STOCHASTIC COMPETENCIES DEVELOPMENT IN GRADE 4

Antonia Wunsch[1], Robert von Hering[2], Henning Sievert[1]

[1]University of Hildesheim, [2]University of Flensburg

Stochastic competencies are crucial in both daily lives of primary school children and their social participation. Hence, developing stochastic competencies is a mandatory goal in mathematics education. However, challenges are apparent in the three subdomains of data, combinatorics, and chance. Despite their growing societal importance, the state of research on primary school stochastic competencies and their development remains insufficient. Accordingly, in this study, we conducted a longitudinal analysis of the development of students' stochastic competencies in Grade 4 by collecting data from 586 students across 42 classes in 18 schools in Germany. The results show developments in all three subdomains as well as significant differences both between schools and classes.

INTRODUCTION

Reading the family calendar, comparing game rules, creating a soccer plan – each of these everyday situations is based on a stochastic context that children are confronted with, even before they start school. On the one hand, this means that primary school students already have some prior knowledge of probability, and on the other hand, that it is important for their everyday lives. Looking at the state of research, we find several studies showing the feasibility of facilitating stochastic competencies in primary school (e.g., Falk et al., 2012; Frischemeier, 2020). However, empirical research also revealed deficits in students' stochastic competencies as well as specific learning barriers in the different subdomains (e.g., Fischbein et al., 1991; Frischemeier & Schnell, 2023; Zapata-Cordona, 2018). Moreover, research on stochastic competencies in primary school is mostly restricted to cross-sectional studies. Accordingly, in this study, we focus on development during the fourth school year by conducting a longitudinal assessment of the competencies of 586 students.

THEORETICAL FRAMEWORK

The following theoretical framework examines the educational content related to the school's topic of *data and chance*, as well as the current state of research on stochastic competencies. There is currently little empirical (longitudinal) evidence available on the current level of competence among students in all three subdomains of stochastic (data, combinatorics, and chance). Both *data literacy* and *probabilistic literacy* are important concepts in teaching stochastics. Ridsdale et al. (2015) define *data literacy* as "the ability to collect, manage, evaluate, and apply data, in a critical manner" (p. 3). *Probability literacy* is defined as understanding probabilistic contexts and questioning random situations (Álvarez-Arroyo et al., 2024).

2026. In H. Krzywacki, M. Hähkiöniemi, E. Haataja, M.S. Hannula & P. Hästö (Eds.). *Proceedings of the 49th Conference of the International Group for the Psychology of Mathematics Education: Research Reports*, pp. 1603-1610. PME.

In accordance with the NCTM, stochastic education should include the following:

> formulat[ing] questions that can be addressed with data and collect, organize, and display relevant data to answer them; select[ing] and us[ing] appropriate statistical methods to analyze data; develop[ing] and evaluat[ing] inferences and predictions that are based on data; understand[ing] and apply[ing] basic concepts of probability (NCTM, 2000, p.48).

In Germany, primary school teachers are generalists, which means that they all teach mathematics, even if not all of them have formal qualifications. The school's topic *data and chance* is a mandatory content area in the German curriculum for primary grades (KMK, 2022). The subdomain *data* covers collecting, reading, interpreting, reflecting, and visualising data (KMK, 2022). In comparison, the NCTM emphasises the PPDAC-cycle (Wild & Pfannkuch, 1999). Unlike the NCTM, the German curriculum includes *combinatorics* as part of the school's topic *data and chance.* Students learn to solve combinatorial problems and apply heuristics; evaluating and comparing random situations and experiments is a content in the subdomain *chance* (KMK, 2022). Both the NCTM and the KMK focus on technical terms and comparing events. However, the NCTM (2000, p. 400) adds "an event can be represented by a number from 0 to 1." Although the curriculum outlines detailed content, it is equally crucial to consider empirical evidence on student competencies in these domains. In Germany, the study TIMSS 2023 reveals that Grade 4 students score slightly below the international average of OECD member states in the topic of *data* (Germany: 520, Average: 528; von Davier et al., 2024).

Deficient statistical competencies are reported, along with difficulties in data interpretation. Young students (aged 5 to 6) tend to focus on single data (*data as case value*) rather than the entire data distribution when working with data (Leavy & Hourigan, 2021), whereas clustering the data is typical for Grade 3 students (*data as classifier*) (Frischemeier & Birk, 2024). Regarding the four levels of different perspectives on data by Konold et al. (2025) (data as *pointer (1)*, *case value (2)*, *classifier (3)*, *aggregate (4)*), Grade 1 students can draw on level 2, and Grade 3 students on level 3. Furthermore, in Grade 3 context knowledge influences data interpretation more than the data presented, so explanations based on the data distributions provided are rarely used (Frischemeier & Birk, 2024). The authors conclude that context knowledge is an important aspect for developing statistical competencies. In addition, Grade 3 students already use metarepresentational competencies and conceptual competencies when drawing diagrams (English, 2014).

Considering difficulties in the subdomain of *combinatorics*, Palmér & van Bommel (2023) report that 14 of 48 Swedish preschoolers use abstract visualisations and a systematic approach to solve combinatorial problems. The drawings are influenced by context and may not correspond to the student's level of abstraction (Palmér & van Bommel, 2023). Further, there is a significant effect of the interaction between the implicit combinatorial model (*selection*, *distribution*, *partition*) and combinatorial operation (*variation*, *permutation*, *combination*, each with or without repetition)

(Batanero et al., 1997). With teaching combinatorics, a distinct improvement has been observed only in the model *selection* and in the operation *variation* (Batanero et al., 1997). Thomas and Pöhler (2025) report difficulties in formulating a word problem for arrangement without repetition ($\frac{6!}{(6-3)!}$). When considering the Cartesian product, students often struggle to combine unequal quantities (Zapata-Cardona, 2018). Moreover, Lookwood (2014) report that post-secondary students prefer "meaningless procedures over conceptual understanding" for solving combinatorial problems (p. 36).

With regard to the subdomain *chance*, students must be familiar with technical terms to express themselves adequately when comparing random events or reasoning about probability situations. Fischbein et al. (1991) report linguistic difficulties with the terms *certain*, *possible*, and *impossible*. According to Martignon (2014), only 11% of Grade 4 students can reason why they would move the more successful token in Ludo to minimise the loss of a token. Nevertheless, five- to eight-year-olds are developing their first initial ideas about probability; active and playful learning environments can help to reinforce and deepen these ideas (Paparistodemou, 2014). Hence, four- to six-year-olds already demonstrate a strong conceptual understanding of *more likely than* and *equally likely* in the subdomain of *chance* (Jaeger, 2024).

However, studies report that it is possible to facilitate stochastic competencies from primary school onwards (Falk et al., 2012; Frischemeier, 2020). Moreover, early stochastic education promotes the long-term development of competencies in secondary education (English, 2014) and thus is the basis for *probability literacy* (Nikiforidou, 2018). Given the increasing availability of diverse data in everyday life, it is essential that students acquire the competencies to interpret and handle such information correctly from an early age (English, 2014). This underscores the importance of developing relevant competencies already in primary school. In this context, the stochastic opportunities to learn vary considerably. Looking at textbooks as indicator of opportunities given in class, Wunsch et al. (2025) point out that Grades 3 and 4 offer the most opportunities to learn. Further, the topics combinatorics and probability tend to be underrepresented (Wunsch et al., 2025). Moreover, significant differences can be observed both between and within textbook series.

RESEARCH QUESTIONS

However, the theoretical framework shows that there are rarely empirical findings that relate to stochastic competency development across all subdomains using a large data set. The question of how stochastic competencies develop remains blurry. Therefore, curriculum requirements and the relevance for societal participation form the starting point for the following research questions:

RQ1: What competencies do students achieve in *data and chance* in Grade 4, and how do the subdomains relate to each other?

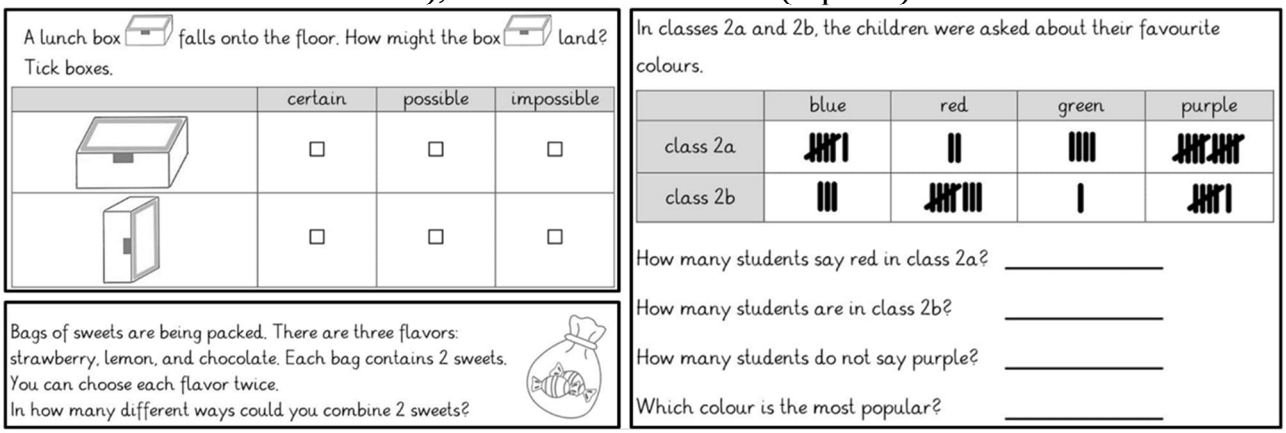
RQ2: Do these competencies differ at school or class level?

RQ3: How do these competencies develop in Grade 4?

METHODOLOGY

To address the research questions, first, a paper-and-pencil test was developed, including 32 items: Part I with 11 items (*data*), Part II with 10 items (*combinatorics*), and Part III with 11 items (*chance*). In part I, the students interpret data presented in tables and bar charts and change the representation. The second part involves solving combinatorial problems and choosing a matching visualisation to determine combinations and frequencies. In part III, making judgments and writing reasons is demanded to qualify random experiments. Students choose the matching term *certain*, *possible*, and *impossible* to evaluate probabilities, and identify the more successful random generator. The test has both open and closed response formats. Most items are coded dichotomously, while nine items are coded using partial credit. Figure 1 shows translated item examples of each part, respectively.

Figure 1: Example items from Part I – *data* (right), Part II – *combinatorics* (bottom left), and Part III – *chance* (top left)



The instrument was tested twice in pilot studies before data collection, with two classes at the end of Grade 3 and Grade 4, respectively, to develop suitable items for the longitudinal survey over the period of Grade 4. A total of 69 students participated in the first pilot study (May/June 2023), and 53 students in the second pilot study (June 2024). For the subsequent main analysis, we collected student data over one year (2024/2025) at two measurement points: at the beginning and at the end of Grade 4. A total of 42 classes from 18 primary schools in Northern Germany participated (N = 586 students). All classes were tested in the second or third week of school (T1) and again in the last month of the school year (T2). The polytomous items of the pretest (T1) were completely double-coded, with a good interrater reliability ($76 \leq \kappa \leq .98$).

The longitudinal polytomous data set was analysed using Item Response Theory (IRT) and Partial Credit Model (PCM) as 2PL Model in RStudio, including measurements at T1 and T2. Person ability parameters were estimated via Weighted Likelihood Estimation (WLE). Both a unidimensional model for the overall test (EAP reliability

= .78) and a three-dimensional model were fitted (Part I – *data*, EAP reliability = .57; Part II – *combinatorics*, EAP reliability = .69; Part III – *chance*, EAP reliability = .70). Model comparisons using information criteria indicated that the three-dimensional model provided substantially better fit than the unidimensional model (AIC: 28922 vs. 29125; BIC: 29160 vs. 29348; CAIC: 29207 vs. 29392). The chi-square test indicated that the three-dimensional model fit the data significantly better than the one-dimensional model ($\chi^2(3) = 208.89$, $p < .001$), providing strong evidence for a multidimensional structure. A two-way repeated-measures ANOVA and a one-way ANOVA were used to analyse whether development differed between dimensions and between schools and classes.

FINDINGS

Table 1 shows the mean values and the standard deviation from the WLE PCM for both the unidimensional model and the three-dimensional model.

Table 1: Mean (standard deviation) of WLE estimates from the PCM models

	Unidimensional Model	**Three-dimensional Model**		
		data	*combinatorics*	*chance*
T1	-.27 (.93)	-.52 (1.07)	-.31 (1.31)	-.30 (1.07)
T2	.22 (.97)	-.15 (.98)	.17 (1.33)	.20 (1.04)

The paired t-test of the unidimensional model shows a significant increase from T1 to T2 ($t(585) = 13.69$, $p < .001$, $d = .57$), with a medium effect size. All three test parts show significant increases from T1 to T2, with small to medium effect sizes: Part I – *data* ($t(584) = 7.51$, $p < .001$, $d = .31$), Part II – *combinatorics* ($t(583) = 8.26$, $p < .001$, $d = .34$), and Part III – *chance* ($t(585) = 11.91$, $p < .001$, $d = .49$).

The two-way repeated-measures ANOVA indicates a significant main effect of time, such that values increased from T1 to T2 ($\beta = .38$, $SE = .05$, $t(2922) = 6.94$, $p < .001$). There was also a significant main effect of dimension: Part II – *combinatorics* was higher than Part I – *data* ($\beta = .21$, $SE = .05$, $t(2922) = 3.79$, $p < .001$) as was Part III – *chance* ($\beta = .23$, $SE = .05$, $t(2922) = 4.19$, $p < .001$). However, the interaction between time and dimension was not significant for either Part II – *combinatorics* ($\beta = .11$, $SE = .08$, $t(2922) = 1.45$, $p = .15$), or Part III – *chance* ($\beta = .12$, $SE = .08$, $t(2922) = 1.50$, $p = .13$). Random intercepts for participants account for individual differences, with a standard deviation of .66, while the residual standard deviation was .93. In addition, differences between 18 schools and 42 classes were analysed. The one-way ANOVA shows significant differences between schools (Table 2) and classes (Table 3).

Table 2: ANOVA between schools

	T1	**T2**
data	$F(17, 567) = 4.39$, $p < .001$	$F(17, 568) = 2.51$, $p < .001$
combinatorics	$F(17, 566) = 2.79$, $p < .001$	$F(17, 568) = 2.52$, $p < .001$
chance	$F(17, 568) = 3.65$, $p < .001$	$F(17, 568) = 3.81$, $p < .001$

Table 3: ANOVA between classes

	T1	**T2**
data	$F(41, 543) = 2.35$, $p < .001$	$F(41, 544) = 2.06$, $p < .001$
combinatorics	$F(41, 542) = 1.85$, $p = .001$	$F(41, 544) = 1.77$, $p = .003$
chance	$F(41, 544) = 2.35$, $p < .001$	$F(41, 544) = 2.38$, $p < .001$

Furthermore, the one-way ANOVA between classes taught by mathematics teachers (MT) and non-specialist teachers (NT) shows varying results across T1 and T2. At T1, significant differences were observed in Part I – *data* and Part II – *combinatorics*, with MT outperforming NT. At T2, a significant difference emerged in Part III – *chance*, with MT outperforming NT. In the other subtests, lower values were also recorded in classes taught by NT.

DISCUSSION

This study investigated stochastic competencies of Grade 4 students and their development over time, as well as differences between schools and classes. To answer RQ1, the mean WLE estimates were higher at T2 than at T1 across all dimensions. The strongest increases occur in Part III – chance, whereas Part I – data and Part II – combinatorics show a smaller increase. However, the non-significant interaction between time and dimensions suggests that the improvement from T1 to T2 is similar across all dimensions, indicating a general developmental trend rather than dimension-specific growth patterns. With regard to RQ2, significant differences between schools and classes point to an influence of mathematics instruction on students' performance. Furthermore, the observed differences between MT and NT suggest that teachers' qualifications may play a role in students' mathematical development. Finally, concerning RQ3, all three test parts show significant increases from T1 to T2 ($.31 \leq d \leq .49$), indicating small to medium effects.

Missing data may have influenced the results. At both T1 and T2, some students did not respond to all items in the paper-and-pencil test, resulting in missing responses at the item level. Since no significant differences between MT and NT were consistently observed across all six measurements, the results suggest that professional knowledge that goes beyond a dichotomous distinction between MT and NT should be taken into account. Hence, interviews were conducted with participating teachers to get deeper insights about teacher knowledge and lesson planning. In addition, an analysis of the

mathematics textbooks used was performed to assess the content and range of possible stochastics instruction in Grade 4. Aiming for long-term development of stochastic competencies, the results can help optimise the teaching and learning of stochastics in primary schools and create conditions for a successful transition to secondary education.

References

Álvarez-Arroyo, R., Batanero, C., & Gea, M. M. (2024). Probabilistic literacy and reasoning of prospective secondary school teachers when interpreting media news. *ZDM - Mathematics Education, 56*(6), 1045–1058.

Batanero, C., Navarro-Pelayo, V., & Godino, J. D. (1997). Effect of the implicit combinatorial model on combinatorial reasoning in secondary school pupils. *Educational Studies in Mathematics, 32*(2), 181–199.

von Davier, M., Kennedy, A., Reynolds, K., Fishbein, B., Khorramdel, L., Aldrich, C., Bookbinder, A., Bezirhan, U., & Yin, L. (2024). *TIMSS 2023 International Results in Mathematics and Science*. Boston College, TIMSS & PIRLS International Study Center.

English, L. D. (2014). Promoting Statistical Literacy Through Data Modelling in the Early School Years. In E. J. Chernoff & B. Sriraman (Eds.), *Probabilistic Thinking. Presenting Plural Perspectives* (pp. 441–457). Springer Science+Business Media.

Falk, R., Yudilevich-Assoulione, P., & Elstein, A. (2012). Children's concept of probability as inferred from their binary choices – revisited. *Educational Studies in Mathematics, 81*(2), 207–233.

Fischbein, E., Nello, M. S., & Marino, M. S. (1991). Factors affecting probabilistic judgements in children and adolescents. *Educational Studies in Mathematics, 22*(6), 523–549.

Frischemeier, D. (2020). Building statisticians at an early age: Statistical projects exploring meaningful data in primary school. *Statistics Education Research Journal, 19*(1), 39–56.

Frischemeier, D. & Birk, L. (2024). Different lenses on distributions and the integration of context knowledge in data-ing processes of primary school students. *ZDM - Mathematics Education, 57*(1), 1–15.

Frischemeier, D. & Schnell, S. (2023). Statistical investigations in primary school – the role of contextual expectations for data analysis. *Math.Ed. Research Journal, 35*, 217–242.

Jaeger, L. (2024). Zur Erfassung des intuitiven Begriffsverständnisses junger Kinder im Kontext des frühen probabilistischen Denkens. [To assess young children's intuitive understanding of concepts in the context of early probabilistic reasoning.] *Stochastik in der Schule 44*(3), 2–7.

Konold, C., Higgins, T., Russell, S. J., & Khalil, K. (2015). Data seen through different lenses. *Educational Studies in Mathematics, 88*(3), 305–325.

Kultusministerkonferenz [KMK] (2022). *Bildungsstandards für das Fach Mathematik Primarbereich. Beschluss der Kultusministerkonferenz vom 15.10.2004.* [*National educational standards in mathematics at primary level*]. KMK.

Leavy, A. & Hourigan, M. (2021). Data Modelling and Informal Inferential Reasoning: Instances of Early Mathematical Modelling. In J. M. Suh, M. H. Wickstrom, & L. D. English (Eds.), *Exploring Mathematical Modeling with Young Learners* (pp. 67–93), Springer Nature.

Lookwood, E. (2014). A set-oriented perspective on solving counting problems, *Educational Studies in Mathematics, 34(2),* 31–37.

Martignon, L. (2014). Fostering Children's Probabilistic Reasoning and First Elements of Risk Evaluation. In E. J. Chernoff & B. Sriraman (Eds.), *Probabilistic Thinking. Presenting Plural Perspectives* (pp. 149–160), Springer Science + Business Media.

National Council of Teachers of Mathematics [NCTM] (2000). *Principles and Standards for School Mathematics*. NCTM

Nikiforidou Z. (2018). Probabilistic Thinking and Young Children: Theory and Pedagogy. In A. Leavy, M. Meletiou-Mavrotheris, & E. Paparistodemou (Eds.), *Statistics in Early Childhood and Primary Education. Supporting Early Statistical and Probabilistic Thinking* (pp. 21–34). Springer Nature.

Palmér, H. & van Bommel, J. (2023). Young Students' Choice of Representation When Solving a Problem-Solving Task on Combinatorics. In H. Palmér, C. Björklund, E. Reikerås, & J. Elofsson (Eds.), *Teaching Mathematics as to be Meaningful – Foregrounding Play and Children's Perspectives. Results from the POEM5 Conference, 2022* (pp. 119–129). Springer Nature.

Paparistodemou, E. (2014). Children's Constructions of a Sample Space with Respect to the Law of Large Numbers. In E. J. Chernoff & B. Sriraman (Eds.), *Probabilistic Thinking. Presenting Plural Perspectives* (pp. 599–612). Springer Science+Business Media.

Ridsdale, C., Rothwell, J., Smit, M., Ali-Hassan, H., Bliemel, M., Irvine, D., Kelley, D., Matwin, S., & Wuetherick, B. (2015). *Strategies and Best Practices for Data Literacy Education: Knowledge Synthesis Report*.

Thomas, C. & Pöhler, B. (2025). Analysis of students' self-written word problems related to arrangement without repetition. In C. Cornejo, P. Felmer, D. M. Gómez, P. Dartnell, P. Araya, A. Peri, & V. Randolph (Eds.), *Proc. of the 48th Conf. of the Int. Group for the Psychology of Mathematics Education* (Vol. 2, pp. 331–338). PME.

Wild, C. J. & Pfannkuch, M. (1999). Statistical Thinking in Empirical Enquiry. *International Statistical Review, 67*(3), 223–265.

Wunsch, A., von Hering, R., & Sievert, H. (2025). Determining textbook quality in primary school's topic data and chance. In B. Pepin, I. Kohanová, & M. B. Langfeldt (Eds.), *Proceedings of the Fifth International Conference on Mathematics Textbook Research and Development* (pp. 84–87). Norwegian University of Science and Technology.

Zapata-Cardona, L. (2018). Supporting Young Children to Develop Combinatorial Reasoning. In A. Leavy, M. Meletiou-Mavrotheris, & E. Paparistodemou (Eds.), *Statistics in Early Childhood and Primary Education. Supporting Early Statistical and Probabilistic Thinking* (pp. 257–272). Springer Nature.