



# Does Concrete Content Help People to Reason Scientifically?

## Adaptation of Scientific Reasoning Scale

Jana Bašnáková<sup>1</sup> · Vladimíra Čavojová<sup>1</sup> · Jakub Šrol<sup>1</sup>

Accepted: 18 February 2021 / Published online: 10 April 2021  
© The Author(s), under exclusive licence to Springer Nature B.V. 2021

### Abstract

In this paper, we explored the scientific literacy of a general sample of the Slovak adult population and examined factors that might help or inhibit scientific reasoning, namely the content of the problems. In doing so, we also verified the assumption that when faced with real-life scientific problems, people do not necessarily apply decontextualized knowledge of methodological principles, but reason from the bottom up, i.e. by predominantly relying on heuristics based on what they already know or believe about the topic. One thousand and twelve adults completed three measures of scientific literacy (science knowledge, scientific reasoning, attitudes to science) and several other related constructs (numeracy, need for cognition, PISA tasks). In general, Slovak participants' performance on scientific reasoning tasks was fairly low and dependent on the context in which the problems were presented—there was a 63% success rate for a version with concrete problems and a 56% success rate for the decontextualized version. The main contribution of this study is a modification and validation of the scientific reasoning scale using a large sample size, which allows for more thorough testing of all components of scientific literacy.

## 1 Introduction

Science has become an integral part of the lives of people in a modern society. Scholars have argued for the importance of science as a vaccination against pseudoscientific beliefs (Fasce & Picó, 2019; Sagan, 1996) and emphasized the role of civic scientific literacy for responsible citizenship and making informed decisions (Lederman et al., 2014; Miller, 1983, 1998, 2004; OECD, 2003; Trefil, 2008). The growing acknowledgment of the importance of science is reflected in educational efforts (Leblebicioglu et al., 2017; Lederman et al., 2014) and inclusion of science literacy in international assessments such as PISA (Eivers & Kennedy, 2006; Harlen, 2005; She et al., 2018). Despite these efforts, studies show that few adults actually understand how science “works”—for example, in 1978, only

---

✉ Vladimíra Čavojová  
vladimira.cavojova@savba.sk

<sup>1</sup> Institute of Experimental Psychology, Centre of Social and Psychological Sciences, Slovak Academy of Sciences, Dúbravská cesta 9, 841 04 Bratislava, Slovakia

14% of adults were able to satisfactorily describe what does it mean to study something scientifically (Miller, 2004), while a Pew Poll in 2019 showed that although the percentage of people understanding science increased considerably, still only 42% of adults with high school education understand the scientific method (Kennedy & Hefferon, 2019). Strikingly, even among those who teach scientific content, there is a misunderstanding of and mistrust in the scientific method—a recent poll among Slovak middle-school teachers showed that almost one-third of them believe that COVID-19 vaccination means the population will be inserted with “nano-chips” in their body, a common conspiracy theory (Gdovinová, 2020). Even though every pupil from primary school onwards gets frequent exposure to scientific content in several subjects, this implicit exposure to scientific knowledge clearly does not lead to widespread understanding on how this knowledge is created, tested, and reviewed. Thus, the main aim of our paper was to examine the scientific literacy of adults in Slovakia with a modified scientific reasoning scale, and to examine factors that help/inhibit scientific reasoning, including whether the content of the problems allows them to reason from concrete examples, or whether they routinely rely on abstract principles.

Scientific literacy is a broad concept, often confused with other related constructs (e.g., engagement with science and technology, science literacy, science understanding, science knowledge, or scientific reasoning) and it is not always clear on what aspect (e.g., intellectual, societal, attitudinal, axiological) of scientific activity researchers focus (Fasce & Picó, 2019). Lederman et al. (2014) discuss the distinction between scientific inquiry (processes by which people acquire scientific knowledge) and the nature of scientific knowledge (“epistemological underpinnings of the activities and products of science,” p. 289). Moreover, while the current pedagogical trends stress the necessity of teaching science as inquiry (Anderson, 2002; Hayes, 2002; McBride et al., 2004), there remains a gap between what students “do” vs. what they “know”—for example, it is one thing to teach students to set up an experiment and perform it, and it is a completely different thing to expect them to grasp the logical necessity of a control group in experimental design implicitly just from performing the activity.

Therefore, it seems that scientific literacy has at least three dimensions (Fasce & Picó, 2019; Miller, 1983, 1998, 2004): (1) knowledge of scientific theories/science vocabulary/conceptual understanding, (2) the understanding of scientific reasoning/scientific method, and (3) trust in science and its values/understanding of science as an organized endeavor/understanding of the importance of science. However, so far all three dimensions of science understanding have not been addressed in a single study. Therefore, it is difficult to distinguish whether poor scientific literacy of adults is brought about by their lack of knowledge, lack of ability to reason scientifically, or by their lack of trust in scientific practices. A clearer understanding of the relationship among these three dimensions of scientific literacy would help educators better address the lacking components; many critics of science education in various countries have pointed out that schools concentrate predominantly on teaching knowledge at the expense of true understanding of the taught concepts or the nature of scientific inquiry (Elakana, 2000; Machamer, 1998; Schwab, 1958). In this paper, we will examine these three components of scientific literacy and their interactions, with predominant focus on scientific reasoning. Even though an intervention to address scientific reasoning is not in the scope of our research, the broader aim is to be eventually able to directly address this set of skills in schools, in addition to teaching scientific content.

In this paper, we broadly define scientific reasoning as the ability to understand methods and principles of scientific inquiry and the skills required to formulate, test, and revise theories and reflect the process of generating evidence; or, more specifically, as the “application of the methods or principles of scientific inquiry to reasoning or problem-solving

situations” (Zimmerman, 2007, p. 173; see also Koslowski, 1996; Kuhn & Franklin, 2007; Wilkening & Sodian, 2005). In this sense, scientific reasoning is a subset of critical thinking skills that help people reason about any complex content, and, despite its name, is not for scientists only. Simply put, scientific reasoning provides us with the instruments to formulate and verify our implicit theories about the world and evaluate them based on reliable evidence. We approach assessing scientific reasoning skills through a test focused on scientific processes rather than scientific facts, which is inspired by the work of Drummond and Fischhoff (2017). Their scientific reasoning scale (SRS) consists of several short scenarios describing fictional research studies that probe participants’ understanding of methodological principles that underlie scientific reasoning, such as that there is a difference between causation and correlation and that one needs a control group for a valid comparison of treatments, confounding variables must be identified and controlled for. To a large extent, these principles are domain-general, as opposed to domain-specific factual knowledge about various scientific domains (Zimmerman, 2000).

Preliminary results using the original SRS in the Slovak population indicated that participants’ ability to solve the tasks was influenced by whether they could imagine an *intuitive* solution to a particular problem (Bašňáková & Čavoјová, 2019). To illustrate, consider the following problem (Drummond & Fischhoff, 2017, p. 36):

*A researcher finds that American states with larger state parks have fewer endangered species. Is the following statement True or False? These data show that increasing the size of American state parks will reduce the number of endangered species.*

The “endangered species” problem probes understanding of causation vs. correlation, but it is its content or context that might help the participant to solve it correctly rather than understanding the methodological principle (*It is unlikely that giving animals more space to live will decrease their population*). At the same time, changing the story (*decreasing park’s surface—fewer species*) might make it more likely for participants to erroneously accept the conclusion as *True*. In other words, while these problems assess the understanding of a methodological principle, the principle is always evaluated in a particular context, including participants’ prior beliefs, which might make it more or less difficult for the participant to give a correct answer even without fully understanding the methodological issue.

In the present study, we addressed this issue via modifying the scale by adding two versions of each item—one using concrete and domain-specific examples as in the original version of the SRS and another using context-free versions illustrating the same methodological principle. This was done to gain a better understanding of how people reason scientifically. On the one hand, they might understand and correctly apply a particular abstract principle regardless of the context in which it is presented; on the other hand, they might simply use contextual cues contained in the scenario to solve the task intuitively. To understand their reasoning process related to context-free and concrete problems in more detail, we asked them to give reasons for their answers and then analyzed these qualitative responses for any systematic errors in scientific reasoning.

## 2 Reasoning with Concrete vs. Context-Free Content

Why should scientific reasoning differ based on how rich the contextual cues contained in the problem are? There are several indications, some of which are general to reasoning and others specifically related to scientific reasoning.

One line of evidence suggesting that reasoning with concrete and more abstract content differs *in general* comes from the psychology of reasoning. In studies using the Wason's selection task (Wason, 1968), participants are shown four cards—each of which has a number on one side and a letter on the other side. The visible faces of the cards show 3, 8, *a*, and *c*. Participants are then asked, which card(s) must one turn over in order to test the truth of the proposition that *if a card shows an even number on one face, then its opposite face contains a vowel*. Numerous studies showed that people usually perform quite poorly at this task since they use mainly confirmatory strategies (choosing 8 and *a*); only a small minority of participants get the right answer (which is 8 and *c*) (Čavojská & Jurkovič, 2017; Rossi et al., 2015; Stanovich & West, 1998). Performance dramatically improves, however, with concrete examples, especially ones about a social contract, such as verifying the rule that in order to purchase an alcoholic drink, the buyer must be over 18 and presenting the participants with cards showing ages 16, 24, beer, and soda (Cheng et al., 1986; Cosmides & Tooby, 1992; Cox & Griggs, 1982; Griggs & Cox, 1982).

In the previous case, the concrete content helped people to reason better because concrete content triggers our intuitions (like a presumably modular system for recognizing the violation of a social contract (Cosmides & Tooby, 1992; Fiddick et al., 2000). However, these intuitions can lead people down the wrong path as well, as is shown in the heuristic and biases research program which uses the so-called conflict problems to study human reasoning. For example, consider a simple base-rate task (taken from De Neys & Glumicic, 2008; adapted from Kahneman & Tversky, 1973): *A group of 1000 people were tested in a study. Among the participants were 995 nurses and 5 doctors. Paul is a randomly chosen participant from this study. Paul is 34 years old. He lives in a beautiful home in a posh suburb. He is well-spoken and very interested in politics. He invests a lot of time in his career. What is most likely? (a) Paul is a nurse (b) Paul is a doctor.* In this case, the content of the task—description of a personality of the participant—exploits existing stereotypes eliciting an intuitive answer (“doctor”) that conflicts with the extreme base-rate probability of Paul being a nurse. In non-conflict tasks, people are able to respond correctly based on the information about base rates. De Neys (2014) argues for the existence of the so-called logical intuition and draws evidence from studies that used abstract versions of the classical conflict problems (e.g., Kahneman & Tversky, 1973), which did not cue an intuitive heuristic response and thus participants were able to answer correctly.

In the following section, we will discuss how these general principles can be reflected in scientific reasoning of adult people and their consequences for science education. The first approach to solving scientific reasoning problems—a one which we will refer to as “top-down reasoning”—would be through the application of (abstract) rules, such as the rule about the difference between causation and correlation. In other words, participants would not be relying on their knowledge of the domain since methodological principles apply across all domains. This is reminiscent of developmental studies by Inhelder and Piaget (1958) who gave children tasks where they could not rely on their prior knowledge while identifying plausible hypotheses about how something works but try out all possible combinations of solutions, e.g. a task where a particular combination of odorless liquids produced a color change.

Another way in which people may proceed when dealing with scientific reasoning problems is by making use of their prior beliefs about a certain domain. Greenhoot et al. (2004) investigated such “bottom-up reasoning” by studying whether prior knowledge and beliefs about a topic will influence how students evaluate scientific evidence. Interestingly, about a fourth of the students did manage to correctly apply methodological concepts to draw valid conclusions when they were asked to evaluate a study in an abstract manner (i.e., make recommendations for a hypothetical experimenter). However, when they were to make the same conclusions for themselves, they failed to consider these principles because they had strong prior beliefs about the topic (e.g., whether TV viewing causes poor language abilities in children). In other words, even if students knew how to critically evaluate evidence, they seemed to suspend this ability when asked about “real-world” problems and were reluctant to revise their prior beliefs in the face of contradictory evidence. This suggests that even if people know how to solve “decontextualized” problems, they might be suppressing this ability and rely on error-prone heuristics when encountering a problem that they know a lot about because such active suppression requires effort and meta-cognitive distancing (Kuhn et al., 1995).

Building on this distinction, we made sure to construct our two versions of the SRS items in a way that one version contained concrete information about the research domain where participants could rely on their prior knowledge, while the other was context-free. For example, an item aimed at testing the difference between correlation and causation was about whether the number of hospitals in a region influences natality of the population in the more concrete version, while in the decontextualized version, it was about the relationship between skill *X* and activity *Y*. We hypothesized that if people reason in a decontextualized, top-down manner, i.e. know the methodological principles and apply them to a particular problem, it will not matter which version of the task they encounter—their performance should be comparable. If, on the other hand, they reason from the bottom up, i.e. by predominantly relying in heuristics based on what they already know or believe about the topic, then there should be a significant difference in their performance on the two versions, as well as in the reasons given for participants’ choices.

## 3 Methods

### 3.1 Participants

Through a market research agency (2Muse), we recruited a sample of 1012 participants, representative of the Slovak general population with respect to age, gender, education, and geographic location, who filled out the questionnaire online. Participants were told that the goal of the study was to validate a new method of testing scientific literacy of the population and that it should take no longer than 45 min. We also encouraged them to provide us with any feedback (including negative) regarding the difficulty, comprehensibility, and/or content of the questions. They were also informed about their right to remain anonymous and withdraw from the study at any time. The questionnaires were administered to participants via the online survey platform Qualtrics, once they had voluntarily agreed

to participate in the research. There were 510 men and 502 women with mean age 39.2 ( $SD=15.6$ ) years. A total of 8.8% of participants had only elementary education (9 years of schooling), 31.7% of participants had vocational education (12 years), 39.4% had full secondary education (13–14 years), 5.4% of participants had a Bachelor's degree, 13.7% of participants had a Master's degree,<sup>1</sup> and 0.9% of participants obtained a post-graduate degree.

### 3.2 Materials<sup>2</sup>

**Scientific Reasoning** In this study, we used a 7-item adaptation of the scientific reasoning scale (Drummond & Fischhoff, 2017), which was based on the results obtained from the sample of professional scientists (Bašnáková & Čavojová, 2019).<sup>3</sup> The 7 final items contained the following validity threats: blinding, causation vs. correlation, confounding variables, construct validity, control group, ecological validity, and random assignment to conditions. We constructed a context-free version of each of the 7 final items so that people could not simply use their knowledge of the topic to arrive at intuitive answers. Context-free versions were created by modifying the concrete versions, to be as closely matched as possible, but leaving out any specific details. For example, the “causation vs. correlation” concrete version was about increasing birth rate (*A researcher wants to find out how to increase natality. He asks for statistical information and sees that there are more children born in cities which have more hospitals. This finding implies that building new hospitals will increase birth rate of a population. Agree/Disagree*). The context-free version just described “ability X” (*A researcher wants to find out how to improve ability X in the population. From the available statistical data, he finds out that more people with ability X regularly engage in activity Y. This result suggests that activity Y improves ability X. Agree/Disagree*).

Just as in the original Drummond and Fischhoff's SRS, participants were asked to indicate their agreement with the conclusion drawn from the scenario after each story. However, we did not ask them to say whether these conclusions are True or False, but rather whether they Agree or Disagree with them. In addition, we also asked about their level of confidence in their choice on a scale from 1 (*not at all certain*) to 10 (*totally certain*) to be later used to detect the rate of guessing. We also included a mandatory explanation of their choice (“Using one sentence, please indicate why you chose ‘agree’ or ‘disagree’:”) The reason for this was to increase the questionnaire's sensitivity beyond the *Agree/Disagree* judgment and detect whether participants truly understood the problems, as well as whether the context-free and concrete explanations systematically differ. Participants answered all SRS items in a randomized order.

<sup>1</sup> In 2000, 9.5% of inhabitants of Slovakia completed a Master's degree. The higher percentage of Master's degrees in comparisons with Bachelor's degrees is a result of the educational system in Slovakia and the low attractiveness of Bachelor's degrees.

<sup>2</sup> All materials are available at <https://osf.io/udgre>.

<sup>3</sup> The original version of the SRS was difficult (only 44% success rate) in our previous research. To distinguish whether the difficulty was related to the particular formulations of the items or to an underlying lack of understanding of the scientific concepts, we conducted a preliminary study, for which we created 3–4 new items measuring each construct, resulting in 34 items. We then piloted them on a sample of researchers ( $N=79$ ), as a proof of principle. Consequently, we selected only those items that had success rate above 70% (50% is chance level) and were the most unambiguous and comprehensible, meaning that some constructs from the original SRS were left out entirely. In the final version, the one scenario/item per construct was selected to mimic the original scale as closely as possible, and also to limit the time participants spend on the questionnaire.

The *scientific knowledge* (SK) was a 9-item true/false questionnaire on basic scientific facts, such as “Antibiotics kills viruses as well as bacteria,” based on National Science Indicators (Miller, 1998; National Science Board, 2010). We used a composite score as an index of public comprehension of science (Allum et al., 2008; Kahan et al., 2012).

*Anti-scientific attitudes* subscale from CART developed by Stanovich et al. (2016) was used to measure anti-scientific sentiments, as the public attitudes about science are not only a matter of understanding how science works but also of trust in scientist and regulatory authorities (Allum et al. 2008). Participants had to indicate their agreement with 13 items on a 6-point scale ranging from 1 (strongly disagree) to 6 (strongly agree), with statements such as “I don’t place great value on ‘scientific facts’, because scientific facts can be used to prove almost anything”. A higher score indicated a stronger anti-scientific attitude, and thus lower trust in science and understanding of the importance of science.

### 3.2.1 Validity Measures

To distinguish scientific literacy from other related constructs (e.g., mathematical ability), we employed several more validity measures. Specifically, we chose to measure numeracy and cognitive reflection to assess the ability to think analytically (the same measures were used by Drummond and Fischhoff to determine the construct validity of the SRS) and need for cognition as a self-reported enjoyment of thinking about difficult problems (as a complement to actively open-minded thinking used by Drummond and Fischhoff), because we expected them to be important correlates of scientific literacy components. Similarly, we also choose to measure the ability to solve PISA tasks based on two main reasons: (1) they are complex tasks which target the participants’ ability to use their knowledge, and (2) it would enable us to compare the results of our adult sample with the performance of the students participating in nationwide PISA testing and draw insights relevant for education.

The *numeracy scale* (NS) contained 2 items from the modified cognitive reflection test (Dudeková & Kostovičová, 2015; Frederick, 2005), where the structure of original items was kept but the content was modified to avoid a common problem with participants recognizing the test items from popular sources (even though it was shown that previous exposure does not significantly affect the results, see Bialek & Pennycook, 2018; Šrol, 2018). One problem with Bayesian reasoning was taken from Zhu and Gigerenzer (2006). The remaining 5 questions were taken from Peters et al. (2006), all on probability.

Program for International Student Assessment (PISA) testing is a triennial test developed by the OECD designed to evaluate schools worldwide by testing a sample of 15-year-old students on key skills and knowledge from several topics. The latest assessment included testing from science, mathematics, reading, collaborative problem-solving, and financial literacy, of which we selected 4 problems on science. Two were focused on understanding which factors were manipulated in an experimental study, one on understanding and computing averages and the last one on understanding the concept behind immunity. We calculated a composite score of all the correct answers (0–4).

We also included 5 items from the *need for cognition scale* (Cacioppo & Petty, 1982, adapted and shortened by Epstein et al., 1996), which measures the tendency of an individual to engage in and enjoy thinking and intellectual problems. Participants had to indicate their agreement on a 6-point scale ranging from 1 (*strongly disagree*) to 6 (*strongly agree*), with statements, such as “I would prefer complex to simple problems”. A higher score indicated a higher need for cognition.

## 4 Results

### 4.1 Quantitative findings

First, we applied item response theory (IRT) model to our data to give us detailed information about the difficulty, discrimination, and guessing parameters of all items. The IRT analysis (presented in full in the [Supplementary materials](#)) showed that item 1 (which dealt with the issue of blinding) in both concrete and abstract versions exhibited a negative correlation with the total scale and was, therefore, dropped from all further analyses. Below, we present analyses pertaining to the concrete and context-free version of SRS while using only the remaining six items for both of the scales.

For the modified SRS, we calculated scores separately for the SRS-concrete and SRS-context-free, as well as the total SRS score. Percentage of correct answers for individual items, as well as mean confidence ratings and weighted scores, are presented in Table 1. Weighted scores were calculated by combining correct/incorrect scores with confidence ratings (Dube et al., 2010; Trippas et al., 2013, 2017). For participants who answered incorrectly, the weighted score was calculated by multiplying their confidence rating by  $-1$  and for participants who answered correctly, the weighted score was calculated by multiplying their confidence rating by  $+1$ . Thus, participants who were very confident (e.g., indicated 10) in their incorrect answer had a weighted score of  $-10$ . On the other hand, participants who gave correct answers and were not very confident (e.g., indicated 5 as a guess) had a weighted score of 5. As a result, positive scores indicate correct answers weighted by confidence, while negative scores indicate incorrect answers weighted by confidence.

For the 12 items of the combined abstract and concrete scientific reasoning scale, the participants on average correctly answered 7.52, or 62.7% ( $SD=2.58$ ).

As expected, participants answered the concrete questions better than the context-free ones. On average, 70% items were answered correctly on the concrete scale ( $M=4.16/6$ ,  $SD=1.46$ ) and 56.2% on the context-free scale ( $M=3.37/6$ ,  $SD=1.48$ ). This difference was statistically significant ( $t(1011)=18.038$ ,  $p<0.001$ ,  $d=0.54$ ). A direct comparison per item showed that all of the items except for one, “control group,” were significantly different (McNemar chi-square test, all  $\chi^2 > 12$ , all  $p$ 's  $< 0.001$ ); however, the differences were not in the same direction (see Fig. 1). Item 6 (random assignment to conditions) had more people scoring correctly on the abstract than on the concrete version. The same pattern, only more pronounced, emerged with the weighted scores. Apart from the “control group” item, all differences between the context-free and concrete version of each item were significant (all  $p$ 's  $< 0.001$ ), but item 6 was in the opposite direction than the rest (context-free  $>$  concrete). In this sense, the weighted scores are a more accurate characterization of the participants' performance than their raw scores.

For the 9 knowledge questions on the SK, participants correctly answered 70.1% of them ( $M=6.36$ ,  $SD=1.52$ ). As for the predictors, our participants' scores on the numeracy scale were slightly above 50% ( $M=3.80/7$ ,  $SD=1.95$ ). Not all items were equally difficult for participants—success rates ranged from as low as 13% for Bayesian task (item 8), to 84% for probabilistic reasoning task (item 4).

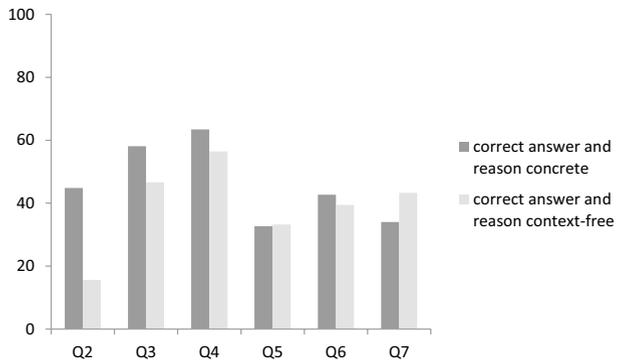
To examine the construct validity of the modified version of SRS, we carried out a correlational analysis of both versions of SRS with measures tapping science knowledge and predictors. As can be seen from Table 2, SRS correlated positively with science knowledge, numeracy, need for cognition, and success on PISA tasks, and negatively with anti-scientific attitudes.

**Table 1** Mean percent correct and confidence ratings for the concrete and abstract versions of each SRS item

Concept	Concrete			Context-free			95% CI for Cohen's <i>d</i>
	% correct	Mean confidence	Weighted score	% correct	Mean confidence	Weighted score	
Item 1. Causation vs. correlation	86%	8.28	6.24	38%	6.83	-2.19	0.945 [0.871, 1.019]
Item 2. Confounding variables	76%	8.15	4.69	55%	7.2	0.90	0.413 [0.348, 0.477]
Item 3. Construct validity	77%	8.38	4.81	71%	7.96	3.70	0.135 [0.073, 0.197]
Item 4. Control group	65%	7.55	2.56	66%	7.54	2.62	-0.008 [-0.070, 0.053]
Item 5. Ecological validity	60%	7.68	1.93	43%	7.67	-0.90	0.281 [0.218, 0.344]
Item 6. Random assignment to conditions	52%	8.17	0.38	64%	7.8	2.44	-0.234 [-0.296, -0.171]
Total score	70%	48.19	20.60	56.2%	44.99	6.56	-0.629 [-0.696, -0.561]

The table contains the mean percentage of correct responses, mean confidence, and average weighted score separately for every item of the two versions of the SRS. The last column shows the effect size of the difference between weighted scores in the concrete and context-free items

**Fig. 1** Percentage of participants per item who made a correct judgment and gave a correct reason for it



The three components of scientific literacy—scientific knowledge, scientific reasoning, trust in science—showed relatively similar patterns of correlations with validity measures (taking into account that we measured negative attitudes indicating distrust toward science, therefore, all correlations with anti-scientific attitudes are in the opposite direction). Somewhat different patterns of relationships were found in case of numeracy scale, where scientific knowledge showed a more pronounced correlation ( $r=0.455$ ) in comparison with scientific reasoning (especially, in the context-free version of the SRS,  $r_{\text{context-free}}=0.337$ ), and anti-scientific attitudes ( $r= -0.174$ ).

### 4.2 Qualitative Findings

To increase the sensitivity of our analysis and understand the participants’ thinking process as well as any differences between concrete and context-free items, we also looked at the reasons they gave to support their *Agree/Disagree* judgments.

A qualitative analysis of the participants’ reasons for choosing a specific answer was carried out in several steps. Firstly, blind to the actual judgments, we read through all of the answers to see whether there are any unifying categories across the items. We then created the following 5 categories: *correct*, *incorrect*, *don’t know/uncertain*, *no explanation/ restatement*, and *no reason/off-topic*.

**Table 2** Correlations between SRS and additional measures or reasoning

	<i>M</i>	<i>SD</i>	$\alpha$	1	2	3	4	5	6	7
1. SciReasoning Concrete	4.16	1.46	0.53							
2. SciReasoning ContextFree	3.37	1.48	0.44	0.55						
3. SciReasoning Scale Full	7.52	2.58	0.67	0.88	0.88					
4. Science knowledge	6.36	1.52	0.35	0.34	0.26	0.34				
5. Anti-scientific Attitudes	43.39	7.57	0.68	-0.18	-0.13	-0.17	-0.23			
6. Numeracy Scale	3.80	1.95	0.68	0.40	0.34	0.42	0.46	-0.28		
7. PISA	3.24	1.14	0.28	0.32	0.22	0.30	0.28	-0.21	0.34	
8. Need For Cognition	19.17	4.39	0.64	0.14	0.15	0.16	0.18	-0.19	0.20	0.12

All correlation coefficients are significant at  $p < 0.001$

The difference between *no explanation* and *no reason* was that in the latter, participants did not write anything or they gave a reason that was obviously off-topic, for example, complained about the length of the test and wrote a joke. In contrast, the former category contained statements where participants were clearly trying to give a good answer but this answer did not contain any *reason why*: “That’s the best option,” “logical,” “common sense,” “I agree,” “my opinion,” or just a restatement of the question, e.g. “The teacher can evaluate their general knowledge”. The rationale behind creating this category and not grouping everything into “incorrect” was that we presumed that inability to give an explicit reason could still signal an *implicit* understanding of the problem. This was also indirectly supported by the fact that this category had on average more correct answers but generally lower confidence ratings than the “incorrect” answers. Note that the correctness or incorrectness only pertains to the reasons, not to the judgments—some correct *Agree/Disagree* judgements had an incorrect explanation and vice versa.

This step was concluded by looking at the judgments associated with ambiguous reasons, which could be correct if the judgment was correct and incorrect if the judgment was incorrect—here, we had to consider the participants’ *Agree/Disagree* judgments. Finally, we grouped all the reasons per item and per category and made another pass through each category, re-categorizing items that did not fit, to correct for any errors and a potential shift in the categories’ boundaries during the process of scoring.

Overall, the number of explicit, well-explained reasons participants gave in support of their judgments was limited and we also accepted reasons that were not well articulated but contained an implicitly correct reason. To check for consistency in participants’ answers, we also look at the number of participants who gave correct reasons for both versions of each item (Figure S.3).

## 5 Discussion

In general, the performance of Slovak participants is consistent with the average results of other tested populations, once again confirming that the ability of the general population to engage in scientific reasoning is fairly low. Given that 50% is the chance level on the test, the average performance of 63% is not particularly high. This is even more apparent on the context-free version of the test—here, the average score was only 56%. Overall, there were marked differences in the difficulty of the items in general—when averaged over the concrete and context-free versions, the percentage correct ranged from 51 to 74%. This suggests that people seem to use different strategies when solving scientific problems embedded in context than when dealing with problems where context cannot facilitate their reasoning. In fact, less than 2% of people got all 12 items correctly.

There were also relatively few items for which participants gave both a correct judgment and a correct reason. Figure S.4 shows that the percentage of correct judgments with correct reasons ranges from 16 to 63% (median 43%). It is evident that for some items (causation and construct validity), people were able to correctly explain their answer less than half of the time. The rest of correct judgments were accompanied by either wrong reasons or no valid reasons at all (e.g., participants restated the conclusion or appealed to “common sense”), or were simply guesses (“don’t know”).

There are several other general results speaking to the validity of the modified SRS scale. As predicted, performance on the adjusted scale showed moderate positive correlations with

people's ability to reason about probabilities and to suppress an intuitive answer when a more reflective one is appropriate (numeracy). There were also positive correlations with knowledge of scientific facts (i.e., understanding some basic facts about the natural world) that were included in our study as one of the three key components of scientific literacy (Miller, 1983, 1998). The fact that this correlation is only in the moderate range (0.34) lends further support to the idea that knowing scientific facts does not automatically imply that one understands the process of acquiring scientific knowledge or is able to evaluate evidence to reach a valid conclusion (e.g., Kuhn, 1993). Also, as predicted, we found that the less people liked to solve complex problems and engage in difficult cognitive endeavors, as measured by the need for cognition scale (Cacioppo & Petty, 1982; Epstein et al., 1996), the worse they performed on the scientific reasoning test. The only negative correlation SRS showed was with anti-scientific attitudes—again, as expected, the less strongly people believe in the scientific approach as a reliable source of knowledge and progress, the lower they score on measures of scientific reasoning. Even though the main focus of our analysis was on scientific reasoning, it is beneficial to closely examine the rather overlooked component of scientific literacy, i.e. attitudes toward science and understanding how science works. Besides scientific reasoning, anti-scientific attitudes correlated negatively with science knowledge, numeracy, and PISA tasks. While our results are correlational and thus it is impossible to distinguish whether poorer knowledge and performance lead people to downplay or reject the importance of science as a way to protect their self-esteem, or whether distrust in science results in having less knowledge and ability to solve real-world problems, studies found that experience with science also improved people's attitudes toward science (Malinowski & Fortner, 2010; Tasdemir et al., 2012). The role of anti-scientific attitudes in accepting pseudoscience and other unfounded beliefs that can have real-life consequences on political and health choices remains understudied. In fact, the importance of trust in science and understanding the importance of science could be one of the key components of scientific literacy. Overall, this pattern of correlations, together with the meticulous process of translating and adjusting the original SRS to the Slovak population, give us confidence that the scale is a valid measure of scientific reasoning.

While the above measures corroborate the test's construct validity and contribute to our aim to explore the components of scientific literacy in Slovak adult population, a moderate positive correlation with a subset of tasks from the PISA testing allows us to go a step further and put forward some hypotheses about the Slovak education system. The results indicate that lack of scientific reasoning is related to poorer PISA scores of adults, which is revealing given that the test is intended for 15-year-old pupils. Clearly, scientific reasoning supports solving problems that are not completely "textbook," i.e. predictable and standard. This suggests that improving other components of scientific literacy besides teaching science knowledge might have beneficial effects extending beyond childhood, and educational settings. There is no formal instruction in scientific reasoning or critical thinking in the Slovak middle or high school curriculum apart from a few specialized institutions and it might be this lack of consistent instruction in scientific thinking that underlies Slovak 15-year-olds' generally poor PISA performance<sup>4</sup> (Miklovičová & Valovič, 2019). In general, many schools concentrate on providing students with facts

<sup>4</sup> In 2018, 29.2% of all tested Slovak 15-year olds scored below level 2, considered a baseline age-appropriate level of science literacy to be reached after completing the 10-year compulsory education. At this level, students can build on their knowledge of basic science content and procedures, identify appropriate explanations, and interpret data (Miklovičová & Valovič, 2019). For comparison, this is 7.3% worse than the OECD average, with best-scoring countries having fewer than 10% of students below this level. In all the rounds of PISA testing so far (except for PISA 2003), Slovak students scored below the OECD average (Miklovičová & Valovič, 2019).

rather than thinking skills, probably presuming that these come naturally. However, developmental research suggests that without specific instructions and practice, this is typically not the case (e.g., Zimmerman, 2000).

## 6 Reasoning About Concrete vs. Context-Free Items

One of the novel aspects of this study concerns the difference in scientific reasoning from concrete vs. context-free examples. A possible reason for this difference is that people in the general population do not seem to be able to reason scientifically based on applying some abstract principles, but rather based on the particularities of each individual case. This reflects a frequent finding from the developmental literature about the inability of children to detect “deep structure” of problems and apply it to other domains (Lai, 2011). Our findings suggest that even adults find it hard to recognize the structure of individual problems and transfer the same solution to other domains.

That is, although we technically compared how people solve concrete and context-free versions of the same scientific reasoning problems, we believe that people do not deal with context-free versions of these problems by applying formal rules at the abstract level at which the problems are presented. Rather, people actively “fill in” the abstract terms with a specific content and then proceed to solve the task nested in this context. There are several lines of evidence supporting this conclusion. First of all, there were significant differences between the number of correct agree/disagree judgments for the two item versions on nearly all items, apart from the “control group” one. Most of them were more difficult in their context-free versions, only “random assignment” differed in the opposite direction. This suggests that participants were not able to see beyond the surface structure of the items to detect the underlying structure, but reasoned from the actual example in front of them. This is indirectly corroborated by the fact that there were few people who got correct reasons on both abstract and concrete versions of a particular item—no item reached more than 50%, and 4 of the 6 items reached less than 25% correct.

A good illustration of this difference between reasoning about concrete or context-free examples is the two stories on construct validity. In the concrete story, a teacher covered three different areas of algebra and geometry, and now wants to use the pupils’ performance on a geometry test as a general measure of their mathematical abilities. In the context-free item, the teacher teaches pupils about domain  $X$  which covers subdomains  $A$ ,  $B$ ,  $C$ , and  $D$  and now wants to use  $D$  as a measure of domain’s  $X$  mastery. Even though there are no key differences between these two stories and their similarity is striking, participants make significantly more (false) assumptions about domain  $X$  than they make about mathematical abilities: for example, that domain  $D$  builds on  $A-C$  or that the teacher knows the pupils’ abilities on  $A-C$  so only needs to test  $D$ . These false assumptions then led to incorrect solutions of the context-free item.

Another good example is causality, which was about a very familiar topic in the concrete version (what is the reason that babies are born) and an unfamiliar topic in the context-free version (the relationship between “ability  $X$ ” and “ability  $Y$ ”). There was a large difference in the frequency of correct solutions between these two versions (86% on the concrete items and only 38% on the context-free item) and the qualitative results suggest that people used very different strategies. In the concrete example, they made a correct judgment simply by knowing that children are a result of sexual intercourse between a man and a woman and therefore it is irrelevant how many hospitals there are. Thus, to reach a correct judgment on the concrete version, they did not need to apply any abstract principles about the difference between

correlation and causation; simply basing their answer on their own world knowledge was sufficient. In contrast, there was no such anchor in the context-free version of the item and the judgment then depended on what kind of example participants came up with for themselves. In fact, in the absence of a clear example to rely on, many participants simply resigned on explaining the flaw in the reasoning behind the conclusion but tried to *justify why the conclusion might be correct* (e.g., the more you practice, the better you are).

However, we are not suggesting that there is a categorical difference between an item being context-free or concrete. Rather, participants seem often to rely on other cues than the methodological principle; or, in other words, they rely on heuristics rather than rules. In general, participants struggled to see the underlying “structure” of an item it was related to a topic they thought they were *familiar* with and possibly had strong beliefs about. Such a topic is weight loss, which we touched upon in concrete items 2 and 6 and context-free item 6. In all of these items, many participants could not hold off from reasoning from what they know about weight loss and not from an abstract principle, e.g. about confounding variables. For example, in the concrete version of item 2 on weight loss, the validity threat was confounding variables (a comparison of two groups of people who ate either carrot or a chocolate bar and then either exercised or meditated). As became evident from the qualitative answers, many participants reached a correct solution—that we cannot be confident about the chocolate bar causing weight loss—not by reasoning about confounding variables but simply by presuming that chocolate cannot lead to weight loss.

Coming back to explanations of the difference in reasoning, there is one more possibility which we cannot rule out, and that is that the reason why people performed more poorly on context-free items. As there was no useful context supplied with the items, they had to come up with their own examples and this contributed to higher variability of contexts (= examples) than in the concrete case. It is also likely that participants came up with examples which were salient for them and therefore were in line with their beliefs, which could have amplified the biasing effect even more. While we have no way of checking this assumption in the present study, this is still in line with the fact that participants were not able to reason *principally*, but got “trapped” in concrete examples. If they had recognized the structure of the problems (e.g., “causation vs. correlation” or “control group”), they should still be able to recognize methodological flaws regardless of the concrete example illustrating it.

Although our results convincingly demonstrate that there is a difference in reasoning between concrete and context-free problems, we are aware of the limitations of our study. Despite trying to keep the level of detail as matched as possible for the concrete and context-free versions of each item, these are not completely identical. To overcome this limitation, the next steps in order to investigate the difference between abstract and concrete examples would be to create a counter-balanced set of items where the background story and the number of concrete details are carefully controlled in a between-subject design.

## 7 Conclusion

In this paper, we examined the scientific literacy of adults in Slovakia with a modified scientific reasoning scale by Drummond and Fischhoff (2017) for the Slovak population and validated it using several other related scales. In doing so, we have indirectly shown that the “missing ingredient” behind our country’s 15-year-old pupils’ poor performance on a worldwide PISA test could be their lack of scientific reasoning skills—and not lack of factual knowledge—as SRS scores were correlated to our participants’ success.

Furthermore, we verified the assumption that people reason from the bottom up, i.e. by predominantly relying on heuristics based on what they already know or believe about the topic. Our results showed that scientific reasoning depends on the context in which people reason. Through varying the extent to which people could rely on situational cues in their reasoning, we have shown that one of the main drivers of poor performance on problems that require scientific reasoning could be their reliance on concrete examples at hand. While this might still lead to a correct solution, it is more prone to error because the reasoning might be biased by a particularly strong cue in the situation at hand or by the participants' prior beliefs (such as in Greenhoot et al., 2004). It seems that without proper training in scientific reasoning, people simply use their general reasoning abilities and get "stuck" on the particularities of each case, using heuristics that can lead them astray. Knowing scientific reasoning principles could give them a structure to be applied to problems that seem different but are in fact structurally the same, and escape from this trap of their own knowledge, beliefs, and assumptions about each particular case. There is no way to test this hypothesis causally on our data as we do not have a group formally trained in scientific reasoning for comparison, but our study brings about indirect evidence pertaining to it.

The education system in Slovakia does not put emphasis on teaching students about how to reason scientifically, how to critically evaluate evidence, or how to check whether all variables are properly controlled. Rather, it places a lot of value on learning facts, i.e. the conclusion of scientific studies, and not on understanding how a certain factual conclusion has been reached and that it has certain limits. In the developmental literature, one of the persistent findings is that children cannot very efficiently abstract the principles behind scientific thinking without getting sufficient training in identifying them explicitly (e.g., Kuhn, 1993). This has also been confirmed for college-level education in the domain of critical thinking (of which scientific reasoning is a subset) (e.g., McLaughlin & McGill, 2017). It seems that the same holds for the adult population in our sample. The different levels of performance on the concrete and context-free items suggest that when teaching the principles of science, it would be beneficial to explain methodological principles on a variety of examples and domains. This could help students to escape the "trap" of reasoning from the particularities of each individual case.

What are the implications of our findings for research on scientific reasoning? First of all, we have gained an important insight into how people who are "naive" to the scientific method solve problems that require scientific reasoning. Thanks to prompting participants for a reason behind their answer, we have learnt that a "correct" answer does not automatically imply understanding, since it is often possible to base one's answer on a real-life example of an intuitive judgment that leads to checking the correct answer without actually understanding the scientific principle. We believe that future measurement tools should either include open-ended answers or ask participants to justify their answer choices. It is also important to make sure the wording of the problems does not lead the participant toward a solution, for example always making the correct option also the intuitively logical one, such as in the example with national park surface and number of endangered species. If the story sounds intuitively incorrect, people seem to be more willing to dig deeper into the structure of the problem, perhaps being more prone to find counterexamples to their intuitions and discover the actual methodological principle. Based on these observations, the wording of the items can make a significant difference to the overall scores and should be carefully piloted for each new population (e.g., children) or language version. Likewise, we recommend to include more than one item for each scientific construct, to increase the generalizability of the results.

The results of our study speak to a growing set of voices in psychology (e.g., Kuhn, 2000) and pedagogy (Moore, 2018) who recognize that teaching scientific thinking is a skill invaluable in a world flooded with information, and that at the application end, we are still failing to do that efficiently.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s11191-021-00207-0>.

**Funding** This study was funded by the Slovak Research and Development Agency by grant scheme no. APVV-16-0153 and by Scientific Grant Agency of the Ministry of Education, Science, Research, and Sport of the Slovak Republic and the Slovak Academy of Sciences no. 2/0053/21. It was also a part of the Social Analysis of Slovakia research project funded by the Institute of Strategic Analyses at the Slovak Academy of Sciences.

#### Declarations

**Conflict of Interest** The authors declare that they have no conflict of interest.

## References

- Allum, N., Sturgis, P., Tabourazi, D., & Brunton-Smith, I. (2008). Science knowledge and attitudes across cultures : A meta-analysis. *Public Understanding of Science*, 17, 35–54. <https://doi.org/10.1077/0963662506070159>.
- Anderson, R. D. (2002). Reforming science teaching: What research says about inquiry. *Journal of Science Teacher Education*, 13(1), 1–12. <https://doi.org/10.1023/A:1015171124982>. Springer Netherlands.
- Bašňáková, J., & Čavojská, V. (2019). Are scientists “scientifically literate”? What is scientific literacy and how to measure it. *Sociálne Procesy a Osobnosť* 2018.
- Bialek, M., & Pennycook, G. (2018). The cognitive reflection test is robust to multiple exposures. *Behavior Research Methods*, 50(5), 1953–1959. <https://doi.org/10.3758/s13428-017-0963-x>.
- Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, 42(1), 116–131. <https://doi.org/10.1037/0022-3514.42.1.116>.
- Čavojská, V., & Jurkovič, M. (2017). Intuition and irrationality. In M. Grežo & M. Sedlár (Eds.), *Sociálne procesy a osobnosť 2016* (pp. 77–83). Ústav experimentálnej psychológie, CSPV SAV. <http://www.spao.eu/pastevents.php>.
- Cheng, P. W., Holyoak, K. J., Nisbett, R. E., & Oliver, L. M. (1986). Pragmatic versus syntactic approaches to training deductive reasoning. *Cognitive Psychology*, 18(3), 293–328. [https://doi.org/10.1016/0010-0285\(86\)90002-2](https://doi.org/10.1016/0010-0285(86)90002-2).
- Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 163–228). Oxford University Press.
- Cox, J. R., & Griggs, R. A. (1982). The effects of experience on performance in Wason’s selection task. *Memory & Cognition*, 10(5), 496–502. <https://doi.org/10.3758/BF03197653>.
- De Neys, W. (2014). Conflict detection, dual processes, and logical intuitions: Some clarifications. *Thinking & Reasoning*, 20(2), 169–187. <https://doi.org/10.1080/13546783.2013.854725>.
- De Neys, W., & Glumicic, T. (2008). Conflict monitoring in dual process theories of thinking. *Cognition*, 106(3), 1248–1299. <https://doi.org/10.1016/j.cognition.2007.06.002>.
- Drummond, C., & Fischhoff, B. (2017). Development and validation of the Scientific Reasoning Scale. *Journal of Behavioral Decision Making*, 30(1), 26–38. <https://doi.org/10.1002/bdm.1906>.
- Dube, C., Rotello, C. M., & Heit, E. (2010). Assessing the belief bias effect with ROCs: It’s a response bias effect. *Psychological Review*, 117(3), 831–863. <https://doi.org/10.1037/a0019634>.
- Dudeková, K., & Kostovičová, L. (2015). Oscary, HDP a CRT: Efekt ukotvenia u finančných profesionálov v kontexte doménovej špecifickosti a kognitívnej reflexie. In I. Farkaš, M. Takáč, J. Rybár, & J. Kelemen (Eds.), *Kognícia a umelý život 2015* (pp. 50–56). Univerzita Komenského.
- Eivers, E., & Kennedy, D. (2006). The PISA assessment of scientific literacy. *The Irish Journal of Education / Iris Eireannach an Oideachais*, 37, 101–119. <https://www.jstor.org/stable/30077514?seq=1>. Accessed 23 Nov 2020.
- Elakana, Y. (2000). Science, philosophy of science and science teaching. *Science and Education*, 9, 463–548.

- Epstein, S., Pacini, R., Denes-Raj, V., & Heier, H. (1996). Individual differences in intuitive-experiential and analytical-rational thinking styles. *Journal of Personality and Social Psychology*, 71(2), 390–405. <https://doi.org/10.1037/0022-3514.71.2.390>.
- Fasce, A., & Picó, A. (2019). Science as a vaccine. *Science & Education*, 1–17. <https://doi.org/10.1007/s11191-018-00022-0>.
- Fiddick, L., Cosmides, L., & Tooby, J. (2000). No interpretation without representation: The role of domain-specific representations and inferences in the Wason selection task. *Cognition*, 77(1), 1–79. [https://doi.org/10.1016/s0010-0277\(00\)00085-8](https://doi.org/10.1016/s0010-0277(00)00085-8).
- Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19(4), 25–42. <https://doi.org/10.1257/089533005775196732>.
- Gdovinová, D. (2020). Tretina učiteľov si myslí, že očkovanie je prípravou na čipovanie. Až polovica učiteľov by sa nedala zaočkovať (prieskum) [One third of teachers thinks that vaccination is preparation for global chipping. Half of teachers would refuse vaccination (survey)]. *Denník N*. <https://dennikn.sk/2095185/tretina-ucitelov-si-mysli-ze-ockovanie-je-pripravou-na-cipovanie-az-polovica-ucitelov-by-sa-nedala-zaockovat-prieskum>. Accessed 19 Oct 2020.
- Greenhoot, A. F., Semb, G., Colombo, J., & Schreiber, T. (2004). Prior beliefs and methodological concepts in scientific reasoning. *Applied Cognitive Psychology*, 18(2), 203–221. <https://doi.org/10.1002/acp.959>.
- Griggs, R. A., & Cox, J. R. (1982). The elusive thematic-materials effect in Wason selection task. *British Journal of Psychology*, 73, 407–420.
- Harlen, W. (2005). The assessment of scientific literacy in the OECD/PISA project. In *Research in science education—Past, present, and future* (pp. 49–60). Kluwer Academic Publishers. [https://doi.org/10.1007/0-306-47639-8\\_5](https://doi.org/10.1007/0-306-47639-8_5).
- Hayes, M. T. (2002). Elementary preservice teachers' struggles to define inquiry-based science teaching. *Journal of Science Teacher Education*, 13(2), 147–165. <https://doi.org/10.1023/A:1015169731478>.
- Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence: An essay on the construction of formal operational structures*. Routledge.
- Kahan, D. M., Peters, E., Wittlin, M., Slovic, P., Ouellette, L. L., Braman, D., & Mandel, G. (2012). The polarizing impact of science literacy and numeracy on perceived climate change risks. *Nature Climate Change*, 2(10), 732–735. <https://doi.org/10.1038/nclimate1547>.
- Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80(4), 237–251. <https://doi.org/10.1037/h0034747>.
- Kennedy, B., & Hefferon, M. (2019). *What Americans know about science*. Pew Research Center. <https://www.pewresearch.org/science/2019/03/28/science-knowledge-appendix-detailed-tables>. Accessed 23 Nov 2020.
- Koslowski, B. (1996). *Theory and evidence: The development of scientific reasoning*. Massachusetts Institute of Technology. <https://doi.org/10.1002/oby.20304>.
- Kuhn, D. (1993). Science as argument: Implications for teaching and learning scientific thinking. *Science Education*, 77(3), 319–337. <https://doi.org/10.1002/sce.3730770306>.
- Kuhn, D., & Franklin, S. (2007). *The second decade: What develops (and how)*. Wiley. <https://doi.org/10.1002/9780470147658.chpsy0222>.
- Kuhn, D., Garcia-Mila, M., Zohar, A., Andersen, C., White, S. H., Klahr, D., & Carver, S. M. (1995). Strategies of knowledge acquisition. *Monographs of the Society for Research in Child Development*, 60(4), i. <https://doi.org/10.2307/1166059>.
- Lai, E. R. (2011). Critical thinking: A literature review research report. *Pearson's Research Reports*, 6, 40–41. <http://www.pearsonassessments.com/research>. Accessed 23 Nov 2020.
- Leblebicioglu, G., Metin, D., Capkinoglu, E., Cetin, P. S., Eroglu Dogan, E., & Schwartz, R. (2017). Changes in students' views about nature of scientific inquiry at a science camp. *Science and Education*, 26(7–9), 889–917. <https://doi.org/10.1007/s11191-017-9941-z>.
- Lederman, J. S., Lederman, N. G., Bartos, S. A., Bartels, S. L., Meyer, A. A., & Schwartz, R. S. (2014). Meaningful assessment of learners' understandings about scientific inquiry—The views about scientific inquiry (VASI) questionnaire. *Journal of Research in Science Teaching*, 51(1), 65–83. <https://doi.org/10.1002/tea.21125>.
- Machamer, P. (1998). Philosophy of science: An overview for educators. *Science & Education*, 7, 1–11.
- Malinowski, J., & Fortner, R. W. (2010). The effect of participation in a Stone Laboratory Workshop (A Place-Based Environmental Education Program) on student affect toward science. *The Ohio Journal of Science*, 110(2), 13–17. <https://kb.osu.edu/handle/1811/52774>. Accessed 23 Nov 2020.
- McBride, J. W., Bhatti, M. I., Hannan, M. A., & Feinberg, M. (2004). Using an inquiry approach to teach science to secondary school science teachers. *Physics Education*, 39(5), 434–439. <https://doi.org/10.1088/0031-9120/39/5/007>.

- McLaughlin, A. C., & McGill, A. E. (2017). Explicitly teaching critical thinking skills in a history course. *Science & Education*. <https://doi.org/10.1007/s11191-017-9878-2>.
- Miklovičová, J., & Valovič, J. (2019). *Národní správa PISA 2018*. NÚCEM.
- Miller, J. D. (1983). Scientific literacy: A conceptual and empirical review. *Daedalus*, 112(2), 29–48. <https://doi.org/10.2307/20024852>.
- Miller, J. D. (1998). The measurement of civic scientific literacy. *Public Understanding of Science*, 7(3), 203–223. <https://doi.org/10.1088/0963-6625/7/3/001>.
- Miller, J. D. (2004). Public understanding of, and attitudes toward, scientific research: What we know and what we need to know. *Public Understanding of Science*, 13(3), 273–294. <https://doi.org/10.1177/0963662504044908>.
- Moore, C. (2018). *Teaching science thinking: Using scientific reasoning in the classroom*. Routledge.
- National Science Board. (2010). Science and engineering indicators 2010. In *Transactions of the American Society of Civil Engineers*. National Science Foundation. <https://doi.org/10.1080/03602458208079655>.
- OECD. (2003). *The PISA 2003 assessment framework—Mathematics, reading, science and problem solving knowledge and skills*.
- Peters, E., Västfjäll, D., Slovic, P., Mertz, C. K., Mazzocco, K., & Dickert, S. (2006). Numeracy and decision making. In *Source: Psychological Science*, 17(5). <http://www.webpages.ttu.edu/jiyang/Joselynsuperlab-paper-2017.pdf>. Accessed 23 Nov 2020.
- Rossi, S., Cassotti, M., Moutier, S., Delcroix, N., & Houdé, O. (2015). Helping reasoners succeed in the Wason selection task: When executive learning discourages heuristic response but does not necessarily encourage logic. *PLoS ONE*, 10(4), e0123024. <https://doi.org/10.1371/journal.pone.0123024>.
- Sagan, C. (1996). *The demon-haunted world: Science as a candle in the dark*. Ballantine Books.
- Schwab, J. J. (1958). The teaching of science as inquiry. *Bulletin of the Atomic Scientists*, 14(9), 374–379. <https://doi.org/10.1080/00963402.1958.11453895>.
- She, H. C., Stacey, K., & Schmidt, W. H. (2018). Science and mathematics literacy: PISA for better school education. In *International Journal of Science and Mathematics Education*, 16(1), pp. 1–5. Springer Netherlands. <https://doi.org/10.1007/s10763-018-9911-1>.
- Šrol, J. (2018). These problems sound familiar to me: Previous exposure, cognitive reflection test, and the moderating role of analytic thinking. *Studia Psychologica*, 60(3), 195–208. <https://doi.org/10.21909/sp.2018.03.762>.
- Stanovich, K. E., & West, R. F. (1998). Cognitive ability and variation in selection task performance. *Thinking and Reasoning*, 4(3), 193–230. <https://doi.org/10.1080/135467898394139>.
- Stanovich, K. E., West, R. F., & Toplak, M. E. (2016). *The rationality quotient*. MIT Press.
- Tasdemir, A., Kartal, T., & Kus, Z. (2012). The use of out-of-the-school learning environments for the formation of scientific attitudes in teacher training programmes. *Procedia - Social and Behavioral Sciences*, 46, 2747–2752. <https://doi.org/10.1016/j.sbspro.2012.05.559>.
- Trefil, J. (2008). *Why science?* Teachers College Press.
- Trippas, D., Handley, S. J., & Verde, M. F. (2013). The SDT model of belief bias: Complexity, time, and cognitive ability mediate the effects of believability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(5), 1393–1402. <https://doi.org/10.1037/a0032398>.
- Trippas, D., Thompson, V. A., & Handley, S. J. (2017). When fast logic meets slow belief: Evidence for a parallel-processing model of belief bias. *Memory & Cognition*, 45(4), 539–552. <https://doi.org/10.3758/s13421-016-0680-1>.
- Wason, P. C. (1968). Reasoning about a rule. *The Quarterly Journal of Experimental Psychology*, 20(3), 273–281. <https://doi.org/10.1080/14640746808400161>.
- Wilkening, F., & Sodian, B. (2005). Scientific reasoning in young children: Introduction. *Swiss Journal of Psychology*, 64(3), 137–139. <https://doi.org/10.1024/1421-0185.64.3.137>.
- Zhu, L., & Gigerenzer, G. (2006). Children can solve Bayesian problems: The role of representation in mental computation. *Cognition*, 98(3), 287–308. <https://doi.org/10.1016/J.COgnITION.2004.12.003>.
- Zimmerman, C. (2000). The development of scientific reasoning skills. *Developmental Review*, 20, 99–149. <https://doi.org/10.1006/drev.1999.0497>.
- Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review*, 27(2), 172–223. <https://doi.org/10.1016/j.dr.2006.12.001>.