"How is this possible?" - Secondary PSTs' conceptions of variability in the context of probabilistic thinking

Arabella Denk
University of Vienna, Austria
arabella.denk@univie.ac.at

Abstract: In the past, the consideration of (data) variability has been repeatedly characterized as a fundamental element of statistical thinking. However, motivated by recent curricular changes in Austrian secondary schools, this study situates the concept within the context of individual ideas about probabilities in random experiments. A thematic analysis of pre-service teachers' answers to a pen-and-paper study in October 2024 gives insights into how variability is conceptualized within the interplay of patterns and deviations in the context of the empirical law of large numbers at the beginning of their tertiary stochastics training.

INTRODUCTION

Traditionally, stochastics teaching in Austria was characterized by a systematic separation of probabilistic and statistical concepts (Schupp, 1982), with a focus on Laplace-modelling and few links to data analysis. The new curricula for stochastics instruction in primary and lower secondary schools include a more combining approach (Biehler, 1994), among others, by highlighting conceptual relations in the context of the empirical law of large numbers (eLLN; Prömmel, 2013). Since (prospective) math teachers have been trained to think about probability theory and statistics in disjoint manners, they may encounter difficulties when attempting to integrate appropriate statistical ideas into their own conceptions of probability and its instruction. In sight of the curricular demands of situating probabilities in random contexts as predictive values for experimental outcomes and observed relative frequencies as possible estimations for underlying probabilities, it is of particular interest, how pre-service teachers (PSTs) conceptualize *(data) variability* when analyzing empirical distributions and how they contextualize probabilistic models within data observations.

Embedded in the national education context, this longitudinal dissertation study draws on prior research on variability (i.e. Schnell, 2014; Watson et al., 2003; 2007) and explores PSTs' individual conceptions of variability at the University of Vienna during their one-year tertiary stochastics instruction. In this paper, two exemplary findings of concept-specific themes (*shared meanings*; Braun & Clarke, 2022) at the beginning of their training will be presented.

THEORETICAL BACKGROUND

The *consideration of (data) variability* has been frequently identified (e.g. Moore, 1990; Shaughnessy, 2007; Wild & Pfannkuch, 1999) as a central element of statistical thinking, and underlies the fundamental idea of conceptualizing raw data within the interplay of identifiable patterns (*variability explained by use of mathematical models*) and deviations (*unexplained variability*), in order to draw conclusions under uncertainty (Eichler & Vogel, 2013, p. 59; Konold & Pollatsek, 2002). Consequently, reasoning about variability in problem solving contexts is linked to individual views on theoretical and empirical distributions (Biehler, 2007; Biehler et al., 2023).

In a statistician's professional context, Wild & Pfannkuch (1999, p. 226) distinguish four components of considering variability: its noticing & acknowledgment, measurement & modelling as well as the application of investigative strategies in relation to variability and its explanation in data analysis processes. Reading and Shaughnessy (2004, p. 203) also highlight the importance of its mathematical description and representation. Based on previous studies in international statistics classrooms (e.g. Makar & Confrey, 2005), an exemplary depiction of targeted learner strategies is given in the epistemological framework of a "deep understanding of variability" (cf. Table 1) by Garfield and Ben-Zvi (2005) and Reading and Reid (2010), incorporating interconnected ideas that are "robust, reflective, i.e., [include] awareness of own boundaries, of limits of applicability, and [are] also operable, i.e., can be used to solve real problems." (Garfield & Ben-Zvi, 2005, p. 93).

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*.

Table 1. An epistemological model by Garfield and Ben-Zvi (2005) and Reading and Reid (2010).

Components	Description
1. Developing intuitive ideas of variability	Acknowledgement of variability as an omnipresent characteristic of a data set
2. Describing and representing variability	Diverse use of graphical representations and statistical measures of central location and dispersion in descriptive analysis
3. Using variability to make comparisons	Consideration of variability within the analysis of differences and similarities in and between groups in informal-inferential tasks
4. Recognizing variability in special types of distributions	Contextualization of (measures of) variability in normal distributions, bivariate data sets, etc.
5. Identifying patterns of variability in fitting models	Situating variability as an indicator for model fit (e.g., the normal curve in empirical distributions)
6. Using variability to predict random samples or outcomes	Reflections on how the expected variability of experimental outcomes and sample statistics depends on the number of experimental trials or the sample size
7. Recognizing sources of variability	Thinking about possible sources of variability linked to the given chance set-up or the population of interest
8. Resolving expectations with observed variability	Understanding differences and relations between expected patterns (e.g. expressed in probabilities of outcomes) and deviations
9. The consideration of variability as an integral part of statistical thinking	Contextualization of and reflections upon data variability in each step of statistical investigation processes

However, stochastics instruction in Austrian schools (BMBWF, 2025) does not stipulate the performance of complete statistical investigations. The new area of competency "data and chance" for the lower secondary level motivates connections between observed data and mathematical theory primarily within the relationship between relative frequencies and underlying probabilities of a chance set-up. Reflections on data variability are therefore situated within the

- the forecast of future experimental outcomes and the estimation of underlying probabilities from observed relative frequencies as well as
- the (qualitative) evaluation of probability assumptions with regard to observed data in experimental settings,

whereby debates on the existence of randomness do not take place (Biehler et al., 2023; p. 222; Schnell, 2014). These objectives relate conceptually to components 1, 4, 5, 6, 7 and 8 in Table 1. Furthermore, the question of how learners considerate data variability within these thematic contexts is therefore linked to their individual probabilistic thinking, eventually incorporating a frequentist (Gal, 2005) or, within applied model-building contexts, a prognostic-hypothetical point of view (Riemer, 2023, p. 17) where they acknowledge "probabilities, especially Laplace's, as human-made models based on experience, which are hypothetical in relation to reality, quantify expectations and aim to make predictions" and "that there are different models for anticipating relative frequencies, which are given different degrees of confidence".

Previous international studies (such as Reading & Shaughnessy, 2004; Torok & Watson, 2000; Watson et al., 2007; Watson & Kelly, 2004) have shown a variety of individual conceptions of (un)expected ranges of experimental outcomes in idealized chance set-ups (i.e. no anticipation of data variation around the model value or ideas about ranges that are too broad or too narrow), as well as the "nature" of random processes in general (such as the conceptualization of random experiments as self-regulating mechanisms). However, they do not explicitly address reflections on the interplay between explained variability (patterns), unexplained variability (deviations) and underlying probabilities within the (empirical) law of large numbers (with exception to Schnell, 2014, who investigated individual ideas of schoolchildren related to this thematic complex within a specific learning environment at the beginning of the lower secondary level) which Biehler and Steinbring (1982) describe as the most important link between mathematical theory and empirical practice; and which plays an important role in the Austrian stochastics curriculum (BMBWF, 2025).

Based on the question in which ways observed stabilizations of relative frequencies can be interpreted mathematically in school settings, Prömmel (2013) formulates *aspects of the empirical law of large numbers*, embedded within different mathematical theories that can be developed at least at a preparatory level in secondary education, such as:

- **Approximity Aspect:** stabilization of relative frequencies with an increasing number of experimental repetitions, convergence of distributions (→Weak Law of Large Numbers, Strong Law of Large Numbers, Central Limit Theorem)
- **Accuracy Aspect**:
  a. For a given fixed range: As the number of repetitions n increases, more and more values fall within the range and fewer outside.
  b. For a given fixed probability: As the number of repetitions n increases, the anticipated range decreases (→ The $\sqrt{n}$ law of dispersion, Chebyshev's inequality, Prediction Intervals) (ibid., p. 77)

Previous studies investigating individual ideas about the eLLN show that learners often only conceptualize the approximity aspect (which in turn can lead to a one-sided understanding of distributions) and make deterministic generalizations such as analytical convergence (Prömmel, 2013). Furthermore, a well-known problem is neglecting the sample size in problem solving contexts, even if this tendency is highly task-dependent (Sommerhoff et al., 2023). Schnell (2014, p. 2) also highlights the importance of the *stochastic context* within the eLLN. This refers, among others, to the prediction of relative frequencies in the 'short term' (meaning the unpredictability of individual outcomes or outcomes in experiments with a small number of repetitions by means of mathematical modelling; unexplained variability) and the 'long term' (the possibility of reliable predictions in experiments with a large number of repetitions; explained variability).

This study builds on these previous works and takes a closer look at how variability is understood by secondary PSTs in the context of observed or anticipated stabilizations of relative frequencies with an increasing number of trials. In particular, it will be investigated how they conceptualize the interplay between explained and unexplained variability within different aspects of the eLLN.

METHOD

Context and participants

The present study was conducted in October 2024 with voluntary secondary PSTs in mathematics (University of Vienna) at the beginning of their *Introductory Course to Probability Theory* (incorporating a brief introduction to statistics). This class is typically scheduled during the seventh semester of the Bachelor of Education program and is succeeded by a course on didactic perspectives on secondary stochastics teaching in the eighth semester (Curriculum BEd, 2022). However, following this pathway is not obligatory and for the purposes of this article, data from students who had already completed a part of the stochastics training were not considered in the analysis (the resulting number of participants after data cleaning will be indicated by n_{surv} and n_{int}, respectively, in the following section). It is important to note that the remaining participants still exhibit a wide range of educational backgrounds, including their mandatory second subject and the extent of prior school experience in stochastics.

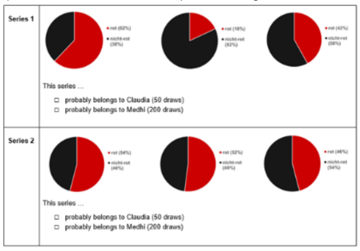
Data collection

The study design comprises a pen-and-paper-survey (135 participants, $n_{\text{surv}} = 101$ without partially completed university training) and a stimulated recall interview study (5 participants, $n_{\text{int}} = 4$) both of which were administered by the author, who was not involved in the teaching of the course. The written survey was conducted during the first exercise class of the Introductory Course to Probability Theory (in 9 different groups with an average of 20 students each, during the first two weeks of October) with an approximate duration of 45 min. However, the questionnaire also incorporated tasks on the handling of data variability in the context of descriptive statistics that will be omitted for the purposes of this paper. At the beginning of the pen-and-paper-survey, the participants were asked to answer as honestly and thoroughly as possible and to refrain from using aids, such as calculators. The open-ended items focusing on thinking about variability in the context of chance set-ups and the eLLN were selected and partially adapted from existing literature (cf. Table 2). Exemplary Criteria of the task selection were:

- The compatibility of the task content with the demands of the national secondary curriculum, as well as with aspects of the epistemological frameworks of Garfield & Ben-Zvi (2005) and Reading & Reid (2010)
- The use of contextualized tasks that encourage reflections on similarities and differences between model-based expectations and observed deviations (Watson et al., 2007, p.1)
- The integration of different aspects of the eLLN (Prömmel, 2013; Schnell, 2014)

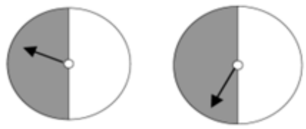
Due to the design and adaptation of the test items, a content validation was carried out with experts, and the tasks were repeatedly tested and revised with target study cohorts in the academic year of 2023 / 2024. The final tasks (differing slightly from the original survey items due to space constraints) are shown in Table 2.

Table 2. Survey Items.

<table>
<tr><th>Task 1</th><th>Task 2</th><th>Task 3</th></tr>
<tr>
<td>Coin Toss:
The probability of obtaining at least 7 heads out of 10 tosses with a fair coin is
• greater than
• equal to
• smaller than
obtaining at least 70 heads out of 100. Explain your choice.
(cf. Sommerhoff et al., 2023)</td>
<td>Urn experiment with 5 red & 5 non-red balls (with replacement):
Which experimental results can be attributed to Medhi (200 draws each), which to Claudia (50 draws each)? Why?


(cf. Biehler & Prömmel, 2013)</td>
<td>Rolling a pencil in math class:
After rolling a real six-sided pencil with one marked side 200 times, the marked side appears 1/3 of the time instead of
$p = 1/6$.

How would you explain the observed differences to your pupils?
How would you react in class?
(cf. Riemer, 2023)</td>
</tr>
<tr><th colspan="2">Task 4</th><th>Task 5</th></tr>
<tr>
<td colspan="2">Randomized Wheels of Fortune (cf. Watson et al., 2007):
• How many successes would you expect when turning the wheels 120 times each? (You win a game if both arrows land in the respective grey area)
Answer: ____ to _____
• How many successes would surprise you? Explain your choices.</td>
<td>Formulations of the eLLN:
The participants comment on two formulations of the eLLN, one stating an analytic convergence (deterministic interpretation of the approximity aspect) and a likely decrease in the range of relative frequencies with a higher number of trials (accuracy aspect), each paired with</td>
</tr>
</table>

	choice-answers "Correct", "Wrong", "I don't know"

Based on the written survey answers as cognitive stimuli, individual stimulated recall interviews (Ericsson & Simon, 1993; duration: 25min – 1h17) have been conducted with a smaller study cohort (n_{int} = 4) a week after the pen-and-paper study. While the questionnaire responses provide an overview of possible thought patterns, the interview data may offer insights into individual motives or lines of argumentation that ultimately led to the written responses but were not mentioned in the questionnaire (as well as into ideas being discarded during the problem-solving process). At the beginning of each interview, participants were instructed to follow a standardized protocol encompassing the silent reading of the task description and their respective written answer, allowing themselves moments of reflection, and then articulating their memories of thoughts occurring during the survey study. In order to reduce bias effects, the interviewer intended to limit herself to probing questions (Díaz, 2022) like *What was going through your mind while working on the task?*. Afterwards, the PSTs' written answers to an item-related self-assessment scale (*Please indicate how confident you felt when completing the respective task*, ranging from 0…*not sure at all*, to 3…*absolutely sure*) at the end of the pen-and-paper survey were used to talk with the participants about possible difficulties, emotions or further cognitive processes that took place during the first part of the study.

Data Analysis

The digitalized survey and interview responses were examined separately by item and in relation to the entire questionnaire, according to an inductive and deductive thematic analysis procedure (Braun & Clarke, 2022). In the first stage of the analysis, a global view was employed on what is being said in the answers and which diverse assumptions in the given task contexts have been formulated (ibid, p. 44). In order to get a deeper insight on how students conceptualize data in the context of explained variability (patterns), unexplained variability (deviations) and underlying probabilities within the (empirical) law of large numbers and its application in (non-)idealized chance set-ups, the subsequent inductive coding procedure was accompanied by emergent guiding questions such as *How do the students describe observed or hypothetical data (such as individual outcomes and accumulated experimental results; with or without relation to assumed probabilities) in the given tasks? Which terms or descriptions are used for observed or hypothetical patterns and deviations? In which ways do PSTs characterize possible relations between relative frequencies and assumed probabilities (or, more generally, data observations and mathematically modelled expectations)?* In addition, the answers were coded deductively according to specific views on the eLLN and its application in experimental contexts categorized by Prömmel (2013; namely the approximation and accuracy aspect) and Schnell (2014). The emerging codes identified during the inductive and deductive procedures were then checked against the data, refined repeatedly and grouped into themes.

RESULTS

With the exception of responses to Task 1 (*Coin Toss*), the majority of students' explanations demonstrated an at least partial understanding of data within the interplay of identifiable patterns (variability that can be explained by mathematical models) and deviations (unexplained variability): Expressed beliefs about the 'arbitrariness' of data throughout the entire survey, in the sense that 'anything can happen anyway' in experimental outcomes without relation to model assumptions, could only be identified in one person's responses. In addition, no participant stated in all (or most) of the given tasks that the relative frequencies must correspond exactly to assumed probabilities (regardless of the number of trials). More specifically, the majority of participants demonstrated at least a partial understanding of the eLLN: Within this group, most of their answers in the questionnaire (excluding Task 4) were consistent with the idea that, with an increasing number of trials, the value of an entity (such as relative or absolute frequencies, probabilities) changes in relation to another entity with respect to the eLLN. This included answers that mentioned, for example, a decrease of deviations of observed relative frequencies from the assumed underlying probability. However, answers stating an increase of anticipated deviations regarding relative frequencies, an expected stabilization of absolute frequencies, or a change of individual probabilities with a higher number of experimental repetitions have been excluded. Additionally, within this group of students with at least a partial understanding of the eLLN, explanations given throughout (at least most) of the survey or even within the same task could often be attributed to both aspects of the eLLN. However, some students found it difficult to conceptualize the dispersion of experimental outcomes in the context of the accuracy aspect. Moreover, the neglect of sample size could only be inferred in Task 1.

Two exemplary findings of the thematic analysis are presented in the following and hint at possible conceptual obstacles within reflections upon variability in the context of probabilistic thinking:

Irrevocability in the context of the eLLN

This may be related to how data is expected to behave in relation to model assumptions. For example, answers in this theme included *deterministic formulations of the eLLN*, such as a certain decrease of deviations of relative frequencies from probabilities with a larger number of trials, without relativizing the given statement in Task 5 regarding their analytic convergence (0X04U0: Task 1 – “The more attempts, the closer it gets to 50 times heads.”, Task 5 [approximity aspect, “Correct”] - “The more trials, the greater the probability of $p = 1/2$.”). In general, particularly in Task 5, most participants did not express any doubts or relativizations about the convergence assumption. Another form of irrevocability in expected data have been attributed to the assumption that *for an infinite number of trials relative frequencies are close or equal to the “true” value* (1N04N1: Task 5 [approximity aspect, “Correct”] - “Not necessarily, but with 1000 attempts, the probability is very likely closer to $p = 1/2$; with infinity, you would arrive at $p = 1/2$. The more attempts, the closer you get to the probability.“). Expressions of irrevocability may also relate to the conceptualization of probabilities: For example, *probabilities in multi-stage random experiments were equated with probabilities in single-stage situations* (with or without mentioning independence), and/ or it was noted that probabilities were dependent of the given relative frequencies in the task, but not specifically on the number of experimental trials (Task 1 - 9A08N0: ”Because with every trial, the probability of tossing heads is 50%.“; 2S07U1: “The probability is always the same when tossing a coin. If you want a certain percentage of all coin tosses to come up heads, it is irrelevant how many tosses are made in total; the only thing that matters is that the relative number of heads must be the same --> equal probability.”). Finally, also attributed to this theme have been answers *not questioning the validity of prior calculated probabilities when experimental results show great deviation* (5S05N0: Task 2 - “The more draws are made, the more accurate the result will be (approaching 50%).”, Task 4 - “How is this possible?”).

Distancing oneself from ideas about the eLLN in the context of probabilities:

This theme has been attributed to participants who answered (almost) all tasks of the survey in relation to the eLLN, but justified their single choice answer in Task 1 by calculations of (parts of) underlying probability distributions (1A03N5: Task 1 [“greater than”] – “For at least 7 times out of 10, the probability is x --> for at least 70 times out of 100, the probability must happen 10 times, so the new probability is x^{10} --> because x is a value less than 1, the “new” x is a value even less than the “old” x.””, Task 2 - “The more draws or ”attempts“, the closer the relative frequency gets to the probability --> Of course, both are possible, but it is more likely that Medhi will have results that are closer to the probability of 50% (5 out of 10) overall with more draws.”). Other participants distinguished in Task 1 between probability concepts in mathematical and experimental situations, whereby their answers to the other tasks were predominantly influenced by valid ideas within the eLLN (Task 5 - 5S05N0: “I chose here that the probability is the same if you toss the coin 7 times (...) out of 10 times or 70 times out of

100 times. (...) By probability, I simply assumed probability here, and yes (...) of course the probability is greater at the bottom that (...) it [gets closer], but the probability should be identical in purely mathematical terms"). Some participants characterized probabilities as both variable entities (possibly meaning relative frequencies) and fixed entities in the context of the eLLN, which in some cases led to inconsistencies in their responses also apart from the first task (0N10U0: Task 1 [„equal to“]– “Because the proportion 7 out of 10 is the same as the proportion 70 out of 100,” Task 5 [approximation aspect, “Correct”]: “According to the law of large numbers, the experimental probability increasingly converges towards the calculated probability as the number of trials increases.”)

DISCUSSION

These first findings give insights in which ways individual views on probabilities may (not) impose epistemological obstacles in the conceptualization of variability within the eLLN, supporting the argument that thinking about variability is highly contextual (Watson et al., 2007): Whereas most of the students demonstrated a basic understanding of (un)explained variability in data, the majority of their responses to Task 1 did not include their available eLLN-ideas, or they rejected them in favor of calculations. Other participants’ explanations were characterized by distinctions between probabilities within doing mathematics (*procedural*, *fixed, unquestioned*) and experimental settings (*variable*). It appears that beliefs about expected solutions for probability problems may hinder applications of existing eLLN-knowledge and that students who considered the probabilities in Task 1 as equal didn’t only do so because of conceptual difficulties with theoretical distributions. The findings also suggest that some PSTs considered probabilities as fixed values to be achieved, while not reflecting upon the changing probability of observed deviations from model assumptions depending on the number of trials or other aspects of the chance set-up. In math class, this could hinder reflections on probabilistic models from a prognostic-hypothetical point of view (Riemer, 2023) as well as meaningful applications of mathematical theory in more advanced statistical contexts (e.g. confidence intervals).

REFERENCES

Biehler, R. (1994). Probabilistic thinking, statistical reasoning, and the search for causes—Do we need a probabilistic revolution after we have taught data analysis? In J. Garfield (Ed.), Research papers from the Fourth International Conference on Teaching Statistics (ICOTS 4) (pp. 20–37). Minneapolis, MN: University of Minnesota.

Biehler, R. (2007). Denken in Verteilungen - Vergleichen von Verteilungen. *Der Mathematikunterricht*, *53*, 3–11.

Biehler, R., & Prömmel, A. (2013). Von ersten stochastischen Erfahrungen mit großen Zahlen bis zum $1/\sqrt{n}$ – Gesetz – ein didaktisch orientiertes Stufenkonzept. *Stochastik in der Schule*, *33*(2), 14–25.

Biehler, R., & Steinbring, H. (1982). Bernoullis Theorem: Eine „Erklärung“ für das empirische Gesetz der großen Zahlen? In H.-G. Steiner (Ed.), *Mathematik-Philosophie-Bildung* (pp. 296–334). Aulis-Verlag Deubner.

Biehler, R., Engel, J., & Frischemeier, D. (2023). Stochastik: Leitidee Daten und Zufall. In R. Bruder, A. Büchter, H. Gasteiger, B. Schmidt-Thieme, & H.-G. Weigand (Eds.). *Handbuch der Mathematikdidaktik* (pp. 243–278). Springer. https://doi.org/10.1007/978-3-662-66604-3_8

BMBWF (2025). *Curriculum for the secondary level (AHS) in Austria*. https://www.ris.bka.gv.at/GeltendeFassung.wxe?Abfrage=Bundesnormen&Gesetzesnummer=10008568

Braun, V. & Clarke, V. (2022). *Thematic analysis: A practical guide (1st ed.). SAGE.*

Díaz, S. (2022). Methodological strategies for using stimulated recall in interpreting research. *Onomázein*, *55*, 55–71. https://revistachilenadederecho.uc.cl/index.php/onom/article/view/51645

Eichler, A., & Vogel, M. (2013). *Leitidee Daten und Zufall : Von konkreten Beispielen zur Didaktik der Stochastik* (2nd ed.). Springer Spektrum. https://doi.org/10.1007/978-3-658-00118-6

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (2nd ed.). MIT Press. https://doi.org/10.7551/mitpress/5657.001.0001

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.

Garfield, J., & Ben-Zvi, D. (2005). A framework for teaching and assessing reasoning about variability. *Statistics Education Research Journal*, *4*(1), 92–99. https://doi.org/10.52041/serj.v4i1.527

Konold, C., & Pollatsek, A. (2002). Data Analysis as the Search for Signals in Noisy Processes. *Journal for Research in Mathematics Education*, *33*(4), 259–289. https://doi.org/10.2307/749741

Makar, K., & Confrey, J. (2005). "Variation-talk": Articulating meaning in statistics. *Statistics Education Research Journal*, *4*(1), 27–54. https://doi.org/10.52041/serj.v4i1.524

Moore, D. S. (1990). Uncertainty. In L. A. Steen (Ed.), *On the shoulders of giants: New approaches to numeracy* (pp. 95–137). National Academies Press.

Prömmel, A. (2013). *Das GESIM-Konzept*. Springer Spektrum. https://doi.org/10.1007/978-3-658-00594-8_2

Reading, C., & Reid, J. (2010). Reasoning about variation: Rethinking theoretical frameworks to inform practice. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS-8)*, Ljubljana, Slovenia. Voorburg, The Netherlands: International Statistics Institute.

Reading, C., & Shaughnessy, J. M. (2004). Reasoning about variation. In D. Ben-Zvi & J. B. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 201–226). Springer. https://doi.org/10.1007/1-4020-2278-6

Riemer, W. (2023). *Statistik unterrichten: eine handlungsorientierte Didaktik der Stochastik*. Klett-Kallmeyer.

Schnell, S. (2014). *Muster und Variabilität erkunden: Konstruktionsprozesse kontextspezifischer Vorstellungen zum Phänomen Zufall*. Springer Spektrum. https://doi.org/10.1007/978-3-658-03805-2

Schupp, H. (1982). Zum Verhältnis statistischer und wahrscheinlichkeitstheoretischer Komponenten im Stochastik-Unterricht der Sekundarstufe I. *Journal für Mathematik-Didaktik*, 3, 207–226.

Shaughnessy, M. (2007). Research on statistics learning and reasoning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 957–1009). Information Age Publishing.

Sommerhoff, D., Weixler, S., & Hamedinger, C. (2023). Sensitivity to sample size in the context of the empirical law of large numbers: Comparing the effectiveness of three approaches to support early secondary school students. *Journal für Mathematik-Didaktik*, *44*(1), 233–267. https://doi.org/10.1007/s13138-022-00213-x

Torok, R., & Watson, J. (2000). Development of the Concept of Statistical Variation: An Exploratory Study. *Mathematics Education Research Journal*, *12*, 147–169. https://doi.org/10.1007/BF03217081

Watson, J. M., Callingham, R. A., & Kelly, B. A. (2007). Students' appreciation of expectation and variation as a foundation for statistical understanding. *Mathematical Thinking and Learning*, *9*(2), 83–130. https://doi.org/10.1080/10986060709336812

Watson, J. M., & Kelly, B. A. (2004). Expectation versus variation: Students' decision making in a chance environment. *Canadian Journal of Math, Science & Technology Education*, *4*(3), 371–396. https://doi.org/10.1080/14926150409556620

Watson, Kelly, B. A., Callingham, R. A., & Shaughnessy, J. M. (2003). The measurement of school students' understanding of statistical variation. *International Journal of Mathematical Education in Science and Technology*, *34*(1), 1–29. https://doi.org/10.1080/0020739021000018791

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