Reflections on diverse and inclusive teaching

Meike Akveld[1]
Department of Mathematics, ETH Zurich
8092 Zurich, Switzerland

Manuel Luethi[2]
Ecole Polytechnique Federale de Lausanne, Institute of Mathematics, EPFL
1015 Lausanne, Switzerland

Abstract

In this paper we will describe our attempt to change a very traditional 2 semester course in Linear Algebra for maths and physics students with a high-stake exam into a more diverse and inclusive experience. We will first describe the background of this course both in light of the Swiss educational system and in the setting of the Maths and Physics Department at the ETH Zürich. In particular, we will address recurring, gender-specific patterns which are common to most STEM subjects taught at ETH. This will be followed by some thorough considerations on the macro and micro planning of the course. We will then describe both the structure of the courses and the changes that were made and how the actual teaching was done and changed and we will finish with a description of the exam. Finally, we will analyse both the exam results and reflect on some observations made over the course of the two years we taught this course.

Description of the course and background

In this paper we will describe two consecutive years of a standard one year linear algebra course for mathematics and physics students at the ETH Zürich. This course is a two semester course, each semester consisting of 14 weeks of 4 hours of lectures and 2 hours of exercise classes. For various reasons to be described below, we decided to make some smaller and some bigger changes to the course and we will reflect here on the reasons for these changes and their effects.

The Swiss educational system

Before diving into the details of the ETH, its curriculum and the general conditions, we first need to explain the nature of the students entering the ETH as first years. The Swiss educational system is determined very much by the federal structure of the country. Switzerland consists of 26 cantons and up to and including secondary education the canton is effectively responsible for the curriculum taught. Although there is a national curriculum, the cantons can decide how to interpret this. In particular neither the number of hours dedicated to a particular subject, nor the total years of schooling, nor the precise contents to be taught nor the way of examining is determined centrally. Nevertheless, Swiss politics demand that every student who achieved the Matura via finishing the so-called Gymnasium (around 23 % of the women and around 16 % of the men, however this varies hugely over the regions, see Wolter et al. 2018) is granted free entry (i.e. no entrance exams or required minimal grades) to any degree he or she decides to pursue at university (with the exception of human and veterinary medicine - for practical reasons). Besides this, it is possible for students to put a special focus, e.g. on maths and physics, during their final years in the Gymnasium but such

1 akveld@math.ethz.ch
2 manuel.luthi@epfl.ch

a focus is never a required prerequisite for entering certain degrees. The result of all this is that there is as much as a factor two between the number of lessons in mathematics that students have had before they enter university. On top of that about a quarter of the students studying maths and physics at the ETH come from abroad (mainly Germany and Austria as teaching in the first year is done in German).

The result of the above mentioned system is that the population that starts a degree at the ETH is enormously diverse and it is not easy for first year lecturers to know where to start and what the right pace should be. How one can deal with this diversity is one of the topics of this paper; the considerations put forth should merely be understood as an attempt to tackle this problem with no claims for correctness or completeness.

The maths and physics department at the ETH and the degree structure

As mentioned before the ETH has no entry requirements besides the Matura for the Swiss students and therefore the exams at the end of the first year are often considered as such. Depending on the degree and the department up to 50 % of the first year students fail these exams at their first try or do not even show up. Therefore, these exams certainly qualify as high stake exams. Up until the summer of 2016 the examination always took place in one block in August which increased the pressure even further. It is important to note that the results for these exams were usually available to the students only a few weeks before the start of the new semester and therefore after the registration deadline in case the student wanted to pursue a different degree.

As in many institutions in Europe the number of female students in maths and physics is relatively low (varying between 20 % and 30 % of the first year population). It has been shown that the best means to predict performance for degrees in mathematics, natural sciences, and engineering sciences is obtained by combining the average grade in high school and the specific grade in mathematics (Trapmann et al. 2007). Nevertheless, an internal evaluation has shown that although women enter the ETH on average with slightly higher school grades (in mathematics) than their male peers, they fail the final exams at the end of the first year significantly more, see also Deiglmayr et al. (2019). This phenomenon is reminiscent of the findings in Rosser (1989), where it was found that high school GPA is only poorly correlated with SAT scores, although high school GPA is a good predictor of college GPA. As stated in Berkowitz et al. (in press), biological factors and differences in basic cognitive abilities may contribute to this phenomenon, but these cannot explain the substantial cross-cultural and historic variability in gender inequality in entry into STEM. Evidence discussed by Gayles & Ampaw (2014) suggests that women who study STEM subjects experience a "cool environment," are more likely to feel discriminated against, try harder, and seek help more often than men.

As a recurring theme to this article, Ballen et al. (2017) found that high stake exams systematically disadvantage women because of the stronger effect that test anxiety has on women’s performance. The high dropout rate after the first year and the underperforming of the women together with other, for this article perhaps less relevant, factors, made the maths and physics department decide to restructure their program. To this end they introduced the so-called “Split Basis Prüfung” meaning that a first block of exams was planned for January / February and the second block in August consisted of the remaining courses. The goals were manifold: to reduce the pressure over the year (less of a high stake exam), to provide feedback to the students earlier in the academic year, and also to reduce the gap in exam performance between the genders, e.g. splitting up the first year exams into two batches should convert the single high stake exam (period) into batches of slightly lower stake exams, thereby reducing the negative impact of high stake examinations on the performance of women.

Planning of the course

Linear Algebra and Analysis are the two main maths courses in the first year for maths and physics students and are traditionally taught by a full professor of the maths department, which in particular implied that they were taught by men. Given the observed impacts of belonging uncertainty (Deiglmayr et al. 2019) the department decided that at least one of the two courses was to be taught by a woman and the first author was asked to lecture Linear Algebra in the academic years 2016 / 2017 and 2017 / 2018 to which she agreed. As she was on sabbatical the semester prior to this, there was plenty of time to prepare the course.

When planning a course like this, you first of all need a red thread. In Linear Algebra one of the crucial decisions to be made at the beginning is the following: Do we start with “vectors and matrices” or with “linear maps and vector spaces”? In engineering courses almost always the first approach is chosen, which absolutely makes sense for a population meeting this material for the very first time and exactly needing matrix computations. However, our population does not consist of future engineers. Vector geometry is part of every high school curriculum, but matrices are done (intensively) by some and are not even mentioned by others. In order to remove the advantage of prior knowledge here, we decided to start with linear maps and vector spaces so that the course was from that start new for all. Another advantage is that we immediately dived into abstract maths and thereby perhaps gave a better representation of the topic for people pursuing a maths or physics degree. In particular, if we started with matrices, then the first exam in January would necessarily mainly have been about matrices and matrix computations; this would not truly reflect the character of the degree, thereby perhaps selecting the wrong people, and it would mainly benefit the students that were lucky to practice these calculations during high school.

During the sabbatical at the UC Berkeley the first author attended a series of lectures on linear algebra by professor Edward Frenkel which served as an inspiration for the style of our lecture and also introduced the lecturer to the book “Linear Algebra” by Friedberg et al. (2003), which was thereafter chosen as the main course reference. This was different from the traditional German literature for this course, which was still on the reading list. The reason for this choice was that it first of all follows the approach mentioned above - starting with linear maps - and secondly its style. In contrast to the classical German literature (Jänich 2004, Fischer 2014), it is a lighter read and has a more colloquial tone of explaining. This does not mean that details are omitted or that the mathematics is not rigorous - on the contrary - but that the language is a little less compact and thereby perhaps a little more accessible for people who are taking an abstract maths course for the first time.

Organisation of the course

When organizing the course, we focused on the following three observations for which we wanted to improve the students’ experience. First, we note that both the structure and format of the first year curriculum have developed historically. In particular, the two main mathematics courses, namely Linear Algebra and Analysis, are taught as classical lectures. Given the growth in population, mobility of students, and a continuous rise of the proportion of high school students obtaining a Matura, this results in a lecture for approximately 500 students. Add to this a reputation of the first year at ETH of being particularly difficult and the association of mathematics with brilliance (Deiglmayr et al. 2019), it is not very surprising to hear that many students do not perceive the classroom atmosphere as welcoming or inclusive. Second, to this day women are underrepresented among faculty at the math department of ETH. Such underrepresentation is often associated with higher levels of reported belonging uncertainty for the women entering such a degree (Deiglmayr et al. 2019). Third, the curriculum for the first year is very densely packed and so is each of the separate courses taught. Therefore, the connection between the separate courses is relatively difficult to spot so early in the career.

In order to emphasize the connection between the (in fact all) mathematical courses, the lecturers decided to teach the first two weeks of the semester jointly, i.e. instead of attending two distinct courses of four and six hours each per week, for the first two weeks of the semester the students were effectively taught (alternating by one or the other lecturer) one single course of ten hours per week, where the common foundations were introduced. After that, the course split up, each part focusing on a different set of tools and the associated type of questions; more algebraic in nature in Linear Algebra and more topological in Analysis. This initial collaboration also made it easier to coordinate other forms of coordination for the rest of the year, as will become apparent below.

Underrepresentation of women, although most pronounced at this level, is not only observed among faculty but already on the level of advanced students. The weekly exercise classes, held among groups of approximately 20 students and one teaching assistant, are usually organized by advanced students that passed the first two years of study with good to excellent grades. In the spirit of rewarding good performance, the teaching assistants were commonly picked according to the grades, which given the lower number of women passing the first year inadvertently resulted in exercise classes being almost exclusively led by male assistants. Of course, it is by no means necessary to have excellent degrees to lead a good exercise class and therefore, in an attempt to cope with belonging uncertainty, we chose the teaching assistants so that approximately half of them were female. There were enough advanced female students to ensure that the teaching assistants chosen still all had good to excellent grades.

Several measures were implemented in order to generate an inclusive and collaborative atmosphere. One measure was the organization of a dedicated Study Center, which was run jointly by the teaching assistants for the exercise classes in Analysis and in Linear Algebra. The Study Center was a place for the students to work on their exercise sheets, with the opportunity to ask the assistants for help. In the case of students asking questions, the assistants were supposed to pair the students with other groups of students working on the same problem and facing the same issues. This way, the students would communicate the inputs obtained among one another and ideally jointly work on the problems. In addition, both lecturers regularly showed up in the Study Center, again showing that they cared.

In a similar spirit, the exercise sheets in Linear Algebra contained marked exercises which were meant to be solved collectively in the exercise classes. The guiding principle for these exercises was that they should be of a more conceptual nature, requiring the students to precisely understand the abstract objects and properties discussed in the lectures. In particular, these exercises lent themselves to explication via examples and counterexamples. It was the duty of the teaching assistant to guide the discussion with the aim at deepening the understanding.

In order to encourage a collaborative and inclusive atmosphere in the exercise classes, we made an effort to provide a collaborative atmosphere among the teaching assistants as well as between them and the lecturer. In particular, there was a weekly meeting between the authors and the teaching assistants where the assistants reported on the difficulties the students were facing and the lecturer informed the teaching assistants about the upcoming material and the intended focus both in class and in the exercises. Moreover, we organized a sequence of peer-visits, i.e. the teaching assistants were visited in their exercise class by another teaching assistant. Besides the direct feedback, the assistants then discussed their observations in groups during the weekly meeting.

Finally, we need to mention that the students are allowed to take a summary of the course to the exam. It has been observed that often summaries consist of a collection of solutions to “typical” exam questions and essentially a transcript of most of the course in extremely small font. This in turn necessitates more and more difficult exam questions and, in addition, has the undesirable effect that every year the questions in the exam have to be essentially new. Both of these make it harder to come up with reasonable exams, as very difficult exam questions in effect force everybody to fail except for a few particularly talented students. In reality, instead of extreme rates of failure one will observe that the necessary score for a passing grade will be extremely low. In order to reduce this impact of the summary and in order to provide a means of participation, we set up an online summary - in the following called eSkript after the platform used - of the whole course and towards the end of the semester offered the students the possibility to vote on what items of the eSkript should be included in the summary, which was then handed out with the exam.

Teaching of the course

When teaching a new course, one always has to put special emphasis on the preparation of the very first lecture. In this lecture, the tone for the remaining lectures - in this case for a full year - will be set, and the expectations and attitude will be communicated to the crowd. Our main goal for the first lecture(s) was to make the students feel safe and fine. This new environment for them should be one characterised by mutual respect, a safe learning environment for all. In order to achieve this, we started with a very brief but personal introduction of ourselves, showing that we are more than a “proving machine” and human beings too. We also emphasized from the beginning that all questions are welcome and that no such thing as a stupid question exists. And then we really tried to live up to this (which is much harder than only stating it in the beginning).

A tool that we have learned to value quite a lot is the so-called EduApp, a Clicker tool developed by the ETH, which can be used on computers and mobile phones. Over the years evaluations have shown that students very much appreciate the use of this app in class. A common application of the EduApp is in the form of a single choice question asked via the app (in this case called a “clicker question”). The app then collects and displays the distribution of the students’ answers. One reason for the popularity of this is certainly that the tool is anonymous and that it not only allows to see the correct answer but also to see that, even if they answered the question wrongly, the students are usually among several students that made the same choice, therefore possibly strengthening a sensation of belonging even in case one could not answer the question correctly. This is in stark contrast to the case where one of the (usually exceptionally strong) students provides the (usually correct) answer after a frighteningly short period of time. We made an attempt to ask one or two clicker questions in every lecture. Such questions could either quickly check whether a definition was well understood, or have the students do a small calculation etc.

Usually this took no more than 3 minutes, sometimes giving time for short peer-to-peer discussion. These Clicker questions first of all give the students a short break - concentrating for 90 minutes is very hard - and also function as a short self assessment. The students see whether they understood the last concepts and also how well they did compared to their peers. At the same time, it gives the lecturer direct feedback on how well his / her teaching has been understood and finally, and this is a very important detail, it shows the students that the lecturer cares about their learning process. We believe that this last effect is much underestimated and deserves to be emphasised. Students are much more willing to work for a course, if they sense that their progress is of importance to the lecturer. Therefore, one should use every opportunity available to show that one genuinely cares.

In what follows we provide an example of such a question. After having introduced the concept of eigenvalues and eigenvectors, the question is whether the following statement is true or false “If a real matrix has one eigenvector, then it has infinitely many eigenvectors”.

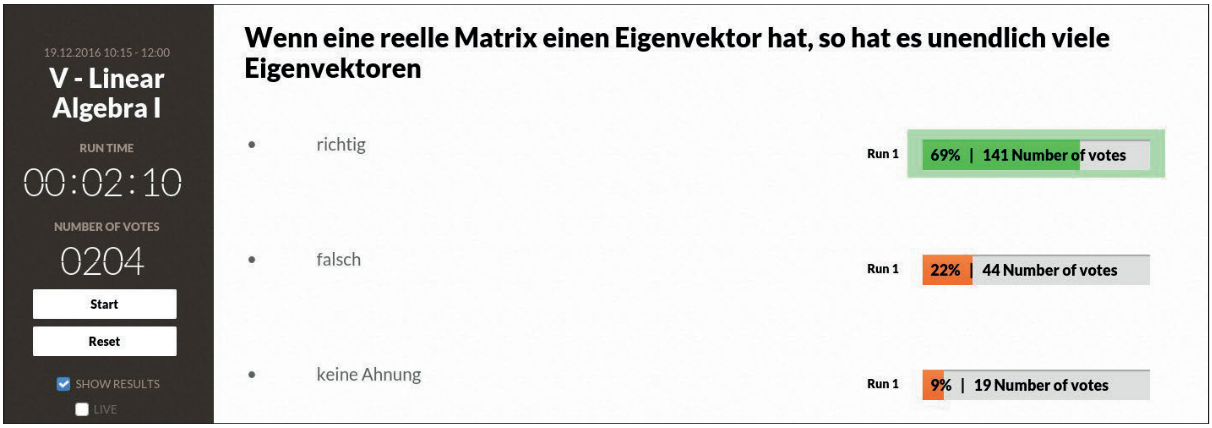
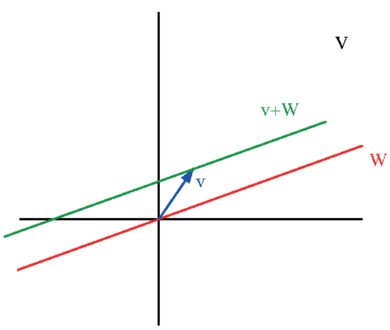


Figure 1: Screenshot from the EduApp Classroom Assessment Tool.

Note that we gave the students slightly more than 2 minutes for this question, as it needs no computation whatsoever. The last option "I don't know" is always there and it is encouraged that students do not guess but choose this option instead. This gives both the lecturer and the students a much more realistic picture about the understanding within the population and, indeed, this option is chosen quite frequently by the students. The question hints at the difference between eigenvalues (usually a finite set) and the concept of an eigenvector and the associated eigenspaces which show up right after this. As can be seen, almost 70% of the students were able to answer the question correctly already, whereas around 30% were still struggling with the concept. And even of the 70% that got the question right, it may have been the question itself that made them realise the validity of the statement. Clicker questions play a very different role than e.g. exam questions. The purposes are manifold and include feedback, deepening understanding, falling in common traps, the feeling of inclusion etc.

For most students a course in linear algebra is the first encounter with abstract mathematics. Therefore, it is of crucial importance that this encounter is smooth. Abstract maths can be scary and one needs to think very carefully about the right approach. We consider it advisable to be gentle in the beginning in order to keep up motivation - there will still be plenty of opportunity for frustration - but not at the cost of rigor so that eventually the students are able to start producing some mathematics, i.e. derive statements and their proofs, on their own. In order to be able to prove a statement, proper understanding is central. Linear algebra is very suitable for this start because a lot of geometric knowledge and intuition can be used in this process. As the lecturer is herself a geometer, she often used the "geometric picture" for clarification. In many cases, after proper clarification, the proof becomes merely a formality. Let us give an example. During the course the following lemma about cosets needed to be proved:

Lemma: *Let V be a vector space and let W a subspace of V. Then v+W is a subspace if and only if v is an element of W.*



Now this in itself is a rather abstract statement. If one, however, considers the picture, then not only the statement seems obvious, but also the proof quickly becomes clear. That's why we started with the following picture (before even stating the lemma) and from there the students actually could guess the statement and the proof became a technical detail. This way we think students will not only understand the statement of this particular lemma in depth but also start to develop some tools for proving. By proving we do not mean prove what one is asked to prove, but to think for oneself about what could be true and then see whether one can prove it. This lemma illustrates this principle well and teaches "doing maths".

And finally, and this may sound completely trivial, we gave a lot of thought to how we wanted to explain things. Of course, as a lecturer one understands the material of the course and one is able to provide a rigorous presentation of the material. But one has to consider how to present and, more to the point, explain it to a novice to the subject. This actually takes up enormous amounts of time and from the student evaluations we learned that this part of the teaching was highly valued (and presumably is not standard). One has to be aware that writing down a series of logically correct steps may be an acceptable proof, but it does not explain the heart of the matter. The following example illustrates this: Matrix multiplication. In traditional engineering courses students are simply told how to multiply matrices and then they do this and they don't question it any more. That's not what we want from maths students. After all, there is a reason why matrix multiplication is done in a specific way. We chose to introduce linear maps first and from the representation of linear maps by matrices and their composition, the definition of matrix multiplication becomes natural.

Exam

The exams consisted of two parts, a single choice part which accounted for 25% of the total of points obtainable and open questions which accounted for the remaining 75% of the total of points obtainable.

Each single choice question consisted of a statement and the student had to indicate whether the statement was true or false. These questions were purposefully kept very simple to the extent that to any mathematician the answers to the questions should be obvious. The total achievable by single choice only was however so small that it was very difficult to obtain a good grade if the student had to actually spend time on these questions. Such a block of single choice questions serves as a light entry into the exam and we think this is a good method to do simple knowledge probing. Moreover, our data shows that the number of points obtained in the single choice part of the exam is very highly correlated with the total number of points obtained for the remaining questions, i.e. excluding the single choice part, therefore supporting the hypothesis that they test a similar set of skills as do the more classical open questions.

A common alternative to the simple single choice questions is a smaller set of more difficult single (or possibly multiple) choice questions offering more points. In case of multiple choice questions, it is customary to award correct answers with a point and subtract a point for wrong answers, as otherwise the only reasonable strategy is to mark every answer on offer. This introduces an element of risk to the examination, which in turn induces elicitation of behavioral patterns. Besides the fact that the goal of the exam is to assess the students' knowledge of the topic instead of behavioral patterns, the introduction of risk in combination with gender-specific risk-assessment might even systematically put one gender at a disadvantage (Schubert et al. 1999, Fehr-Duda 2006, Schubert 2006, Marin & Rosa-Garcia 2012, Baldiga 2014). Therefore, the decision for single choice over multiple choice. Our decision to choose a large set of simple questions was made in light of the findings that high stake exams tend to systematically "disadvantage women because of the stronger effect that test anxiety has on women's performance" (Ballen et al. 2017).

The remaining part of the exam consisted of open questions, which were organized in accordance with the material of the course, i.e. six (respectively five in the second iteration) questions each covering one of the main subjects of the course; the total number of questions in the second iteration was reduced because we noticed that in the first year the exam was perceived to be much too long. Every open question was split into three to four smaller parts of increasing difficulty to provide the students with a relatively comfortable entry to the topic and a means to warm up. Moreover, partial points were awarded and there was always the possibility to solve a later part of a question even if one was unable to solve the earlier parts.

Analysis of the exam performance

As a result of having taught the course over a total of two semesters, we had data for four examinations available. In particular, we have observations of individual performance not only on the level of the exam, i.e. individual grade, but on the level of exercises within the exam.
We could not observe any qualitative effect of the above changes on the performance of women on average. The patterns are in line with the previous observations on the aggregated level, cf. the results reported in (Rütsche et al. 2019). On average, both on the level of the exams and (in most cases) on the level of individual exercises, the female participants were not performing better than the male participants. This is independent of whether the exercises were of single choice type, i.e. closer to knowledge probing, or open questions. Nor could we identify from this data any gender-specific factors that could influence overall performance, e.g. specific strategies to solving the exam. In the end, even though the students explicitly appreciated the atmosphere in the course as well as many of the measures introduced, the data on relative performance is very much in line with the data for the whole year and similarly with the results from previous years.

Conclusion and further reflections

As described throughout the article, we introduced several measures simultaneously and therefore it is not possible to judge the effectiveness of each measure on its own. Overall, we conclude that the combination of these measures does not affect the discrepancy in failure rates in a significant way. In particular, we conclude that the gender gap has to be addressed separately and might require much more substantial examination besides a few tweaks in the classroom.

In view of the evaluation of the course, we still recommend most of the measures introduced for the course, not with an intent of influencing the performance gender gap but rather with the aim of generating an inclusive classroom atmosphere. The collective generation of a summary was not appreciated among the students and there is doubt that this was truly perceived as a collaboration among students and the organizers of the course. In view of the resource intensity of this measure, it seems advisable to look for a different way to attack the problem with the individual summaries. With respect to the other measures, the students explicitly mentioned the welcoming atmosphere, the use of the clicker-questions, and the apparent interest of the lecturer as motivating factors for the course.

Acknowledgements

We would like to thank Daniela Hansen for pointing out several references of interest. We are indebted to Bruno Rütsche for several informal discussions and for providing us with some results of their report at an early stage. Many thanks to Renate Schubert, whose vast work on and experience with gender-specific risk assessment helped both for the planning of the course and the interpretation of the results. ML would like to thank Reto Zihlmann for several discussions regarding statistical methods to identify examination strategies from exam responses.

Bibliography

Baldiga, K. (2014). Gender Differences in Willingness to Guess. *Management Science*, 60(2), pp. 434-448. Online: https://doi.org/10.1287/mnsc.2013.1776

Ballen, C. J., Salehi, S. & Cotner, S. (2017). Exams disadvantage women in introductory biology. *PLoS ONE*, 12(10), e0186419.

Berkowitz, M., Stern, E., Hofer, S. & Deiglmayr, A. (in press). Girls, boys and schools: On gender (in)equalities in education. In F. M. Cheung & D. F. Halpern (eds.). The Cambridge International Handbook on Psychology of Women. Cambridge: Cambridge University Press.

Deiglmayr, A., Stern, E. & Schubert, R. (2019). Beliefs in "brilliance" and belonging uncertainty in male and female STEM students. *Frontiers in Psychology*, 10, No. 1114.

Fehr-Duda, H., de Gennaro, M. & Schubert, R. (2006). Gender, Financial Risk, and Probability Weights. *Theory and Decision*, 60, pp. 283-313.

Fischer, G. (2014). Lineare Algebra. Berlin: Springer.

Friedberg, S. H., Insel A. J. & Spence, L. E. (2003). Linear Algebra. London: Pearson.

Gayles, J. G. & Ampaw, F. (2014). The Impact of College Experiences on Degree Completion in STEM Fields at Four-Year Institutions: Does Gender Matter? *The Journal of Higher Education*, 85(4), pp. 439-468.

Jänich, K. (2004). Lineare Algebra. Berlin: Springer.

Marin, C. & Rosa-Garcia, A. (2012). Gender bias in risk aversion: evidence from multiple choice exams. MPRA Paper No. 39987, posted 10. July 2012.

Rosser, P. (1989). The SAT Gender Gap: Identifying the Causes. Washington, D.C.: Center for Women Policy Studies.

Rütsche, B., Altherr, W., Deiglmayr, A., Dittmann Domenichini, N., Kapur, M., Kortemeyer, G., Schlienger-Merki, C., Schubert, R., Stern, E. & Vaterlaus, A. (2019). Interner Bericht zur Pilotphase mit geteilter Basisprüfung: Evaluation basierend auf den Eintrittsjahrgängen 2016 bis 2018. ETH Zürich.

Schubert, R. (2006). Analyzing and Managing Risks – On the Importance of Gender Differences in Risk Attitudes. *Managerial Finance*, 32(9), pp. 706-715.

Schubert, R., Brown, M., Gysler, M. & Brachinger, H.W. (1999). Financial Decision Making: Are Women Really More Risk Averse? Gender and Economic Transactions. *The American Economic Review Papers and Proceedings*, 89(2), pp. 381-385.

Trapmann, S., Hell, B., Weigand, S. & Schuler, H. (2007). Die Validität von Schulnoten zur Vorhersage des Studienerfolgs – eine Metaanalyse. *Zeitschrift für Pädagogische Psychologie*, 21/1, S. 11-27.

Wolter, S. C., Cattaneo, M. A. & Denzler, S. (2018). Bildungsbericht Schweiz 2018. Schweizerische Koordinationsstelle für Bildungsforschung (SKBF).