The Impact of Prior Knowledge in Narrative-Based Learning on Understanding Biological Concepts in Higher Education

Samuel Tobler (samuel.tobler@gess.ethz.ch)
Professorship for Learning Sciences and Higher Education, ETH Zurich
Zurich, Switzerland

Tanmay Sinha (tanmay.sinha@gess.ethz.ch)
Professorship for Learning Sciences and Higher Education, ETH Zurich
Zurich, Switzerland

Katja Köhler (koehler@imsb.biol.ethz.ch)
Department of Biology, ETH Zurich
Zurich, Switzerland

Ernst Hafen (hafen@imsb.biol.ethz.ch)
Department of Biology, ETH Zurich
Zurich, Switzerland

Manu Kapur (manukapur@ethz.ch)
Professorship for Learning Sciences and Higher Education, ETH Zurich
Zurich, Switzerland

Abstract

Fundamental concepts in biology are often challenging to understand. More strikingly, studies also report incorrect or incomplete understanding of such concepts for undergraduate natural science students even after instruction. Recent research suggests that embedding conceptual information in a narrative could support students' learning process and facilitate conceptual change. Therefore, we designed learning materials covering complex concepts in biology either in the form of a narrative presenting the to-be-learned concepts in a historical context or as an expository text as control. We then assessed conceptual understanding and potential learning mechanisms. Results indicate that students learned from narrative texts and expository texts to a similar extent. However, if the prior knowledge was higher, the effect on learning was bigger in the narrative group than in the expository group. Moreover, the narrative led to better enjoyment and a higher germane cognitive load than the expository text material.

Keywords: Narrative; Conceptual Understanding; Biology; Higher Education; University; Prior Knowledge

Introduction

Concepts in biology are often taught focusing on facts and absolute knowledge, while the actual complexity of the underlying concepts is neglected (Brumby, 1984; Fiedler, Tröbst, & Harms, 2017). Consistently, empirical studies conducted at universities report that undergraduate students in natural sciences frequently do not fully understand fundamental concepts in biology and, more strikingly, that these students also often even fail to grasp these concepts after instruction (Champagne Queloz, Klymkowsky, Stern, Hafen, & Köhler, 2016; Fiedler et al., 2017).

Studies showed that the two major concepts of random processes in biological systems and enzyme energetics repeatedly challenge students' understanding (Champagne Queloz et al., 2016; Fiedler et al., 2017). One potential reason for these difficulties includes students' teleological mindset (Coley & Tanner, 2015). A teleological mindset describes the approach of explaining natural phenomena by ascribing a causal reason to a random process. For example, bacteria do not evolve because they are treated with antibiotics, but those resistant to this antibiotic prior to the treatment survive.

In the light of a conceptual change framework (Vosniadou, Vamvakoussi, & Skopeliti, 2008), students arrive in lectures with their naïve ideas of how nature works (i.e., a teleological mindset) and strive to integrate newly presented knowledge fragments into their pre-existing knowledge network. For example, when teaching the random movement of molecules, students will try to combine the new information with their already existing conception. Thereby, there is no guarantee that students will reach the conceptual understanding that is consistent with the scientifically accepted theory. Therefore, it is crucial to investigate how biological knowledge can be conveyed to students to overcome their teleological self-explanations and ultimately aid them in fully grasping the concepts.

Imparting theoretical information through a narrative has been proposed as one potential solution to make students aware of their incorrect beliefs and lead them to an improved understanding of the taught concepts (Dahlstrom, 2014). Furthermore, neuroscientific studies that compare brain activation upon narrative or non-narrative description report a greater activation of brain regions, including the temporoparietal junction, the posterior superior temporal

In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society.*

sulcus, or the posterior cingulate cortex, that are associated with memory functions (Yuan, Major-Girardin, & Brown, 2018). These results indicate greater activation of prior knowledge and thus provide a more extensive knowledge framework to integrate new information.

Quantitative studies in educational settings provide support for these conjectures and findings. Several studies report increased conceptual understanding upon reading a narrative, compared to a non-narrative instruction (see Dai, Williams, Witucki, & Rudge (2021) or Emmons, Smith, and Kelemen (2016) for recent examples). Additionally, these and other studies reported several potential learning mechanisms (yet prevalently only qualitatively assessed), contributing to the positive effects of narratives on understanding. These mechanisms comprise effects on the cognitive load (Cooper, Corley, & Underwood, 2013; Fisch, 2000) and motivation and interest (Arya & Maul, 2012).

Nonetheless, the impact on the understanding after having read the concepts in a narrative text, compared to an expository text (i.e., a standard textbook version about this concept), is not as well investigated at the university level compared to a lower level of education. In fact, Wolfe and Mienko (2007) even report a lower performance for students who read a narrative text compared to an expository one with the same contents. Additionally, they showed that the amount of prior knowledge affected the effectiveness of the intervention. Similarly, other scholars report a positive influence of less prior knowledge as, for example, assessed in tests prior to the intervention on the impact of the narrative on understanding (see Emmons et al. (2016), Hopkins and Weisberg (2021), or Reuer (2012) for examples).

In this study, we investigate the effects on undergraduate natural science students' conceptual understanding when embedding the concepts in a historical narrative compared to presenting the same concepts in an expository text. We targeted concepts based on previously reported difficulty understanding them (Champagne Queloz et al., 2016) and then investigated the effects on understanding in an immediate and a delayed post-test. Furthermore, we assessed the putative learning mechanisms. Accordingly, we posed the following research questions:

1. What is the impact of narratives on the understanding of biological concepts compared to expository texts in higher education?
2. What is the impact of prior knowledge on the effects of learning with narrative texts?
3. How do narratives affect the retention of the concepts over time compared to expository texts?
4. Which learning mechanisms are differently affected, comparing narrative with expository text materials?

Thereby, we hypothesize that learning materials designed as a narrative will lead to a greater conceptual understanding than expository text materials (H1). Taking the educational major in high school as an indicator of prior knowledge, we hypothesize that prior knowledge impacts the effect of the intervention in the narrative condition (H2a). Furthermore, we conjecture that relative to expository texts, students with less former biology education and consequently lower prior knowledge will profit more from the narrative text materials (H2b). Also, students who read narratives would better retain the learned concepts over time (H3). Finally, we hypothesize that students who have worked with the narrative materials will report higher enjoyment (H4a), motivation, and interest towards the topic (H4b) and a higher germane cognitive load (H4c). In contrast, we expect the extraneous cognitive load between the conditions to remain similar as the narrative should not distract students from learning the contents (H4d).

Methods

Participants

The participants of this study consisted of 74 first-year undergraduate medicine (13.5%) and health science and technology (86.5%) students from a highly ranked Swiss university. The participants were 19.7 ± 1.6 years old (mean ± SD), whereby 71.6% of the students were female. 62.2% of the participants majored in STEM-related subjects in high school. Of this initial set of participants, 27 students also participated in the delayed part of the study (81.5% female). The ethics committee of the university approved the study prior to its conductance. Participants were recruited from an introductory biology course by the lecturer. Participation in the study was voluntary and there was no compensation.

A priori power analysis for a balanced ANOVA suggested a sample size of at least 33 participants per condition necessary for detecting an effect size of Cohen's d = 0.5 (estimation based on Tobler, Sinha, Köhler, Hafen, & Kapur, 2022) with a power level of 80% and a significance level of .05 using the *pwr* package (v. 1.3-0; Champely, 2020) in the R software environment (v. 1.4.1717; R Core Team, 2021).

Text materials used in the intervention

The selected concepts of enzyme energetics and randomness (Champagne Queloz et al., 2016) were historically framed in a narrative text or presented in an expository text, similar to a biology textbook.

The narrative text materials were designed following the embedded narrative theory (Norris, Guilbert, Smith, Hakimelahi, & Phillips, 2005). Thereby, the to-be-learned concepts were integrated into the historical context of their discovery. An exemplary text passage reads as follows:

> *Narrative:* While looking through a microscope, Robert observed tiny particles floating in the water and moving irregularly. Brown believed he found the presumed life force in plants. Today, we know that what he observed is equivalent to the random heat motion of molecules.
> *Expository:* This is made possible by the so-called Brownian motion. The irregular motion of molecules in water is caused by the heat energy and results in the random movement of molecules in the solution.

The text materials were assessed for content correctness and comparability with biology and literature sciences experts.

Table 1: Descriptive statistics of the subscales of the questionnaire. n: number of items per subscale.

Subscale	n	Exemplary item	Cronbach's α
Germane cognitive load	3	Reading the text material was challenging.	α = 0.91
Extraneous cognitive load	5	I found it easy to distinguish important from unimportant information.	α = 0.73
Enjoyment	7	I would like to read more scientific texts written like this.	α = 0.88
Interest and motivation	6	My interest in the topic has increased because of the text in this study.	α = 0.91

Text analysis indicated overall similar text characteristics comparing lexical complexity (i.e., type-token ratio (TTR)) following Richards (1987) (narrative: TTR = 0.39 ± 0.12; expository: TTR = 0.39 ± 0.16) and readability following Flesch (1948) (narrative: 37.02, predicate difficult; expository: 41.56, predicate difficult).

The narrative text was longer than the expository one (narrative: 108 sentences; expository: 73 sentences). In conclusion, different potential effects on learning with either of the two texts are likely to result from content-intrinsic characteristics of the instruction materials.

Test Materials and Questionnaires

Participants' conceptual understanding of the topics enzyme energetics and randomness was tested through 16 multiple-choice questions (9 for randomness, 7 questions for enzyme energetics) in both, the immediate post-test (α = 0.65) and the delayed post-test (α = 0.69). 4 of these questions were adapted from previously published concept inventories (Fisher, Williams, & Lineback, 2011; Klymkowsky, Garvin-Doxas, & Zeilik, 2003). The remaining questions were self-developed and analyzed with biology experts. A sample question from the test was:

Question: In plant cells, sucrose molecules are continuously synthesized from one fructose and one glucose monomer. Which statement about sucrose synthesis is true?

a. Since the synthesis reaction occurs at an enzyme, it can occur even if the reaction itself is unfavorable.
b. The energy needed to drive the unfavorable reaction and thus synthesize sucrose is produced during photosynthesis.
c. Sucrose can only be synthesized if an energetically favorable reaction is coupled to it. *(correct answer)*
d. The synthesis of sucrose is an energetically favorable reaction because the order in the system increases.

The learning mechanisms were assessed in self-developed questionnaires using a five-point Likert scale (from 1 to 5, whereby the value 1 indicates strong disagreement and 5 strong agreement). Examples of the various subscales and descriptive statistics are shown in Table 1. Additionally, students' educational background (i.e., major in high school) and language fluency were assessed in single self-report questions.

Procedure

The participants were randomly assigned to one of two experimental groups: they either read the narrative or the expository text, followed by the immediate post-test and the learning mechanism questionnaires. Three months later, participants took a delayed post-test. One participant was excluded due to early submission, one participant was excluded due to a lack of effort, and 6 participants were excluded due to self-reported disfluency in the intervention language. A meta-analysis by Melby-Lervåg and Lervåg (2014) supports the latter, showing that reading comprehension in second-language learners is lower than in first-language learners. This resulted in the final sample size of 67 participants (narrative: n = 32; expository: n = 35) who took the immediate post-test and 24 participants (narrative: n = 10; expository: n = 14) who took the delayed post-test.

The post-test results for each topic were analyzed separately using ANOVA analyses, followed by simple main effect calculations, or using a repeated-measures ANOVA to investigate performance over time. Additionally, effect sizes were determined. For the subscales of the questionnaire, the average answer per participant was calculated, and Kruskal-Wallis tests were performed for significance testing on the ordinal scale. For the extraneous cognitive load, an independent group Welch's equivalence test was performed. Bayes factors (BF) were calculated for the different comparisons if applicable to provide evidence favoring the null or alternative hypotheses.

Results

Immediate Post-Test

The ANOVA analyses of the post-test performance with the educational background and intervention group as fixed factors did not reveal a significant main effect of the intervention group on the performance in the immediate post-test (randomness: $F(1,63) = 0.009$, $p = .925$; enzyme energetics: $F(1,63) = 1.505$, $p = .224$). In contrast, a significant main effect was found for the participants' educational background in high school (i.e. STEM-major) in the enzyme energetics test ($F(1,63) = 10.517$, $p = .002$). No effect was found for the background in the randomness test ($F(1,63) = 1.784$, $p = .187$). The interaction effects of educational background and intervention group were not significant (randomness: $F(1,63) = 0.038$, $p = .846$; enzyme energetics: $F(1,63) = 0.339$, $p = .563$). Descriptive statistics of the immediate post-test are shown in Table 2.

Table 2: Descriptive statistics of the immediate post-test performance.
Max: maximal number of achievable points; SE: standard error; n: sample size.

		Narrative			*Expository*		
Immediate post-test	Max	Mean	SE	n	Mean	SE	n
Together							
Randomness	9	4.03	0.07	32	4.09	0.06	35
Enzyme energetics	7	2.53	0.03	32	2.86	0.03	35
STEM major							
Randomness	9	4.37	0.12	19	4.33	0.15	14
Enzyme energetics	7	2.95	0.04	19	3.14	0.08	14
Non-STEM major							
Randomness	9	3.54	0.18	13	3.71	0.10	21
Enzyme energetics	7	1.92	0.09	13	2.32	0.06	21

The effects of the different educational backgrounds in the two intervention conditions were separately investigated by analyzing simple main effects. With the intervention group as moderator factor, the comparisons for the randomness test revealed a small effect size for the background in both groups (expository group: d = 0.20, F(1,63) = 0.687, *p* = .41; narrative group: d = 0.27, F(1,63) = 1.14, *p* = .29) favoring students with STEM majors.

Similar, but more pronounced results were found for the enzyme energetics test: in the expository group, a medium effect size and a non-significant difference were found (d = 0.48, F(1,63) = 3.76, *p* = .06). In the narrative group, there were significant performance differences of a large effect size between the participants with and without STEM-major in high school (d = 0.66, F(1,63) = 7.10, *p* = .01). In conclusion, we found no evidence for hypothesis H1 that students who have read the narrative materials understand the materials better than students of the other group. However, there is strong evidence in favor of the hypothesis H2a that prior knowledge (i.e., amount of prior STEM education) influences the effects of the intervention on performance. This effect on learning is only significant in the narrative group.

We further analyzed STEM and non-STEM majors separately. For that, we compared simple main effects of the intervention condition with educational background as moderator factor. For students with a STEM-background, no significant differences and a null effect were found by contrasting the performance in the randomness test in the two experimental conditions (d = 0.01, F(1,63) = 0.003, *p* = .96). Likewise, no differences were found for non-STEM students when comparing the performance of the two experimental groups (d = -0.05, F(1,63) = 0.044, *p* = .83). Similarly, for the enzyme energetics test, the performance comparison of the two experimental groups when only considering STEM-background students did not reveal significant differences, but a small effect size disfavoring students in the narrative intervention group was found (d = -0.14, F(1,63) = 0.334, *p* = .57). A comparable result was obtained for the intervention group comparison for only those students without a STEM background (d = -0.31, F(1,63) = 1.51, *p* = .22).

In conclusion, no evidence was found favoring hypothesis H2b that students with no STEM background and thus lower prior knowledge in the covered topics profit more from the narrative text than from the expository text.

Delayed Post-Test

The descriptive statistics of the delayed post-test are shown in Table 3. No significant group differences were detectable for both concepts (randomness: d = 0.09, 95% CI [-0.76, 0.94], t(21.5) = -0.22, *p* = .41; enzyme energetics: d = -0.18, 95% CI [-1.04, 0.67], t(20.04) = 0.45, *p* = .67). However, as only a subset of participants volunteered in both parts of the study (n = 24), comparing only their performance at the two timepoints revealed more insightful results than looking at the results from the delayed post-test in isolation.

To do so, a repeated-measures ANOVA was conducted for both concepts individually to investigate *between-subject* effects of the intervention condition. Thereby, no significant

Table 3: Descriptive statistics of the post-test performance of those participants who participated in the immediate and the delayed post-test. Max: maximal number of achievable points; SE: standard error; n: sample size.

		Narrative			*Expository*		
Subgroup analysis	Max	Mean	SE	n	Mean	SE	n
Immediate post-test							
Randomness	9	5.55	0.22	11	4.92	0.16	13
Enzyme energetics	7	3.64	0.14	11	4.31	0.12	13
Delayed post-test							
Randomness	9	4.27	0.16	11	4.08	0.19	13
Enzyme energetics	7	4.00	0.12	11	4.23	0.09	13

effects were found for both, the randomness concept ($F(1,1) = 0.159$, $p = .695$) and the enzyme energetics concept ($F(1,1) = 1.870$, $p = .187$). Therefore, there is no evidence in favor of the hypothesis H3 that participants who have read the narrative text better retained the conceptual understanding of the covered topics over time than those from the expository group.

Learning Mechanisms

In comparison to the participants of the expository text condition, there was strong evidence for a higher enjoyment when reading the materials in the narrative group ($t(64.3) = -1.57$, $p = .06$, $BF_{10} = 15.26$). Furthermore, there was strong evidence for a higher germane cognitive load in the narrative group ($t(64.88) = -1.54$, $p = .06$, $BF_{10} = 14.11$). Correlating the average self-reported scores for germane cognitive load with the total score in the immediate post-test revealed a significant and positive correlation for the expository group ($R = 0.34$, $p = .044$) but negligible non-significant correlation for the narrative group ($R = -0.08$, $p = .65$) (Figure 1).

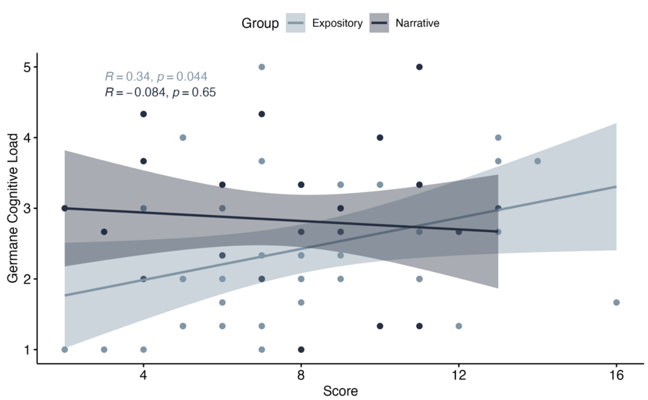

Figure 1: Group-wise correlation of the total score of the immediate post-test and the average germane cognitive load. Each point indicates one participant. Lines indicate the regression line of the correlation; transparent areas correspond to a 95% confidence interval.

There was no evidence in favor of the hypothesis H4b regarding higher interest and motivation in the narrative group compared to expository group ($t(64.95) = -0.64$, $p = .26$, $BF_{10} = 2.69$). Lastly, the equivalence test results for the extraneous cognitive load demonstrate that the observed effect is statistically not different from zero, and thus the groups are equivalent ($t(178.32) = 1.459$, $p = .146$).

In summary, we found evidence supporting the hypotheses regarding the enjoyment (H4a), the germane cognitive load (H4c), and the extraneous cognitive load (H4d). However, there was no evidence in favor of the hypothesis regarding interest and motivation (H4b). Descriptive statistics of the learning mechanisms are shown in Table 4.

Table 4: Descriptive statistics of learning mechanisms. ΔMean (*Nar-Exp*): Mean rating difference between narrative and expository group; SE: propagated standard error; CL: cognitive load; n = 67.

Questionnaire Subscale	ΔMean *(Nar-Exp)*	SE *(Nar-Exp)*
Enjoyment	0.32	0.02
Germane CL	0.38	0.02
Extraneous CL	0.11	0.03
Interest & Motivation	0.18	0.03

Discussion

The present study showed that a more differentiated view on learning from narratives is needed since they do not promote conceptual understanding more than learning from expository texts in all circumstances.

Testing the understanding upon reading either a narrative or an expository text addressing theoretical aspects of the concepts randomness and enzyme energetics revealed that first-year university students did not perform better when having read these concepts embedded in a narrative. These findings are consistent with earlier results by Wolfe and Woodwyk (2010). In agreement with their findings, our empirical results also showed that prior knowledge is an essential predictor of performance (i.e., students with higher prior knowledge performed better in the immediate post-test). Additionally, we showed the effects of prior knowledge on learning are more pronounced in the narrative group.

As we conducted the intervention during the participants' second week at university, they had not yet learned much at university despite being undergraduate students. Therefore, a significant part of their prior knowledge comes from their previous education and thus is strongly dependent on whether the students chose a STEM major in high school. Comparing then the two high school biology curriculums (STEM vs. non-STEM) of one of Switzerland's biggest high schools revealed that a) the concept of enzyme energetics may only be briefly mentioned in high school chemistry classes in STEM majors and not in non-STEM majors, and that b) the concept of randomness could have been already partially and indirectly covered.

Therefore, it would be expectable that the participants have higher prior knowledge of the randomness concept and none or low prior knowledge of the enzyme energetics concept. Reconsidering the results of the immediate post-test, the findings indicate that with higher prior knowledge, the negative effect of the narrative on understanding becomes smaller when compared to expository text materials. Taking into account other studies that report the benefits of narratives on students' understanding compared to expository texts (e.g., Dai et al., 2021), these results suggest that students could benefit more from a narrative than from an expository text in those cases in which the prior knowledge is sufficient to learn the new contents.

In contrast, when the prior knowledge is insufficient, students could struggle to extract the essential conceptual elements from the narrative. A reason for why learning with narratives could hinder conceptual understanding could be that reading a narrative text forces the reader to continuously switch between concentrating on the conceptual or the story elements (Fisch, 2000; Jetton, 1994). Still, the questionnaire results concerning the extraneous cognitive load demonstrate that the two groups are equivalent and thus that participants of the narrative group were not more distracted by text-intrinsic elements than those in the expository group.

However, descriptively, it appears that those students who performed below average in the immediate post-test reported a higher germane cognitive load if they were reading the narrative compared to the expository text (Figure 1). While there was no significant correlation between performance and germane cognitive load in the narrative group, a worse performance significantly correlated with a lower germane cognitive load in the expository group. Thus, narratives could help keep the germane cognitive load higher, especially for students who struggle in class. The absence of a correlation between high germane cognitive load and better performance in the narrative group could be explained by the Likert scale intrinsic limitations as the choice of a mid-point answer in case of response uncertainty (Nadler, Weston, & Voyles, 2015). Regression to the mid-point of the germane cognitive load scale is also detectable for the expository group with higher performance.

Concomitant, the results of this study also indicate that the enjoyment when reading the narrative text was higher than when reading the expository text, thus revealing a potential reason why students had a higher germane cognitive load in the narrative group as they could focus better on the task.

Nevertheless, the present findings do not show significant differences in motivation or interest towards the topic upon reading the narrative text compared to the expository one, indicating that these factors might play a minor role in explaining the performance.

Limitations and Future Work

Even though the students were made aware of the delayed post-test component of the study, the recruitment remained one of the biggest challenges, resulting in a relatively low sample size for the delayed post-test. Moreover, higher participation in the first part of the study would have allowed a better quantification of the group differences.

Furthermore, it would be interesting to measure not only the students' understanding of the concept but also assess their ability to recall the contents. This could be achieved, for example, in a first part in which both groups would be tested using recall questions on the concepts, and in a second part in which only the narrative group would be asked to recall elements of the historical parts. Like that, the content recall in the two groups could be compared. Moreover, correlation analyses could be performed to study interactions between off-concept content recall and understanding or domain-specific concept recall.

Ultimately, in an ensuing study, students with higher prior knowledge should be targeted to see whether the narrative group outperforms the expository group in this case, as prior knowledge seems to be an essential factor concerning the effectiveness of narratives in education.

Conclusion

This study yielded valuable insights on the effects of prior knowledge on learning with narratives to reach conceptual understanding in biology. Using narratives to design learning materials may only be favorable when students have enough prior knowledge to follow the contents. Nevertheless, even in cases in which narratives are as effective as expository texts for learning, narratives increase the germane cognitive load and lead to better enjoyment of the topic.

References

Arya, D. J., & Maul, A. (2012). The role of the scientific discovery narrative in middle school science education: An experimental study. In *Journal of Educational Psychology* (Vol. 104, Issue 4, pp. 1022–1032). American Psychological Association. https://doi.org/10.1037/a0028108

Brumby, M. N. (1984). Misconceptions about the concept of natural selection by medical biology students. *Science Education*, *68*(4), 493–503. https://doi.org/10.1002/sce.3730680412

Champagne Queloz, A., Klymkowsky, M. W., Stern, E., Hafen, E., & Köhler, K. (2016). Debunking Key and Lock Biology: Exploring the prevalence and persistence of students' misconceptions on the nature and flexibility of molecular interactions. *Matters Select*. https://doi.org/10.19185/matters.201606000010

Champely, S. (2020). *pwr: Basic Functions for Power Analysis*. https://cran.r-project.org/package=pwr

Coley, J. D., & Tanner, K. D. (2015). Relations between intuitive biological thinking and biological misconceptions in biology majors and nonmajors. *CBE Life Sciences Education*, *14*(1). https://doi.org/10.1187/cbe.14-06-0094

Cooper, M. M., Corley, L. M., & Underwood, S. M. (2013). An investigation of college chemistry students' understanding of structure–property relationships. *Journal of Research in Science Teaching*, *50*(6), 699–721. https://doi.org/10.1002/tea.21093

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 LP – 13620. https://doi.org/10.1073/pnas.1320645111

Dai, P., Williams, C. T., Witucki, A. M., & Rudge, D. W. (2021). Rosalind Franklin and the Discovery of the Structure of DNA. *Science & Education*, *30*(3), 659–692. https://doi.org/10.1007/s11191-020-00188-6

Emmons, N., Smith, H., & Kelemen, D. (2016). Changing Minds With the Story of Adaptation: Strategies for

Teaching Young Children About Natural Selection. *Early Education and Development*, *27*(8), 1205–1221. https://doi.org/10.1080/10409289.2016.1169823
Fiedler, D., Tröbst, S., & Harms, U. (2017). University Students' Conceptual Knowledge of Randomness and Probability in the Contexts of Evolution and Mathematics. *CBE Life Sciences Education*, *16*(2), ar38. https://doi.org/10.1187/cbe.16-07-0230
Fisch, S. (2000). A Capacity Model of Children's Comprehension of Educational Content on Television. *Media Psychology - MEDIA PSYCHOL*, *2*, 63–91. https://doi.org/10.1207/S1532785XMEP0201_4
Fisher, K. M., Williams, K. S., & Lineback, J. E. (2011). Osmosis and diffusion conceptual assessment. *CBE Life Sciences Education*, *10*(4), 418–429. https://doi.org/10.1187/cbe.11-04-0038
Flesch, R. (1948). A new readability yardstick. *The Journal of Applied Psychology*, *32*(3), 221–233. https://doi.org/10.1037/h0057532
Hopkins, E. J., & Weisberg, D. S. (2021). Investigating the effectiveness of fantasy stories for teaching scientific principles. *Journal of Experimental Child Psychology*, *203*, 105047. https://doi.org/https://doi.org/10.1016/j.jecp.2020.105047
Jetton, T. L. (1994). Information-driven versus story-driven: What children remember when they are read informational stories. *Reading Psychology*, *15*(2), 109–130. https://doi.org/10.1080/0270271940150203
Klymkowsky, M. W., Garvin-Doxas, K., & Zeilik, M. (2003). Bioliteracy and teaching efficacy: what biologists can learn from physicists. *Cell Biology Education*, *2*(3), 155–161. https://doi.org/10.1187/cbe.03-03-0014
Melby-Lervåg, M., & Lervåg, A. (2014). Reading comprehension and its underlying components in second-language learners: A meta-analysis of studies comparing first- and second-language learners. *Psychological Bulletin*, *140*(2), 409–433. https://doi.org/10.1037/a0033890
Nadler, J. T., Weston, R., & Voyles, E. C. (2015). Stuck in the Middle: The Use and Interpretation of Mid-Points in Items on Questionnaires. *The Journal of General Psychology*, *142*(2), 71–89. https://doi.org/10.1080/00221309.2014.994590
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/doi:10.1002/sce.20063
R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. https://www.r-project.org/
Reuer, M. D. (2012). *Backroads to learning : the use of narratives in high school biology*. 1–124. https://scholarworks.montana.edu/xmlui/handle/1/2123
Richards, B. (1987). Type/Token Ratios: what do they really tell us? *Journal of Child Language*, *14*, 201–209. https://doi.org/10.1017/S0305000900012885
Tobler, S., Sinha, T., Köhler, K., Hafen, E., & Kapur, M. (2022). Effects of narratives on undergraduate student understanding of fundamental concepts in biology. *[Accepted for Publication at the 16th International Conference of the Learning Sciences (ICLS 2022), Hiroshima, Japan]*.
Vosniadou, S., Vamvakoussi, X., & Skopeliti, I. (2008). The framework theory approach to the problem of conceptual change. *International Handbook of Research on Conceptual Change*, 3–34.
Wolfe, M., & Mienko, J. (2007). Learning and memory of factual content from narrative and expository text. *The British Journal of Educational Psychology*, *77*, 541–564. https://doi.org/10.1348/000709906X143902
Wolfe, M., & Woodwyk, J. (2010). Processing and memory of information presented in narrative or expository texts. *British Journal of Educational Psychology*, *80*(3), 341–362. https://doi.org/10.1348/000709910X485700
Yuan, Y., Major-Girardin, J., & Brown, S. (2018). Storytelling Is Intrinsically Mentalistic: A Functional Magnetic Resonance Imaging Study of Narrative Production across Modalities. *Journal of Cognitive Neuroscience*, *30*(9), 1298–1314. https://doi.org/10.1162/jocn_a_01294