International Journal of Mathematics and Statistics
ISSN (Online): 2208-2212
Volume 10 Issue 04 December 2024
DOI: https://doi.org/10.69980/rpsstb49

CROSS-CLASSIFIED LOGISTIC MODELING OF CORRECT RESPONSE PROBABILITY IN HIGHER-EDUCATION MATHEMATICS ASSESSMENT

Dr. Jonathan P. Whitaker[1*], Prof. Maria L. Schneider[2], Dr. Kenji R. Nakamura[3]

[1*]*Department of Educational Statistics and Measurement, Northbridge University, Edinburgh, United Kingdom*
[2]*Institute of Mathematical Education and Assessment, Rhine-Westphalia University, Cologne, Germany*
[3]*Department of Statistical Learning and Educational Research, Pacific Institute of Technology, Vancouver, Canada*

Abstract:

Assessment data in higher education mathematics courses are frequently characterized by complex dependence because individual students answer several questions while the same questions may be answered by multiple students. The analysis included 9,546 responses from 372 students across 833 mathematics questions. Overall, 46.8% of responses were correct. Some topic-specific differences were detected, with lower odds of correct answers for Differentiation, Probability, Real Functions of a Single Variable, and Fundamental Mathematics compared to Linear Algebra, while Set Theory had higher odds. The proportion of latent variance due to student-level heterogeneity was substantially larger than that due to question-level heterogeneity. The cross-classified model showed better model fit, apparent discrimination, and probabilistic accuracy than traditional logistic regression and single-cluster alternatives, as evidenced by lower WAIC, log loss and Brier scores, and a higher AUC. Sensitivity analyses indicated that the main topic effects and variance structure were robust to changes in the specification. The results are congruent with cross-classified modeling as a systematic method to analyze mathematics assessment data and partition sources of variation at the learner- and item-level.

Keywords: *cross-classified logistic regression; mathematics assessment; higher education; student heterogeneity; response probability*

1. Introduction:

Mathematics is pivotal in higher education, serving as the foundation for analytical thinking, quantitative decision-making, scientific literacy, and problem-solving in a myriad of fields. Although an important topic, performance in mathematics at the university level is very variable, influenced by students' prior preparation and their adaptation to the new, more challenging mathematics, as well as by their cognitive engagement and confidence. The shift to university mathematics is especially significant because they are expected to learn and engage in more abstract mathematical concepts, procedural complexity, and demands for independent study which may affect their persistence and success (Geisler et al., 2023).

Content knowledge is not the only factor that affects individual differences in mathematics performance. Affective factors may have significant impact on pupils' interactions with quantitative content, especially when it is being formally assessed. In the context of the university, for instance, mathematics anxiety can impact negatively on students' academic achievement by diminishing confidence and compromising problem solving processes (Khasawneh et al., 2021). The same issues have been found in statistics education: instructional design and learning environment can be important in reducing anxiety and enhancing students' willingness to work with mathematically challenging tasks (Bjälkebring, 2019). These results indicate that the observed response patterns could be due to student characteristics and/or characteristics of the items in the assessments.

Another major source of variation in performance is the variation from topic to topic in mathematics. Students can demonstrate relatively strong knowledge in some areas while experiencing greater difficulty in others, depending on the cognitive demands of the topics, their previous exposure, and their conceptual understanding. Recent studies have demonstrated that prior knowledge is an important factor in predicting future mathematical performance, and that it should be assessed at the content domain level rather than viewing mathematics achievement as a monolithic construct (Alreshidi, 2023). The level of difficulty should also be taken into account because, depending on the nature and complexity of specific items, performance on the assessment may differ.

Recent years have seen an increase in use of digital technologies in mathematics education, which now allows for capturing detailed student-response data in various topics and at different levels of difficulty. The digital learning environment enables adaptive practice, online assessment, automated feedback, and multiple interactions with mathematical content, resulting in data that can be analysed at a finer level of resolution (Engelbrecht & Borba, 2024). Similarly, fully online mathematics instruction has shown how essential it is to have technology-mediated interactions and learning community structures in undergraduate education (Trenholm & Peschke, 2020).

These have all helped to foster the wider adoption of learning analytics in higher education. While there is much descriptive work on learning analytics, it is also possible to systematically investigate students' behaviour and performance based on large and complex datasets from learning processes (Viberg et al., 2018). Hence, the need to look for ways to better link analytic outputs with substantive learning processes and statistically appropriate models of educational data has been raised (Guzmán-Valenzuela et al., 2021). In today's assessment landscape, there is a growing need for methods that can handle dependency structures, repeated observations, and learner and item heterogeneity (Caspari-Sadeghi, 2023).

It may not be appropriate to use conventional logistic regression if there are multiple responses from the same students and the same questions answered by multiple students. In these situations, cross-classified models are more appropriate models because they can separate the variation due to students and the variation due to questions, as observations are nested within two categories of non-hierarchical clustering (Vera, 2022). Accordingly, the present study was guided by the following objectives:

1. To examine the associations of mathematical topic and question difficulty with the probability of a correct response using cross-classified mixed-effects logistic regression.
2. To quantify student-level and question-level heterogeneity in correct-response probability through variance-component estimation.
3. To evaluate the statistical performance and robustness of the cross-classified model through comparison with simpler logistic specifications and sensitivity analyses.

2. Methodology

2.1 Study Design and Data Source

A quantitative secondary data design was adopted in order to investigate the phenomenon of variation in higher education mathematics assessment in the probability of correct responses. Data were collected from a publicly available database of student answers to mathematics assessment questions (Azevedo et al., 2024). The analysis involved 9,546 records of responses to 833 mathematics questions from 372 students. Data collected included correctness of the response, difficulty level of the question, mathematical content area of the question, student identity, and question identity.

The response structure was cross-classified due to the fact that the same students may have responded to more than one item, and the same items may have been answered by more than one student. Therefore, statistical analysis was done taking into account dependence at student and question levels.

2.2 Study Variables

The primary outcome was response correctness with scoring of 1 for correct response and 0 for incorrect response. Question difficulty level and mathematical topic were the two main explanatory variables.

Question difficulty was coded as Basic or Advanced (using Basic for reference). Mathematical topic comprised 14 mathematical domains; Linear Algebra was chosen as the reference category since it had the highest number of responses. Due to repeated observations of students across the assessment items and of the assessment items across students, student ID and question ID were both included as clustering variables.

2.3 Descriptive Analysis

Descriptive statistics were used to describe the distribution of responses over the levels of question difficulty and mathematical topics. For categorical variables, frequencies and percentages were calculated and proportions of correct responses reported with 95% Wilson score confidence intervals.

The repeated-response structure was also explored by observing the unique student–question combinations. These summaries were used to give an initial description of variation in mathematics performance prior to fitting the cross-classified model.

2.4 Cross-Classified Logistic Regression

A cross-classified mixed-effects logistic regression model was used for estimating the probability of a correct response, taking into account the clustering of students and questions. The dependent variable was response correctness, while question difficulty and mathematical topic were included as fixed effects.

Student ID and Question ID were added as random intercepts to account for differences in the likelihood of a correct response that remained unexplained by the variables in the model, across students and items on the assessment. Fixed-effect coefficients were converted to the odds-ratio scale and model estimation was done in a Bayesian framework. The results have been presented as adjusted odds ratios (aORs) and 95% credible intervals (CrIs).

2.5 Variance Components and Model Comparison

Student-level and item-level random-effect variances were also estimated showing the extent of variability in the model that is not accounted for by the question characteristics and mathematical topic. Variance proportions were computed on the latent logistic scale, based on the conventional residual variance of $\pi^2/3$. Median odds ratios were also computed to provide an estimate of the size of the between-cluster heterogeneity on the odds-ratio scale.

The cross-classified model was contrasted with three alternative models: a conventional logistic regression model without random effects; a random-intercept model for the questions only; and a random-intercept model for the students only. The fit of the models was assessed using the Widely Applicable Information Criterion (WAIC) and better model fit is represented by a lower value. Additional summary measures of apparent discrimination and probabilistic accuracy were the area under the receiver operating characteristic curve (AUC), log loss, and Brier score.

2.6 Sensitivity Analysis

Sensitivity analyses were performed using combinations of student–question records to determine whether repeated student–question records had a material impact on the findings. The cross-classified model was re-estimated, leaving out exactly repeated records. Second, a different weighting specification was used to ensure an equal overall weighting of each unique student–question combination regardless of the number of responses recorded.

Robustness was assessed by examining the direction and size of the adjusted ORs and student- and question-level random-effect estimates in the primary and alternative specifications.

3. Results

3.1 Distribution of Responses Across Question Levels and Mathematical Topics

The analysis consisted of 9,546 responses from 372 students across 833 mathematics questions, of which 4,470 (46.8%) were correct and 5,076 (53.2%) were incorrect. These responses represented 6,782 distinct student–question combinations, indicating the presence of repeated observations.

In total, there were 7,844 responses to the Basic-level questions, of which 46.1% were correct, and 1,702 responses to the Advanced questions, of which 50.1% were correct. Correct-response rates were also subject to a wide range of variation by mathematical topic, as can be seen in Table 1. Linear Algebra was the most extensive domain (n = 5,726), out of which 49.0% of the responses were correct. The smallest response rates were in the topics of Differentiation (34.2%), Real Functions of a Single Variable (35.4%), Probability (37.5%), Optimization (38.5%) and Numerical Methods (38.7%). Comparatively, higher rates were found for Differential Equations (53.7%), Graph Theory (58.2%) and Set Theory (64.3%).

The between-topic variation in observed performance is further illustrated in Figure 1. Although Set Theory and Graph Theory exhibited the highest observed correct-response proportions, their confidence intervals were comparatively wide because of the smaller number of responses within these domains.

Table 1. Distribution of responses and observed correct-response rates by question level and mathematical topic

Category	Responses, n	Correct responses, n	Correct responses, %	95% CI
Question level				
Basic	7,844	3,617	46.1	45.0–47.2
Advanced	1,702	853	50.1	47.7–52.5
Mathematical topic				
Linear Algebra	5,726	2,807	49.0	47.7–50.3
Fundamental Mathematics	818	381	46.6	43.2–50.0
Complex Numbers	592	269	45.4	41.5–49.5
Differentiation	579	198	34.2	30.4–38.2
Analytic Geometry	358	175	48.9	43.7–54.0
Statistics	340	163	47.9	42.7–53.2
Numerical Methods	310	120	38.7	33.5–44.2
Optimization	182	70	38.5	31.7–45.7

Real Functions of a Single Variable	164	58	35.4	28.5–42.9
Integration	144	64	44.4	36.6–52.6
Probability	128	48	37.5	29.6–46.1
Differential Equations	108	58	53.7	44.3–62.8
Graph Theory	55	32	58.2	45.0–70.3
Set Theory	42	27	64.3	49.2–77.0

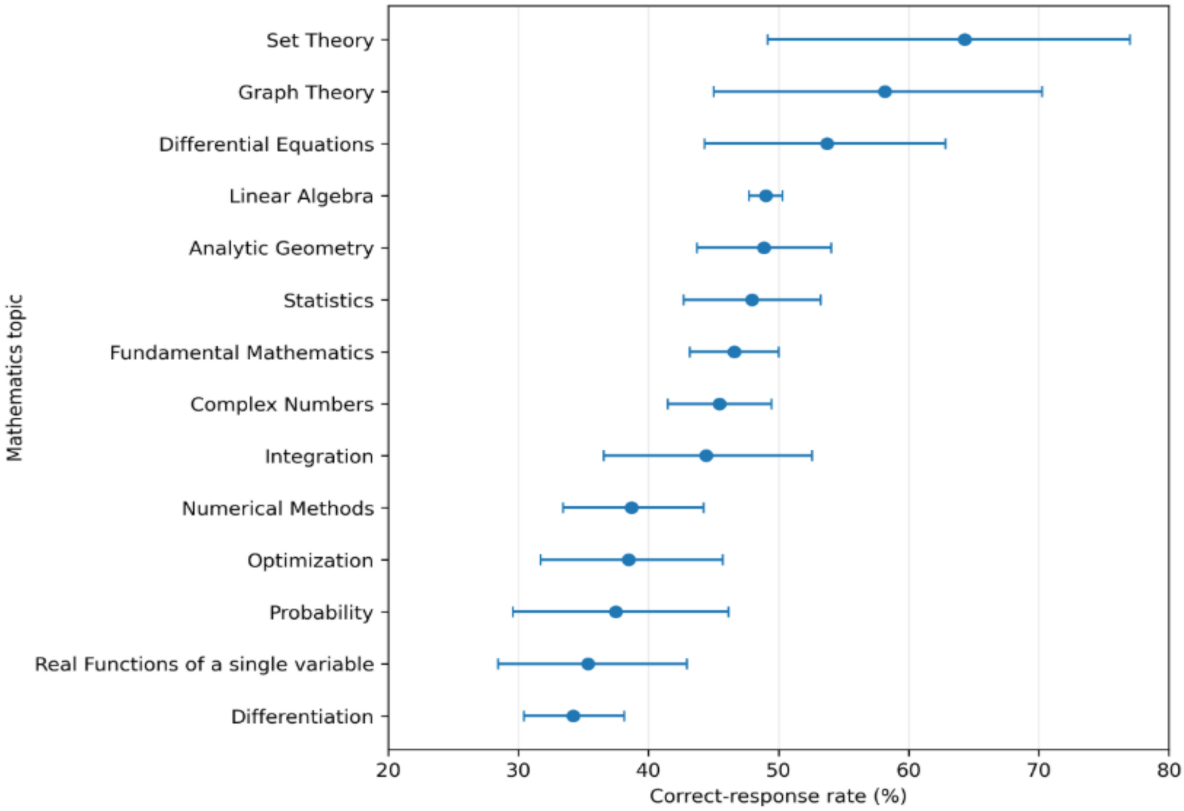

Figure 1. Topic-Specific Variation in Correct-Response Rates Across Higher-Education Mathematics Assessments

3.2 Fixed Effects on Correct-Response Probability

The cross-classified mixed-effects logistic model incorporated random intercepts for both students and questions, with Basic difficulty and Linear Algebra specified as the reference categories. After adjustment for topic and the crossed random-effect structure, Advanced questions were associated with modestly higher odds of a correct response than Basic questions (aOR = 1.16, 95% CrI: 1.04–1.29).

Substantial topic-specific differences remained after adjustment. As presented in Table 2, Differentiation was associated with the largest reduction in correct-response odds relative to Linear Algebra (aOR = 0.42, 95% CrI: 0.35–0.51). Lower odds were also identified for Probability (aOR = 0.50, 95% CrI: 0.34–0.74), Real Functions of a Single Variable (aOR = 0.64, 95% CrI: 0.46–0.88), and Fundamental Mathematics (aOR = 0.74, 95% CrI: 0.64–0.85). Complex Numbers showed a weaker negative association, with the interval extending to unity (aOR = 0.83, 95% CrI: 0.69–1.00).

Set Theory showed the strongest positive association with correct-response probability (aOR = 2.72, 95% CrI: 1.35–5.47), although the precision of this estimate was lower because of the comparatively small number of Set Theory observations. The adjusted estimates for Analytic Geometry, Differential Equations, Graph Theory, Integration, Numerical Methods, Optimization, and Statistics did not show clear evidence of differences from Linear Algebra.

The magnitude and uncertainty of the adjusted associations are displayed in Figure 2, which highlights the particularly strong negative effects associated with Differentiation and Probability and the positive estimate for Set Theory.

Table 2. Fixed-effect estimates from the cross-classified logistic model

Predictor	Adjusted OR	95% CrI
Question level		
Basic	1.00	Reference
Advanced	1.16	1.04–1.29
Mathematical topic		
Linear Algebra	1.00	Reference
Analytic Geometry	1.00	0.78–1.27

Complex Numbers	0.83	0.69–1.00
Differential Equations	1.10	0.72–1.66
Differentiation	0.42	0.35–0.51
Fundamental Mathematics	0.74	0.64–0.85
Graph Theory	1.69	0.96–2.96
Integration	0.81	0.56–1.16
Numerical Methods	0.92	0.71–1.20
Optimization	0.83	0.60–1.14
Probability	0.50	0.34–0.74
Real Functions of a Single Variable	0.64	0.46–0.88
Set Theory	2.72	1.35–5.47
Statistics	1.02	0.81–1.29

OR, odds ratio; CrI, credible interval.

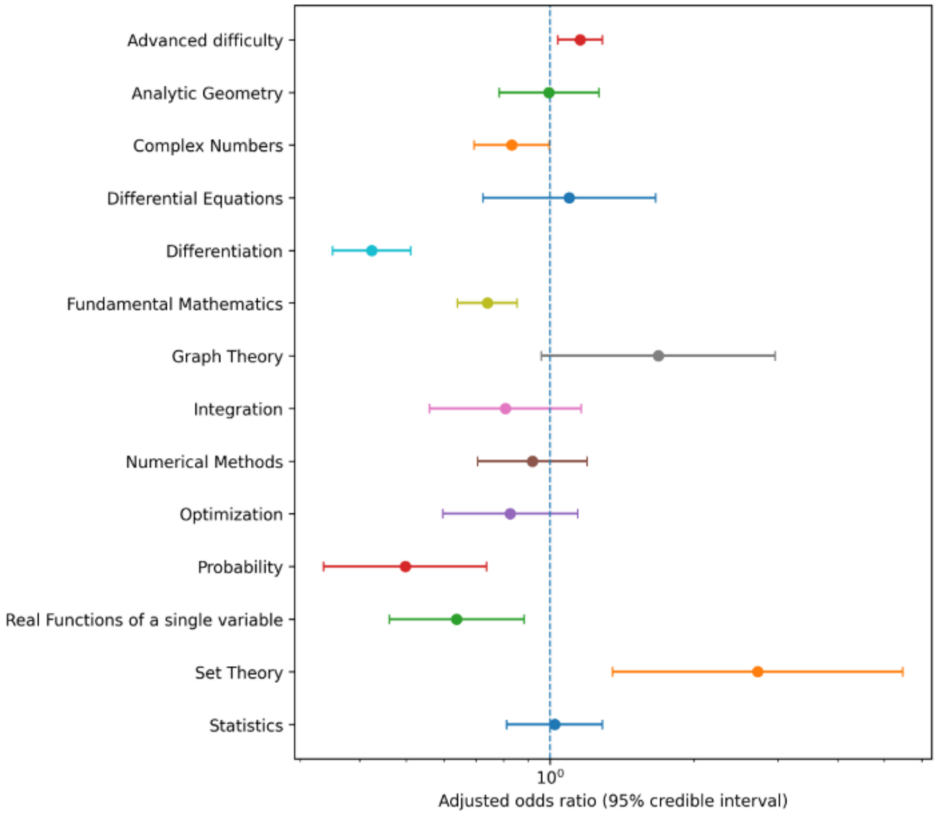

Figure 2. Adjusted Odds Ratios for Correct Responses From the Cross-Classified Logistic Regression Model

3.3 Student- and Question-Level Heterogeneity

Substantial unexplained heterogeneity in correct-response probability remained after adjustment for question level and mathematical topic. The estimated student-level random-intercept standard deviation was 0.951, corresponding to a variance of 0.904, whereas the question-level standard deviation was 0.366, corresponding to a variance of 0.134.

As summarized in Table 3, approximately 20.9% of the latent residual variability was attributable to differences between students and 3.1% to differences between questions. Together, the two clustering components accounted for approximately 24.0% of the total latent variance. The student-level component was therefore substantially larger than the question-level component.

The corresponding median odds ratios were approximately 2.48 at the student level and 1.42 at the question level, further demonstrating that differences between students contributed more strongly to heterogeneity in correct-response probability than unexplained differences between individual questions.

Table 3. Random-effect variance components of the cross-classified logistic model

Variance component	SD	Variance	95% CrI for variance	Proportion of latent variance
Student	0.951	0.904	0.783–1.043	20.9%
Question	0.366	0.134	0.122–0.147	3.1%
Logistic residual	1.814	3.290	—	76.0%
Student + question	—	1.038	—	24.0%

3.4 Comparative Model Performance

Model fit improved substantially when the cross-classified response structure was incorporated. The conventional logistic model produced a WAIC of 13,131.7, whereas inclusion of question-level random effects reduced the WAIC to 13,061.0. A considerably greater improvement was observed after incorporating student-level random effects, with the WAIC decreasing to 11,824.1.

As shown in Table 4, the full cross-classified model produced the lowest WAIC (11,758.4), representing an improvement of 65.6 WAIC units over the student-only model and 1,373.2 units over conventional logistic regression. This pattern supported simultaneous modeling of student- and question-level heterogeneity.

The cross-classified specification also yielded the strongest apparent discrimination (AUC = 0.772) and the lowest log loss (0.572) and Brier score (0.195). The student-only model showed the next-best performance, whereas the conventional logistic model exhibited substantially weaker discrimination and probabilistic accuracy. Collectively, the model-comparison measures favored the crossed student–question specification over the simpler alternatives.

Table 4. Comparative fit and performance of alternative logistic model specifications

Model	WAIC	ΔWAIC	AUC	Log loss	Brier score
Conventional logistic regression	13,131.7	1,373.2	0.544	0.686	0.247
Question random intercept	13,061.0	1,302.6	0.640	0.666	0.237
Student random intercept	11,824.1	65.6	0.745	0.592	0.204
Cross-classified student–question model	11,758.4	0.0	0.772	0.572	0.195

Lower WAIC, log loss, and Brier score indicate better model fit or probabilistic performance; higher AUC indicates stronger discrimination.

3.5 Sensitivity Analysis

Of the 6,782 unique student–question combinations, repeated observations occurred in 1,457 combinations. The stability of the principal estimates was therefore checked following the removal of exactly repeated records and using an alternative specification under which each student–question combination had equal weight.

There was a high degree of consistency in the main topic effects between specifications. In the primary model, the adjusted OR for Differentiation was 0.42, following removal of repeated records it was 0.40, and under equal student–question weighting it was 0.39. The corresponding estimates for Probability were 0.50, 0.47, and 0.45, and those for Real Functions of a Single Variable were 0.64, 0.63, and 0.60. The positive association for Set Theory was also maintained, with corresponding estimates of 2.72, 2.59, and 2.57.

The association between question difficulty and the probability of a correct response was less stable across specifications. In the main analysis, the adjusted OR for Advanced versus Basic questions was 1.16 (95% CrI: 1.04–1.29); with equal weighting of student–question combinations, the OR was 1.17 (95% CrI: 1.03–1.32); and following removal of exactly repeated observations, the OR was 1.11 (95% CrI: 0.99–1.25).

The random-effects structure was also stable across sensitivity specifications. The standard deviations for the primary, repeated-record-removed, and equally weighted models were 0.951, 0.834, and 0.921 at the student level, respectively, and 0.366, 0.275, and 0.337 at the question level, respectively. In each specification, variation across students remained substantially greater than variation across questions, and the principal topic associations were retained.

4. Discussion

The current analysis revealed that the probability of correct responding on higher education mathematics assessment items is influenced by both item content and item characteristics, as well as substantial variation among students. The cross-classified logistic model was able to better represent the structure of the response, compared to conventional logistic regression and models with a single clustering component. The pattern suggests that mathematics assessment results should not be viewed as a simple product of a learner's response to individual assessment items, but rather as an expression of the simultaneous contribution of learner- and item-level sources of variation. Furthermore, the size of the student-level heterogeneity in the results after accounting for the mathematical topic and question difficulty indicates that there is additional variability in student achievement that is not accounted for by these factors, suggesting that other factors (such as prior preparation, mathematical proficiency, learning experience, and cognitive engagement) are still important contributors to the success of student responses.

The large differences between students in the model are similar to what has been reported in the literature that previous mathematical background is an important factor in determining success in the early years of University. Higher education students have many different levels of knowledge of mathematics and the ways that they have organized and made accessible to them varies significantly, which can affect how they understand and solve problems in university-level mathematics. The size of the student level random effect in the current analysis therefore seems reasonable, as the model accounted for the residual level of heterogeneity that was not accounted for by the topic and difficulty measures. The relatively large median odds ratio at the student level also shows that two students with otherwise similar assessment situations may have considerably different chances of giving a correct answer due to differences between the students (Rach & Ufer, 2020).

The greater impact of student-level heterogeneity than question-level heterogeneity also has implications for the interpretation of university mathematics performance. There are differences between the requirements of abstraction, proof expectations, conceptual depth and the level of independent learning in the transition from secondary to university mathematics. Students are likely, therefore, to respond differently to seemingly similar assessment tasks based on past learning experiences and their capacity to respond to the demands of the tasks. This could help to account for the fact that differences between students remained substantial even after adjusting for mathematical topic and question difficulty. These findings confirm the perspective that mathematics achievement in the university is the result of an interplay between cognitive, affective, contextual and educational aspects and not only of the difficulty of the assessment items (Di Martino et al., 2023).

It was also found that there were marked differences between mathematical topics. Differentiation, Probability, Real Functions of a Single Variable and Fundamental Mathematics were found to have lower adjusted odds of correct response compared to Linear Algebra; whereas Set Theory had higher odds of correct response. These contrasts indicate that performance in mathematics cannot be assumed to be uniform throughout the mathematics curriculum. Certain domains may be more demanding conceptually or in terms of representational demands, or may be more congruent with students' knowledge or instructional exposure. The comparatively low adjusted odds for Differentiation are particularly notable as this topic involves the students coordinating symbolic procedure, conceptual understanding, functional reasoning and interpretation of change. Probability, for instance, can involve applying formal rules along with probabilistic reasoning, thereby adding to the cognitive load for students. These differences highlight the need to look at characteristics of items and content and not just at overall mathematics scores.

The topic variation also concurs with the general principle that the range of difficulty of the assessment items is not only nominal but also in the extent to which they discriminate between different levels of student achievement. Mathematics items can be cognitively challenging or not, can have varying representational structures, can be varying degrees of procedural difficulty, and can be varying degrees of conceptual dependence. Consequently, two questions classified under the same broad difficulty category may still differ substantially in their empirical response patterns. This finding is corroborated by the non-negligible question-level random effect that was identified, suggesting that measured difficulty categories are not sufficient to fully capture item complexity and therefore there is a need to model the residual item variability explicitly (Gyamfi & Wren, 2022).

The adjusted odds of a correct response were slightly greater for Advanced than for Basic questions in the primary model, but this relationship was less consistent and less robust across sensitivity analyses. The credible interval for the estimate widened to include 1 after removal of exactly repeated records suggesting that the apparent benefit of Advanced questions should be interpreted with caution. This may be attributed to self-selection, which means that pupils who choose to study more challenging material may be a stronger or more-confident group of pupils. Alternatively, it may be that general difficulty labels are not comprehensive in their description of item complexity. The positive association is relatively small and the relationship between difficulty classification and actual empirical response probability is not necessarily one to one, so it would be incorrect to assume that a positive association means that advanced mathematics questions are necessarily easier.

Instructional strategies used for each topic might also explain the differences in achievement. Research on conceptual learning in university mathematics has shown that carefully structured activities can improve understanding even in domains traditionally regarded as difficult. Studies with exact differential equations have shown that instructional strategies that allow students to work with mathematical relationships in a more active way could enhance students' conceptual understanding instead of merely relying on routine procedure (Rezvanifard et al., 2023). The lower response probabilities found in several domains in the current analysis may thus yield insights into where instructional design, formative assessment, and targeted practice might be especially beneficial.

Results from the comparative model also confirm that statistical methods which acknowledge the crossed nature of educational response data should be used. The complete model had the smallest WAIC and better AUC, log loss and Brier score than the other models. The improvement compared with conventional logistic regression was substantial and showed that disregarding clustering might lead to an understatement of the presence of important sources of heterogeneity and, at the same time, to less model adequacy. Both person and item dependency are common with educational assessment data and modeling both of these simultaneously can help with both the statistical interpretation and the model adequacy, discrimination, and probabilistic accuracy. That is because, evidence has indicated that models that consider aspects of both learners and assessment items can offer a more informative representation of response behaviour than models that consider observations as independent (Park et al., 2023).

Sensitivity analyses reinforced confidence in the primary findings. The major topic effects remained directionally and quantitatively similar after the removal of exactly repeated records and under equal weighting of distinct student–question combinations, while student-level heterogeneity remained greater than question-level heterogeneity. The fact that this stability was not due to repeated observations alone is a sign of stability. However, due to the lack of attempt order, timing of attempts, previous achievement, instructional exposure, and background characteristics at the student level, causal conclusions cannot be made. Consequently, the model does not detect causal relationships, but structured associations in correct-response probability.

The results as a whole indicate that there is substantial amount of variation both between students and between assessment items that is attributed to higher education mathematics performance and that the larger source of variation is student-level. The better fit with the cross-classified model demonstrates that these various dependencies need to be addressed simultaneously in the statistical analysis of mathematics assessment. This gives a more stringent foundation for identifying challenging content areas, understanding learner variation, and creating databased assessment and instructional opportunities.

5. Conclusion

The present study illustrates the usefulness of cross-classified logistic modeling in the analysis of higher education mathematics assessment data where the responses are clustered both within the student and within the item. Results indicated that there was meaningful variation in the probability of a correct response across mathematical topics, together with substantial unexplained variation at the student and question levels. The variability between students was considerably greater than that of the questions, suggesting that variation between students was a major source of variability even after taking into account the mathematical topic and question difficulty.

Differences within topics also emerged from the model. Differentiation, Probability, Real Functions of a Single Variable, and Fundamental Mathematics were associated with lower adjusted odds of a correct response compared with Linear Algebra, while Set Theory was associated with higher adjusted odds of a correct response. The effect of question difficulty was, in comparison, not as large and was not as consistent across sensitivity specifications. The results indicate that overall difficulty ratings cannot be used to give a complete picture of mathematics assessment performance.

The cross-classified model achieved a substantially better fit than conventional logistic regression and models with a single clustering component. Its better WAIC, discrimination, and probabilistic accuracy confirm the relevance of accounting for learner- and item-level dependence in educational response data. Results of the main findings were also stable across sensitivity analyses, adding further support to the robustness of the results.

Overall, the results indicate that statistically rigorous modeling of mathematics assessment should recognize the joint influence of student heterogeneity and question characteristics. These can form a better foundation for determining challenging areas of mathematics, the design of assessments, and more focused instructional and learning strategies in the higher education environment.

References

1. Alreshidi, N. A. K. (2023). Enhancing topic-specific prior knowledge of students impacts their outcomes in mathematics. Frontiers in Education, 8, Article 1050468.
2. Azevedo, B. F., Pacheco, M. F., Fernandes, F. P., & Pereira, A. I. (2024). Assessing mathematics learning in higher education [Data set]. UCI Machine Learning Repository.
3. Bjälkebring, P. (2019). Math anxiety at the university: What forms of teaching and learning statistics in higher education can help students with math anxiety? Frontiers in Education, 4, Article 30.
4. Caspari-Sadeghi, S. (2023). Learning assessment in the age of big data: Learning analytics in higher education. Cogent Education, 10(1), 2162697.
5. Di Martino, P., Gregorio, F., & Iannone, P. (2023). The transition from school to university in mathematics education research: New trends and ideas from a systematic literature review. Educational Studies in Mathematics, 113(1), 7-34.
6. Engelbrecht, J., & Borba, M. C. (2024). Recent developments in using digital technology in mathematics education. ZDM–Mathematics Education, 56(2), 281-292.
7. Geisler, S., Rolka, K., & Rach, S. (2023). Development of affect at the transition to university mathematics and its relation to dropout—Identifying related learning situations and deriving possible support measures. Educational Studies in Mathematics, 113(1), 35–56.
8. Guzmán-Valenzuela, C., Gómez-González, C., Rojas-Murphy Tagle, A., & Lorca-Vyhmeister, A. (2021). Learning analytics in higher education: a preponderance of analytics but very little learning? International Journal of Educational Technology in Higher Education, 18(1), 23.
9. Gyamfi, A., & Wren, D. G. (2022). Determining the difficulty and discrimination parameters of a mathematics performance-based assessment. Creative Education, 13(11), 3483–3489.
10. Khasawneh, E., Gosling, C., & Williams, B. (2021). What impact does maths anxiety have on university students? BMC Psychology, 9(1), 37.
11. Park, J. Y., Dedja, K., Pliakos, K., Kim, J., Joo, S., Cornillie, F., Vens, C., & Van den Noortgate, W. (2023). Comparing the prediction performance of item response theory and machine learning methods on item responses for educational assessments. Behavior Research Methods, 55(4), 2109–2124.
12. Rach, S., & Ufer, S. (2020). Which Prior Mathematical Knowledge Is Necessary for Study Success in the University Study Entrance Phase? Results on a New Model of Knowledge Levels Based on a Reanalysis of Data from Existing Studies. International Journal of Research in Undergraduate Mathematics Education, 6(3), 375-403.
13. Rezvanifard, F., Radmehr, F., & Rogovchenko, Y. (2023). Advancing engineering students' conceptual understanding through puzzle-based learning: A case study with exact differential equations. Teaching Mathematics and Its Applications: An International Journal of the IMA, 42(2), 126-149.
14. Trenholm, S., & Peschke, J. (2020). Teaching undergraduate mathematics fully online: a review from the perspective of communities of practice. International Journal of Educational Technology in Higher Education, 17(1), 37.
15. Vera, J. F. (2022). Distance-based logistic model for cross-classified categorical data. British Journal of Mathematical and Statistical Psychology, 75(3), 466-492.
16. Viberg, O., Hatakka, M., Bälter, O., & Mavroudi, A. (2018). The current landscape of learning analytics in higher education. Computers in Human Behavior, 89, 98-110.