Who benefits from an exam retake option? A case study in Norwegian STEM higher education on test anxiety and performance

Ruben S. Thormodsæter*[1], Thomas Gjesteland [2], Ragnhild Gya[1] og Sehoya Cotner[1]

[1] Department of Biological Sciences, University of Bergen, Norway
[2] Department of physics and applied mathematics, University of Agder, Norway

*Presenting

Format:
Oral Presentation

Summary: Globally, we are not training enough scientists and mathematicians to meet pressing societal needs, patterns echoed in Norwegian higher education. For example, at one university in Norway, attrition between years one and two of the bachelor program in mathematics exceeds 30%. We see that introductory math courses are for many students a bottleneck for continuing STEM. As an effort to reduce failure, decrease negative emotions like test anxiety, and increase retention overall, low stakes testing with retakes was implemented in a large (n>400) introductory math course in Norway. The course changed from a single high-stakes exam to multiple low-stakes exams with up to 4 retakes for each. Here, we focus on the retake option, specifically, we tested how retakes affected overall performance and level of test anxiety. We used linear mixed effects models to test the impact of having multiple retakes, its effect on students' performance per exam, test anxiety and final grade. The failure rate went from a historic range of ~35% to 12%. Further, we saw test anxiety go down (estimate -0.44, $p<0.01$), and that high achieving students used fewer attempts, but that the students using more attempts did improve their score (Estimates = 1.23, $p < 0.001$).

Key Words:
STEM Education, Summative Assessment, Introductory Mathematics, Test Anxiety, Exam Retakes, Quantitative Research

1 Introduction

Globally, we are not training enough scientists and mathematicians to meet pressing societal needs, patterns echoed in Norwegian higher education. For example, at the University of Bergen, attrition between years one and two of the bachelor program in mathematics exceeds 30%. We see that introductory math courses are for many students a bottleneck for continuing STEM (Ellis et al., 2016; Jiang et al., 2020; Seo et al., 2019). This phenomenon cannot be attributed simply to lack of interest (de Brey et al., 2019; Medicine et al., 2016), prior knowledge (Valencia, 1997; Webb & Paul, 2023) or learning rate (Koedinger et al., 2023). Instead, lower rates of STEM retention appears to be driven by, among other things, the reliance on high-stakes exams (Stanger-Hall, 2012; Zakariya et al., 2022). One of the culprits for lower performance on high-stakes exams can be the increase of test anxiety. Where students commonly feel stress and anxiety prior to an academic evaluation, a higher academic and social pressure on the assessment can increase the amount of negative arousal, and higher test anxiety has been linked to lower performance (Ballen et al., 2017; Cotner et al., 2020; Ewell et al., 2022; Silaj et al., 2021). Here, we define test anxiety as a feeling of 'concern about possible negative consequences of failure on a test or other evaluation situations' (from Zeidner, 1998). In the current research, we examine how assessment methods, specifically the change from one single high-stake exam to multiple low-stakes exams with retakes relate to performance and test anxiety in STEM, with particular emphasis on an introductory-level math course.

As part of an effort to reduce failure and increase retention overall, we asked the following research questions: (1) how do retakes affect performance and level of test anxiety, and (2) is this type of intervention beneficial by increasing performance and decreasing test anxiety for students entering university with different math skill level?

2 Methods

In Norway, we studied a large (~400 students) introductory math course that changed from a single high-stakes summative assessment to multiple low-stakes exams with up to 4 retakes for each exam. All the students received, at the beginning, middle and end of the semester a survey through the class course webpage (Canvas), but it was volunteer to answer. The surveys asked for consent to collect their high-school grades, alongside the scores and grade they were getting in the course. To measure their test anxiety, we used items that were drawn from the Motivated Strategies for Learning Questionnaire (MSLQ; Pintrich & Others, 1991)). More specifically, we used four items that has been used and validated in several other studies (Cotner et al., 2020; Ewell et al., 2022; Papageorgiou, 2022; Silaj et al., 2021). Sample items, to which students could indicate their level of agreement on a 7-point scale, include "I am so nervous during a test that I cannot remember facts I have learned", and "I worry a great deal about tests". Students were grouped, based on their high school math grade, into low-performance (Norwegian equivalent of a score between F and E), medium-performance (score between D and C) and high-performance (score between B and A). Using the high-school math grades as a controlling factor, we then tested for correlations between retakes and

performance. We were also able to get course grades and high school grades for students that took the course back in 2019.

To answere the research questions, we used linear function models to test the impact of using multiple retakes. When comparing the data for 2019 to this year's (2023) we used a Chi-Squared test to test for significant differences. For investigating the test anxiety levels throughout the semester (start, mid, end), we built 4 linear mixed-effect models as growth models, with 4 different functional forms (nonlinear, linear, quadratic and log-linear). Test Anxiety was the response variable for all the models, with student-specific random-effect. Model selection was conducted using Akaike's Information Criterion (AIC (McElreath, 2018). The model with the lowest AIC value was considered the best-supported model given the data, this was the log-linear model (*Test Anxiety ~ 1 + log((Time-1) + 1) + (1|ID)*). All analysis for this study was done using R, version 4.3.2 (R Core Team, 2023).

3 Results and discussion

Participants

In total, 93 students gave us their consent and answered all the questions about test anxiety at the start of the semester. Of these 93 students, 13 students entered with a high school grade categorized as *low*, 57 with *middle* and 23 with *high* performance grade. For the dataset comparing between years (ignoring test anxiety), we had data from 141 students in 2019 (*N* low high school performance group = 44, *N* Middle = 70 and *N* High = 27), and 186 students in 2023 (*N* low = 65, *N* Middle = 91, *N* High = 30).

How did retakes affect performance, overall and for different groups of students?

High school grades were positive and significant on final grades in this course (Estimates = 0.63, $p < 0.001$). Overall, the failure rate in the low-stakes math class went from a historic range of ~35% to 12%. Further looking at the different performance groups, we saw that the ratio of students was approximately the same in 2019 and 2023, but the ratio of students performing better in 2023 was significantly, and positively (X^2=13.367 with *df* = 2, *p*-value = 0.001) changed, with fewer low-performance students, and more middle and high-performance students (figure 1).

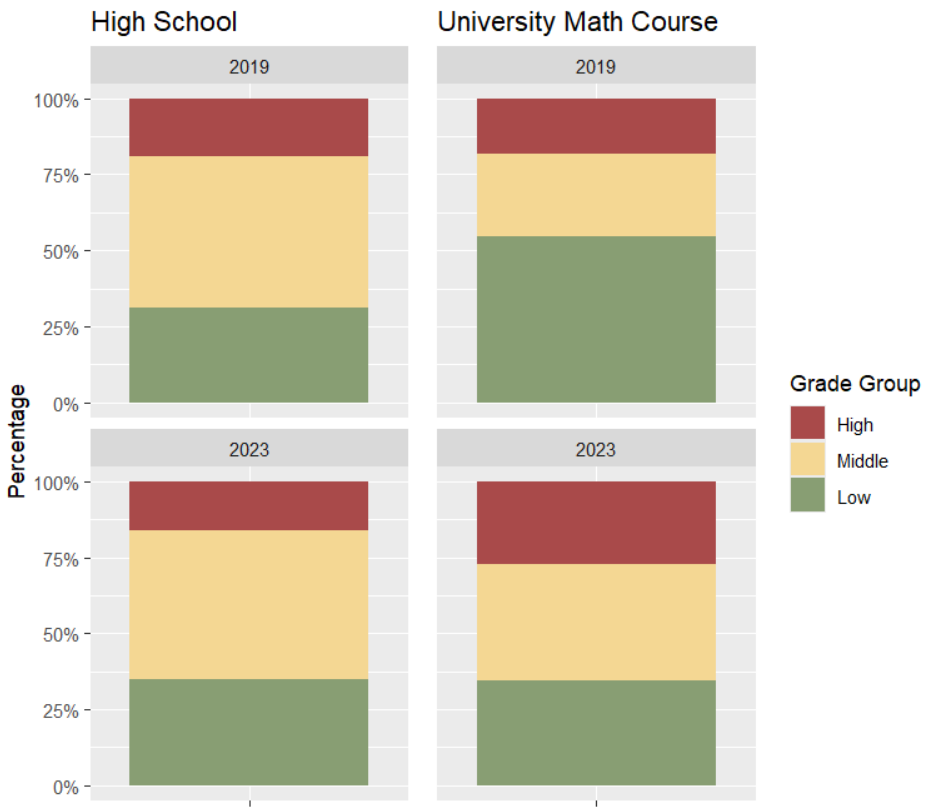

Figure 1 - Students grouped based on either high school-grades (left column) or grades they got in this course (right column) at two different years (first row being 2019, second row 2023). The groups are "Low" (F to E), "Middle (D to C) and "High" (B to A).

Looking at the different performance groups (based on high school math grade), we saw an expected outcome of high achieving students using fewer attempts compared to the other groups (Table 1).

Table 1 - Average exam score and number of retakes per exam for each performance group based on high school math grade (Low = F to E, Middle = D to C, High = B to A)

Exam number	High Performance - Average Exam score (average number of retakes)	Middle Performance - Average Exam score (average number of retakes)	Low Performance - Average Exam score (average number of retakes)
1	14.2 (2.52)	13.1 (3.05)	11.6 (3.26)
2	13.8 (1.92)	12.9 (3.00)	8.30 (2.88)
3	13.9 (2.81)	13.0 (2.93)	11.2 (2.94)
4	13.9 (2.43)	11.2 (2.53)	9.28 (1.62)

Crucially, the high-performance students did get on average higher scores than the other groups (High avg. = 13.95, Middle avg. = 12.55, Low avg. = 10.09). But students that did use more attempts did improve their score (Estimates = 1.23, $p < 0.001$). Critically, the retakes intervention appears to have benefited the middle- and lower-performing students more, with a higher slope on performance per retake, especially in the first exams, with the exception being the last exam (Figure 2).

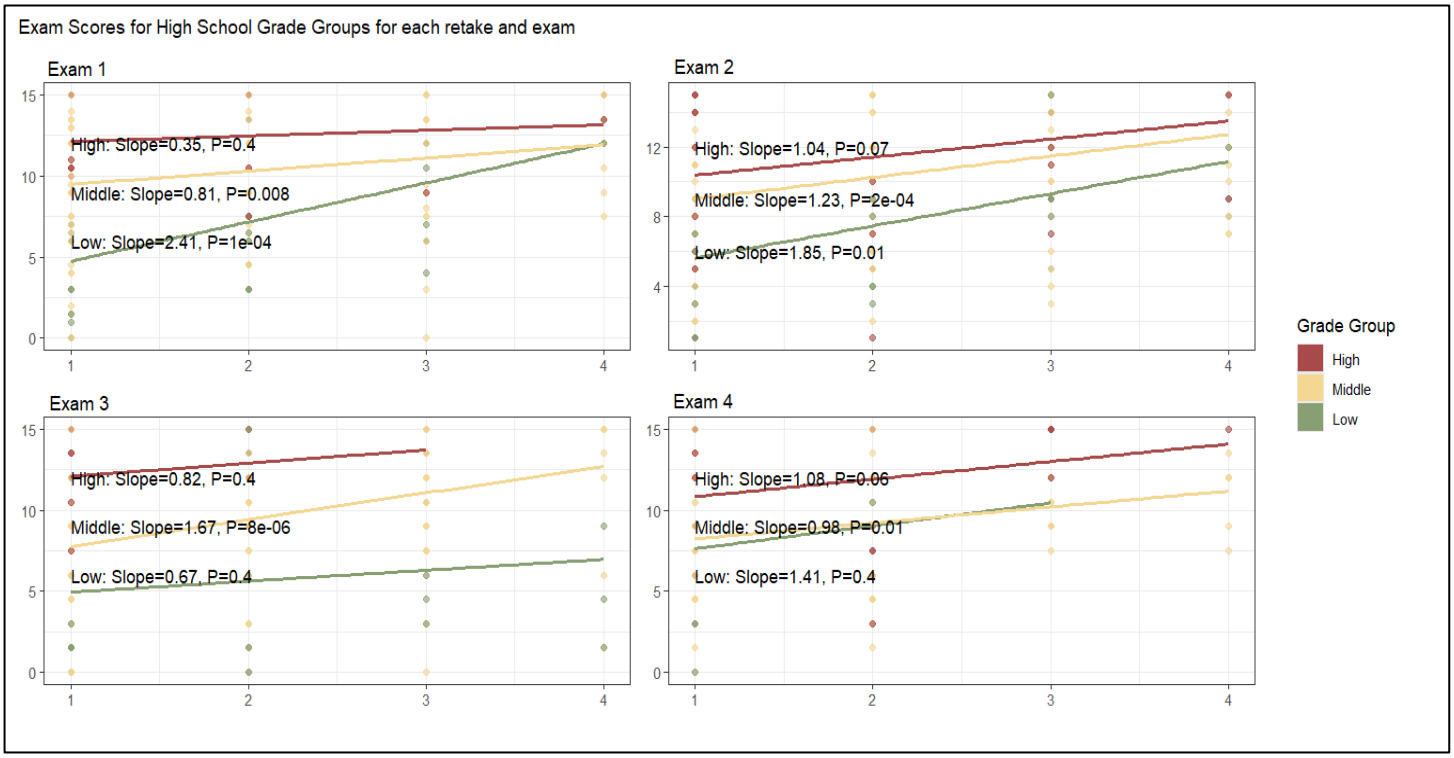

Figure 2 - Exam score (y-axis) for each high school grade group for each retake (x-axis) at the 4 exams during the semester. Green line representing students with low performance from High School (grade between F and E), yellow line representing students with middle grades (D to C), red line representing students with high grades from High School (B to A). The slope for each group at all the exams are postive, meaning that the group increased its score when retaking the exam. Low performance group has the highest slope in the first two exams, being significantly positive, with middle group being second most positve (also significant). High performance group has the flattest slope, positive but not significant. For the third exams it's the middle group that has the steapest alope, and the only group with a significant increase between each retake. The fourth exam has the High group with marginally significant increase, and Middle group with the only statistically significantly positve slope.

How did test anxiety affect performance, retakes and change over the course of the semester?

We were not able to produce any significant results regarding test anxiety and retakes, or the relationship between test anxiety and performance, contradictive to other studies on test anxiety (Cotner et al., 2020; Harris et al., 2019; Silaj et al., 2021). This might be due to the nature of low-stakes exams with retakes "takes away" the source of the negative arousal (high-stakes tests), and therefore do not have a clear impact on performance. We did see a significant and negative impact of high school grade on level of Test Anxiety.

And we did see that test anxiety goes down during the semester. To see this relationship between time and test anxiety, we can plot the predicted mean profile based on the log-linear model's fixed effects (Figure 3). Test Anxiety has a negative slope (and its significant, $p < 0.001$) when going from the start of the semester to the end of the semester.

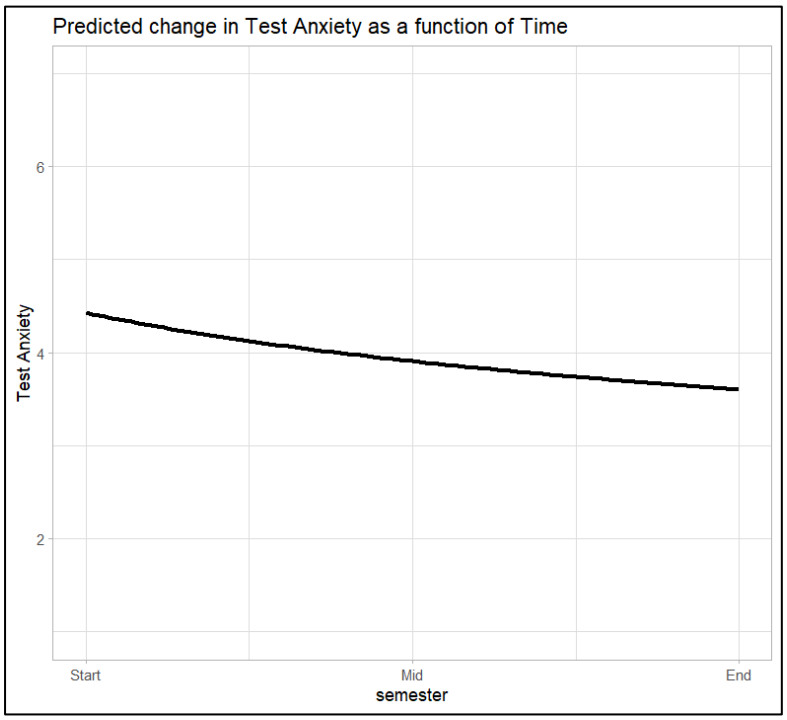

Figure 3 - Predicted change in Test Anxiety as a function of Time, with student-specific random-effect. Slope is significantly negative from Start to mid, and mid to end of the semester).

4 Limitations

Out of 65 students from the low-performance group, only 13 answered the test anxiety questions, leaving a large portion of the group we would expect had the highest amount of test anxiety. Further studies might evaluate this student group better, and more in depth.

At high-school in Norway, you can choose different math type, the recommended math to have for this course is called "R-math". For this early analysis, and for better comparison, we only looked at students that had R-math in high school, excluding students with other types of math because they must take a summer seminar before entering university, thus having a different pathway before entering uni. This group is smaller than the one we looked at, but still a group that should be included in further analysis.

5 Conclusion

These findings have clear implications for curriculum design and assessment choices. Our research contributes to our understanding of how low-stakes assessment but specifically retakes have a positive effect by addressing structural barriers. Students entering university with lower math grades from high school were able to increase their overall score when using the retakes option. With the results presented here, we can assume that with a low-stakes assessment course, without the retakes option would produce larger gaps between students at different high school math levels. As the current analysis stands, we are not able to further investigate test anxiety. We were not able to produce any significant results regarding test anxiety and retakes, or the relationship between test anxiety and performance, but we did see that test anxiety did go down during the semester. It is important to note that out of 65 students from the low-performance group, only 13 answered the test anxiety questions, leaving a large portion of the group we would expect had the highest amount of test anxiety.

The take home message from this study is that:

1) Students with better high school math grades do better at the university introductory math course and use fewer retakes.
2) Students with lower high school math grades do use more retakes but increase their score (on average) every time they do.
3) Test anxiety goes down during the semester, different from what we would expect in a course where the assessment method is based on higher stakes, without the option to retake.

This 'intervention' (going from a single high-stake to multiple low stakes with the option for retakes) aligns with broader efforts to create more inclusive and equitable practices in STEM education.

Citations

Adesola, S., & Li, Y. (2018). The Relationship between Self-regulation, Self-efficacy, Test Anxiety and Motivation. *International Journal of Information and Education Technology*, *8*, 759–763. https://doi.org/10.18178/ijiet.2018.8.10.1135

Agarwal, P. K., Nunes, L. D., & Blunt, J. R. (2021). Retrieval practice consistently benefits student learning: A systematic review of applied research in schools and classrooms. *Educational Psychology Review*, *33*(4), 1409–1453. https://doi.org/10.1007/s10648-021-09595-9

Ballen, C. J., Salehi, S., & Cotner, S. (2017). Exams disadvantage women in introductory biology. *PLOS ONE*, *12*(10), e0186419. https://doi.org/10.1371/journal.pone.0186419

Broks, V. M. A., Dijk, S. W., Van den Broek, Walter. W., Stegers-Jager, K. M., & Woltman, A. M. (2024). Self-regulated learning profiles including test anxiety linked to stress and performance: A latent profile analysis based across multiple cohorts. *Medical Education*, *58*(5), 544–555. https://doi.org/10.1111/medu.15283

Cotner, S., & Ballen, C. J. (2017). Can mixed assessment methods make biology classes more equitable? *PLOS ONE*, *12*(12), e0189610. https://doi.org/10.1371/journal.pone.0189610

Cotner, S., Jeno, L. M., Walker, J. D., Jørgensen, C., & Vandvik, V. (2020). Gender gaps in the performance of Norwegian biology students: The roles of test anxiety and science confidence. *International Journal of STEM Education*, *7*(1), 55. https://doi.org/10.1186/s40594-020-00252-1

de Brey, C., Musu, L., McFarland, J., Wilkinson-Flicker, S., Diliberti, M., Zhang, A., Branstetter, C., & Wang, X. (2019). Status and Trends in the Education of Racial and Ethnic Groups 2018. NCES 2019-038. In

National Center for Education Statistics. National Center for Education Statistics. https://eric.ed.gov/?id=ED592833

Ellis, J., Fosdick, B. K., & Rasmussen, C. (2016). Women 1.5 Times More Likely to Leave STEM Pipeline after Calculus Compared to Men: Lack of Mathematical Confidence a Potential Culprit. *PLOS ONE*, *11*(7), e0157447. https://doi.org/10.1371/journal.pone.0157447

Ewell, S. N., Josefson, C. C., & Ballen, C. J. (2022). Why Did Students Report Lower Test Anxiety during the COVID-19 Pandemic? *Journal of Microbiology & Biology Education*, *23*(1), e00282-21. https://doi.org/10.1128/jmbe.00282-21

Fournier, K. A., Couret, J., Ramsay, J. B., & Caulkins, J. L. (2017). Using collaborative two-stage examinations to address test anxiety in a large enrollment gateway course. *Anatomical Sciences Education*, *10*(5), 409–422. https://doi.org/10.1002/ase.1677

Hilpert, J. C., Stempien, J., van der Hoeven Kraft, K. J., & Husman, J. (2013). Evidence for the Latent Factor Structure of the MSLQ: A New Conceptualization of an Established Questionnaire. *Sage Open*, *3*(4), 2158244013510305. https://doi.org/10.1177/2158244013510305

Jiang, S., Simpkins, S. D., & Eccles, J. S. (2020). Individuals' math and science motivation and their subsequent STEM choices and achievement in high school and college: A longitudinal study of gender and college generation status differences. *Developmental Psychology*, *56*(11), 2137–2151. https://doi.org/10.1037/dev0001110

Koedinger, K. R., Carvalho, P. F., Liu, R., & McLaughlin, E. A. (2023). An astonishing regularity in student learning rate. *Proceedings of the National Academy of Sciences*, *120*(13), e2221311120. https://doi.org/10.1073/pnas.2221311120

Lenski, S., Zinke, N., Merkt, M., Reich-Stiebert, N., Stürmer, S., & Schröter, H. (2024). Early Indicators of Study Delay and Dropout: Test Anxiety and its Link to Exam Participation and Performance. *Journal of College Student Retention: Research, Theory & Practice*, 15210251241266547. https://doi.org/10.1177/15210251241266547

McElreath, R. (2018). *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. CRC Press.

Medicine, N. A. of S., Engineering, and, Engineering, N. A. of, Affairs, P. and G., Workforce, B. on H. E. and, Education, D. of B. and S. S. and, Education, B. on S., & Degrees, C. on B. and O. in C. 2-Y. and 4-Y. S. (2016). *Barriers and Opportunities for 2-Year and 4-Year STEM Degrees: Systemic Change to Support Students' Diverse Pathways*. National Academies Press.

Papageorgiou, E. (2022). Self-regulated learning strategies and academic performance of accounting students at a South African university. *South African Journal of Higher Education*, *36*(1), Article 1. https://doi.org/10.20853/36-1-4546

Pintrich, P. R., & Others, A. (1991). *A Manual for the Use of the Motivated Strategies for Learning Questionnaire (MSLQ)*. https://eric.ed.gov/?id=ED338122

R Core Team. (2023). *R: A Language and Environment for Statistical Computing* (Version R version 4.3.2 (2023-10-31 ucrt)) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Salehi, S., Cotner, S., & Ballen, C. J. (2020). Variation in incoming academic preparation: Consequences for minority and first-generation students. *Frontiers in Education*, *5*, 552364.

Seo, E., Shen, Y., & Alfaro, E. C. (2019). Adolescents' Beliefs about Math Ability and Their Relations to STEM Career Attainment: Joint Consideration of Race/ethnicity and Gender. *Journal of Youth and Adolescence*, *48*(2), 306–325. https://doi.org/10.1007/s10964-018-0911-9

Silaj, K. M., Schwartz, S. T., Siegel, A. L., & Castel, A. D. (2021). Test Anxiety and Metacognitive Performance in the Classroom. *Educational Psychology Review*, *4*(33), 1809–1834.

Stanger-Hall, K. F. (2012). Multiple-Choice Exams: An Obstacle for Higher-Level Thinking in Introductory Science Classes. *CBE—Life Sciences Education*, *11*(3), 294–306. https://doi.org/10.1187/cbe.11-11-0100

Valencia, R. R. (Ed.). (1997). *The Evolution of Deficit Thinking: Educational Thought and Practice*. Routledge. https://doi.org/10.4324/9780203046586

Varasteh, H., Ghanizadeh, A., & Akbari, O. (2016). The role of task value, effort-regulation, and ambiguity tolerance in predicting EFL learners' test anxiety, learning strategies, and language achievement. *Psychological Studies*, *61*(1), 2–12. https://doi.org/10.1007/s12646-015-0351-5

Webb, D. J., & Paul, C. A. (2023). Attributing equity gaps to course structure in introductory physics. *Physical Review Physics Education Research*, *19*(2), 020126. https://doi.org/10.1103/PhysRevPhysEducRes.19.020126

Zakariya, Y. F., Midttun, Ø., Nyberg, S. O. G., & Gjesteland, T. (2022). Reforming the Teaching and Learning of Foundational Mathematics Courses: An Investigation into the Status Quo of Teaching, Feedback Delivery, and Assessment in a First-Year Calculus Course. *Mathematics*, *10*(13), Article 13. https://doi.org/10.3390/math10132164

Zeidner, M. (Ed.). (1998). The Origins and Development of Test Anxiety. In *Test Anxiety: The State of the Art* (pp. 145–170). Springer US. https://doi.org/10.1007/0-306-47145-0_6