and causal links, connected to the human character. The math narrative text is more focused on the history and features of notation than on operations, and richer in characters and settings than in causal relations. Given the lack of difference with the expository text, the text features from the math narrative text may have been less helpful in making sense of information compared with the expository structure. The chemistry texts' more sequential and time-bounded structure may have proved particularly helpful to make sense of the science content, as compared to the expository text. This is because in low-knowledge domains temporal and causal scaffolding may be especially critical, as readers lack the prior knowledge base needed to generate bridging inferences autonomously ([Kintsch, 1998](#)). The chemistry narrative provided exactly this scaffolding; the math narrative, while temporally structured, was richer in characters and settings than causal relations, offering less inferential support.

These facts point to an alternative explanation to the narrative advantage. It may be that the chemistry content was more "narrativizable" than the math one. Because text type and topic are necessarily entangled in the chosen within-participants design, it is difficult to ascertain whether the observed advantage reflects the narrative structure, the scientific content, or their interaction. Future studies should

include multiple topic domains per text type to disentangle specificities of particular texts from text type effects, enabling stronger generalization and external validity.

The math narrative text also had more subordinate clauses per paragraph than any other text, a feature that has been suggested to increase complexity and processing difficulty (e.g., [Lord, 2002](#); [Wang, 1970](#)). The absence of a narrative advantage in the math topic could thus be due to syntactic factors. However, the impact of subordination varies according to the specific type of subordinate clause, as well as other text factors (e.g., [Costa, 2005](#); [Gayraud & Martinie, 2008](#)). The idea that subordination per se necessarily makes processing more difficult has therefore been questioned (e.g., [Baten & Håkansson, 2015](#); [de Ruiter et al., 2020](#)). In fact, comprehension was overall higher for math than for chemistry, but the possibility that the higher number of subordinate clauses contributed to the absence of a narrative text advantage cannot be completely excluded.

Additionally, despite our best design efforts, the chemistry narrative text was overall better evaluated than its expository counterpart. A better equivalence was achieved for math, as the narrative text was only rated as more interesting. This introduces a potential confound, as it may imply that we are witnessing a coherence advantage, not necessarily tied to text type. However, the mediation analyses are informative in this regard: both difficulty and interest only partially mediated the text type effect on chemistry comprehension, with a significant direct effect remaining. This suggests that narrative structure contributed to the comprehension advantage beyond what can be accounted for by these evaluative differences.

Critically, participants had substantially less prior knowledge of the chemistry content and rated it, accordingly, as more difficult and as less interesting. Participants held more prior knowledge on the math texts' topics. A previous study also observed differences between text types only with the most difficult content ([Cervetti et al., 2009](#)) and a few studies have observed that narrative texts are particularly beneficial for participants with lower prior knowledge (e.g., [Wolfe & Woodwyk, 2010](#)).

This content-specific pattern actually offers a theoretically coherent account that weakens the content-specific explanation presented above: the narrative advantage emerged precisely in the domain where prior knowledge was near floor and was absent where prior knowledge was substantially higher. This pattern is consistent with the construction-integration model's predictions, regardless of topic. This result suggests that narrative structures are most helpful when readers have limited domain knowledge: when familiar content provides sufficient scaffolding for integration, the structural benefits of narrative may be less necessary.

Converging with this interpretation, the comprehension-knowledge slope was negligible in chemistry ($b = 0.003$) but significantly positive in math ($b = 0.117$). This suggests that in the more familiar domain, individual differences in prior knowledge were more predictive of performance – consistent with readers drawing on existing knowledge to integrate text content, independently of how that content was structured. These patterns are consistent with the construction-integration model's prediction that prior knowledge facilitates integration when text content is familiar ([Kintsch, 1998](#)), and with the RAND framework's view that reader, text, and activity characteristics interact dynamically to shape comprehension outcomes ([Snow, 2002](#)). This convergence makes a purely content-based explanation less parsimonious, though it cannot be fully ruled out (as testing if knowledge moderates the text-type effect is beyond the scope of this paper).

The effects of the level of comprehension and of the retention interval provide further important information on activity features. Although we predicted narrative texts would be more beneficial at higher representational levels, the narrative advantage in chemistry emerged uniformly across all levels. These results stand in contrast with some reports from the literature ([Wannagat et al., 2022](#); [Wolfe & Woodwyk, 2010](#)) but are consistent with [Hong and Lin-Siegler's](#)

(2012) finding of broad narrative benefits. Results further suggest that when some prior knowledge is present, the text types may not significantly differ at any item level, as in the math content. In contrast, when prior knowledge is sufficiently low, narrative scaffolding may support comprehension processes at every level – from proposition building to situation model transfer. While it is possible that the comprehension items were insufficiently sensitive to detect subtler level-specific differences, the items did reliably differentiate between comprehension levels within each topic. This result suggests that the items were sensitive to representational differences, even if text-type by level interactions did not emerge.

The variability across item levels offers further insight into the cognitive processes underlying comprehension, independently of text type. Increasing item elaboration does not necessarily correspond to increasing difficulty, as L1 scores (i.e., recall item) were higher with the math content but not with the chemistry content, in which L2 scores (i.e., inference items) were higher. It is possible that chemistry L1 items were more challenging because they required retrieving more technical terms (e.g., atomic weight; polarized light; solubility) than the math L1 items (e.g., empty space; 1 and 10; 59).

As for the effect of retention interval, regardless of text type, comprehension was significantly better in the first session. The decrease in comprehension scores was more evident with the chemistry topic than with the math topic, which may be related to the fact that the former was perceived as more difficult and yielded overall lower scores. Contrary to our hypothesis – and to previous arguments that narrative texts promote better retention (e.g., Mandler, 1984; Mar et al., 2021) – the interaction between text type and session was not significant, suggesting that comprehension of both text types decayed comparably across the week.

Previous studies reported a wide range of findings, including losses in the information retained from narrative texts (e.g., Prins et al., 2017; Wolfe & Woodwyk, 2010). A single brief reading, as used here, may not provide sufficient initial encoding depth for text type to differentially influence memory consolidation over one week (Squire et al., 2015). Studies reporting longer-term narrative advantages typically involved multiple exposures or extended texts (e.g., Hong & Lin-Siegler, 2012). Moreover, we cannot exclude the possibility of memory interference beyond mere decay. The counterbalanced presentation order of texts and the use of matched but distinct item sets across sessions were intended to minimize interference effects. Yet, both interference between information from the two texts read in the first session and interference between sessions (due to reading experiences during the intervening week) may have occurred.

4.1. Limitations

We note that our power analysis provides only an indicative estimate of statistical sensitivity. This is because it was based on a multiple regression approximation rather than a simulation-based approach tailored to mixed-effects models. This means that power estimates for interaction effects may have been overoptimistic, and some null interaction results should be interpreted with appropriate caution. Future research should employ simulation-based power analyses for more precise design optimization. Also, the students who chose to participate were from a relatively homogeneous sociodemographic and socioeconomic background. These aspects raise issues concerning the difference between statistical and educational significance that should be kept in mind when interpreting the present findings. Likewise, the experimental nature of our study, while providing control and internal validity, limits the ecological validity of the findings. More ecologically valid interventions can better investigate how narrative texts' benefits translate to applied contexts.

The measures of prior knowledge, text difficulty, and text interest used were all based on participants' self-reports, capturing perceived rather than objective properties of reader–text–activity interactions. While subjective ratings of this kind are ecologically valid – reflecting

the experience of reading as it unfolds – they may induce response biases. Future work might complement these self-report indices with objective assessments, such as pretested domain knowledge scores or independent ratings of text difficulty, to better isolate the mechanisms underlying the observed patterns. Even so, the comprehension differences observed were in line with the differences in perceived prior knowledge, and our focus was not conceptual change, but text comprehension processes.

The mediation results should be interpreted as evidence consistent with an experiential pathway, not as definitive proof of causality, because post-reading mediators can still be subject to unmeasured mediator–outcome confounding.

Finally, although several effects were small in magnitude, the whole pattern of results aligns with prior research showing that effect sizes in science comprehension and discourse processing are typically modest but theoretically meaningful, reflecting the multifactorial nature of these cognitive phenomena (e.g., Soares et al., 2023).

4.2. Implications for education and research

Narrative framing can offer a measurable comprehension advantage for unfamiliar science content – and this advantage is partly driven by subjective experience: when a text feels less difficult and more interesting, comprehension is facilitated.

The complexity of text comprehension observed here helps specify who, and under which conditions, can particularly benefit from narrative science materials. This understanding helps reveal specific narrative effects that may emerge under that specific configuration of characteristics. Importantly, it can provide clues for future research on science narratives, as well as on how educators can more meaningfully integrate these materials in science education in formal and non-formal education.

The present results indicate that people's proclivity to temporal structures, cause and effect, and human context (Bruner, 1986; Newkirk, 2014) may benefit science comprehension even among young adults (i.e., from 18 years onward), albeit less knowledgeable ones. The fact that narrative texts supported comprehension even in young adults with well-developed expository processing strategies suggests they can serve as a useful tool for lifelong learning. Like previous similar reports (e.g., Wolfe & Mienko, 2007), the present study examined specific narrative features, namely differences in how temporality is built and woven with characters and settings.

The comprehension differences produced by the two science topics are also a relevant hint, as authors had previously proposed that some contents may lend themselves more to be conveyed by narratives than others (Frisch, 2010) and that a narrative advantage might only be observed in the most challenging content (Cervetti et al., 2009). Future studies should continue to disentangle the comprehension processes associated with the comprehension of different text types and science topics and their specific contents.

Narrative texts may thus be particularly useful for tackling unknown content, serving as an entry point during its first stages of learning. Additionally, subjective reading experience and prior knowledge have a differential role: in low-familiarity contexts, perceived difficulty and interest are stronger correlates of comprehension success, whereas prior knowledge becomes a more meaningful predictor when content is more familiar. This indicates that activity characteristics do not influence comprehension uniformly – their effects are contingent on the familiarity of the content and the subjective experience of the reader. The post-reading ratings (text difficulty and interest in text) may themselves be mechanisms through which text type affects comprehension outcomes: reduced perceived difficulty could enable deeper processing, while increased interest may sustain attention and motivation. The content-specific pattern observed in the main analysis – where narrative texts showed comprehension advantages only in chemistry – may be partly mediated by these differential subjective experiences.

Both difficulty and interest emerged as partial mediators of the chemistry narrative advantage, consistent with the interpretation that subjective reading experience is both a consequence of successful comprehension and a potential mechanism facilitating it. Instructional designers targeting unfamiliar STEM content may gain more from narrative framing that increases engagement and reduces perceived difficulty.

Narrative texts may promote interest even in the absence of measurable comprehension gains – which in itself is a reason to consider their use. While it might not necessarily translate into increased comprehension in the short term, as seen with the math content, this engagement might plant seeds for comprehension gains in the longer term (a question also raised by (McQuiggan et al., 2008)). Longitudinal designs can test whether initial interest elicited by narrative framing translates into sustained engagement and future comprehension gains.

Narrative texts can be used to increase the variety of science educational materials, contributing to the development of more diversified language types (May et al., 2020; Newkirk, 2014) and to the tightening between literacy and science comprehension processes. Narrative texts are not proposed as a substitute for expository texts; rather, they can be a powerful complement as a scaffolding tool, while more dense expository texts continue to fulfill a role in more advanced learning stages.

In the specific context of this line of research, the RAND framework offers a useful lens for framing research questions and designing instructional tasks, given its explicit attention to reader, text, and activity as interacting sources of variability in comprehension outcomes. Similarly, the construction-integration model provides a theoretically motivated basis for developing and evaluating comprehension items in science comprehension research, as its distinction between text-based and situation model representations maps directly onto the kinds of comprehension differences that narrative and expository text types may differentially support.

Rather than asking whether narrative science texts support comprehension, future research can now ask more precisely which do, when, for whom, and through which mechanisms – questions that the present findings begin to address by situating narrative effects within the dynamic interaction between text, reader, and activity features.

5. Conclusion

Science education requires understanding when, why, and how different texts support comprehension. Narrative texts can benefit comprehension of unfamiliar science content across different comprehension levels, and this advantage operates partly through subjective reading experience – reduced perceived difficulty and heightened interest – a mechanism that appears most consequential when domain

knowledge is low. Together, these results advance our understanding of science comprehension by showing that text-type advantages are modulated by reader knowledge and mediated by experiential responses to the activity. They also offer educators a principled basis for integrating narrative texts as scaffolding tools when introducing unfamiliar content in science curricula.

Author note

Declaration of interest: none. This research was funded by Fundação para a Ciência e a Tecnologia (PD/BD/135345/2017; COVID/BD/152481/2022), Lisboa, Portugal. This study was not preregistered. At the time of data collection and first draft redaction, R. Kolinsky was Research Director of the Fonds de la Recherche Scientifique-FNRS (FRS-FNRS), Belgium. She is now Emeritus of the FRS-FNRS and Professor at ULB. We are immensely grateful to Ana Moreira, João Lopes and Ricardo Francisco for their invaluable help and involvement in material development, data collection and pretest data analysis. We also thank Tiago Hirth, David Felismino and Jorge Baptista for their help in the development of the materials and José Morais for his crucial input during the conceptualization of the study.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used Perplexity to assist with structuring and drafting parts of the manuscript. Outputs were reviewed and verified by the authors during use. After using this tool, all authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

CRediT authorship contribution statement

Sara Soares: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Visualization, Writing – original draft. **João O. Santos:** Data curation, Formal analysis, Writing – review & editing. **Cláudia Simão:** Conceptualization, Data curation, Formal analysis, Methodology, Writing – review & editing. **Matilde Gonçalves:** Conceptualization, Methodology, Resources, Writing – review & editing. **Raquel Barata:** Conceptualization, Methodology, Resources, Writing – review & editing. **Rita Jerónimo:** Conceptualization, Formal analysis, Funding acquisition, Methodology, Project administration, Supervision, Writing – review & editing. **Régine Kolinsky:** Conceptualization, Formal analysis, Funding acquisition, Methodology, Project administration, Supervision, Writing – review & editing.

Appendix F. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.learninstruc.2026.102436>.

Appendix A. Text Evaluation Scales Values Averaged as a Function of Text Type

After reading each assigned text, we also asked participants to rate it in terms of novelty of the information, cohesion, and local coherence; this allowed us to check for further significant differences on these features between texts (see Table A1). The fact that narrative texts were rated as more novel than expository texts might be because these texts presented scientific information but also personal/historical information.

Table A1
Averages (and Standard Deviations) of the Remaining Evaluations Made by Participants of each Text Type Read

Parameter	Question	NT		ET		<i>t</i>
		M	SD	M	SD	
Novelty of information	On a scale of 1 (not novel at all) to 5 (totally novel) please rate how novel the contents were for you	4.07	0.99	3.92	1.02	−0.05 **
Cohesion (sentences)	On a scale of 1 (totally disperse) to 5 (totally related) please rate how well connected the sentences were among themselves	3.98	0.93	3.62	1.04	3.11 **
Local Coherence (ideas)	On a scale of 1 (totally disperse) to 5 (totally related) please rate how well connected the ideas were among themselves	4.05	0.87	3.76	1.01	3.04 **

Note. The *df* were 123. ET = expository text, NT = narrative text.

Appendix B. Flowchart of the Procedure

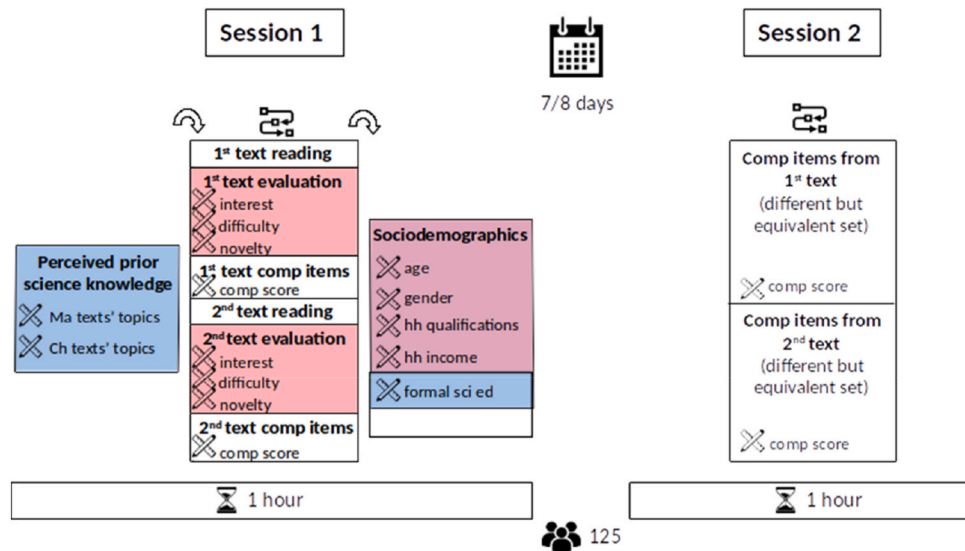


Fig. B1. Flowchart of the Procedure Representing the Timing of the Sessions, the Number of Participants, and the Variables and Timing of the Applied Tasks
Note. Ma: math, Ch: chemistry, comp: comprehension, hh: household, sci: science, ed: education. As indicated by the icons on top of each task block, all tasks were presented in sequential order.

Appendix C. Analysis 1

Participant Ratings of Text Characteristics

Participants' ratings of text characteristics were analyzed using a mixed-effects model with measure (difficulty, interest, knowledge, novelty), text type (expository text, narrative text), and topic (chemistry, math) as fixed effects, and random intercepts and slopes for topic within participant. Ratings of difficulty, interest, and novelty were collected after reading; perceived prior knowledge was collected before. Consequently, post-reading measures may reflect not only inherent text properties but also participants' subjective experiences shaped by the text type manipulation.

Omnibus Effects

Type III tests (Kenward-Roger) revealed significant evidence of a measure $\times$ text type $\times$ topic interaction, $F(3, 3738) = 6.55, p < .001, \eta_p^2 = 0.005$. The main effect of measure was the only large effect in the model, $F(3, 3738) = 1444.90, p < .001, \eta_p^2 = 0.537$. There was no significant evidence of main effects of text type, $F(1, 123) = 1.03, p = .312, \eta_p^2 = 0.008$; or topic, $F(1, 123) = 2.25, p = .136, \eta_p^2 = 0.018$. There was significant evidence of a two-way interactions of measure $\times$ topic, $F(3, 3738) = 173.49, p < .001, \eta_p^2 = 0.122$, and measure $\times$ text type, $F(3, 3738) = 57.42, p < .001, \eta_p^2 = 0.044$. There was no significant evidence of a text type $\times$ topic interaction, $F(1, 123) = 0.52, p = .474, \eta_p^2 = 0.004$.

Decomposing the Three-Way Interaction

Primary Decomposition: Text Type Effects Within Each Measure and Topic. Table C1 presents text type comparisons for each measure within each topic — the theoretically focal decomposition addressing whether expository texts and narrative texts differ on each characteristic in each topic.

Difficulty. There is significant evidence that expository texts elicited higher difficulty ratings than narrative texts in both topics (see Table C1). In chemistry, where perceived prior knowledge was low, $M = 1.10$, 95% CI [1.00, 1.20], expository ratings were near the scale midpoint while narrative ratings fell below it. In math, where prior knowledge was higher, $M = 1.82$, 95% CI [1.73, 1.92], both text types were rated as less difficult than in chemistry, though the text type difference persisted.

Interest. There is significant evidence that narrative texts elicited higher interest ratings than expository texts in both topics (see Table C1). Narrative text ratings were above the scale midpoint in both topics; Expository text ratings were near or below it.

Prior Knowledge. There is no significant evidence that text type affected perceived prior knowledge ratings in either topic (see Table C1). As these ratings were collected before reading, this indicates that participants assigned to expository texts and narrative texts for each topic were comparable in prior knowledge at baseline.

Novelty. In chemistry, there is no significant evidence that text type affected novelty ratings, with both versions rated near the scale maximum (see Table C1). In math, there is significant evidence that narrative texts elicited higher novelty ratings than expository texts.

Secondary Decomposition: Topic Effects Within Each Text Type and Measure. Table C2 presents topic comparisons for each measure within each text type. There is significant evidence that chemistry was rated as more difficult, more novel, and lower in prior knowledge than math for both text types. Conversely, there is significant evidence that math was rated as more interesting than chemistry for both text types (see Table C2).

Tertiary Decomposition: Measure Comparisons Within Each Text Type and Topic. Table C3 presents ratings across the four measures for each text type $\times$ topic combination. In the chemistry expository text, difficulty and interest ratings converged near the scale midpoint, each about 0.35 from the midpoint on opposite sides. In the chemistry narrative texts, there is significant evidence that interest exceeded difficulty (see Table C3). This reversal illustrates the three-way interaction: the chemistry expository text was experienced as both difficult and uninteresting, whereas the chemistry narrative text was experienced as easy and interesting. In both math conditions, difficulty was low and interest was relatively high. Knowledge ratings were consistently lowest across all combinations, and novelty ratings were consistently highest, except in the math narrative text where interest and novelty did not significantly differ (see Table C3).

Table C1
Text Type Comparisons for Each Measure Within Topic Domains

Measure	Chemistry			Math		
	ET	NT	<i>t</i>	ET	NT	<i>t</i>
Difficulty	3.35 [3.21, 3.50]	2.73 [2.59, 2.87]	6.16***	2.76 [2.63, 2.90]	2.48 [2.35, 2.62]	2.81**
Interest	2.65 [2.50, 2.79]	3.27 [3.13, 3.41]	-6.16***	3.03 [2.90, 3.17]	3.52 [3.38, 3.65]	-4.89***
Knowledge	1.08 [0.94, 1.23]	1.12 [0.98, 1.26]	-0.38	1.90 [1.76, 2.03]	1.75 [1.61, 1.89]	1.48
Novelty	4.45 [4.31, 4.59]	4.44 [4.30, 4.58]	0.07	3.40 [3.26, 3.53]	3.69 [3.56, 3.83]	-2.99**

Note. ET: expository text; NT: narrative text. Values are estimated marginal means [95% CI]. *df* = 391 for chemistry, 421 for math. ****p* < .001, ***p* < .01.

Table C2
Topic Domain Comparisons for Each Measure Within Text Types

Measure	Expository Texts			Narrative Texts		
	Chemistry	Math	<i>t</i>	Chemistry	Math	<i>t</i>
Difficulty	3.35 [3.21, 3.50]	2.76 [2.63, 2.90]	5.91***	2.73 [2.59, 2.87]	2.48 [2.35, 2.62]	2.46*
Interest	2.65 [2.50, 2.79]	3.03 [2.90, 3.17]	-3.85***	3.27 [3.13, 3.41]	3.52 [3.38, 3.65]	-2.46*
Knowledge	1.08 [0.94, 1.23]	1.90 [1.76, 2.03]	-8.10***	1.12 [0.98, 1.26]	1.75 [1.61, 1.89]	-6.25***
Novelty	4.45 [4.31, 4.59]	3.40 [3.26, 3.53]	10.52***	4.44 [4.30, 4.58]	3.69 [3.56, 3.83]	7.49***

Note. Values are estimated marginal means [95% CI]. *df* = 769 for ET, 770 for NT. ****p* < .001, **p* < .05.

Table C3
Measure Ratings Within Each Text Type $\times$ Topic Combination

Combination	Difficulty	Interest	Knowledge	Novelty
Chemistry ET	3.35 [3.21, 3.50]	2.65 [2.50, 2.79]	1.08 [0.94, 1.23]	4.45 [4.31, 4.59]
Chemistry NT	2.73 [2.59, 2.87]	3.27 [3.13, 3.41]	1.12 [0.98, 1.26]	4.44 [4.30, 4.58]
Math ET	2.76 [2.63, 2.90]	3.03 [2.90, 3.17]	1.90 [1.76, 2.03]	3.40 [3.26, 3.53]
Math NT	2.48 [2.35, 2.62]	3.52 [3.38, 3.65]	1.75 [1.61, 1.89]	3.69 [3.56, 3.83]

Note. ET: expository text; NT: Narrative text. Values are estimated marginal means [95% CI]. All pairwise comparisons among measures within each combination are significant (*p* < .001; Tukey-adjusted) except for the math narrative text comparison interest vs novelty (*t* = -2.26, *p* = .11).

Appendix D. Analysis 2

Relationships Between Comprehension Performance and Text Characteristic Ratings

To examine whether the relationships between comprehension performance and text characteristic ratings varied by measure and topic, we fitted a mixed-effects model with measure, topic, and comprehension (standardized comprehension performance) as fixed effects, including all interactions, and random intercepts and slopes for topic within participant.

Omnibus Effects

Type III tests using the Kenward-Roger method revealed significant evidence of a three-way interaction of measure $\times$ topic $\times$ comprehension (scaled), $F(3, 3736.08) = 10.02$, *p* < .001, $\eta_p^2 = 0.008$, indicating that the relationship between comprehension performance and ratings varied across measures and topics. The largest effect was the main effect of measure, $F(3, 3736.08) = 1474.59$, *p* < .001, $\eta_p^2 = 0.54$. There was significant evidence of two-way interactions: measure $\times$ comprehension (scaled), $F(3, 3736.08) = 74.08$, *p* < .001, $\eta_p^2 = 0.06$; and measure $\times$ topic, $F(3, 3736.08) = 171.96$, *p* < .001, $\eta_p^2 = 0.12$. There was no significant evidence of main effects of topic, $F(1, 123.67) = 2.35$, *p* = .128, $\eta_p^2 = 0.02$, comprehension (scaled), $F(1, 2973.65) = 1.07$, *p* = .301, $\eta_p^2 < 0.001$; or topic $\times$ comprehension (scaled) interaction, $F < 1$.

Decomposing the Three-Way Interaction: Slopes by Measure and Topic

To interpret the three-way interaction, we examined the slopes (relationship between comp_scaled and ratings) for each measure $\times$ topic combination using estimated marginal trends. Table D1 presents the complete pattern of slopes.

Note. Slopes represent the change in rating (1-5 scale) per 1-SD increase in comprehension performance. Negative slopes indicate that higher comprehension is associated with lower ratings; positive slopes indicate higher comprehension is associated with higher ratings.

Difficulty Ratings

There is significant evidence that higher comprehension performance was associated with lower difficulty ratings in chemistry, $b = -0.29$, 95% CI $[-0.36, -0.21]$, and no significant evidence of such a relationship in math, $b = -0.08$, 95% CI $[-0.17, 0.01]$. There is significant evidence that these slopes differed between topics, $t(3860) = -3.52$, $p < .001$.

Interest Ratings

There is significant evidence that higher comprehension performance was associated with higher interest ratings in both chemistry, $b = 0.43$, 95% CI $[0.35, 0.50]$, and math, $b = 0.25$, 95% CI $[0.16, 0.33]$. There is significant evidence that this relationship was stronger in chemistry than math, $t(3860) = 3.10$, $p = .002$.

Prior Knowledge Ratings

In chemistry, there is no significant evidence that comprehension performance related to prior knowledge ratings, $b = 0.00$, 95% CI $[-0.07, 0.08]$. In math, there is significant evidence of a positive relationship, $b = 0.12$, 95% CI $[0.03, 0.20]$. Evidence for a difference between slopes across topic was weak, $t(3860) = -1.96$, $p = .050$.

Novelty Ratings

There is significant evidence that higher comprehension performance was associated with lower novelty ratings in both chemistry, $b = -0.11$, 95% CI $[-0.19, -0.03]$, and math, $b = -0.18$, 95% CI $[-0.26, -0.09]$. There is no significant evidence that these slopes differed between topics, $t(3860) = 1.17$, $p = .244$.

Illustrative Predicted Means at Representative Comprehension Levels

Table D2 presents predicted ratings at low (-1 SD), mean (0), and high ($+1$ SD) comprehension performance to illustrate the pattern of slopes.

Table D1

Relationship Between Comprehension Performance and Ratings by Measure and Topic

Measure	Chemistry Topic			Math Topic		
	Slope [95% CI]	SE	<i>t</i>	Slope [95% CI]	<i>t</i>	SE
Difficulty	-0.28 [-0.36, -0.21]	0.039	-7.29 ***	-0.08 [-0.17, 0.01]	-1.81	0.044
Interest	0.43 [0.35, 0.50]	0.039	10.932 ***	0.25 [0.16, 0.33]	5.63 ***	0.044
Knowledge	0.00 [-0.07, 0.08]	0.039	0.07	0.12 [0.03, 0.20]	2.68 **	0.044
Novelty	-0.11 [-0.19, -0.03]	0.039	-2.80 **	-0.18 [-0.26, -0.09]	-4.07 ***	0.044

Note. Slopes represent the change in rating (1-5 scale) per 1-SD increase in comprehension performance. Negative slopes indicate that higher comprehension is associated with lower ratings; positive slopes indicate higher comprehension is associated with higher ratings. *** $p < .001$, ** $p < .01$.

Table D2

Predicted Ratings by Comprehension Level, Measure, and Topic

Measure	Chemistry			Math		
	Low (-1 SD)	Mean	High ($+1$ SD)	Low (-1 SD)	Mean	High ($+1$ SD)
Difficulty	3.32 [3.19, 3.44]	3.03 [2.93, 3.13]	2.75 [2.62, 2.87]	2.71 [2.58, 2.84]	2.63 [2.53, 2.72]	2.55 [2.42, 2.67]
Interest	2.54 [2.42, 2.67]	2.97 [2.87, 3.07]	3.40 [3.27, 3.53]	3.02 [2.89, 3.15]	3.27 [3.17, 3.36]	3.51 [3.38, 3.64]
Knowledge	1.10 [0.98, 1.23]	1.10 [1.01, 1.20]	1.11 [0.98, 1.23]	1.70 [1.57, 1.83]	1.82 [1.72, 1.92]	1.94 [1.81, 2.07]
Novelty	4.55 [4.43, 4.68]	4.44 [4.35, 4.54]	4.34 [4.21, 4.46]	3.73 [3.60, 3.86]	3.55 [3.45, 3.65]	3.37 [3.24, 3.50]

Note. Values are estimated marginal means [95% CI]. Low = -1 SD below mean comprehension, High = $+1$ SD above mean comprehension.

Appendix E. Analysis 3

Mediation analysis for the chemistry topic in the first session

Table E1

Domain-Specific Mediation Effects of Difficulty and Interest for Chemistry

Domain	Mediator	Effect	Estimate	95% CI	<i>p</i>
Chemistry	Difficulty	ACME	0.15	[0.04, 0.30]	0.003
		ADE	0.43	[0.07, 0.76]	0.019
		Total Effect	0.58	[0.23, 0.90]	<0.001

(continued on next page)

Table E1 (continued)

Domain	Mediator	Effect	Estimate	95% CI	p
Interest		Prop. Mediated	0.26	[0.07, 0.74]	0.003
		ACME	0.25	[0.08, 0.44]	0.005
		ADE	0.32	[0.02, 0.62]	0.040
		Total Effect	0.58	[0.24, 0.649]	0.001
		Prop. Mediated	0.44	[0.17, 0.94]	0.006

Note. ACME = average causal mediation effect (indirect effect); ADE = average direct effect. N = 125 observations per domain.

Table E2

Domain-Specific Mediation Effects of Difficulty and Interest for Math

Domain	Mediator	Effect	Estimate	95% CI	p
Math	Difficulty	ACME	0.02	[-0.03, 0.10]	0.502
		ADE	-0.15	[-0.51, 0.20]	0.405
		Total Effect	-0.13	[-0.49, 0.22]	0.466
		Prop. Mediated	-0.15	[-1.88, 1.76]	0.791
	Interest	ACME	0.15	[0.02, 0.31]	0.022
		ADE	-0.28	[-0.60, 0.04]	0.091
		Total Effect	-0.13	[-0.48, 0.22]	0.473
		Prop. Mediated ^a	-1.18	[-12.56, 11.09]	0.492

Note. ACME = average causal mediation effect (indirect effect); ADE = average direct effect. N = 125 observations per domain.

^a Prop. Mediated values are undefined when the total effect is non-significant.

References

- Adam, J. M. (2011). *The narrative sequence: History of a concept and a research area* [paper presentation]. 1er Colloque International organisé par le Réseau Romand de Narratologie (RRN). Switzerland: Fribourg.
- American Psychological Association. (2018). Retention measure. In *APA dictionary of psychology*. <https://dictionary.apa.org/retention-measure>.
- 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>
- Avraamidou, L., & Osborne, J. (2009). The role of narrative in communicating science. *International Journal of Science Education*, 31, 1683–1707. <https://doi.org/10.1080/00207179.2009.333826>
- Barth, A. E., Tolar, T. D., Fletcher, J. M., & Francis, D. (2014). The effects of student and text characteristics on the oral reading fluency of middle-grade students. *Journal of Educational Psychology*, 106(1), 162–180. <https://doi.org/10.1037/a0033826>
- 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(2), 517–547. <https://doi.org/10.1017/S0272263114000552>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B*, 57(1), 289–300. <http://www.jstor.org/stable/2346101>
- Bolkan, S., Goodboy, A. K., & Kromka, S. M. (2020). Student assessment of narrative: Telling stories in the classroom. *Communication Education*, 69(1), 48–69. <https://doi.org/10.1080/03634523.2019.1622751>
- 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. (1986). *Actual minds, possible worlds*. Harvard University Press.
- Bruner, J. (1996). *The culture of education*. Harvard University Press.
- 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). Ethnolinguistics and identity: black students representing science identities in the making. *Science Education*, 104(2), 193–221. <https://doi.org/10.1002/scs.21558>
- Costa, M. A. (2005). Processamento de frases em Português Europeu: Aspectos cognitivos e linguísticos implicados na compreensão da língua escrita. *Fundação Calouste Gulbenkian*.
- Dahlstrom, M. F. (2014). Using narratives and storytelling to communicate science with nonexpert audiences. *Proceedings of the National Academy of Sciences*, 111 (Supplement 4), 13614–13620. <https://doi.org/10.1073/pnas.1320645111>
- 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, Article 104130. <https://doi.org/10.1016/j.cognition.2019.104130>
- 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>
- 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>
- 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>
- 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>
- 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>
- 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>
- Goldman, S. R., Britt, M. A., Brown, W., Cribb, G., George, M., Greenleaf, C., Lee, C. D., & Shanahan, C. (2016). Disciplinary literacies and learning to read for understanding: A conceptual framework for disciplinary literacy. *Educational Psychologist*, 51(2), 219–246. <https://doi.org/10.1080/00461520.2016.1168741>
- Golke, S., Hagen, R., & Wittwer, J. (2019). Lost in narrative? The effect of informative narratives on text comprehension and metacomprehension accuracy. *Learning and Instruction*, 60, 1–19. <https://doi.org/10.1016/j.learninstruc.2018.11.003>
- Golke, S., & Wittwer, J. (2024). Informative narratives increase students' situational interest in science topics. *Learning and Instruction*, 93, Article 101973. <https://doi.org/10.1016/j.learninstruc.2024.101973>
- Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-metrix: Providing multilevel analyses of text characteristics. *Educational Researcher*, 40(5), 223–234. <https://doi.org/10.3102/0013189x11413260>
- 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>
- Hannon, B. (2022). Children's comprehension of narrative and expository texts: The contributions of word decoding and knowledge integration skills vary within and between text types. *Journal of Education and Training Studies*, 10(4), 1–15. <https://doi.org/10.11114/jets.v10i4.5465>
- Hattan, C., Alexander, P. A., & Lupo, S. M. (2023). Leveraging what students know to make sense of texts: What the research says about prior knowledge activation. *Review of Educational Research*, 94(1), 73–111. <https://doi.org/10.3102/00346543221148478>
- 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/a0026224>
- Kendeou, P., van den Broek, P., Helder, A., & Karlsson, J. (2014). A cognitive view of reading comprehension: Implications for reading difficulties. *Learning Disabilities Research & Practice*, 29(1), 10–16. <https://doi.org/10.1111/ldrp.12025>
- 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., & Mangalath, P. (2010). The construction of meaning. *Topics in Cognitive Science*, 3(2), 346–370. <https://doi.org/10.1111/j.1756-8765.2010.01107.x>
- Kloser, M. (2013). Exploring high school biology students' engagement with more and less epistemologically considerate texts. *Journal of Research in Science Teaching*, 50 (10), 1232–1257. <https://doi.org/10.1002/tea.21109>
- 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>

- Kulesz, P. A., Francis, D. J., Barnes, M. A., & Fletcher, J. M. (2016). The influence of properties of the test and their interactions with reader characteristics on reading comprehension: An explanatory item response study. *Journal of Educational Psychology*, 108(8), 1078–1097. <https://doi.org/10.1037/edu0000126>
- 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>
- Lord, C. (2002). Are subordinate clauses more difficult? In J. L. Bybee, & M. Noonan (Eds.), *Complex sentences in grammar and discourse* (pp. 223–234). John Benjamins.
- Maley, C. C., & Seyedi, S. (2022). The life history theory of the lord of the rings: A randomized controlled trial of using fact versus fiction to teach life history theory. *Evolution*, 15(1), 2. <https://doi.org/10.1186/s12052-022-00160-8>
- Mandler, J. M. (1984). *Stories, scripts, and scenes: Aspects of schema theory*. Erlbaum.
- Mar, R. A., Li, J., Nguyen, A. T., & Ta, C. P. (2021). Memory and comprehension of narrative versus expository texts: A meta-analysis. *Psychonomic Bulletin & Review*, 28, 732–749. <https://doi.org/10.3758/s13423-020-01853-1>
- 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. (2010). Computational methods to extract meaning from text and advance theories of human cognition. *Topics in Cognitive Science*, 3(1), 3–17. <https://doi.org/10.1111/j.1756-8765.2010.01117.x>
- McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*, 14(1), 1–43. https://doi.org/10.1207/s1532690xci1401_1
- McNamara, D. S., Louwerse, M. M., McCarthy, P. M., & Graesser, A. C. (2010). Coh-matrix: Capturing linguistic features of cohesion. *Discourse Processes*, 47(4), 292–330. <https://doi.org/10.1080/01638530902959943>
- McNamara, D. S., & Magliano, J. P. (2009). Differential competencies contributing to children's comprehension of narrative and expository texts. *Learning and Instruction*, 19(3), 228–242. <https://doi.org/10.1016/j.learninstruc.2008.06.005>
- McQuiggan, Scott W., Rowe, Jonathan P., Lee, Sunyoung, & Lester, James C. (2008). Story-Based Learning: The Impact of Narrative on Learning Experiences and Outcomes. *Lecture Notes in Computer Science* (pp. 530–539). Berlin, Heidelberg: Springer Berlin Heidelberg, ISBN 9783540691303. https://doi.org/10.1007/978-3-540-69132-7_56
- Mensink, M. C., Kendeou, P., & Rapp, D. N. (2021). Do different kinds of introductions influence comprehension and memory for scientific explanations? *Discourse Processes*, 58(5–6), 491–512. <https://doi.org/10.1080/0163853X.2021.1904754>
- Meyer, B. J. (1975). *The organization of prose and its effects on memory*. I. North-Holland Publishing Company.
- Morais, J., & Kolinsky, R. (2016). Literacia científica: leitura e produção de textos científicos. *Educar Em Revista*, (62), 143–162.
- Negrete, A., & Lartigue, C. (2010). The science of telling stories: Evaluating science communication via narratives (RIRC method). *Journal of Media and Communication Studies*, 2, 98–110. <https://doi.org/10.5897/JMCS.9000080>
- Neri, S., Bernholt, S., & Schreiber, C. (2023). Strategies for expository and narrative texts: Students' perspectives on cognitive and metacognitive reading processes. *L1: Educational Studies in Language and Literature*, 23(2), 109–135. <https://doi.org/10.17239/L1ESLL-2023.23.02>
- Newkirk, T. (2014). *Minds made for stories: How we really read and write informational and persuasive texts*. Heinemann.
- 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. <https://doi.org/10.1002/sce.20063>
- Peterson, S. E. (1991). The cognitive functions of underlining as a study technique. *Reading Research & Instruction*, 31(2), 49–56. <https://doi.org/10.1080/19388079209558078>
- Plavén-Sigray, P., Matheson, G. J., Schiffler, B. C., & Thompson, W. H. (2017). The readability of scientific texts is decreasing over time. *eLife*, 6, Article e27725.
- 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>
- 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>.
- Schank, R. C., & Berman, T. R. (2002). The pervasive role of stories in knowledge and action. In M. C. Green, J. J. Strange, & T. C. Brock (Eds.), *Narrative impact: Social and cognitive foundations* (pp. 287–313). Lawrence Erlbaum Associates Publishers.
- Singmann, H., & Kellen, D. (2019). An introduction to mixed models for experimental psychology. In *New methods in cognitive psychology* (pp. 4–31). Routledge.
- Snow, C. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. Rand Corporation.
- Soares, S., Gonçalves, M., Jerónimo, R., & Kolinsky, R. (2023). Narrating science: Can it benefit science learning, and how? A theoretical review. *Journal of Research in Science Teaching*, 60(9), 2042–2075. <https://doi.org/10.1002/tea.21851>
- Spencer, M., Gilmour, A. F., Miller, A. C., Emerson, A. M., Saha, N. M., & Cutting, L. E. (2019). Understanding the influence of text complexity and question type on reading outcomes. *Reading and Writing: An Interdisciplinary Journal*, 32(3), 603–637. <https://doi.org/10.1007/s11145-018-9883-0>
- Squire, L. R., Genzel, L., Wixted, J. T., & Morris, R. G. (2015). Memory consolidation. *Cold Spring Harbor Perspectives in Biology*, 7(8), Article a021766. <https://doi.org/10.1101/cshperspect.a021766>
- Stuckey, M., Hofstein, A., Mamlok-Naaman, R., & Eilks, I. (2013). The meaning of 'relevance' in science education and its implications for the science curriculum. *Studies in Science Education*, 49(1), 1–34. <https://doi.org/10.1080/03057267.2013.802463>
- Thiede, K. W., Anderson, M., & Theriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology*, 95(1), 66.
- Tingley, Dustin, Yamamoto, Teppei, Hirose, Kentaro, Keele, Luke, & Imai, Kosuke (2014). mediation: RPackage for Causal Mediation Analysis. *Journal of Statistical Software*, 59(5). <https://doi.org/10.18637/jss.v059.i05>. <http://www.jstatsoft.org/v59/i05/>
- Tobler, S., Sinha, T., Köhler, K., Hafen, E., & Kapur, M. (2022). The impact of prior knowledge in narrative-based learning on understanding biological concepts in higher education. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 44(44).
- Tobler, S., Sinha, T., Köhler, K., Hafen, E., & Kapur, M. (2024). Impact of narrative versus expository instruction in science education on recall, understanding, and transfer: A meta-analysis. *EdArXiv*. <https://doi.org/10.3929/ethz-b-000696792>
- Tobler, S., Sinha, T., Köhler, K., & Kapur, M. (2025). Teaching biology with narratives: Examining the impact of affective and cognitive variables on undergraduate student learning. *Instructional Science*, 53(4), 921–948. <https://doi.org/10.1007/s11251-025-09723-w>
- Wang, M. D. (1970). The role of syntactic complexity as a determiner of comprehensibility. *Journal of Verbal Learning and Verbal Behavior*, 9(4), 398–404. [https://doi.org/10.1016/S0022-5371\(70\)80079-2](https://doi.org/10.1016/S0022-5371(70)80079-2)
- Wannagat, W., Steinicke, V., Tibken, C., & Niding, G. (2022). Same topic, different genre: Elementary school children's mental representations of information embedded in narrative and expository texts. *Learning and Instruction*, 80, Article 101559. <https://doi.org/10.1016/j.learninstruc.2021.101559>
- Webb, P. (2010). Science education and literacy: Imperatives for the developed and developing world. *Science*, 328(5977), 448–450. <https://doi.org/10.1126/science.1182596>. PMID: 20413487.
- 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>
- Zabel, L., & Averdunk, M. (2024). Narrative and expository texts in educational contexts: Cognitive and linguistic distinctions. *Text & Discourse Studies*, 15(1), 112–130. <https://doi.org/10.1234/tads.2024.01501>
- Zhang, Y., Liu, X., & Wang, H. (2023). Cognitive and motivational predictors of reading comprehension across narrative and expository text genres: A longitudinal study. *Journal of Educational Psychology*, 115(1), 47–67. <https://doi.org/10.1037/edu0000756>
- Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, 123(2), 162. <https://doi.org/10.1037/0033-2909.123.2.162>

Update

Learning and Instruction

Volume 106, Issue , December 2026, Page

DOI: <https://doi.org/10.1016/j.learninstruc.2026.102453>



Corrigendum to “Tell me how: Teaching science with narration vs exposition” [Learning and Instruction 106 (2026) 102436]

Sara Soares^{a,b,*}, João O. Santos^c, Cláudia Simão^d, Matilde Gonçalves^e, Raquel Barata^{f,g}, Rita Jerónimo^a, Régine Kolinsky^{b,h}

^a Iscte-Instituto Universitário de Lisboa, CIS-Iscte, Lisboa, Portugal

^b Unité de Recherche en Neurosciences Cognitives (Unescog), Center for Research in Cognition and Neurosciences (CRCN), Université Libre de Bruxelles (ULB), Belgium

^c ISPA - Instituto Universitário de Ciências Psicológicas, Sociais e da Vida, Portugal

^d Católica-Lisbon School of Business and Economics, Universidade Católica Portuguesa, Lisboa, Portugal

^e Centro de Linguística da Universidade NOVA de Lisboa, Faculdade de Ciências Sociais e Humanas, Universidade NOVA de Lisboa, Lisboa, Portugal

^f Jardim Botânico, Museu Nacional de História Natural e da Ciência, Universidade de Lisboa, Lisboa, Portugal

^g Centre for Ecology, Evolution and Environmental Changes, Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal

^h Fonds de la Recherche Scientifique-FNRS, Belgium

The authors regret that the following funding acknowledgement was inadvertently omitted from the published article, as it was not communicated in time for inclusion in the final version of the manuscript:

"The research of Matilde Gonçalves is supported by Portuguese national funding through the FCT - Portuguese Foundation for Science and

Technology, I.P., as part of the project UID/03213/2025 - Research Center for Linguistics at NOVA University Lisbon (CLUNL) (<https://doi.org/10.54499/UID/03213/2025>)."

This correction concerns the funding acknowledgement only and does not affect the scientific content or conclusions of the article.

The authors would like to apologise for any inconvenience caused.

DOI of original article: <https://doi.org/10.1016/j.learninstruc.2026.102436>.

* Corresponding author. Iscte-Instituto Universitário de Lisboa, CIS-Iscte, Lisboa, Portugal.

E-mail address: sapss@iscte-iul.pt (S. Soares).

<https://doi.org/10.1016/j.learninstruc.2026.102453>

Available online 13 August 2026

0959-4752/© 2026 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.