Are student outcomes pre-determined?

– A study on the importance of prior knowledge in higher education mathematics

Ted Kronvall, Anna Lindgren, and Johan Lindström

***Abstract*—This work addresses the effect prior knowledge has on future learning in higher education; in particular how domain-specific knowledge (prior courses in mathematics) can predict the test scores in mathematical statistics. This observational study analyzes prior knowledge and test scores of 2576 engineering students registered for the basic course in mathematical statistics during a three-year period using linear, ordinal logistic and binary logistic regression. This work shows that: 1) a passing or failing grade can be predicted using prior knowledge alone, 2) the need for formal prerequisites is justified, 3) the learning outcomes might need to be updated for some engineering programs, and 4) some engineering programs show differences that cannot be explained by prerequisites alone.**

***Index Terms*—student learning, prior knowledge, prerequisites, mathematical statistics, regression, engineering, higher education mathematics**

I. What is prior knowledge?

The term *learning* is often defined as the cognitive process of linking new knowledge to prior knowledge [1]. In line with that definition, this work tries to validate the well-established result that prior knowledge plays an important role for learning outcomes (see, e.g., [2] and the references therein).

Prior *knowledge* is a broad term, potentially encompassing the entire wealth of knowledge relevant for an upcoming learning task. One may outline several dimensions of prior knowledge (see, e.g., [3]), one of which is the distinction between *domain-specific* and *domain-transcending* prior knowledge. While the former refers to specific knowledge within the same or similar subject matter; including declarative (what) and procedural (how) knowledge [4], domain-transcending knowledge has a wider scope, including e.g., *conditional* knowledge (when and why), *strategic* knowledge (for example study skills), and *knowledge-of-self.*

Knowing students' prior knowledge gives teachers the ability to adapt their teaching, but teachers should also use students' prior knowledge to determine whether they have the capacity of reaching the course's learning outcomes. Thus, prior knowledge may also be a factor for choosing learning outcomes, or prerequisites that limit student access to the course.

Measuring prior knowledge is, however, difficult as one needs to know, for every student, what knowledge is relevant for the learning task [5]. Domain-specific knowledge can be measured in several ways, e.g., in-class assessments done early in the course [6], but domain-transcending knowledge is more elusive to quantify. This work focuses on measuring prior knowledge using formal academic prerequisites, herein being student grades from previous courses in mathematics. Ideally, grades should reflect a student's domain-specific knowledge for the subject matter assessed, but in practice, as previous courses themselves also require domain-transcending knowledge, grades might also quantify knowledge in the wider sense.

While classic research of student pre-understanding describes the entire retention from learning as *knowledge*, newer research rather describes these outcomes as *knowledge, skills, judgement, behaviors, and abilities* [7]. While the latter distinction is more in line with the verbs of learning used in Bloom's revised taxonomy [8], this paper uses the term prior knowledge in the more inclusive meaning.

II. The course under study

The introductory course in mathematical statistics (*the Course*) is compulsory for most engineering programs at the Faculty of Engineering at Lund University (LTH) and is typically taken by students during year two or three. The Course is given in five variants across the different engineering programs, but all variants have a formal prerequisite of having passed at least one prior course in calculus. The purpose of this prerequisite is to set a minimum bar on the domain-specific knowledge required for a student to be able to achieve the learning outcomes of the course.

The primary aim of this study is to get a better understanding of how the student's prior knowledge, as measured using their results on the previous courses in mathematics (*the Prior Courses*), vary across engineering programs, and how this affects the results on the different variants of the Course.

III. Data and methodology

The basis of this work is an *observational study* of prior knowledge and test scores from 2576 students registered for one of the five variants of the Course during the three academic years 2022/23–2024/25. These variants service 16 out of the 18 engineering programs at LTH Campus Lund, see Table 1.

We collected all first-time registered students on the five course variants given during the study period. We excluded any student who, 1) studied another course variant than specified for their program, 2) studied FMSF50 at another time than specified; it is given twice a year for

Manuscript submitted November 6, 2025.

Ted Kronvall (corresponding author; email: ted.kronvall@matstat.lu.se), Anna Lindgren (email: anna.lindgren@matstat.lu.se), and Johan Lindström (johan.lindstrom@matstat.lu.se) are with the Division of Mathematical Statistics, Centre for Mathematical Sciences at Lund University, Box 118, SE 221-00 Lund, SWEDEN.

different programs, 3) already had a Bachelor of Science in Engineering from Helsingborg, 4) transferred to LTH from foreign university, 5) studied their Prior Courses at another Swedish university but where the original grades were not noted on the course transfer.

For the selected students, we also collected data on, 1) the reported grades (U, 3, 4, or 5) for the Prior Courses; two courses in univariate calculus, one course in linear algebra, and one course in multivariate calculus, 2) the number of points awarded on the final exam in the Course (0–70), and the corresponding grade (no-show, U, 3, 4, or 5). In addition, the Course has at least one, at most three, computer-based aptitude tests, that act as both formative and summative assessment during the Course; and the dataset includes whether these were passed on time, serving as a proxy for student procrastination.

Course variant	Programs	Min.prereq.	Color
FMSF20	D, E	B2	Black
FMSF80	I, F, Pi	B2	Light gray
FMSF50	L, R, V, BR, C, M	B1	Dark gray
FMSF75	W	B1	Mid gray
FMSF70	B, BME, K, N	B1	White

Table 1. The five variants of the Course and their minimum prerequisites. B1: at least 6 ECTS in uni- or multivariate calculus. B2: at least 6 ECTS in uni- or multivariate calculus including integrals. The color scheme is used in Figure 1.

The relation between prior knowledge and learning outcomes is modelled using a multiple linear regression model for test scores, as well as an ordinal logistic regression model for the corresponding grade, where the explanatory variables include the Prior Course grades, the aptitude test (pass/fail on time), and dummy variables for Course variant and engineering program.

IV. Results and analysis

A. *Differences in prior knowledge*

The differences in prior knowledge between different programs is illustrated using the average number of higher grades (4 or 5) on the Prior Courses, see Figure 1 (*Top*).

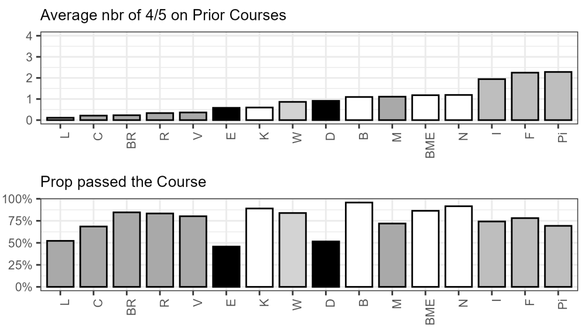

Figure 1. *Top*: Prior knowledge as the average number of grades 4 or 5 on the four Prior Courses, from lowest to highest. *Bottom*: Proportion of students that passed the Course. For gray scales, see Table 1.

It should be noted that for all Prior Courses, students have to answer additional questions to get grade 4 or 5. This results in grade 3 being by far the most common passing grade (70%), with fewer 4:s (17%) and 5:s (13%).

In general, programs F, Pi, and I have significantly better results on the Prior Courses than all the other programs, both in the number of courses they have passed, and in the grades achieved on them. It should also be noted that program C does not have a course in multivariate calculus, thus they are limited to three Prior Courses and, potentially, have a lower prior knowledge than the other programs.

B. *Differences in learning outcomes*

The proportion of students who pass the exam, see Figure 1 (*Bottom*), shows that despite a large difference in prior knowledge, most programs' students do well on the Course. Here, all exam questions are used to determine the grade, resulting in a more even distribution of passing grades with 3 (30 %), 4 (36 %), 5 (34 %). The notable exceptions are E and D who, despite mid-range prior knowledge, have the lowest proportions of passing grades. In comparison, both L and C, despite having the lowest prior knowledge among the programs, perform better than E and D.

C. *Modeling test scores using prior knowledge*

The estimated model for test scores is presented in Table 2, which shows parameter values and 95 % confidence intervals for the different explanatory variables (unit: exam points). The model's reference is a student of Course variant FMSF50 (L, R, V, BR, C, M) with minimal prerequisites; passing prior course B1 (see Table 1) with grade 3 without taking the aptitude test(s) on time. Such a student is expected to get on average 25.7 points on the exam (Pass is 35 out of 70). The expected exam points can be increased by taking the aptitude test(s) on time, having passed additional Prior Courses, and having a higher average grade on the Prior Courses.

After taking this prior knowledge into account, there remain differences between the Course variants, reflecting the varying difficulty levels relative to the students' prior knowledge. The "easiest" variant is thus FMSF70 while FMSF80 and FMSF20 are more "difficult". The only programs that diverge from the overall pattern within their Course variants are 1) L – which underperforms compared to the other programs on FMSF50 by, on average, 3.5 points, given the same prior knowledge, and 1) C – which overperforms by, on average, 5.0 points.

Variable	Estimate	95% CI
Reference: FMSF50, has one Prior Course with grade 3 and did not take the test(s) on time	25.7	(23.4, 28.1)
Took test(s) on time (Yes)	4.6	(3.1, 6.2)
Additional Prior Courses (per course)	4.7	(4.0, 5.4)
Average grade on the Prior Courses (per grade above 3)	10.5	(9.6, 11.4)
Course variant:		
-FMSF20: D, E	−9.6	(−11.0, −8.2)
-FMSF80: F, I, Pi	−7.2	(−8.5, −5.8)
-FMSF75: W	2.4	(0.5, 4.4)
-FMSF70: B, BME, K, N	5.0	(3.7, 6.4)
Program: L	−3.5	(−5.7, −1.3)
Program: C	5..0	(2.4, 7.7)

Table 2. Parameter estimates for linear regression of Course exam points among those who participated in the exam.

Together, these factors explain 37.8 % of the variability in the exam points of those who wrote the exam. Since the probability of writing the exam also depends on these same factors a binary logistic regression model was used (not shown), for predicting whether the student would write the exam or not. This model achieves a specificity of 73.4 %

(proportion of students not taking the exam who are correctly identified). A similar model, instead predicting whether the student would write and pass the exam achieved an accuracy of 74.2 %, which is a slight but significant ($p < 0.001$) improvement on the observed rate of 71 %. An ordinal logistic regression model was also used (not shown), for predicting the grade, including fail due to not writing the exam, predicting the correct grade (U, 3, 4, or 5) with an accuracy of 45.9 %.

V. Conclusions and future research

A. Conclusions

The main conclusion from the results in this work is that prior knowledge is indeed important for learning. More specifically, the following conclusions may be inferred:

- Accounting for prior knowledge, the level of FMSF20 is set too high for D and E students. For example, the average exam points are almost one full grade unit (10 exam points) lower than for a corresponding student on FMSF50, see Table 2. An E student with minimum prerequisites only has a 9 % chance to pass the exam, while a similar K student has a 49 % chance. Taking prior knowledge into account, a D or E student should on average only do slightly worse on the exam than an F, I or Pi student (−2.4 points). However, since, in practice, the latter student cohort has much stronger prior knowledge; passing more Prior Courses and with higher grades, their pass rate on the Course is much better than D and E.
- We see that, for the Course, continuous examination has a positive impact on the exam score, as passing the aptitude test(s) on time adds, on average, 4.7 exam points (half a grade unit).
- A reason for the under-performance of L students, 3.5 points on average when taking prior knowledge into account, might be that they study the Course during year two, immediately after the last of the Prior Courses, leaving little time for the prior knowledge to settle.
- The over-performance of C students, 5.0 points when taking prior knowledge into account, is similar to the contribution of passing an additional Prior Course. This compensation for not studying multivariate calculus might be a lurking effect of other courses giving similar domain-transcending knowledge, e.g., EDAA75 Discrete Structures.
- The students' grade distribution for the Course differs considerably from the grade distribution of the Prior Courses. The Prior Courses have a much higher proportion of the lower passing grades, while the Course has almost equal proportions. One possible reason for this is that the assessment for grade 4 or 5 is not elective in the Course. It can also be noted that increasing the average grade of the Prior Courses by one unit approximately corresponds to an expected increase in grade by one step for the Course, see Table 2. One may thus conclude that the higher average grade on Prior Courses for F, I, and Pi students compensates for the negative contribution of their course dummy (by being more "difficult").

B. Discussion points

Finally, we outline some discussion points for taking the discourse further:

- In this study, we see that the variants of the Course are compensating for the differences in prior knowledge between engineering programs. Is this an appropriate practice, given the different educational goals of the different programs?
- We see that for many programs, the inferred probability to pass the Course is small given the minimum formal prerequisite. How should we prioritize when setting the formal prerequisite, should we give students a chance or should we try to ensure some minimum level of prior knowledge?
- The models herein account program-specific contributions for C and L. Are these effects an indication of domain-transcending prior knowledge? Are there other lurking factors? For instance, it would be interesting to investigate if choosing to take the additional assessment for higher grades in the Prior Courses affects grades on the Course.
- Should the insights gained from this work be shared with the students, and if so, how? In particular, the binary logistic model infers that poor prior knowledge increases the risk of students not taking the exam. How can these students be better supported during their studies?

References

[1] Piaget, J. (1950). The Psychology of Intelligence (D.E. Berlyne, & M. Piercy, Trans.; 1st ed.). Routledge. https://doi.org/10.4324/9780203164730

[2] Hailikari, Telle, Anne Nevgi, and Erkki Komulainen. 2007. "Academic Self-beliefs and Prior Knowledge as Predictors of Student Achievement in Mathematics: A Structural Model." Educational Psychology 28 (1): 59–71. doi:10.1080/01443410701413753.

[3] Dochy, Filip J. R. C., and Patricia A. Alexander. "Mapping Prior Knowledge: A Framework for Discussion among Researchers." *European Journal of Psychology of Education*, vol. 10, no. 3, 1995, pp. 225–42.

[4] Alexander, Patricia A.. "Domain knowledge: Evolving themes and emerging concerns." *Educational Psychologist* 27 (1992): 33-51.

[5] Hailikari T, Nevgi A, Lindblom-Ylänne S. Exploring alternative ways of assessing prior knowledge, its components and their relation to student achievement: a mathematics based case study. Studies in Educational Evaluation. 2007;33(3-4):320-337. doi: 10.1016/j.stueduc.2007.07.007

[6] Angelo, T. A., & Cross, K. P. (1993). *Classroom assessment techniques: A handbook for college teachers* (2nd ed.). San Francisco, CA: Jossey-Bass Publishers

[7] Stanny, C. J. (2016). Reevaluating Bloom's Taxonomy: What Measurable Verbs Can and Cannot Say about Student Learning. *Education Sciences, 6*(4), 37

[8] Krathwohl, David R. 2002. "A Revision of Bloom's Taxonomy: An Overview." Theory Into Practice 41 (4): 212–18.