



HAL
open science

Reimagining Statistics and Probability Education in the Age of Data and AI: An introduction to the work of TWG05 Leader

Aisling Leavy, Daniel Frischemeier, Sibel Kazak, Orlando Rafael González, Luca Lamanna, María Magdalena Gea

► **To cite this version:**

Aisling Leavy, Daniel Frischemeier, Sibel Kazak, Orlando Rafael González, Luca Lamanna, et al.. Reimagining Statistics and Probability Education in the Age of Data and AI: An introduction to the work of TWG05 Leader. Proceedings of the Fourteenth Congress of the European Society for Research in Mathematics Education (CERME14), Free University of Bozen-Bolzano; ERME, Feb 2025, Bozen-Bolzano, Italy. <hal-05330897>

HAL Id: hal-05330897

<https://hal.science/hal-05330897v1>

Submitted on 25 Oct 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Copyright - All rights reserved

Reimagining Statistics and Probability Education in the Age of Data and AI: An introduction to the work of TWG05

Leader: Aisling Leavy (*Ireland*)

Co-leaders: Daniel Frischemeier (*Germany*), Sibel Kazak (*Türkiye*), Orlando Rafael González (*Thailand*), Luca Lamanna (*Italy*), and María Magdalena Gea (*Spain*)

Introduction

This report presents a summary of the activities and key outcomes of Thematic Working Group 05 (TWG05): Statistics and Probability Education at CERME14, held in Bolzano from February 4–8, 2025. The group brought together 30 participants representing 13 countries, who contributed through 18 paper presentations and 6 poster sessions. Guided by CERME's core principles of Communication, Cooperation, and Collaboration, TWG05 cultivated a dynamic and inclusive atmosphere that encouraged meaningful dialogue and constructive exchange throughout the week.

The CERME14 TWG05 team was led by Aisling Leavy (Ireland), with co-leaders Daniel Frischemeier (Germany), Orlando Gonzalez (Thailand), Sibel Kazak (Türkiye), María Magdalena Gea (Spain), and Luca Lamanna (Italy).

The week began with an ice-breaker activity, "TWG05 Colleague BINGO," aimed at encouraging participants to connect and form discussion groups. Sessions were structured to promote active engagement, with each paper allocated 10 minutes for presentation followed by 5 minutes for questions. Broader discussions then took place in small groups, each facilitated by a co-leader and including the presenter. The responsibility for chairing sessions was shared among the co-leaders throughout the week, spanning multiple sessions from Tuesday to Friday, including a dedicated poster session.

Key subthemes identified for discussion across all submissions included:

- I. data science
- II. visibilizing statistical concepts
- III. teacher education
- IV. probabilistic reasoning and risk literacy

The identification of these subthemes was the result of a careful examination of the accepted papers and their thematic focus. This mapping process was significantly informed by the themes identified as emerging from CERME13 (see Primi et al., 2024), which included "teacher education," "reasoning about data," "statistical and probabilistic thinking and reasoning," and "data science." Furthermore, it drew insights from two recent and influential research publications: Batanero and Álvarez-Arroyo (2024), and Burrill and Pfannkuch (2024). Batanero and Álvarez-Arroyo (2024) synthesized prominent research in probability education, highlighting areas such as children's probabilistic knowledge, the impact of visualizations on solving conditional probability problems, teacher

education, and probability modelling. Their work also emphasized the growing interest in the concept of risk and its pedagogical link to probability, and called for more research into teachers' mathematical and didactic knowledge of probability. Complementing this, Burrill and Pfannkuch (2024) identified four major emerging trends in statistics education research: data science, visualizing statistical concepts, social statistics, and new contexts for learning. They underscored the importance of fostering statistical reasoning, argumentation, and the critical evaluation of data-based evidence, including aspects of risk, along with enhancing the accessibility of statistical concepts through interactive visualizations. Thus, the chosen CERME14 subthemes reflect these convergent and critical areas of contemporary research in statistics and probability education. Guiding questions were prepared to facilitate discussions on these subthemes, with groups assigning note-takers to document their discussions on task cards. An online shared space, Task Cards, was utilized to systematically document the content and focus of each session's discussions. This comprehensive documentation was critical for a successful and substantial concluding discussion on the final day of CERME14, where all insights gathered during the week were collectively summarized.

Data science

During the TWG05 discussions on Data Science, several key learnings and open questions emerged from the group conversations. Participants emphasized the importance of introducing data work and data literacy at an early age, acknowledging that children are already exposed to tools that use data. The consensus was that data literacy should be embedded within inquiry-based or problem-based learning, where skills emerge through answering questions. The group discussed the use of different types of data, agreeing that both "dirty" (messy) and "clean" real-world data are important for different learning goals, as real data incorporates a sense of "reality". It was suggested that learners could start by formulating questions and then be presented with existing data, and any disconnects could lead to modifying their questions. Familiar contexts, such as social media or sustainability, were deemed important for relevance, along with using examples like recommender systems to help students understand how their own data can be used for predictions, fostering an awareness of the world. Critical thinking and reflection were highlighted as essential components of Data Science activities. Regarding tools, it was noted that while non-coding applications and technologies can provide support, programming skills (e.g., Python) are important for understanding the "inner workings" of models and fostering algorithmic thinking.

The discussion also outlined the desired knowledge for an "ideal" 16-year-old learner in Data Science. Such a learner should: Be aware of algorithms' impact on their lives (e.g., in insurance); Understand that data-driven decision-making systems can have biases; Recognize that probabilities underpin these systems; Know that models and data are simplifications of the real world and that elements can be omitted; Possess critical thinking skills, asking questions about the data and what is presented to them.

Several open questions and areas needing further investigation were identified:

- **Institutional Constraints and Teacher Support:** How to overcome institutional limitations and address the vulnerability felt by teachers regarding Data Science education. The need to support teachers in using real-world data, perhaps by providing sample datasets and guidelines, was acknowledged, recognizing the significant responsibility placed on mathematics teachers.
- **The "Black Box" Concept:** When and how deeply to "scratch" or open the "black box" of algorithms. This includes understanding the underlying mathematics, acknowledging that some models are not fully explainable, and fostering algorithmic thinking through insights into the black box. It was noted that there might be both "pedagogical black boxes" (intentional simplifications for learning) and "real black boxes" (complex systems not entirely transparent).
- **Distinguishing Data Science from Statistics Education:** Clarifying the differences, particularly the role of context. In statistics, the context and potential models often guide data collection, whereas in Data Science, data and context are often pre-existing, and the task is to formulate questions based on them. The group emphasized the importance of teaching both perspectives.
- **Teacher Education:** How to effectively equip teacher education students with the necessary knowledge and skills to support Data Science thinking and learning. This includes rethinking the organization of Data Science lessons, potentially promoting more interdisciplinary teaching, and addressing constraints like the need for longer project work blocks.
- **Tools and Activities:** Developing and making available specialized tools that facilitate understanding, while also considering the benefits of general-purpose tools like Python for understanding inner workings. The challenge of ensuring learners gain insights into models when using simulations was also raised, along with the need for project-based work and activities that encourage reflection.

Visibilizing statistical concepts

Several key learnings and open questions emerged from the group conversations during the TWG05 discussions on Visibilizing Statistics Concepts. Participants emphasized the importance of fostering the ability to represent the same concept or data in different ways (i.e., transnumeration), using different types of representations, in both figural (e.g., scatterplots, tabular displays, decision trees) and non-figural form (e.g., verbal or narrative models, reflective statements, color schemes). Engaging learners in the creation or use of these visualizations, representations, or models of concepts must be done to fit the purpose of making visible previously unknown statistical concepts by the learners (like in the "modeling for" approach, which facilitates a deeper understanding of the underlying mechanisms and abstract ideas that the model represents), instead of adopting a correct/incorrect approach. The creation of visual models or graphical representations by learners, as well as their use of interactive visual data analysis tools (e.g., TinkerPlots, computer simulations), were reported by participants as effective ways to elicit and reveal students' intuitive reasoning, conceptual processes, and informal notions of statistical concepts such as inference, sampling distribution, sampling variability, central tendency, correlation, enumeration, and uncertainty. Moreover, the participants agreed that learners' interpretations of visualizations of concepts are socio-cultural and context-dependent, which allow teachers and researchers to delve into and understand more about not only about students' intuitive reasoning and conceptions, but also about the

experiences and socio-cultural frameworks and schemas that they bring to visualizations (e.g., students' underlying values, norms and semiotic conventions), which are rooted in the learners' socio-cultural background and context.

From the discussion, the new learner's needs and roles in relation to data, and how Visibilizing Statistics Concepts can be instrumental in meeting those needs and roles, were outlined. Today's learners are *data consumers*, individuals who frequently interact with happenstance data or data and visual representations produced by someone else's efforts, usually using digital tools, Software platforms, interactive visual data analysis tools, and meaningful contexts. They are also *data producers*, because they frequently engage in empirical investigations, represent and interpret their own data, and communicate their findings and conclusions, usually making them visible. In addition to these roles, today's learners are also *data prosumers*, because they often have to engage in activities in which they should act as both data producers and consumers.

The participants also highlighted the important role currently played by technology-rich visual learning environments and software platforms that allow students to create, make sense and analyze visual models (e.g., TinkerPlots) in the learning of statistics. Participants also identified a high degree of interconnectedness between activities on Visibilizing Statistics Concepts and the following elements: data science, technology, teacher education, curriculum, affect, context (authenticity), and technology. The interconnectedness between Visibilizing Statistics Concepts and these elements opens new avenues for researchers to explore and gain better understanding of students' reasoning, actions, and conceptual development in learning statistics, while engaged in activities in which statistical concepts are made visible to them.

From the discussions, participants were able to pose several open questions, from which areas needing further investigation are identified:

- How can the interconnectedness of Visibilizing Statistics Concepts with different elements (e.g., technology, context) and domains (e.g., data science, mathematics) be included in the curriculum, in order to make it possible for students to be introduced much earlier to statistical and probabilistic ideas?
- How can the use of technology tools (e.g., TinkerPlots, simulations, case blocky programming with Scratch) be used to visibilize statistical and probabilistic concepts while enhancing both the computational, statistical and probabilistic thinking of students?
- How can AI and machine learning methods (e.g., Support Vector Machine [SVM]) be used to visibilize key mathematical, statistical and probabilistic ideas while contributing to AI literacy and demystification?
- How can simulations and modeling activities be used to make students transition from “modeling of” to “modeling for”, while promoting students' reasoning with informal statistical models and probability modeling activities using technology-rich visual learning environments?
- How can researchers, teacher educators, and curriculum developers keep the pace of advances in technology, in order to be able to quickly respond to the current and future developments that will be relevant to students' needs and roles in relation to visibilizing data?

Teacher education

During the TWG05 group conversations on the theme of Teacher Education, one of the main insights was that, in order to engage meaningfully and make interesting use of statistics and probability, teachers should develop deeper contextual knowledge, e.g., in connection with other fields such as politics, medicine, or climatology - an aspect that clashes with actual, mostly outdated, curricula across the countries that do not offer much space for the exploration of real-world applications of both statistics and probability. Many of the presentations and discussions in TWG05 highlighted, on the other hand, concerns that most teachers possess limited statistical content knowledge and pedagogical content knowledge in statistics, largely because they themselves have likely not experienced high-quality teaching of statistics or probability. In connection with this, mathematics textbooks often fail to support interdisciplinary teaching - or, in general, link with other fields - and digital tools are still not widely available or easily integrated into real classroom settings.

A shared understanding emerged during the discussions that data science cannot be meaningfully implemented without the use of (digital) tools. This led to the recognition that, beyond statistical content knowledge and statistical pedagogical content knowledge, teachers also need technological pedagogical content knowledge (TPACK) to effectively incorporate digital tools into their teaching.

Another critical point concerned the need to shift teachers' beliefs about probability and statistics - from viewing them as isolated procedures to be memorized, toward understanding them as ways of thinking about data and context.

All in all, the need for more work towards the above-mentioned directions of research was highlighted, aiming at investigating the teachers' perspectives and knowledge, together with all the institutional setting and constraints, aiming towards a more meaningful teacher training and, in general, teaching of probability and statistics.

Several open questions and areas for further exploration were identified, both for practice and research. In terms of future research directions, there is a clear need for systematic investigation into the development of teachers' specialized content knowledge (SCK), pedagogical content knowledge (PCK), and technological pedagogical content knowledge (TPACK) - in both statistics and probability - building, in particular, on the foundational work of the ICMI study by Batanero et al. (2011). From a constructive and design-based research perspective, efforts should focus on developing textbooks and teaching materials that meaningfully integrate contextual knowledge, data, probability, and professional development. The goal is to create resources that support both classroom practice and in-service teacher learning. Additionally, affective aspects must be addressed, as inspiring teachers to rethink and transform their approach to teaching statistics and probability remains a key challenge - one that calls for further targeted research.

Finally, the following three fields were identified as future issues and points of contact for ways forward:

- Emerging curricular trends, such as data science and STEAM;
- Citizen and democracy education;
- Risk literacy.

Probabilistic reasoning and risk literacy

During the TWG05 group conversations on this theme, the key learnings centered around the topics, such as conceptualization of risk, recurring themes across different papers, and connection of probabilistic reasoning and risk literacy with the other subthemes. As part of stochastic education at school, risk literacy has become an emerging area. Conceptualization of risk from socio-constructivist and realist perspectives was highlighted as useful in fostering risk literacy, as they were found to be interwoven in students' reasoning. There was an agreement that risk literacy would aim at informed-decision making and assessing probabilities would be a crucial part of this process. It was noted that simulations can be used for inquiry-based learning to foster both risk literacy and probabilistic reasoning. When discussing the papers related to "probabilistic reasoning and risk literacy", the group emphasized the links with other topics, including data science (e.g., data generation through sampling or simulation), teacher knowledge to manage risk literacy in the mathematics class (e.g., open discussions, curricula constraints/possibilities), visualization/representation to identify and argue in (with) probabilistic concepts, and relevance of the context.

Several open questions and areas needing further investigation were identified for practice and research:

- Two questions, which are practice-related, were raised with the recognition that subjective interpretation of probability is left out since it is seen as non-calculable and risk literacy is new to mathematics education: How can we address subjective probabilities in school? How can we bring risk literacy into the school curriculum?
- Various suggestions were made in terms of research. Local theories for instruction on risk (e.g., levels of risk literacy and reasoning about risk) can be developed through design-based research. Specifying the conceptualization of probabilistic reasoning and risk literacy in terms of where they overlap and where they are different is needed. Designing meaningful tasks to foster risk literacy was discussed as a way to enable teachers to reinterpret mathematical concepts (e.g., expected value, conditional probabilities, and compound probabilities) already available in the curriculum through a risk lens by connecting them to decision-making under uncertainty. The group also emphasized addressing the emotional influences in students' perception of risk.

Conclusion

The Thematic Working Group 05 (TWG05) at CERME14 served as a dynamic and inclusive platform for exploring critical aspects of Statistics and Probability Education, uniting 30 participants from 13 countries for meaningful dialogue and constructive exchange.

The insights gained from these discussions underscore the imperative to equip learners with essential skills for the modern world. This includes fostering early data literacy and critical thinking within data science contexts, emphasizing the use of both "dirty" and "clean" real-world data, and understanding how data is used for predictions. Discussions also highlighted the importance of enabling learners to interpret and create diverse statistical visualizations, recognizing them as data

consumers, producers, and even "prosumers". A consistent focus was placed on teacher education, acknowledging the pressing need for educators to develop deeper contextual knowledge and technological pedagogical content knowledge (TPACK) to effectively incorporate digital tools and navigate these complex topics. Furthermore, the discussions on probabilistic reasoning and risk literacy reinforced their growing significance for informed decision-making in an increasingly uncertain world.

As we stand at the threshold of a new era profoundly shaped by data science, artificial intelligence, and advanced technological tools, this is indeed an exciting time for TWG05 and the broader field of probability and statistics education. While the challenges are substantial—including overcoming institutional constraints, providing adequate teacher support, and adapting curricula to rapid technological advancements—the possibilities for fostering deeper statistical and probabilistic thinking, enhancing AI literacy, and preparing learners to critically engage with data-driven realities are immense. We eagerly anticipate tackling these challenges and harnessing these possibilities to further advance the teaching and learning of probability and statistics.

References

- Batanero, C., & Álvarez-Arroyo, R. (2024). Teaching and learning of probability. *ZDM–Mathematics Education*, 56(1), 5-17. <https://doi.org/10.1007/s11858-023-01511-5>
- Batanero, C., Burrill, G., & Reading, C. (2011). *Teaching statistics in school mathematics-Challenges for teaching and teacher education. A Joint ICMI/IASE Study: The 18th ICMI Study*. Springer.
- Burrill, G., & Pfannkuch, M. (2024). Emerging trends in statistics education. *ZDM–Mathematics Education*, 56(1), 19-29. <https://doi.org/10.1007/s11858-023-01501-7>
- Primi, C., Leavy, A., Kazak, S., Frischemeier, D., González, O. R., & Andre, M. (2024). An introduction to TWG5: Probability and statistics education. In P. Drijvers, C. Csapodi, H. Palmér, K. Gosztonyi, & E. Kónya (Eds.), *Proceedings of the Thirteenth Congress of the European Society for Research in Mathematics Education (CERME13)* (pp. 1-5173). Alfréd Rényi Institute of Mathematics and ERME. HAL ID: hal-04430809.