The integration of probability-based arguments in risk-related contexts

Theresa Büchter[1], Karin Binder[2] and Andreas Eichler[1]
[1]University of Kassel, Germany
[2]Paderborn University, Germany
tbuechter@mathematik.uni-kassel.de

Probabilistic reasoning is key for decision making, especially in risk-related contexts. For example, erroneous HIV self-test results pose risks that must be evaluated when considering public approval. This requires incorporating probabilities (e.g., for not being infected even though a positive test result is given) into the decision-making process. We examine how upper secondary school students use such probability-based arguments in risk-related contexts before and after an intervention on conditional probabilities. A qualitative content analysis classifies their arguments as (i) mathematical, (ii) context-related, (iii) transitional (between mathematical and context-related), (iv) affective, or (v) based on anticipated personal experience. A central result of the analysis is that after the intervention, students used mathematical arguments more frequently than prior to the intervention on conditional probabilities.

INTRODUCTION AND THEORETICAL BACKGROUND

Probabilistic reasoning in risk-related contexts

Risk-related contexts are integral in everybody's life (Schenk et al., 2019). This can include examples of individual risks such as rushing onto a metro train even though there was no time left to buy a ticket (Martignon & Hoffrage, 2019) but also situations with an impact on a larger group of people such as decisions regarding social distancing during the Covid pandemic (Harman et al., 2021). In this contribution, we follow Aven (2009) in understanding risk as an event with a possibly negative outcome and its corresponding probability. Thus, as Aven (2023) points out, probabilities are measures which quantify risk. Therefore, it is widely accepted that probabilistic reasoning is required for decision-making in risk-related contexts (Martignon & Hoffrage, 2019). Further, for risk literacy, i.e., the adequate understanding and evaluation of risks for decision making (Aven, 2023), probabilistic and statistical competencies are (among others) considered a necessary precondition (Eichler & Vogel, 2015; Martignon & Monti, 2010; Radakovic, 2015; Hansen & Hammann, 2017; Aven & van Kessenich, 2020). Beyond the mathematical aspects, risk literacy according to Martignon & Hoffrage (2019) also includes (1) to identify risks, (2) weigh risks and benefits against each other and (3) decide and act in risk-related contexts. A specific example for decision-making in a risk-related context is the question of whether or not HIV self-tests should be approved for public use. Mathematical considerations for evaluating the risks in this context include, for example, estimations of the likelihood to be not infected given a positive test result.

Bayesian reasoning as a specific type of probabilistic reasoning in risk-related contexts

The specific example on HIV self-tests relates to a situation with two binary events (i.e., positive/negative test result and infection/no infection), and hence a so-called Bayesian situation (Zhu & Gigerenzer, 2006). Reasoning in these situations is understood as Bayesian reasoning. A typical task used to study Bayesian reasoning in such situations is given in Table 1.

Table 1. Bayesian reasoning task in the situation on HIV self-tests.

Given information	The probability for a randomly chosen person in Germany to be infected with HIV infection is 0.1%. (*base rate*) If a randomly chosen person is infected with HIV, the probability is 99.5% that this person actually receives a positive test result. (*sensitivity*) If a randomly chosen person is not infected with HIV, the probability is 0.2% that this person erroneously receives a positive test result. (*false-positive rate*)
Question	What is the probability that a randomly chosen person is actually not infected given a positive test result?
Answer	≈ 66.7%

In: Birk, L., Loth, G., Jotzo, L., Binder, K., & Frischemeier, D. (Eds.) (2026). *Statistics and Data Science Education in STEAM. Proceedings of the Satellite Conference of the International Association for Statistical Education (IASE), September/October 2025, Münster, Germany*. © ISI/IASE.

As evident in Table 1, the conditional probability that a randomly chosen person is *not* infected with HIV given a positive test result is 66.7%. This is often perceived as surprising given the high sensitivity and low false-positive rate (Gigerenzer et al., 2007). In the past, when HIV infections had more serious consequences than today, false-positive test results were even correlated to suicides (Stine, 1996). Still, even today, such a false-positive test result certainly leads to unnecessary worrying, particularly if people are unaware that a positive test result does not necessarily indicate an infection. Therefore, when considering whether HIV *self*-tests should be approved for public use, it is essential to keep in mind, among other things, the probability from Table 1. Moreover, in this context it is concerning that on average, Bayesian reasoning tasks as in Table 1 can only be solved by about 5% of the people (McDowell & Jacobs, 2017) and the probability is often underestimated (Prinz et al., 2015). This could increase unnecessary worrying based on false-positive test results.

Performance in Bayesian reasoning tasks can be improved through a presentation of the given information in so-called natural frequencies (e.g., 995 in 1,000 infected people correctly receive a positive test result) instead of the given probabilities in Table 1 (Gigerenzer & Hoffrage, 1995; McDowell & Jacobs, 2017). Moreover, representing the Bayesian reasoning task with a natural frequency-based visualization can further increase performance (Binder et al., 2015) with the 2×2 table outperforming other natural frequency-based visualizations in previous studies with German students (Binder et al., 2020; Böcherer-Linder & Eichler, 2019). The supportive effect of natural frequency-based visualizations has also been observed for instruction (Steib et al., 2025). Additionally, simulations may support a better understanding of conditional probabilities in the context of Bayesian reasoning (Budgett & Pfannkuch, 2018; Engel & Martignon, 2015; Schulze & Hertwig, 2022).

In previous intervention studies on Bayesian reasoning, the focus was on studying effects on Bayesian reasoning tasks such as in Table 1 (Sedlmeier & Gigerenzer, 2001; Feufel et al., 2023; Kim et al., 2024). In addition to these previous studies, it would be interesting to analyse if the effects of an intervention on Bayesian reasoning also affect the arguments in the reasoning within these risk-related contexts. To the best of our knowledge, this has not yet been studied.

Arguments in risk-related Bayesian situations

Arguments in risk-related contexts have, to the best of our knowledge, not yet been analysed in Bayesian situations. However, prior research from science and mathematics education did study how students reason in other risk-related contexts. Thereby, Pratt et al. (2011) identified four different types of arguments used in risk-related contexts which illustrate the complexity of decision-making in risk-related contexts: arguments have been observed to be based on (1) personal experiences (Christenson et al., 2011; Kolsto, 2006; Pratt et al., 2011), (2) affective responses (Christenson et al., 2011; Pratt et al., 2011), (3) understanding the problem context (Pratt et al., 2011; Prodromou, 2015) and (4) mathematical considerations of uncertainty and probability (Christenson et al., 2011; Pratt et al., 2011). This complexity of arguments involved in real-world risk-related contexts can be a barrier to understanding the statistical concepts of a risk-related context according to Pratt et al. (2011). Nevertheless, Pratt et al. (2011) argue for finding settings and tools with which statistical concepts can be learned *within* real-world risk-related contexts.

Research question

In the project siMINT-Risk, our goal is to develop an intervention on conditional probabilities within a risk-related context and to study secondary school students' reasoning in the context of HIV self-tests and ask the following questions:

- RQ1: Which categories of arguments do students use when reasoning about the approval of HIV-self-tests?
- RQ2: Do the categories of arguments in the reasoning about the approval of HIV-self-tests change through an intervention on conditional probabilities?

METHOD

In our study, secondary school students from three different mathematics courses in grade 13 (final grade of secondary school) voluntarily participated in the study as part of their regular mathematics lessons. Two of the three courses were advanced level, the other course was regular level.

The students carried out a pre-test prior to a 3-hour intervention on conditional probabilities. On the same day as the third lesson of the intervention, they also carried out a post-test. Prior to participation in the study, the students were not familiar with conditional probabilities. The tests were administered digitally and the use of calculators was allowed. In our analysis we only included students who were present in the pre-test, the intervention and the post-test. This resulted in a sample of $N = 42$ students.

Intervention

The 3-hour intervention began with exploring the problem-context including an explanatory video about HIV and the opportunity to use HIV self-tests (about 10 minutes). Afterwards, the students in class identified potential risks when carrying out such HIV self-tests and the teacher directed the discussion, making sure to also identify the following two risks (about 10 minutes):

- Probability for actually *not* being infected given a positive test result
- Probability for actually being infected given a negative test result

Afterwards, students explored these probabilities through a simulation (about 60 minutes). In the simulation the students could draw samples, vary the base rate, sensitivity and false-positive rate. The frequencies of the simulated samples as well as the estimations for both probabilities based on the gathered data were displayed to the students (see following link for the original German version of the simulation: http://bayesianreasoning.de/simint_Sim). The observations of the students were discussed in class. On the next day with a mathematics lessons, the students then received instruction on how to calculate the conditional probabilities, which had previously been simulated. For that we have designed a worked-example (Renkl, 2014) making use of the strategies of natural frequency-based 2×2 tables. Figure 1 displays the completely filled out and natural frequency-based 2×2 table which represents the given information of the Bayesian reasoning task in Table 1. The structure of the worked-example is similar to the ones described in Büchter et al. (2022) and Steib et al. (2025) and can be accessed in the original German version (http://bayesianreasoning.de/simint_WE). The teachers received detailed instruction on how to use the material (i.e., worked-example and simulation) to ensure comparability between the three courses.

people	infected	not infected	total
positive test result	995	1,998	2,993
negative test result	5	997,002	997,007
total	1,000	999,000	1,000,000

Figure 1. Natural frequency-based 2×2 table representing the given information in Table 1.

Materials

In each test (i.e., pre-, post- and follow-up-test), students first answered Bayesian reasoning tasks (as in Table 1) and afterwards reasoned about whether or not HIV-tests should be admitted to the public as self-tests (Task: “HIV self-tests can be purchased without restriction at a drugstore or ordered online. Take a position on this unrestricted availability of HIV self-tests. Weigh the possible advantages and disadvantages against each other”). Additionally, they answered typical questions about conditional probabilities (e.g., their calculation). In order to ensure that working on the tests was finished within the time of the lessons, the questions were time-limited (pre-test: one out of 3 questions in 8 minutes; post-test: 3 minutes for the question).

Analysis of the answers

The answers to the Bayesian reasoning tasks (as in Table 1) were coded as correct, if the exact value was provided. For analyzing the arguments in the reasoning about the admittance of HIV-self-tests, we used the categories from Pratt et al. (2011) as a structure but added the subcategories inductively through a qualitative content analysis. The resulting categories are displayed in Figure 2.

Beyond these inductively derived categories of arguments, we also coded for each argument whether it was expressed as an advantage or a disadvantage. The coding was carried out by one of the authors.

Category	Sub-category covers arguments about…	Code
mathematical	Probability to be infected given a positive test result	11
	Probability not to be infected given a negative test result	12
	Influence of the base rate on the probabilities	13
	Probabilities of the test parameters (i.e., sensitivity & specificity)	14
	Joint probability for false-positive test results	15
	Joint probability for false-negative test results	16
transitional (between mathematical and context-related)	Possibility of false-positive test results	21
	Possibility of false-negative test results	22
	The potential of test results (e.g., as a hint vs. evidence for an infection)	23
understanding the problem-context	Severity of a possible infection	31
	Characteristics of the test (e.g., potential cross reactions with other antibodies, diagnostic window for testing, etc.)	32
	Biologically positive consequences of testing (e.g., earlier detection of an infection, earlier treatment, fewer dissemination of the virus)	33
	Biologically negative consequences of testing (e.g., later detection of an infection, later treatment, higher dissemination of the virus)	34
affective	Positive affective consequences (e.g., higher feeling of security, more anonymity, reduced shame)	41
	Negative affective consequences (e.g., unnecessary worrying, incorrect feeling of security, shame)	42
anticipated personal experiences	Anticipated necessity for clarification (e.g., lack of knowledge about infection or test results, need/possibility for follow-up doctors appointment)	51
	Anticipated procedure for performing such a self-test (e.g., no doctors-appointment, no waiting for an appointment)	52
	Anticipated organizational advantages (e.g., it's easy, anyone can buy a test, results are known faster, costs are low, uncomplicated accessibility)	53
	Anticipated organizational disadvantages (e.g., high costs, complicated accessibility)	54
	Anticipated consequences of non-regulated access to self-tests (e.g., more people testing themselves, lower threshold)	55
	Anticipated consequences of self-testing (e.g., errors in performing the test, lower/higher cautionary measures)	56

Figure 2. Inductively derived categories for arguments in the reasoning about the admittance of HIV-self-tests.

RESULTS

Use of different arguments (RQ1)

Figure 3 displays three examples of students reasoning about the admittance of HIV-self-tests and their corresponding categorization into the categories displayed in Figure 1. The first example is an answer from the pre-test, while the second and third answer are from the post-test. Beyond the coded categories which are already highlighted in Figure 2, it can be seen, that only in the second answer it is clearly stated that the advantages outweigh the disadvantages which was hence coded as a decision in favor of the admittance of HIV self-tests. The other two answers did not come to a decision about the admittance of HIV self-tests.

Student ID	Student's answer	Coding
33 (pre-test)	Advantages: - quickly accessible Disadvantages: - some people probably don't know how to deal with a positive test result, so it would be better if they were with a doctor or someone similar - incorrect use can lead to false results and therefore to unnecessary anxiety	 53 51 56 / 21&22 / 42
28 (post-test)	Advantages: clarity about HIV infection worries can be resolved prevents people with HIV from spreading the virus further Disadvantages: false results can lead to unsubstantiated fears or a false sense of security, which might not stop potential transmission In my opinion, the advantages outweigh the disadvantages.	 23 41 33 21&22 / 42 34
14 (post-test)	Positive aspects: - the tests are relatively reliable when they show a negative result - anyone can freely buy these self-tests and, in case of a positive result, go to a doctor to have it confirmed Negative aspects: - the tests are not very reliable when they show a positive result - if someone is actually HIV-positive but receives a negative test result, they may have no concerns	 12 53 / 51 11 42

Figure 3. Examples of students' reasoning about the admittance of HIV-self-tests and their corresponding categorization into the categories displayed in Figure 1.

Change in the use of arguments (RQ2)

First, in order to check whether the intervention did in fact improve Bayesian reasoning, we compare the performance in the Bayesian reasoning tasks from pre- and post-test. From our perspective this is a precondition for analyzing a change in the students' arguments (RQ2), particularly with regard to the use of mathematical arguments. None of the students was able to solve any of the Bayesian reasoning tasks prior to the intervention (as expected, as the students were not familiar with conditional probabilities). Afterwards performance in the Bayesian reasoning tasks was 51%. Thus, it can be assumed that the intervention was successful for increasing understanding of conditional probabilities.

Figure 4 displays the use of different arguments among the advantages and disadvantages expressed. In Figure 4 (but also in Figure 3), from a descriptive perspective it can be seen that in the pre-test hardly any mathematical arguments (i.e., relating to probabilities in the situation) were used in the students' answers. Yet, after the intervention, mathematical arguments were used among advantages as well as disadvantages. Thus, students seem to have been more sensitive to include probability-related arguments after instruction, yet non-mathematical arguments still outweigh probability-based arguments after instruction.

For inferential analysis regarding RQ2, we carried out a Cochran's Q test to assess whether the expression any of the five categories varied from pre- to post-test for each of the five categories separately. This test is chosen based on the binary outcome variable (category is expressed or not) and the repeated measures (McCrum-Gardener, 2008). The Cochran Q test shows a significant change in the use of mathematical (Q=14.0, df=1, p<0.01) and affective arguments (Q=10.9, df=1, p<0.01) from pre- to post-test but not for the other categories. This implies that the frequency for expressing mathematical or affective arguments varied statistically significantly between the pre- and post-test.

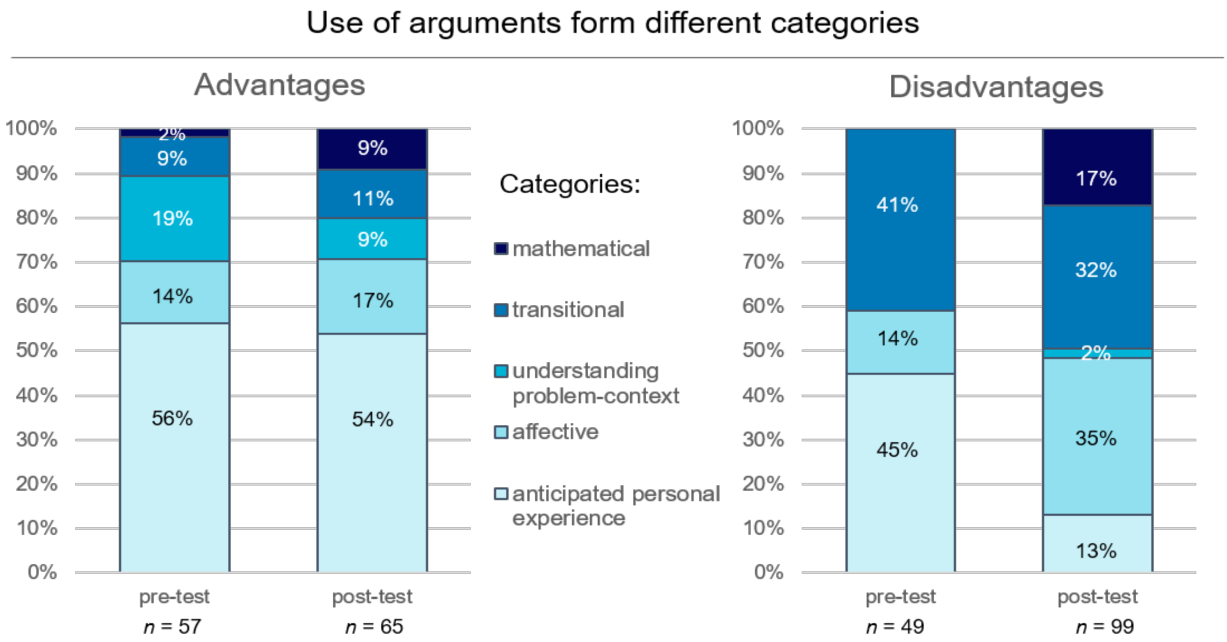

Figure 4. Comparison of the use of different arguments as advantages and disadvantages in the pre- and post-test.

DISCUSSION AND IMPLICATIONS

Pratt et al. (2011) have previously analyzed students' arguments in risk-related contexts and observed arguments based on (i) mathematical reasoning, (ii) understanding the problem context, (iii) affective reactions and (iv) personal experiences. The results of our study first show that it was possible to use these categories by Pratt et al. (2011) as a structure for our analysis of the students' arguments. Based on this structure, we inductively derived sub-categories tailored to the specific problem context of the approval of HIV self-tests.

The aim of the presented article was to analyze these arguments in a risk-related Bayesian situation and study how an intervention tailored to Bayesian reasoning affected the students' reasoning. The intervention on conditional probabilities affected two categories of arguments: mathematical and affective arguments. Other categories of arguments (i.e., for understanding the problem-context and anticipated personal experience) were not significantly affected by the intervention. The observed change in the use of arguments, particularly in the use of *mathematical* arguments, is a central result. The change in mathematical arguments likely goes back to the learning with regard to the relevant mathematical concepts (e.g., conditional probabilities) as part of the instruction between pre- and post-test. It is encouraging that students seem to have applied these acquired mathematical skills to decision making, which suggests that mathematics lessons in school can change, e.g., the critical stance as a dispositional element of probability literacy (Gal, 2005) or their recognition for the need of data (Wild & Pfannkuch, 1995).

Our results are limited as we are lacking a control group. Thus, it cannot be excluded that the observed effects go back to repeated testing. The presented results from the three classes are the pilot study for the main study in the project siMINT-Risk (http://bayesianreasoning.de/en/br_simint_en.html). In the main study with a larger sample size and three different intervention groups we will be able to address future questions such as: does a change in arguments depend on the intervention? How do arguments differ between those with correct and incorrect answers in the Bayesian reasoning tasks? Further, a more detailed analysis of the structure of arguments could be interesting by analyzing how the different arguments are interrelated (e.g., are certain mathematical arguments more often linked to specific affective arguments than others?).

FUNDING

The project receives funding from the Federal Ministry of Education and Research under the grant 16MF1070C – siMINT.

REFERENCES

Aven, T., & Renn, O. (2009). On risk defined as an event where the outcome is uncertain. *Journal of Risk Research*, *12*(1), 1–11. https://doi.org/10.1080/13669870802488883.

Aven, T. (2023). Risk literacy: Foundational issues and its connection to risk science. *Risk Analysis: An Official Publication of the Society for Risk Analysis.* Advance online publication. https://doi.org/10.1111/risa.14223.

Aven, T., & van Kessenich, A. M. (2020). Teaching children and youths about risk and risk analysis: what are the goals and the risk analytical foundation? *Journal of Risk Research*, *23*(5), 557–570. https://doi.org/10.1080/13669877.2018.1547785.

Binder, K., Krauss, S., & Bruckmaier, G. (2015). Effects of visualizing statistical information–an empirical study on tree diagrams and 2×2 tables. *Frontiers in Psychology, 6*, 1186. https://doi.org/10.3389/fpsyg.2015.01186.

Binder, K., Krauss, S., & Wiesner, P. (2020). A New Visualization for Probabilistic Situations Containing Two Binary Events: The Frequency Net. *Frontiers in Psychology*, *11*, 750. https://doi.org/10.3389/fpsyg.2020.00750.

Böcherer-Linder, K., & Eichler, A. (2019). How to improve performance in Bayesian inference tasks: a comparison of five visualizations. *Frontiers in Psychology, 10*, 267. https://doi.org/10.3389/fpsyg.2019.00267.

Büchter, T., Eichler, A., Steib, N., Binder, K., Böcherer-Linder, K., Krauss, S., & Vogel, M. (2022). How to train novices in Bayesian reasoning. *Mathematics*, *10*(9), 1558. https://doi.org/10.3390/math10091558.

Budgett, S., & Pfannkuch, M. (2019). Visualizing chance: tackling conditional probability misconceptions. *Topics and trends in current statistics education research: International perspectives*, 3-25. https://doi.org/10.1007/978-3-030-03472-6_1.

Christenson, N., Chang Rundgren, S. N., & Höglund, H. O. (2012). Using the SEE-SEP model to analyze upper secondary students' use of supporting reasons in arguing socioscientific issues. *Journal of Science Education and Technology*, *21*, 342-352. https://doi.org/10.1007/s10956-011-9328-x.

Díaz, C., & Batanero, C. (2009). University students' knowledge and biases in conditional probability reasoning. *International Electronic Journal of Mathematics Education, 4*(3), 131-162. https://doi.org/10.29333/iejme/234.

Engel, J., & Martignon, L. (2015). Dynamisch-interaktive Visualisierung elementarer Konzepte zu Daten und Wahrscheinlichkeiten. *Lernen und Lernstörungen, 4*, 139–145. https://doi.org/10.1024/2235-0977/a000101.

Feufel, M. A., Keller, N., Kendel, F., & Spies, C. D. (2023). Boosting for insight and/or boosting for agency? How to maximize accurate test interpretation with natural frequencies. *BMC Medical Education, 23*(1), 75. https://doi.org/10.1186/s12909-023-04025-6.

Gal, I. (2005). Towards "probability literacy" for all citizens: Building blocks and instructional dilemmas. In G. A. Jones (Ed.), *Exploring Probability in School: Challenges for Teaching and Learning* (pp.39–63). Springer. https://doi.org/10.1007/0-387-24530-8_3.

Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., & Woloshin, S. (2007). Helping doctors and patients make sense of health statistics. *Psychological science in the public interest, 8*(2), 53-96. https://doi.org/10.1111/j.1539-6053.2008.00033.x.

Hansen, J., & Hammann, M. (2017). Risk in Science Instruction. *Science & Education*, *26*(7-9), 749–775. https://doi.org/10.1007/s11191-017-9923-1.

Harman, J. L., Weinhardt, J. M., Beck, J. W., & Mai, I. (2021). Interpreting time-series COVID data: Reasoning biases, risk perception, and support for public health measures. *Scientific Reports*, *11*(1), 15585. https://doi.org/10.1038/s41598-021-95134-z.

Kim, S., Kim, S., Choi, Y.-J., & Do, Y. K. (2024). Natural frequency tree- versus conditional probability formula-based training for medical students' estimation of screening test predictive values: A randomized controlled trial. *BMC Medical Education, 24*(1), 1207. https://doi.org/10.1186/s12909-024-06209-0.

Kolstø, S. D. (2006). Patterns in students' argumentation confronted with a risk-focused socio-scientific issue. *International Journal of Science Education*, *28*(14), 1689-1716. https://doi.org/10.1080/09500690600560878.

Martignon, L., & Hoffrage, U. (2019). *Wer wagt, gewinnt? Wie Sie die Risikokompetenz von Kindern und Jugendlichen fördern können* (1. Auflage). Hogrefe.

Martignon, L., & Monti, M. (2010). Conditions for risk assessment as a topic for probabilistic education. In *Proceedings of the Eighth International Conference on Teaching Statistics. Lublijana, Slovenia: International Statistical Institute and International Association for Statistical Education.*

McCrum-Gardner, E. (2008). Which is the correct statistical test to use? *British Journal of Oral and Maxillofacial Surgery, 46*(1), 38–41. https://doi.org/10.1016/j.bjoms.2007.09.002.

McDowell, M., & Jacobs, P. (2017). Meta-analysis of the effect of natural frequencies on Bayesian reasoning. *Psychological Bulletin, 143*(12), 1273. https://doi.org/10.1037/bul0000126.

Operskalski, J. T., & Barbey, A. K. (2016). Risk literacy in medical decision-making. *Science, 352*(6284), 413–414. https://doi.org/10.1126/science.aaf7966.

Pratt, D., Ainley, J., Kent, P., Levinson, R., Yogui, C., & Kapadia, R. (2011). Role of context in risk-based reasoning. *Mathematical Thinking and Learning*, *13*(4), 322-345. https://doi.org/10.1080/10986065.2011.608346.

Prodromou, T. (2015). Adults' perceptions of risk in the Big Data Era. *The Mathematics Enthusiast*, *12*(1), 364-377. https://doi.org/10.54870/1551-3440.1353.

Radakovic, N. (2015). Pedagogy of risk: Why and how should we teach risk in high school math classes? *The Mathematics Enthusiast. 12*(1), 307–329. https://doi.org/10.54870/1551-3440.1350.

Renkl, A. (2014). The Worked Examples Principle in Multimedia Learning. In R. E. Mayer (Ed.) *The Cambridge Handbook of Multimedia Learning*, Cambridge University Press: Cambridge, UK, 391–412. https://psycnet.apa.org/doi/10.1017/CBO9781139547369.020.

Schenk, L., Hamza, K. M., Enghag, M., Lundegård, I., Arvanitis, L., Haglund, K., & Wojcik, A. (2019). Teaching and discussing about risk: seven elements of potential significance for science education. *International Journal of Science Education*, *41*(9), 1271–1286. https://doi.org/10.1080/09500693.2019.1606961.

Schulze, C., & Hertwig, R. (2022). Experiencing statistical information improves children's and adults' inferences. *Psychonomic Bulletin & Review, 29*(6), 2302–2313. https://doi.org/10.3758/s13423-022-02075-3.

Sedlmeier, P., & Gigerenzer, G. (2001). Teaching Bayesian reasoning in less than two hours. *Journal of Experimental Psychology: General, 130*(3), 380–400. https://doi.org/10.1037//0096-3445.130.3.380.

Steib, N., Büchter, T., Eichler, A., Binder, K., Krauss, S., Böcherer-Linder, K., Vogel, M., & Hilbert, S. (2025). How to teach Bayesian reasoning: An empirical study comparing four different probability training courses. *Learning and Instruction*, *95*, 102032. https://doi.org/10.1016/j.learninstruc.2024.102032.

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International statistical review*, *67*(3), 223-248. https://doi.org/10.1111/j.1751-5823.1999.tb00442.x.

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.