Can. J. Sci. Math. Techn. Educ. (2026) 26:69
https://doi.org/10.1007/s42330-026-00477-2

Advancing Statistical Inference: A Comprehensive Exploration of Curriculum Development, Teacher Seminars, and Teacher Perceptions in Hungarian Secondary Schools

Péter Fejes-Tóth[1,2] · Manfred Borovcnik[3] · Ödön Vancsó[4,5]

Received: 13 September 2024 / Accepted: 11 March 2026

Abstract Our research project aims to integrate inferential statistics into the Hungarian secondary school curriculum. Between 2019 and 2023, we developed and tested an experimental curriculum incorporating simulations and Excel-based calculations. This approach addresses broader challenges in understanding statistical inference and presents our strategies for designing an experimental seminar for in-service teachers. The underlying principles of the curriculum are inspired by Complex Mathematics Education, a long-term initiative rooted in the ideas of Tamás Varga. Using a pre-post-test design and semi-structured teacher interviews, we evaluated the curriculum's feasibility. Feedback indicates that our revised approach and materials align well with the current Hungarian curriculum reform. Findings across the project's phases suggest that, with adequate technical support and sufficient practice tasks, even complex statistical concepts can be successfully taught to beginners. This approach supports diverse learner profiles and accommodates varying levels of prior knowledge and ability across schools. The article presents key insights from our teacher-training efforts over three phases, involving a total of 32 educators.

✉ Péter Fejes-Tóth
fejes.toth.peter@uni-mate.hu

Manfred Borovcnik
manfred.borovcnik@aau.at

Ödön Vancsó
vancso.odon@ttk.elte.hu

1 Institute of Mathematics and Basic Science, Dpt. of Applied Statistics, Hungarian University of Agriculture and Life Sciences, Villányi Út 29-43, H-1118 Budapest, Hungary

2 Department of Mathematics, Corvinus University of Budapest, Budapest, Hungary

3 Institute of Statistics, University of Klagenfurt, Klagenfurt, Austria

4 Centre for Didactics of Mathematics, Eötvös Loránd University, Budapest, Hungary

5 Alfréd Rényi Institute of Mathematics, Budapest, Hungary

Springer

Published online: 17 May 2026

Résumé Notre projet de recherche vise à intégrer la statistique inférentielle dans le programme scolaire secondaire hongrois. Entre 2019 et 2023, nous avons élaboré et testé un programme expérimental qui intègre des simulations et des calculs réalisés à l'aide d'Excel. Cette approche aborde les défis généraux liés à la compréhension de l'inférence statistique et nous permet de présenter nos stratégies pour la conception d'un séminaire expérimental destiné aux enseignants exerçant leur profession. Les principes fondamentaux du programme s'inspirent de l'enseignement des tâches complexes en mathématiques, une initiative à long terme fondée sur les idées de Tamás Varga. À l'aide d'un protocole d'évaluation fondé sur un test de style avant-après et des entrevues semi-structurées menées auprès des enseignants, nous avons évalué la faisabilité du programme. La rétroaction reçue indique que notre approche et nos supports pédagogiques révisés s'inscrivent parfaitement dans le cadre de la réforme en cours du programme scolaire hongrois. Les résultats obtenus au cours des différentes phases du projet suggèrent qu'avec un soutien technique adéquat et suffisamment d'exercices pratiques, même des concepts statistiques complexes peuvent être enseignés avec succès à des débutants. Cette approche s'adapte à des profils d'apprenants variés et sied à différents niveaux de connaissances et de compétences préalables dans les écoles. Nous présentons dans l'article les principales conclusions tirées de nos efforts de formation auprès des enseignants menés sur trois phases, impliquant au total 32 d'entre eux.

Keywords Bayesian inference · Curriculum reform · Design-based teaching · Statistical hypothesis testing · Teacher beliefs · Task systems in mathematics education

Introduction and Theoretical Background

Since 2021, Hungary has undergone a major reform of the national curriculum and final exams, with ongoing revisions to fine-tune the changes. Amid this reform, a key question arises: How can inferential statistics be taught feasibly? In response, we designed a curriculum and developed seminar materials on inferential statistics aligned with the reform. To contextualise our vision, we outline the historical development of statistics education and explore curricular design options (section "Introducing Statistical Inference into Secondary School Curricula"). Common conceptual challenges in statistical inference, as highlighted in the literature, help clarify our approach (section "Empirical Studies Unveiling Learner Challenges in Inferential Statistics"). We then present the study's goals and research questions (section "Goal of the Study and Research Questions").

Introducing Statistical Inference into Secondary School Curricula

Probability and statistics are vital for students' future careers, as reflected in projected job growth rates of 35% for data scientists and 32% for statisticians (Bureau of Labor Statistics, n.d.), and broader discussions such as Alpana (2017). Yet, statistical inference receives little attention in secondary education (Ministry of Education of Hungary, 2021). This article explores strategies for integrating inference into the secondary school curriculum. Globally, probability education has roots in combinatorics (Dinges & Rost, 1982). In the 1960 s, the New Math movement used probability to exemplify axiomatic structures. By the 1970 s, however, New Math had faded, replaced by a "back to basics" push (Klein, 2003). Rising interest in real-world applications led to a curricular shift from probability to descriptive statistics. Freudenthal's (1973) *Realistic Mathematics Education* (RME) may have influenced this shift. Truran (1995) cites a pragmatic reason: descriptive statistics is considerably easier to teach. The introduction of *inference*, however, generated didactic disagreement. Some (e.g. Cobb, 2007) argue that inference can emerge entirely from data analysis, while others (e.g. Borovcnik, 2011) maintain that probability is essential for understanding statistical inference.

Despite its importance, inference — introduced into curricula in the 1990 s — posed challenges due to its complexity, especially when extended to hypothesis testing, *p*-values, and confidence intervals (Thompson et al., 2007). These methods have also been widely debated (e.g. Hubbard & Bayarri, 2003). Attempts to simplify instruction during the 1990 s reflected educators' awareness of students' difficulties (e.g. Biehler, 1991). At the same time, interest in the Bayesian perspective grew, raising questions about its integration and decision-theoretic aspects (Witmer et al., 1997). Yet both Bayesian and classical inference faced curricular obstacles (Freedman, 1997; Albert, 1995). In response, alternative approaches emerged, including simulation-based methods (Batanero et al., 2005) and contextualised scenarios that give concepts natural — if limited — meaning. As Kang et al. (2019) note, contextual framing shapes understanding and promotes more direct learning.

Despite differing perspectives, the importance of teaching statistical inference is widely acknowledged. Yet, balancing complexity and comprehensibility remains a core challenge in curriculum development. Our project addresses:

- Designing a feasible approach to inference.
- Revising materials based on teacher feedback.
- Identifying educators' needs for effective instruction.

The curriculum was expanded and refined through input from teachers and students. The 2019 seminar piloted the programme with four teachers. The 2022 seminar, delayed by the pandemic, involved nine teachers and used pre-teaching questionnaires. In 2023, 19 teachers explored a parallel approach to classical and Bayesian inference. We aim to show the curriculum board the value of including inferential statistics.

Empirical Studies Unveiling Learner Challenges in Inferential Statistics

Understanding inferential statistics presents significant challenges for learners, as shown in multiple studies (Batanero et al., 2020; del Már López-Martín et al., 2019; Lugo-Armenta & Pino-Fan, 2021). Case and Jacobbe (2018) note that students often perceive "[p-values as] a tool for making decisions about the null hypothesis or a way to quantify the strength of evidence, but lack an integrated conceptual understanding of what the p-value represents" (Aquilonious & Brenner, 2015, p. 1).

Even with targeted instruction, difficulties persist (delMas et al., 2007). Lane-Getaz (2017) categorises these into issues with terminology, conceptual relationships, inferential logic, and *p*-value interpretation — while noting that type II errors are often ignored. Nilsson (2020) observes students' preference for contextual over probabilistic reasoning. Didactic literature discusses when and how context or assumptions about the parent distribution distract students from judging sampling variability. Burrill and Biehler (2011) emphasise "the relation between samples and the population and the essence of deciding what to believe from how data are collected to drawing conclusions with some degree of certainty" (p. 63), without clarifying this "degree of certainty". They cite Harradine et al. (2011), who focus on "understanding [the roles of] the null and alternative hypotheses" (p. 240), referencing Vallecillos (1999), who highlights the specific issue of misinterpreting the "significance level as simple [unconditional] probability of the null hypothesis" (p. 2) in significance testing.

First, Burrill and Biehler (2011) misleadingly describe the correctness of decisions as a "degree of certainty", a notion incompatible with classical inference. In the classical framework, hypotheses do not carry probabilities; rather, the probability refers to the long-run performance of the method when applied repeatedly. If they intended to reference subjective credibility in a Bayesian sense, their argument would align with Bayesian inference — which they do not pursue. Second, a chain of references attempts to explain the logic of inference, but none addresses its core conceptual difficulties in depth.

What remains missing is a clear articulation of the key elements of inferential logic, along with structured learning paths that support understanding while avoiding common misconceptions. Although students may misunderstand the evidentiary role of hypothesis testing, a correct view of that role does not ensure comprehension of conditional error probabilities. Misunderstandings — such as the fallacy of the transposed conditional, often transferred to *p*-values — persist unless students grasp conditional probabilities (Batanero et al., 2015, p. 7). Yet, such components represent only isolated fragments of inferential logic. Concepts like type I and II errors and their interplay illustrate deeper structural relationships within the logic of inference. Most studies omit Bayesian reasoning. Our curricular reform is driven by Rossman and Chance's (1999) "Top Ten" recommendations. In the long term, we advocate for a parallel introduction of classical and Bayesian inference to enhance teacher preparation (Vancsó, 2009). While Bayesian methods are taught at the university level (Albert, 1997, 2002; Berry, 1997; Hoegh, 2020; Berg & Hawila, 2021), their integration into secondary education remains largely unexplored. Migon and Gamerman (1999) offer a *parallel* framework for university education. Wickmann (1991) provides a Bayes-only approach to inference at school level accompanied with a programme for outsourcing the required calculations – highly innovative and ahead of the digitalisation of teaching in the early 2000's.

The parallel approach to inference employs both classical and Bayesian methods to address the same problem, aiming to enhance understanding of core concepts, underlying assumptions, and methodological limitations. We do not seek to synthesise the approaches, as this would risk conceptual incoherence. Rather, the comparison highlights distinct inferential frameworks and their boundaries (see Borovcnik, 2021). Instead of engaging deeply with foundational controversies, we illustrate key critiques of classical inference through selected quotations:

- No conditioning on observed data:

  > The most fundamental limitation of standard frequentist inference is that it does not condition on the observed data. The resulting paradoxes have sparked a philosophical debate that statistical practitioners have conveniently ignored (Wagenmakers et al., 2008, p. 181).

- Confusion from mixed frameworks:

  > […] this mixing has resulted in widespread confusion over the interpretation of [Fisherian] p values and [Neyman-Pearson] α levels. (Hubbard & Bayarri, 2003, p. 171).

- Ignoring uncertainty in alternatives:

  > This type of analysis does not take into account uncertainty for the alternative value. [...] The replication probability, which is related to power, is the probability that rejection of the null hypothesis would be achieved (or not) if data collection were conducted a second time (Kruschke, 2010, p. 296).

The longstanding debate between classical and Bayesian inference has spotlighted issues like computational demands and differing uses of conditional probability (Moore, 1997; Witmer et al., 1997). Classical inference is conceptually complex, whereas Bayesian methods offer conceptual clarity but computational intensity (cf. Myung et al., 2008, p. 312). Proposed solutions include Wickmann's (1990) Bayesian-only approach, simulation-based inference (Cobb, 2007), context-embedded strategies (Batanero & Borovcnik, 2016; Vancsó et al., 2021), and connecting classical tests to Bayes' theorem (Wolpers & Götz, 2002). Following Barnett (1999), we integrate subjective probability and employ Excel and GeoGebra (Vancsó et al., 2018), alongside simplification strategies from Gigerenzer's work (Hoffrage et al., 2015; Mellers & McGraw, 1999).

Goal of the Study and Research Questions

Teacher education in Hungary is heavily theory-oriented. In probability, prospective teachers study discrete (hypergeometric, binomial, uniform) and continuous distributions (uniform, exponential, normal), as well as simplified versions of the Law of Large Numbers and the Central Limit Theorem (Bernoulli's version), typically in a two-hour lecture and seminar. Combinatorics is taught within discrete mathematics. However, inferential statistics was introduced to mathematics-teacher education only within the past decade. It is taught as part of advanced mathematics, covering confidence intervals and standard tests (Gauss, *t*-, chi-squared, *F*-test). Didactical aspects of probability and statistics are generally excluded from university teacher education. Concerning didactics in general, teacher students delve into it for four semesters, covering curriculum development, evaluation, lesson planning and didactic material (details may be found in Dobos et al., 2001; Gordon Györi et al., 2020; Novotná et al., 2021). In the fifth semester, they undergo a six-month practicum at a specialised school. Collaborating with a senior teacher, they teach 15 classes and observe 45, engaging in extensive discussions about the classroom dynamics. They may address topics related to probability and descriptive statistics, but not inferential statistics, as this is not yet part of the high-school curriculum.

This study investigates whether a newly developed curriculum on inferential statistics is feasible within Hungary's educational context, considering its institutional structure and culturally rooted teacher beliefs. Despite the complexity of the long-term project, clear interconnections exist among stakeholders (teacher educators, experimental teachers, in-service teachers, and students).

Our core research questions are:

1. *To what extent is the suggested curriculum for teaching statistical inference acceptable and feasible for teachers?* Is the proposed curriculum — including content, task systems, media, and teaching methods — feasible for Hungarian teachers introducing inferential statistics at the high-school level? Can inferential statistics be taught meaningfully without a full mathematical background? Can teachers be convinced that it is a valuable topic for the final examination?
2. *Is the Bayesian way in parallel to classical inference useful to understand concepts of either side?* Does teaching Bayesian inference alongside classical methods improve conceptual understanding? Does it help convince teachers that inference is feasible and pedagogically legitimate?
3. *What additional support do teachers need for teaching statistical inference?* What additional support is required for teachers to adopt and implement the curriculum?

This study explores both teachers' and learners' challenges and presents principles guiding our curriculum-design research. We focus on the impact of the material and seminars on teachers' skills and attitudes. Our twofold aim includes (1) a theoretical rationale — examining how introducing a new curriculum affects teachers and students, and identifying key curriculum features that influence success; and (2) a policy rationale — providing evidence for decision-makers and evaluating whether the curriculum can be scaled across schools and contexts, based on teacher feedback and student performance. Rather than seeking simple yes/no answers, the study examines degrees of acceptance, epistemological alignment, and the conditions under which curriculum innovation becomes both teachable and scalable.

Design, Content, and Evolution of the Teacher Seminars

We provide background on the design of our seminars (Section "General Background of the Seminars") and present task systems that highlight interrelations between classical and Bayesian inference (Section "The Parallel Approach Towards Statistical Inference"). The training initially introduced hypothesis testing through examples like the chi-squared test. Following teacher feedback, we incorporated more inference problems solvable by combinatorial reasoning and extended the programme to include Bayesian methods taught in parallel with classical approaches. This parallel structure supports a deeper understanding of both.

General Background of the Seminars

We outline the curriculum design and its evolution across seminar phases, consistent with design-research principles. The seminars included three segments: an introduction, core instructional content, and technical aspects of measuring and evaluating learning outcomes. The introductory segment covered the rationale for experimental teaching, including a brief history of statistics education in Hungary and the introduction of hypothesis testing in schools. The final part detailed the experimental curriculum, supplementary materials, and tools for evaluating student understanding. The central segment focused on instructional content and methods, offering a comprehensive overview of hypothesis testing: null and alternative hypotheses, type I and type II errors, significance levels, decision criteria using p-values or critical values, and interpretation of decisions. We introduced examples solvable via combinatorial methods and taught the chi-squared test for independence and goodness of fit. To broaden understanding, we addressed all three interpretations of probability — classical, frequentist, and subjective (Carranza & Kuzniak, 2008) — through the "medical-test paradox" (Lu, 2020), also presenting Bayesian alternatives to classical methods.

Phase 1 (2019), held in Budapest, introduced statistical inference and the chi-squared goodness-of-fit test (Fejes-Tóth, 2020). Phase 2, with nine teachers, expanded to an 8-h format for deeper engagement. It added combinatorial problems (Fejes-Tóth et al., 2022) and introduced the Bayesian formula in the context of medical diagnostics.[1] Rather than solving problems, teachers explored scenarios described by the formula (Gigerenzer, 2003) and learned the interpretation of related conditional probabilities (Falk, 1989). While we addressed Bayesian inference conceptually, no problem-solving was required. Feedback emphasised the need for more mathematical background and time. Teachers also requested a wider variety of tests to understand the range of inferential methods.

Phase 3 preparations began in spring 2023. Seminars in August (11 teachers in Békéscsaba, eight in Miskolc) extended to 16 h based on previous feedback. We added more detail on classical inference, including one- and two-sample t-tests, F-tests (Freedman et al., 2007), and ANOVA (Johnson, 1989), aiming to enrich background knowledge not for classroom use. A Bayesian extension of the classical coin-flipping problem was also introduced. Teachers examined philosophical differences between classical and Bayesian inference and explored prior and posterior distributions, focusing on their interpretation. The concept of a subjective prior — representing prior knowledge about the probability of Heads — was discussed (Albert, 1995). A consistent takeaway across all phases was the importance of digital tools. Teachers unanimously agreed that software like Excel is indispensable for teaching inference.

The Parallel Approach Towards Statistical Inference

In the third phase of the project, we extended the seminar duration to allow for broader delivery of the envisioned curriculum, including parallel instruction in classical *and* Bayesian inference. We focused on a key example where a prior distribution is defined over a few discrete values for the probability of Heads in a coin toss. While classic problems like Lady Tasting Tea (Fisher, 1971) illustrate significance

[1] Hereby, we focused on accurately interpreting the posterior probability of having the disease under scrutiny when the diagnosis is positive (which means it indicates the presence of this disease).

testing well, they do not readily support Bayesian modelling. Instead, we used a *task system* based on two coin-tossing problems to highlight core differences between classical and Bayesian approaches. We also included a medical diagnostics scenario to illustrate parallel teaching of inference in greater detail.

Classical Solution

A coin is flipped 100 times, resulting in 63 Heads and 37 Tails. This raises the question: Is the coin fair? The null hypothesis H_0 states the coin is fair, with equal probabilities for Heads and Tails. The alternative H_1 negates the null: The coin is biased. We investigate two different ways to answer the question.

Utilising the Binomial Distribution to Depict the Coin's Behaviour Under the Null Following Fisher's (1971) significance test, we compute the "discrepancy" between the observed result and the null. Under the assumption that the null hypothesis holds, the sampling distribution of the number of Heads follows a binomial distribution with $p = 0.50$. Consequently, the probability of obtaining 63 *or more* Heads in 100 tosses is 0.0120 (two-tailed, see Fig. 1(a and b)).

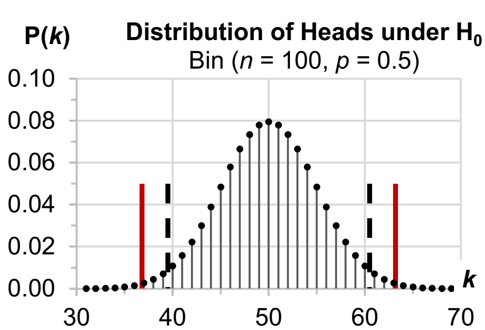
The likelihood of obtaining such a different result from the expected outcome on the basis of the null is very low. Thus, at the 0.05 significance level, we reject the null that the coin is fair (the p-value is 0.012):

$$p = 2\sum_{k=0}^{37}\binom{100}{k}0.50^k \cdot 0.50^{100-k} = \frac{2}{2^{100}}\sum_{k=0}^{37}\binom{100}{k} \approx 0.0120.$$

Goodness-of-Fit Test The chi-squared test assesses the disparity between observed frequencies (f_i) and those to be expected under the null ($n\,p_i$) using a "standardised" Euclidean distance (Fig. 2(a)):

$$X^2 = \sum_{i=1}^{k}\frac{(f_1 - np_i)^2}{np_1} = \frac{(63-50)^2}{50} + \frac{(37-50)^2}{50} = 6.76.$$

If the observed χ^2 exceeds the critical threshold of 3.84 (Fig. 2(b)), we infer — at the 0.05 significance level — that the coin is not fair. While the binomial model is more accessible for the school settings, the chi-squared distribution poses greater conceptual demands. Nevertheless, the statistics are simple to compute, and the null distribution can be effectively approximated through simulation (Batanero et al., 2005), providing empirical estimates to the critical values shown in Fig. 2(b).



a

Critical values for the exact test

k	$P(x = k)$	one-tailed $P(x \le k)$	two-tailed $2 \times P(x \le k)$
36	0.0016	0.0033	0.0066
37	0.0027	0.0060	**0.0120**
38	0.0045	0.0105	0.0210
39	0.0071	0.0176	0.0352
40	0.0108	0.0284	**0.0569**
41	0.0159	0.0443	0.0886

b

Fig. 1 (**a**) Distribution of Heads under the null. (**b**) Critical values for the exact test

Frequencies	Heads	Tails	All
Observed	63	37	100
Expected	50	50	100

a

Critical values of chi-sqared distribution

df \ α	0.001	0.01	0.05	0.10
1	10.83	6.63	3.84	2.71
2	13.82	9.21	5.99	4.61
3	16.27	11.34	7.81	6.25
4	18.47	13.28	9.49	7.78
5	20.52	15.09	11.07	9.24

b

Coin	Prior(Coin)	P(Head| Coin)	Binomial distribution P(X ≤ 37| Coin)	Multiplication Rule Prior(Coin) × P(X ≤ 37| Coin)	Bayes rule Posterior(Coin)
A	1/3	0.55	0.00022	0.00007	0.0031
B	1/3	0.50	0.00602	0.00201	0.0844
C	1/3	0.45	0.06507	0.02169	0.9125
	1	–	–	0.02377	1
			Cond. Prob. of Data	Total probability of Data	

c

Fig. 2 (**a**) Observed vs expected frequencies. (**b**) Critical values of chi-squared distribution. (**c**) Scheme for the calculation of the posterior probability of the coins A – C from their prior and the binomial probability of the data: $Data = (X_{100} \leq 37)$

Bayesian Approach

In a Bayesian framework we consider three coins — A, B, and C — where B represents the fair coin with a 0.50 probability for Heads. A and C differ slightly:

$$P(Heads|Coin\ A) = 0.55,\ P(Heads|Coin\ B) = 0.50,\ P(Heads|Coin\ C) = 0.45$$

We adopt a simpler prior than that used by Albert (1995), embedding the problem into the following experiment: one of the three coins is randomly selected, each with a probability of 1/3, and flipped 100 times. “If we observe 37 Heads, which coin is most likely flipped?” Intuitively, *Coin* C seems most plausible. “How sure can we be about this?” Initially, each coin had an equal prior probability. After observing the outcome, the posterior probability for each coin is updated using Bayes’ rule. For *Coin* B (the fair coin), we obtain (see Fig. 2(c), last column):

$$P(Coin\ B|Data) = \frac{P(Data|Coin\ B) \times P(Coin\ B)}{P(Data)} = \frac{0.00201}{0.02377} = 0.0844$$

For the biased *Coins* A and C, we get:

$$P(Coin\ A|Data) = 0.0031 \text{ and } P(Coin\ C|Data) = 0.9125$$

Consequently, *Coin* C is by far the most likely, with a posterior probability exceeding 0.91. Our intuition is confirmed and quantified. The conclusion also aligns with the classical approach: the coin is *likely not fair*. However, the Bayesian model still assigns a conditional probability of 0.0844 to the fair coin, prompting valuable classroom discussion: The classical approach does not assign probabilities to hypotheses (i.e. parameter values) and results in a binary decision — reject or fail to reject the null. In contrast, the Bayesian method incorporates prior probabilities and yields a direct probability for each hypothesis. This contrast highlights a core dilemma. The classical approach requires designating a specific parameter value as the null, which may be problematic. The Bayesian method avoids this by distributing credibility across possibilities but introduces subjectivity through the choice of prior. What reconciled participants was the shared outcome: both approaches identified *Coin* C as most plausible. Yet, the Bayesian solution provides more nuanced information, inviting richer interpretation.

Methodology and Construction Principles for the Curriculum

This section outlines the core principles guiding curriculum construction and our methodology for evaluating seminar effectiveness through structured feedback loops. We implemented the curriculum in diverse school settings to assess its feasibility and acceptance across different contexts. Our iterative design process consisted of the following key stages (see Fig. 3):

1. Initial pilot and iteration: The curriculum was tested in a small number of classrooms, followed by two additional seminars involving experimental classes — forming two feedback loops.
2. Expanded pilot testing: The pilot was broadened to include more schools, beginning with teacher seminars (Fig. 2(a)), followed by implementation by experimental teachers in classrooms (Fig. 2(b)).
3. Feedback mechanisms: We collected structured feedback from teachers (and students) throughout all phases to inform ongoing adjustments. Teachers could seek advice during planning and teaching, which provided valuable insights for refining the curriculum.

Following the analysis, we will develop a detailed strategy for scaling the curriculum and formulating evidence-based policy recommendations.

Research Design and Methods

We applied a design-based research methodology (Cernusca & Ionas, 2014; Majgaard et al., 2011) to evaluate the curriculum's viability and effectiveness for both students and teachers. Conducted in three phases (Table 1), the study evolved through iterative adjustments based on feedback. Evaluation focused on teacher and student attitudes, perceptions of the curriculum, and student learning outcomes (Fig. 3). Programme impact on skills and competencies was assessed using a pre-post summative test design.

We assessed student feasibility and acceptance through a questionnaire capturing their opinions and attitudes toward the material. For teachers, semi-structured interviews were conducted post-training and post-teaching in all phases and analysed narratively. Questionnaire responses were statistically summarised. Within the research team, we reviewed statement relevance to our research questions, focusing on feasibility, seminar communication, misunderstandings, and suggestions for improvement, including topic extensions. Clarifications were obtained during interviews when needed. In phases 2 and 3, one full feedback cycle per phase included an evaluative teacher survey, with summary statistics used for analysis.

To ensure representativeness, we went beyond the usual elite gymnasiums[2] in Budapest. Two additional cities were chosen to balance urban and rural demographics. The overarching goal was to develop a curriculum accessible to all societal levels, acknowledging Hungary's educational disparities. Schools from various tiers were recruited for the experiment.

Main Principles for the Construction of the Curriculum

Our core principles build on Varga's ideas and the earlier national curriculum reform (1960–1978), marked by *Complex Mathematics Education* (CME). Varga introduced probability through task systems and experiments that linked intuitive ideas—before formal definitions of probability—to concepts in statistical inference. Similarly, we aim to develop a curriculum that helps students understand inference, including Bayesian methods, despite their complexity. To this end, our curriculum and teacher training

[2] Public schools in Hungary vary considerably in level, infrastructure and financial resources, and class size.

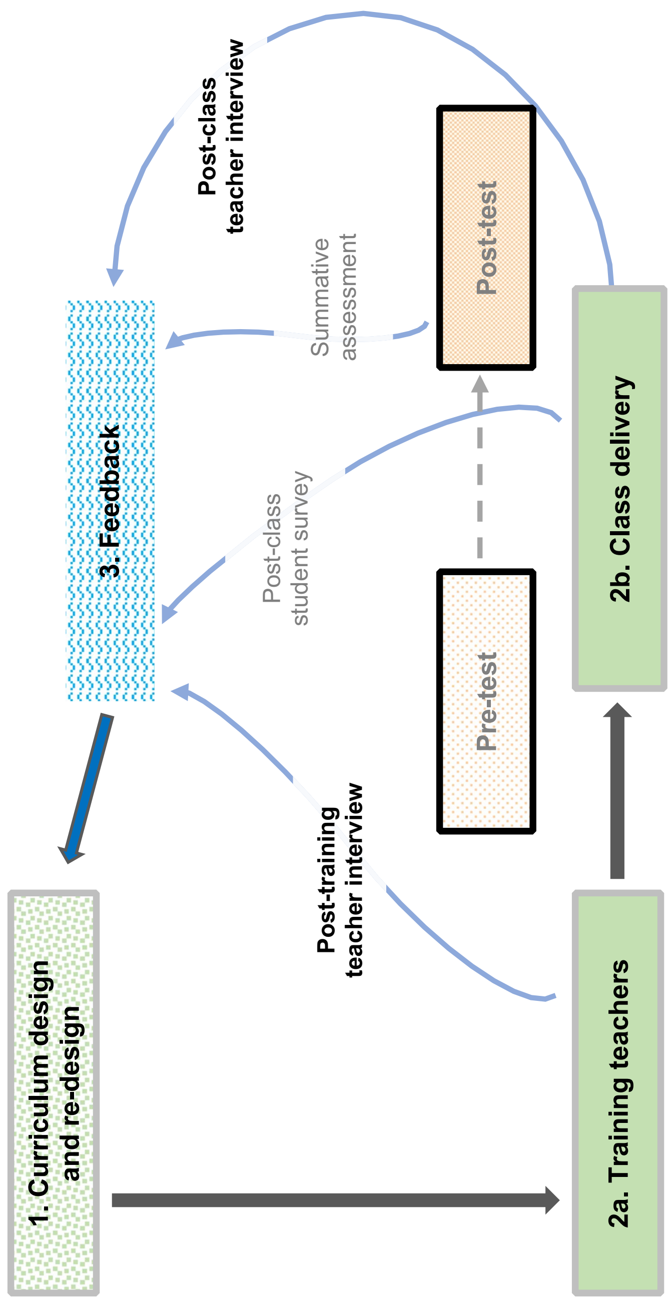


Fig. 3 Feedback on teacher seminars and experimental classes — two full loops for feedback and revision — while student feedback was essential for revisions, it is not discussed in this paper

Table 1 Data of the training phases

Characteristics	Phase 1	Phase 2	Phase 3
Teachers' seminar	2019/Dec	2022/Dec	2023/Aug
Location	Budapest	Budapest	Békéscsaba/Miskolc
Participating teachers	4	9	19
Duration (45-min units)	4	8	16
Class delivery	2020/spring	2023/spring	2023/2024
Participating students	64	85	142
Lessons (45 min)	6–10	4–6	6–10

draw on diverse approaches in mathematics didactics, which also underpin CME (see Gosztonyi, 2023; Varga, 1983).

1. *Flexible curriculum structure*: We distinguish between suggested and compulsory topics, allowing for dynamic adjustments as students' understanding evolves. The flexibility also extends to the teaching methods: Teachers should be able to switch between different methods to meet the needs of the students.
2. *Dialogic teaching and learning*: Grounded in Bruner's (1961) work, our approach embraces dialogic teaching, fostering inquiry-based learning (Artigue & Blomhøj, 2013; Mercer & Littleton, 2007). It aligns with CME (the revived long-term project) principles, nurturing interconnected mathematical thinking across domains (see Kiss, 2022).
3. *Task systems*: Our experiments are structured around task systems, guiding students towards mathematical insights (Artigue, 2002; Stein et al., 1996). Tasks are arranged to promote understanding and build on prior knowledge (Kieran et al., 2015). Task systems also make it possible to pursue two of Pöhler's (2020) four principles, namely *conceptual understanding before procedures* and *integrative rather than additive learning*. Connections between tasks form learning paths that convey the essential features of the involved concepts.
4. *Teaching by games and playful learning*: To enhance engagement and conceptual understanding, we include games in our approach (Deterding et al., 2011; Gee, 2003; Papp, 2017). Games promote critical thinking and provide real-world contexts for mathematical concepts. Yet, it requires "a skilful teacher […] creating teacher-student interactions that focus on the mathematical ideas underlying the game […]" (Heshmati et al., 2018, p. 792).
5. *Simulation for insight*: Simulation is crucial for grasping complicated mathematical relationships, particularly in probability and sampling theory (SERC portal for educators, n.d.). We use simulation to simplify tasks and illustrate sophisticated concepts, enhancing comprehension (Batanero et al., 2005; delMas, 2017).
6. *Realistic problems*: Our experiments align with *Realistic Mathematics Education*, emphasising realistic, contextual problems to foster deeper understanding (Freudenthal, 1973, 1981; Gravemeijer & Doorman, 1999). Realistic problems provoke curiosity and connect mathematics to daily life.
7. *Hands-on activities*: Teachers utilise hands-on activities and simulations to enhance understanding, bridging the gap between abstract concepts and real-world applications (Clements & Battista, 1992; Van de Walle et al., 2019). These activities promote conceptual growth and provide tangible connections to mathematical ideas.
8. *Tasks connecting probability and statistical inference*: In line with Varga (1972, 1975, 1983), our tasks tightly connect probability to statistical inference, engaging students in playful discovery. Students explore real problems, enabling comprehension and application of statistical concepts.

In line with Maaß and Artigue (2013), our goal was to support teachers to shift from traditional content delivery to facilitating student learning. Our approach integrates the key principles outlined above.

Results and Redesign

We analyse the results of various feedback loops from teachers and students and the ensuing redesign of curriculum and materials. The specific impact of the parallel approach and the use of digital technology are discussed in the light of the feedback.

General Feedback from Teachers

Feedback from phase 1 was broadly positive across all dimensions, yet some shortcomings emerged. In response, we redesigned the seminar and revised materials following phase 2. Two recurring issues were identified: the need for more practical examples and a clearer explanation of the mathematical foundations underlying the methods.

To address the first, we added practice problems with detailed solutions and introduced real-life examples across diverse fields. Recognising that teachers in phase 1 had limited engagement with Bayesian methods, largely due to time constraints and lack of confidence, we added a two-lesson introduction in phase 2, expanded to four lessons in phase 3. Moreover, we recommended 6 h of classroom implementation.

Clarifying the mathematical background proved more challenging. Although methods like the chi-squared test appear simple, their theoretical basis is complex — often beyond the reach of students and, in some cases, teachers. Still, as emphasised by Freedman et al. (2007), understanding why methods work is crucial, not just how to apply them. Therefore, we prioritised problems solvable using combinatorial reasoning, which students already possess — particularly through binomial and hypergeometric distributions (Fejes-Tóth et al., 2022). This grounding supports a justified use of chi-squared tests. Tables 2 and 3 summarise teacher responses to two key questions from the final questionnaires in phases 2 and 3.

Table 2 Positive feedback from phases 2 and 3 teachers (selection)

Describe in three sentences what was positive about the training for you!
"… a new perspective with philosophical considerations, making me contemplate subjective decisions. The loaded-die experiment was particularly interesting."
"I liked the multi-perspective approach to statistical thinking. This material is closer to … students than a significant part of the curriculum. The world is multi-coloured, even [probability] use the [0, 1] interval …, not just 0 and 1; I like the two-valued logic …" [a]
"…very inspired by everything [especially] the relationship between the learning process and Bayesian statistics. … important that students become familiar with."
"… now I understand it much better. … possible to prepare interesting tasks and projects for students." "[In Bayesian …, we] develop a variety of competencies. Useful training."
"guides thinking, prior knowledge, and scientific attitude of future biologists, engineers, …, and facilitates understanding of … more serious theories. For average students, it shows a comprehensive and complete analytical examination of everyday phenomena. …"

[a] Reflecting that classical testing ends up with accepting or rejecting the null, while the Bayesian assigns a posterior probability to the hypothesis, which is a degree of credibility. See Carranza and Kuzniak (2008) for their critique of confusing frequentist and subjectivist probability throughout the curriculum

Table 3 Areas to improve based on feedback from teachers of phases 2 and 3 (selection)

Describe in three sentences what was negative about the training for you!
"I believe that 'accepting this formula' without further explanation is a problem in Hungarian mathematics education. We have to go beyond this, …"
"I accept that this is the formula. … I don't know why it applies; I have to live with that. I feel unprepared to answer students' questions why it really works." [b]
"… not sure I would be able to give good answers or correct children's solutions … they would not accept ready-made formulas and methods without explanation."

[b]This resonates with thoughts we noticed when teachers ask for mathematical completeness of the argument in the presentation. This seems to provide them some safe space for teaching in class

In response to phase 2 feedback, we added material with clearer explanations focused on points that teachers identified as confusing. From discussions about what qualifies as valid mathematical reasoning — particularly scepticism about simulation being "real" mathematics — we expanded the use of combinatorial-based probability and strengthened the link between hypothesis testing and combinatorics. For a small group of teachers — typically those with a stronger background in pure mathematics — the preferred resolution of these concerns would be a deeper engagement with the full formal mathematical derivations underlying the methods. However, for most secondary school teachers, pursuing the complete theoretical foundations of inferential procedures would neither be realistic within the constraints of teacher preparation nor necessary for effective classroom implementation.

Our aim is therefore not to replace formal mathematical understanding, but to articulate a form of justification appropriate to the secondary school level. Simulation-based reasoning, combinatorial arguments, carefully structured analogies, and digital tools can provide mathematically meaningful forms of explanation that support conceptual understanding, even when full formal proofs are beyond scope. In this sense, the challenge is not merely technical but epistemological: broadening what is recognised as legitimate mathematical reasoning in the context of teaching statistical inference.

Table 4 presents selected teacher perceptions on the seminar's usefulness, using a 7-point Likert scale. Overall, *teachers found the theoretical background clear*, with minimal variation between cities (6.45–6.57). Practice problems were rated even higher in *clarity* (6.75–6.86) and were widely considered *interesting*: Budapest (6.43), Békéscsaba (6.73), Miskolc (7.00). However, ratings for the question "Can the practice problems be taught at the secondary school level?" were lower and more variable: Budapest (6.14), Miskolc (6.13), Békéscsaba (5.36). This likely reflects the school types: most participants in Budapest and Miskolc were from gymnasiums, while those in Békéscsaba taught at general high schools.

In summary, perceptions of the seminar content were consistent across teaching backgrounds. Yet, judgements on whether the material could be effectively taught to students varied by school type, with nearly a 1-point difference on the 7-point scale. The following tables present mean responses (*M*), with SD in brackets, from key questions in the final questionnaires for phases 2 and 3.

Table 4 Teachers' evaluation of the seminar on a 7-point Likert scale — phases 2 and 3 — *M* (SD)

Impact of the seminar	Budapest $N=7$	Békéscsaba $N=11$	Miskolc $N=8$
Theory was clear	6.57 (0.49)	6.45 (0.50)	6.50 (0.50)
Practice problems were clear	6.86 (0.35)	6.82 (0.39)	6.75 (0.43)
were interesting	6.43 (0.73)	6.73 (0.45)	7.00 (0.00)
can be taught at secondary-school level	6.14 (0.64)	5.36 (1.72)	6.13 (0.60)

The comparison across locations is not intended as an evaluative ranking but as an analytical tool to identify context-sensitive variations in teacher attitudes and implementation conditions. Given that large-scale adoption will necessarily occur across heterogeneous school types, understanding this variability is essential for interpreting feasibility and scalability.

Preparing Teachers to Teach the Parallel Approach

This section examines how teacher confidence in teaching Bayesian inference increased with more comprehensive seminar coverage (Table 5). Since one goal of our approach is to deepen understanding of classical hypothesis testing by incorporating the Bayesian view, we asked teachers what additional support they need to teach it (Table 6). Interestingly, teachers rated their understanding of the material higher than their ability to teach it, and even lower were their expectations for student comprehension (Table 7). Another key factor in teaching inference is the integration of digital tools (Table 8).

Table 5 Preparedness to teach classical and Bayesian statistics after phase 2–3 seminar — *M* (SD) of judgements

Teacher feels prepared to teach the basics of	Budapest $N=7$	Békéscsaba $N=11$	Miskolc $N=8$
Classical hypothesis testing	5.86 (0.64)	5.18 (1.11)	6.25 (0.66)
Bayesian statistics	4.17 (1.34)	5.09 (0.90)	5.75 (0.66)

Table 6 Needs for teaching (phase 2–3 selection; numbers represent similar answers; $N=26$)

What else do you need to feel prepared to teach...	
classical hypothesis testing?	basics of Bayesian statistics?
"More practice; independent work." (26)	"Additional practice; sample problems." (11)
"Solving practical tasks on my own, … get routine."	"Lesson plans, more time." (5)
"Knowledge of [mathematical] background." (5)	"Theoretical background." (5)
"to discuss my thoughts and experience before and during teaching …" (5)	"Theoretical and practical deepening."
"… a variety of examples from different fields could be useful." (8)	"reach a routine … not only understand other's solutions …"
"time … to go more deeply into theory, tasks and terminology."	"Since I have not encountered this before, it takes some time to accept it."
"… several problems interesting to students, not requiring too long calculations."	"To be 'kids playing' through tasks … was not possible in the tight framework. Lots of homework …"
"detailed lesson plans; a task repository."	"Better preparedness of students in probability; computer labs."

Table 7 Self-report on teaching hypothesis testing on a 7-point Likert scale — phase 3 — *M* (SD)

Categories of understanding	Békéscsaba $N=11$	Miskolc $N=8$
I understood the material	5.91 (0.67)	6.25 (0.66)
I became able to teach statistical hypothesis testing	5.45 (1.08)	6.13 (1.05)
I think my students are capable to learn the material	4.55 (1.97)	6.00 (0.71)

Table 8 Teachers' perceptions of Excel — post-test results in phase 2/3 — mean (SD) of judgements

Statements about Excel	Budapest $N=9$	Békéscsaba $N=10$	Miskolc $N=9$
I am experienced in Excel	4.86 (1.12)	4.64 (1.82)	5.50 (1.87)
Excel provided in training was understandable	6.00 (0.82)	5.73 (1.66)	6.38 (0.86)
It is a good idea to use Excel in class	5.86 (1.12)	5.55 (1.23)	6.38 (0.70)

Confidence to Teach Bayesian Inference Increases with Greater Coverage

As the materials evolved and seminar conditions varied — duration, structure, teacher profiles, and class size — we interpret the differences rather than pool results. Key differences between phase 2 (Budapest) and phase 3 (Békéscsaba and Miskolc) include seminar length (8 vs. 16 h) and the depth of Bayesian content. Phase 2 introduced the Bayesian formula within medical diagnostics, but did not explicitly link it to hypothesis testing. In phase 3, these links were made clear: classical test problems were reformulated and solved using Bayesian methods. Extended seminar time, additional examples, and supporting materials enabled more thorough instruction.

Table 5 summarises feedback on preparedness to teach classical and Bayesian inference. Confidence in teaching classical inference was highest in Miskolc (6.25), followed by Budapest (5.86) and Békéscsaba (5.18). Confidence in teaching Bayesian inference also increased: 5.09 in Békéscsaba and 5.75 in Miskolc, compared to 4.17 in Budapest. Teacher backgrounds may explain the variation. In Budapest, five of seven respondents had prior exposure to classical hypothesis testing. Still, they struggled to connect the Bayesian formula with statistical tests, suggesting that effective teaching of Bayesian methods requires not only conceptual explanations, but also experience, examples, and explicit links to classical concepts.

The data suggests that sufficient time and explicit exploration of Bayesian problems (this differs Budapest from the other two) lead the teachers to feel as confident teaching Bayesian topics as classical inference, regardless of their level of expertise.

Additional Needs of Teachers to Feel Empowered to Teach Classical and Bayesian Inference

We illustrate the pattern of responses from the post-seminar questionnaire (Table 6).

Teacher reflections after the Bayes module (Table 6) show that strong conceptual understanding is key to confident teaching. They stressed the need for clear, gap-free explanations and noted feeling more confident addressing student questions in other mathematics areas. In statistics, uncertainty arises when formulas rely on simulations rather than clear reasoning. Teachers were unsure where to find this knowledge, e.g. books, colleagues, or other sources. The following statements affirm that including the Bayesian approach contributes to a deeper understanding of classical methods:

> I like teaching Bayes' theorem because it brings subjectivity to the formula. If the initial condition changes, you get a different result. It sheds light on why people make different decisions in similar cases. They have different initial impressions, different inputs throughout their lives, and accordingly evaluate the situation differently. The Bayes formula seeks to capture this (Teacher in the final plenum discussion at the Miskolc seminar).
>
> I always interpreted the classical result wrongly because I thought the confidence interval contained the estimated parameter with the given probability, […] I have understood at last what it really means. […] I *really* like the Bayesian method because I saw for the first time why people have different opinions in many cases. (Pre-service teacher students from an earlier study; Vancsó, 2009, p. 199)

Different Levels of Teachers' Confidence: Understanding, Teaching, and Belief of Students' Understanding

We investigated teachers' needs and perceived barriers to teaching inferential statistics, which will shape the final curriculum. Table 7 presents three phase 3 questions (not asked in Budapest), where Békéscsaba scores were consistently lower than Miskolc's: *understanding the material* (5.91 vs. 6.25), *ability to teach the topic* (5.45 vs. 6.13), and *students' perceived comprehension* (4.55 vs. 6.00). These differences likely reflect the school types: most Békéscsaba teachers taught at average high schools, while Miskolc participants were from top gymnasiums.

Disregarding city-specific differences, the trend across categories is striking: teachers consistently rate their own understanding of the material higher than their confidence in teaching it. Most notably, they assess their students' ability to grasp the content considerably lower still.

$$\textit{I understood the material} \succ \textit{I am able to teach} \succ \textit{My students are capable to learn}$$

The systematic decrease from self-understanding to teaching confidence and further to perceived student capability suggests that teachers may partly project their own epistemological difficulties onto their students. Statistical inference differs from traditional school mathematics in that it relies not exclusively on deductive closure but also on modelling, simulation, contextual reasoning, and analogy. For teachers socialised in a proof-oriented mathematical culture, such reasoning may appear conceptually incomplete. Students, however, may not experience these epistemological tensions in the same way and may more readily accept simulation-based or analogy-driven explanations. Thus, the gap in perceived student capability may reflect teachers' disciplinary expectations rather than actual learner limitations. When analysing student feedback (Fejes-Tóth, 2025; Fejes-Tóth et al., 2022), the data show that, in general, students are capable and willing to master the topic.

Technology Enhances Statistical Teaching Competence

We introduced a method for teaching inferential statistics using software and gathered teacher feedback on its adoption. Table 8 shows that Miskolc teachers rated the approach about one point higher on the 7-point Likert scale, likely due to the high number of informatics-trained teachers (four of eight had computer science as a second subject). Despite varying Excel proficiency (4.64–5.50), teachers found the seminar content accessible (5.73–6.38) and viewed Excel use in mathematics classes positively (5.55–6.38) — ratings roughly one point above their self-assessed Excel skills.

Some teachers expressed a need for support, as they were unfamiliar with creating spreadsheets. However, after a short introduction, they felt more confident using the materials in class. One teacher from Miskolc remarked: "The supporting materials provided (especially Excel) are very useful even for students. I could not have prepared them myself." This comment, from the youngest participant, highlights a broader issue: preservice training often fails to equip even recent graduates with software skills. Compared to other subjects, mathematics lags behind — biology teachers, for instance, routinely use Excel in university-level statistics. We therefore recommend incorporating software-based training into teacher education.

Following the phase 3 seminars, we distributed a questionnaire and conducted semi-structured interviews. Results showed that even Excel presented difficulties, leading us to expand training. A small but notable group — often mathematically strong — remains unconvinced that teaching inference requires omitting detailed mathematical proofs.[3] Engaging these teachers is vital. We propose bridging gaps

[3] Teaching inference requires the use of simulation technique for illustrating essential features of the concepts (and methods), we have to justify concepts, e.g. by analogy of other intuitive arguments, we cannot develop the full mathematical interrelations. As useful as these approaches are, not all teachers acknowledge them as mathematical arguments. Notably, teachers who are more mathematically inclined (gifted) follow such belief systems. We have to convince these teachers so that they finally will teach inference with the same enthusiasm as other areas of mathematics. This might not be only a Hungarian specialty.

through meta-knowledge, analogies, simulations, and digital tools, which can support understanding by prioritising core concepts over complex calculations.

The findings suggest the need for a structured two-level approach: (1) compulsory foundational training in educational software within initial teacher education, and (2) dedicated coursework on pedagogical integration of digital tools for simulation-based and concept-oriented instruction. Strengthening digital competence is not merely a technical matter but supports teachers in accepting non-proof-based reasoning approaches central to statistical inference.

Summary, Discussion, and Conclusions

Our research aims to open a new era in Hungarian education by incorporating inferential statistics into secondary school curricula. We summarise our approach, which includes curriculum development, teacher seminars, feedback iterations, and educational impact.

Research Methodology and Curriculum Development

Our curriculum aimed to make inferential statistics accessible and engaging for novice teachers through hands-on experiments, Excel-supported simulations, and supplementary materials. Initially, four teachers and 64 students participated, with improvements made based on feedback. Semi-structured interviews offered qualitative insights that informed revisions. Post-implementation interviews revealed that teachers requested more practical examples, tutorials, and longer seminars. Notably, comprehensive texts like Freedman et al. (2007) dedicate substantial space to exercises, underscoring the value of applied practice.

Teacher Seminars, Perceptions, and Challenges

The second seminar in 2022, attended by nine teachers, featured an expanded curriculum and improved Excel-based simulations based on phase 1 feedback. Questionnaire evaluations offered critical insights that guided revisions. We learned that teaching inference through a parallel classical–Bayesian path requires time, structured materials, and detailed task discussions. The third phase (2023) focused on analysing prior feedback to improve in-class implementation. This iterative process ensures the curriculum evolves to meet teacher and student needs.

Teacher perceptions revealed varied levels of mathematical background. Teachers from vocational schools in Békéscsaba and grammar schools in Budapest differed in software experience. Still, all found the content understandable and engaging. Time constraints in phase 2 limited Bayesian coverage, with focus on hypothesis testing. In phase 3, we devoted more time to a task system that highlights relations between classical and Bayesian inference. Teachers affirmed the value of teaching both approaches, given sufficient support and clear, worked-out examples. Most prefer introducing classical methods first and adding Bayesian concepts later. Well-structured tasks are essential for clarifying both approaches. Although few teachers had formal training in inference, their openness to teaching classical tests was promising. Differences in familiarity with digital tools affected teachers' readiness to use them, with less experienced users requesting support. Simulations and hands-on activities were well received. Future seminars will train teachers to mentor peers, promoting a cascade for widespread curriculum adoption.

Conclusion and Future Steps

Based on teacher feedback from interviews and seminar discussions, we now have a clearer understanding of:

- Where teachers remain unconvinced about including statistical inference.
- What supplementary materials they require for instruction.
- How to design Excel tools that are practical and classroom ready.
- How to provide accessible background knowledge in place of overly complex mathematics.

What Do Teachers Need to Teach Statistical Inference Effectively?

- Teachers can teach statistical inference, but they need support (Table 6).
- A textbook suited to the high school level and teacher education—advanced texts alone are insufficient.
- Teaching materials that present concepts in a clear, teachable format.
- A broader range of tasks from diverse contexts to demonstrate the value and applicability of statistical methods.

What Are the Limitations of Our Current Approach?

- Increasing the diversity of task contexts.
- Providing fully worked, annotated solutions to tasks.

Research Question 1: To What Extent Is the Suggested Curriculum for Teaching Statistical Interference Acceptable and Feasible for Teachers?

General Acceptance and Feasibility Teachers generally responded positively to statistical inference, the experimental seminars, and the materials presented (Table 2). With sufficient time and resources, they consider statistical inference feasible to teach, even without prior background. However, they emphasised the need for more materials, especially varied contextual tasks.

Statistics and Mathematics in School Culture In the Hungarian context, teaching statistics without full mathematical rigour remains difficult. Teachers request more background and fully worked, annotated solutions (Table 6). Yet, current textbooks offer little support. Meta-level tools such as analogies, simulations, and paradigmatic task systems may address this gap, as Table 3 suggests.

Convincing Teachers of Exam Suitability Teachers must be convinced not only of the topic's value, but also its accessibility for students. Our curriculum emphasises simulations over traditional analytic methods, diverging from teachers' expectations. Table 2 shows overall support; Table 3 highlights concerns over missing mathematical justifications; Table 4 shows strong endorsement for teaching the topic at the secondary level.

Research Question 2: Is the Bayesian Way in Parallel to Classical Inference Useful to Understand Concepts of Either Side?

Explicitly addressing classical hypothesis testing alongside Bayesian inference makes both approaches more accessible to teachers. Contexts such as coin tossing and medical diagnosis effectively illustrate key elements of hypothesis testing and their relation to Bayes' theorem. This parallel approach enhances conceptual understanding from both perspectives. Table 2 reflects generally positive teacher feedback on integrating Bayesian ideas into instruction. In particular, discussions following the coin-tossing task system significantly clarified core concepts for many teachers.

Research Question 3: What Additional Support Do Teachers Need for Teaching Statistical Inference?

Teachers emphasised the need for more detailed and explicit materials to build the confidence required to teach statistical inference effectively (Table 6). A suitable textbook in Hungarian is lacking. They also expressed the need for software support. Even those experienced in software valued Excel sheets that are practical and ready for classroom use. These positive findings should be understood within the context of Hungarian education, where teaching statistical inference requires a shift in instructional culture away from the traditionally mathematics-centred approach.

What Implications Can Be Drawn from Our Findings?

- Teacher education must evolve. Mathematics-focused courses alone are insufficient. Probability and statistics — especially inference — should be embedded in a supportive teaching–learning environment. Future teacher training must also go beyond simplified statistics courses, enabling teachers to teach interactively using digital support.
- Teacher reflections after the Bayes module (Table 6) show that strong conceptual understanding is key to confident teaching. They stressed the need for clear, gap-free explanations and noted feeling more confident addressing student questions in other mathematics areas. In statistics, uncertainty arises when formulas rely on simulations rather than clear reasoning. Teachers were unsure where to find this knowledge—books, colleagues, or other sources.
- Their feedback affirms that introducing the Bayesian perspective helps deepen understanding of classical inference, supporting our aim to enhance teacher competence through parallel instruction.
- Teachers' belief systems need broadening. Their views on what constitutes a valid basis for teaching statistical inference must expand to include a wider range of conceptual approaches.
- Inference can be taught in an interdisciplinary way, progressing from real-world problems to model construction and contextual interpretation.
- The medical-diagnostic context, explored through both classical and Bayesian lenses, supports a deeper understanding of the concepts and their limitations.
- Task systems linking combinatorics, probability, and inference help develop coherent learning trajectories. These systems enhance teachers' conceptual understanding and demonstrate their practical teaching value.
- Excel tools are most effective when integrated into teacher seminars with proper documentation, explaining didactical purpose, required inputs, and output interpretation. This ensures they serve as direct and practical teaching aids.

A key lesson from Varga's reform is to introduce changes gradually, supported by teacher training. We propose expanding training through a 30-h accredited programme, matching teachers' capacity and learning pace. Emphasis is placed on strengthening combinatorial reasoning in hypothesis testing and connecting classical with Bayesian inference. Although our study was conducted in Hungary, the designed-oriented approach to curriculum development has broader relevance. Local adaptation remains crucial due to curricular traditions and attitudes. Our findings support strong pre-service teacher education and offer a model for improving statistics education internationally.

Acknowledgements This study was partly funded by the *Research Programme for Public Education Development* of the *Hungarian Academy of Sciences* (KOZOKT2021-16).

Funding Open access funding provided by Hungarian University of Agriculture and Life Sciences.

Declarations

Competing Interests The authors declare no competing interests.

References

Albert, J. (1995). Teaching inference about proportions using Bayes and discrete models. *Journal of Statistics Education, 3*(3), 1-16. https://doi.org/10.1080/10691898.1995.11910494

Albert, J. (1997). Teaching Bayes' rule: A data-oriented approach. *American Statistician, 51*(3), 247–253. https://doi.org/10.2307/2684896

Albert, J. (2002). Teaching introductory statistics from a Bayesian perspective. In B. Phillips (Ed.), *Developing a statistically literate society: Proceedings of the Sixth International Conference on Teaching Statistics (ICOTS6).* ISI/IASE. https://icots.info/documents/papers/icots6/3f1_albe.pdf?1402524960

Alpana, S. (2017). Role of statistics in different fields. *Research Journal of Science and Technology, 9*(1), 118–122. https://doi.org/10.5958/2349-2988.2017.00018.3

Aquilonious, B. C., & Brenner, M. E. (2015). Students' reasoning about p-values. *Statistics Education Research Journal, 14*(2), 7–27. https://doi.org/10.52041/serj.v14i2.259

Artigue, M. (2002). Learning mathematics in a CAS environment: The genesis of a reflection about instrumentation and the dialectics between technical and conceptual work. *International Journal of Computers for Mathematical Learning, 7*(3), 245–274. https://doi.org/10.1023/A:1022103903080

Artigue, M., & Blomhøj, M. (2013). Conceptualizing inquiry-based education in mathematics. *ZDM–Mathematics Education, 45*(6), 797–810. https://doi.org/10.1007/s11858-013-0506-6

Barnett, V. (1999). *Comparative statistical inference* (3rd ed.). Wiley. https://doi.org/10.1002/9780470316955

Batanero, C., Biehler, R., Maxara, C., Engel, J., & Vogel, M. (2005). Using simulation to bridge teachers' content and pedagogical knowledge in probability. In Proceedings of the 15th ICMI Study Conference: The Professional Education and Development of Teachers of Mathematics (pp. 1–6). https://www.mathunion.org/fileadmin/ICMI/files/Conferences/ICMI_studies/Study15/BataneroCarmen_ICMI15_prop.doc

Batanero, C., Begué, N., Borovcnik, M., & Gea, M. (2020). Ways in which high school students understand the sampling distribution for proportions. *Statistics Education Research Journal, 19*(3), 32–52. https://doi.org/10.52041/serj.v19i3.55

Batanero, C., & Borovcnik, M. (2016). *Statistics and probability in high school*. Sense Publishers. https://brill.com/display/title/37568

Batanero, C., Díaz, C., & Sánchez, E. (2015). Prospective teachers' semiotic conflicts in computing probabilities from a two-way table. *International Electronic Journal of Mathematics Education, 10*(1), 3−16. https://doi.org/10.29333/iejme/288

Berg, A., & Hawila, N. (2021). Introducing Bayesian inference with the taxicab problem. In A. A. Bilgin & S. Budgett (Eds.), *Proceedings of the Tenth Australian Conference on Teaching Statistics* (pp. 55–60). Statistical Society of Australia and The New Zealand Statistical Association. https://www.statsoc.org.au/Past-Conferences

Berry, D. A. (1997). Teaching elementary Bayesian statistics with real applications in science. *The American Statistician, 51*(3), 241–246. https://doi.org/10.2307/2684895

Biehler, R. (1991). Computers in probability education. In R. Kapadia & M. Borovcnik (Eds.), *Chance encounters: Probability in education*. (pp. 169–211). Kluwer. https://doi.org/10.1007/978-94-011-3532-0_6

Borovcnik, M. (2011). Strengthening the role of probability within statistics curricula. In C. Batanero, G. Burrill, & C. Reading (Eds.), Teaching statistics in school mathematics – Challenges for teaching and teacher education. Springer, (pp. 71–83). https://doi.org/10.1007/978-94-007-1131-0_11

Borovcnik, M. (2021). The meaning of probability from a foundational perspective. *Revista Sergipana de Matemática e Educação Matemática, 6*(3), 42–76. https://doi.org/10.34179/revisem.v6i3.14904

Bruner, J. S. (1961). The act of discovery. *Harvard Educational Review, 31*(1), 21–32. Reprinted as chapter in J. S. Bruner (2006). *In search of pedagogy, The selected works of Jerome Bruner, 1957-1978*. https://doi.org/10.4324/9780203088609

Bureau of Labor Statistics (n.d.). Fastest growing jobs. *Occupational Outlook Handbook*. https://www.bls.gov/ooh/fastest-growing.htm

Burrill, G., & Biehler, R. (2011). Fundamental statistical ideas in the school curriculum and in training teachers In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics – Challenges for teaching and teacher education. New ICMI Study Series* (Vol 14, pp. 57–69). Springer. https://doi.org/10.1007/978-94-007-1131-0_10

Carranza, P., & Kuzniak, A. (2008). Duality of probability and statistics teaching in French education. In C. Batanero, G. Burrill, C. Reading, & A. Rossman (Eds.), *Joint ICMI/IASE study: Teaching statistics in school mathematics – Challenges for teaching and teacher education.* IASE https://icots.info/documents/papers/rt2008/T1P2_Carranza.pdf?1402524989

Case, C., & Jacobbe, T. (2018). A framework to characterize student difficulties in learning inference from a simulation–based approach. *Statistics Education Research Journal, 17*(2), 9–29. https://doi.org/10.52041/serj.v17i2.156

Cernusca, D., & Ionas, I. G. (2014). Design-based research as a form of action research. In J. W. Willis, & C. L. Edwards (Eds.), *Action research: Models, methods, and examples* (pp. 195–220). Information Age Publishing.

Clements, D. H., & Battista, M. T. (1992). Geometry and spatial reasoning. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 420–464). Macmillan.

Cobb, G. W. (2007). The introductory statistics course: A Ptolemaic curriculum? *Technology Innovations in Statistics Education, 1*(1), 1–15. https://doi.org/10.5070/T511000028

del Már López-Martín, M., Batanero, C., & Gea, M. M. (2019). Conocen los futuros profesores los errores de sus estudiantes en la inferencia estadística? [Do prospective teachers know their students' errors in statistical inference?]. *Bolema, Rio Claro (SP), 33* (64), 672–693. https://doi.org/10.1590/1980-4415v33n64a11

delMas, R. (2017). A 21st century approach towards statistical inference – Evaluating the effects of teaching randomization methods on students' conceptual understanding [Paper presentation]. *61st International Statistical Institute World Statistics Congress*, Marrakesh, Morocco.

delMas, R., Garfield, J., Ooms, A., & Chance, B. (2007). Assessing students' conceptual understanding after a first course in statistics. *Statistics Education Research Journal, 6*(2), 28–58. https://doi.org/10.52041/serj.v6i2.483

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining "gamification". *In Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments* (pp. 9–15). Association for Computing Machinery. https://doi.org/10.1145/2181037.2181040

Dinges, H., & Rost, H. (1982). *Prinzipien der Stochastik*. Teubner.

Dobos, S., Ocskó, E., & Vásárhelyi, E. (2001). National Presentation Hungary. *Reference levels in school mathematics education in Europe*. European Mathematical Society, Committee on Mathematics Education. http://www.emis.de/

Falk, R. (1989). Inference under uncertainty via conditional probabilities. In R. Morris (Ed.), *Studies in Mathematics Education,* (Vol. 7): *The teaching of statistics* (pp. 175–184). UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000097820

Fejes-Tóth, P. (2020). Inferential statistics – Research on the introduction of a new topic in the Hungarian high-school curriculum. In G. Ambrus, J. Sjuts, Ö. Vancsó, & É. Vásárhelyi (Eds.), *Komplexer Mathematikunterricht: Die Ideen von Tamás Varga in aktueller Sicht* (pp. 157–180). WTM. https://doi.org/10.37626/GA9783959871648.0.10

Fejes-Tóth, P. (2025). Inferential statistics in Hungarian secondary schools – Results of an experiment. PhD dissertation. Eötvös Loránd University Budapest. https://doi.org/10.15476/ELTE.2025.012

Fejes-Tóth, P., Vancsó, Ö., & Borovcnik, M. (2022). Combinatorial thinking as key for introducing hypothesis testing – Evaluation of the planned secondary-school reform in Hungary. In S.A. Peters, (Ed.), *Bridging the gap: Empowering and educating today's learners in statistics*. International Association for Statistical Education. https://icots.info/icots/11/proceedings/pdfs/ICOTS11_229_FEJESTTH.pdf?1669865538

Fisher, R. A. (1971). *The design of experiments* (9th ed.). Oliver & Boyd. (Original work published 1935). https://home.iitk.ac.in/~shalab/anova/DOE-RAF.pdf

Freedman, D. A. (1997). Discussion of "Bayes for Beginners? Some Reasons to Hesitate". *The American Statistician, 51*(3) 266–268 https://doi.org/10.2307/2684901

Freedman, D., Pisani, R., Purves, R. (2007). Statistics (4th ed.). W.W. Norton.

Freudenthal, H. (1973). *Mathematics as an Educational Task.* Reidel.

Freudenthal, H. (1981). Major problems of mathematics education. *Educational Studies in Mathematics, 12*(2), 133–150. https://doi.org/10.1007/BF00305618

Gee, J. P. (2003). What video games have to teach us about learning and literacy. *Computers in Entertainment, 1*(1), 20–20. https://doi.org/10.1145/950566.950595
Gigerenzer, G. (2003). *Calculated risks: How to know when numbers deceive you*. Simon & Schuster.
Gordon Győri, J., Fried, K., Köves, G., Oláh, V., & Pálfalvi, J. (2020). The traditions and contemporary characteristics of mathematics education in Hungary in the post-socialist era. In A. Karp (Ed.), *Eastern European mathematics education in the decades of change.* Springer. https://doi.org/10.1007/978-3-030-38744-0_3
Gosztonyi, K. (2023). The New Math in Hungary: Tamás Varga's Complex Mathematics Education Reform. In D. De Bock (Ed.), *Modern mathematics: History of mathematics education* (pp. 285–301). Springer. https://doi.org/10.1007/978-3-031-11166-2_14
Gravemeijer, K., & Doorman, M. (1999). Context problems in realistic mathematics education: A calculus course as an example. *Educational Studies in Mathematics, 39*(1–3), 111–129. https://doi.org/10.1023/A:1003749919816
Harradine, A., Batanero, C., & Rossman, A. (2011). Students and teachers' knowledge of sampling and inference. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics–challenges for teaching and teacher education.* (pp. 235–246). Springer. https://doi.org/10.1007/978-94-007-1131-0_24
Heshmati, S., Kersting, N., & Sutton, T. (2018). Opportunities and challenges of implementing instructional games in mathematics classrooms: Examining the quality of teacher-student interactions during the cover-up and un-cover games. *International Journal of Science and Mathematics Education, 16*(4), 777–796. https://doi.org/10.1007/s10763-016-9789-8
Hoegh, A. (2020). Why Bayesian ideas should be introduced in the statistics curricula and how to do so. *Journal of Statistics Education, 28*(3), 222–228. https://doi.org/10.1080/10691898.2020.1841591
Hoffrage, U., Krauss, S., Martignon, L., & Gigerenzer, G. (2015). Natural frequencies improve Bayesian reasoning in simple and complex inference tasks. *Frontiers in Psychology, 6*, 1473. https://doi.org/10.3389/fpsyg.2015.01473
Hubbard, R., & Bayarri, M. J. (2003). Confusion over measures of evidence (p) versus errors (α) in classical statistical testing. *The American Statistician 57*(3), 171–182. https://doi.org/10.1198/0003130031856
Johnson, D. E. (1989). An intuitive approach to teaching analysis of variance. *Teaching of Psychology, 16*(2), 67–68. https://doi.org/10.1207/s15328023top1602_5
Kang, J., Keinonen, T., Simon, S., Rannikmäe, M., Soobard, R., & Direito, I. (2019). Scenario evaluation with relevance and interest (SERI): Development and validation of a scenario measurement tool for Context-Based Learning. *International Journal of Science and Mathematics Education, 17*(7), 1317–1338. https://doi.org/10.1007/s10763-018-9930-y
Kieran, C., Doorman, M., & Ohtani, M. (2015). Frameworks and principles for task design. In A. Watson & M. Ohtani (Eds.), *Task design in mathematics education. New ICMI Study Series* (pp. 19–81). Springer International. https://doi.org/10.1007/978-3-319-09629-2_2
Kiss, A. (2022). Complex Mathematics Education: An integrated and inquiry-based mathematics teaching method. *Canadian Journal of Science, Mathematics and Technology Education, 22*(4), 758–772. https://doi.org/10.1007/s42330-022-00250-1
Klein, D. (2003). A brief history of American K–12 mathematics education in the 20th century. In J. M. Royer (Ed.), *Mathematical cognition (pp. 177–225).* Information Age Publishing.
Kruschke, J. K. (2010). What to believe: Bayesian methods for data analysis. *Trends in Cognitive Sciences, 14,* 293–300. https://doi.org/10.1016/j.tics.2010.05.001
Lane-Getaz, S. J. (2017). Is the p-value really dead? Assessing inference learning outcomes for social science students in an introductory statistics course. *Statistics Education Research Journal, 16*(1), 357–399. https://doi.org/10.52041/serj.v16i1.235
Lu, C. (2020). Channels' confirmation and predictions' confirmation: From the medical test to the Raven Paradox. *Entropy 22*(4), 384. https://doi.org/10.3390/e22040384
Lugo-Armenta, J. G., & Pino-Fan, L. R. (2021). Inferential reasoning of secondary school mathematics teachers on the Chi-Square statistic. *Mathematics, 9*(19), 2416. https://doi.org/10.3390/math9192416
Maaß, K., & Artigue, M. (2013). Implementation of inquiry-based learning in day-to-day teaching: a synthesis. *ZDM–Mathematics Education, 45*, 779–795. https://doi.org/10.1007/s11858-013-0528-0
Majgaard, G., Misfeldt, M., & Nielsen, J. (2011). How design-based research and action research contribute to the development of a new design for learning. *Designs for Learning, 4*(2), 8–27. https://doi.org/10.16993/dfl.38
Mellers, B. A., & McGraw, P. (1999). How to improve Bayesian reasoning: Comment on Gigerenzer and Hoffrage (1995). *Psychological Review, 106*(2), 417–424. https://doi.org/10.1037/0033-295X.106.2.417
Mercer, N., & Littleton, K. (2007). *Dialogue and the development of children's thinking: A sociocultural approach*. Routledge. https://doi.org/10.4324/9780203946657
Migon, H. S., & Gamerman, D. (1999). *Statistical inference: An integrated approach*. Chapman & Hall/CRC.
Ministry of Education of Hungary (2021). *National curricula in mathematics at secondary level*. Budapest.
Moore, D. S. (1997). Bayes for beginners? Some reasons to hesitate. *The American Statistician, 51*(3), 254–261. https://doi.org/10.1080/00031305.1997.10473972
Myung, J. I., Karabatsos, G., & Iverson, G. J. (2008). A statistician's view on Bayesian evaluation of informative hypotheses. In H. Hoijtink, I. Klugkist, & P. A. Boelen (Eds.), *Bayesian evaluation of informative hypotheses* (pp. 309–328). Springer. https://doi.org/10.1007/978-0-387-09612-4_15

Nilsson, P. (2020). Students' informal hypothesis testing in a probability context with concrete random generators. *Statistics Education Research Journal, 19*(3), 53–73. https://doi.org/10.52041/serj.v19i3.56

Novotná, J., Gosztonyi, K., Hošpesová, A., Fried, K., Moraová, H., Gordon Győri, J. (2021). Mathematics teacher education in the Czech Republic and Hungary: Commonalities and differences. *ZDM–Mathematics Education, 53*,1541–1553. https://doi.org/10.1007/s11858-021-01278-7

Papp, T. A. (2017). Gamification effects on motivation and learning: Application to primary and college students. *International Journal for Cross-Disciplinary Subjects in Education, 8*(3), 3193–3201. https://doi.org/10.20533/ijcdse.2042.6364.2017.0428

Pöhler, B. (2020). Role of facilitators in supporting teacher collaboration during PD courses on language responsive mathematics teaching. In H. Borko & D. Potari (Eds.), *ICMI Study 25: Teachers of mathematics working and learning in collaborative groups* (pp. 516–523). University of Lisbon. https://www.mathunion.org/fileadmin/ICMI/Publications/ICMI%20Study%2025/updated%20DD/201114%20ICMI25Proceedings6.13.2020.pdf

Rossman, A. J., & Chance, B. L. (1999). Teaching the reasoning of statistical inference: A "top ten" list. *The College Mathematics Journal, 30*(4), 297–305. https://doi.org/10.1080/07468342.1999.11974074

SERC portal for educators (n.d.). Why teach with simulations. *Pedagogy in Action*. Science Education Resource Center at Carleton College. https://serc.carleton.edu/sp/library/simulations/why.html

Stein, M. K., Grover, B. W., & Henningsen, M. (1996). Building student capacity for mathematical thinking and reasoning: An analysis of mathematical tasks used in reform classrooms. *American Educational Research Journal, 33*(2), 455–488. https://doi.org/10.2307/1163292

Thompson, P. W., Liu, Y., & Saldanha, L. A. (2007). Intricacies of statistical inference and teachers' understandings of them. In M. Lovett & P. Shaw (Eds.), *Thinking with data* (pp. 207–231). Erlbaum

Truran, J. (1995). *Probability & statistics: A tale of uncertainty*. Paper presentation. AARE Annual Conference, Hobart, Australia. Australian Association for Research in Education. https://www.aare.edu.au/data/publications/1995/trurj95331.pdf

Vallecillos, A. (1999). Some empirical evidences on learning difficulties about testing hypotheses. In *Proceedings of the 52nd Session of the International Statistical Institute*. International Statistical Institute. https://icots.info/documents/papers/isi52/vall0682.pdf?1402524975

Vancsó, Ö. (2009). Parallel discussion of classical and Bayesian ways as an introduction to statistical inference. *International Electronic Journal of Mathematics Education, 4*(3), 181–212. https://doi.org/10.29333/iejme/242

Vancsó, Ö, Beregszászi, E., Burian, H., Emese, G., Stettner, E., & Szitányi, J. (2018). Complex Mathematics Education in the 21st century: Improving combinatorial thinking based on Tamás Varga's heritage and recent research results. In E. W. Hart & J. Sandefur (Eds.), *Teaching and learning discrete mathematics worldwide: Curriculum and research* (pp. 111–134). Springer. https://doi.org/10.1007/978-3-319-70308-4_8

Vancsó, Ö., Borovcnik, M., & Fejes-Tóth, P. (2021). A complex concept about statistical inference and a planned school experiment based on it. In *Proceedings of the 63rd World Statistics Congress*. International Statistical Institute. https://isi-web.org/sites/default/files/import/pdf/154-day3-ips078-a-complex-concept-about-statis.pdf

Van de Walle, J. A., Karp, K. S., & Bay-Williams, J. M. (2019). *Elementary and middle school mathematics: Teaching developmentally* (10th ed). Pearson. https://www.pearson.com/nl/en_NL/higher-education/subject-catalogue/Education/Elementary-and-Middle-School-Mathematics-Teaching-Developmentally-Global-Edition-10E.html

Varga, T. (1972). Logic and probability in the lower grades. *Educational Studies in Mathematics, 4*(3), 346–357. https://doi.org/10.1007/BF00302583

Varga, T. (1975). *Komplex matematikatanítás [Complex mathematics teaching]*. PhD dissertation. The Hungarian Academy of Sciences.

Varga, T. (1983). Statistics in the curriculum for everybody: How young children and how their teachers react. In D. R. Grey, P. Holmes, V. Barnett, & G. M Constable (Eds.), *Proceedings of the First International Conference on Teaching Statistics* (Vol. 1, pp. 71–80). Teaching Statistics Trust. https://icots.info/documents/papers/icots1/ICOTS1.pdf?1454729073

Wagenmakers, E. J., Lee, M., Lodewyckx, T., & Iverson, G. J. (2008). Bayesian versus frequentist inference. In H. Hoijtink, I. Klugkist, & P. A. Boelen (Eds.), *Bayesian evaluation of informative hypotheses* (pp. 181–207). Springer. https://doi.org/10.1007/978-0-387-09612-4_9

Wickmann, D. (1991). Bayes-Statistik: Einsicht gewinnen und entscheiden bei Unsicherheit. Bibliographisches Institut. https://de.book-info.com/isbn/3-411-14671-0.htm

Witmer, J., Short, T. H., Lindley, D. V., Freedman, D. A., & Scheaffer, R. L. (1997). Teacher's corner. Discussion of papers by D. A. Berry, J., Albert, & D. S. Moore. *The American Statistician, 51*(3), 262–274. https://doi.org/10.1080/00031305.1997.10473973

Wolpers, H., & Götz, S. (2002). *Mathematikunterricht in der Sekundarstufe II* (Vol. 3). *Didaktik der Stochastik*. Vieweg. https://doi.org/10.1007/978-3-322-83144-6