Self- and Peer-Evaluation of Individual Project Work: An Innovative Course Assessment Method to Increase Student Motivation

Marcel Frehner

Abstract Besides teaching itself, which most university lecturers enjoy, evaluating and particularly grading students is one of the biggest challenges. While modern educational developments transform university courses from pure knowledge-oriented to skill-oriented, the way students are evaluated and graded has not kept up with these developments. Classical written or oral examinations often only assess factual knowledge, and not the desired skills. Here, an innovative evaluation procedure is presented that applies self- and peer-evaluation (and -grading) of the students themselves. The procedure is used to evaluate and grade end-of-semester projects in a Master's level course. After finishing their end-of-semester project work, students are asked to evaluate and grade their own work (self-evaluation) and the work of one fellow student (peer-evaluation). For both evaluations, they have to give a mark and write a one-page evaluation report that justifies the mark they give. This mark is not modified in any way by the teacher and counts one-to-one for the final mark; hence, the students have to take responsibility. The final mark is then averaged between the self- and peer-evaluation, as well as the mark from the teacher. The outcome of three years of experience with this evaluation procedure shows that the students can handle the given responsibility very well and take both the self- and peer-evaluation very seriously. Compared to the teacher's mark, they do not evaluate themselves or their colleagues overly positive or negative. If anything, the self-evaluation is slightly more negative and the peer-evaluation slightly more positive than the teacher's evaluation. On average, the final mark is exactly the same as the teacher's mark. Nevertheless, the evaluation procedure allows training several soft-skills that are crucial for the student's future career, such as having an opinion, taking responsibility, or being objective and honest with oneself and colleagues.

Keywords Assessment method · Project work · Self-evaluation · Peer-evaluation Soft skills

M. Frehner (✉)
Geological Institute, Sonneggstrasse 5, 8092 ETH Zurich, Switzerland
e-mail: marcel.frehner@erdw.ethz.ch; mf@marcelfrehner.ch

S. Mukherjee (ed.), *Teaching Methodologies in Structural Geology and Tectonics*, Springer Geology, https://doi.org/10.1007/978-981-13-2781-0_2

1 Introduction

Evaluating and grading students in semester courses is a very crucial task for every university lecturer. However, while most lecturers enjoy teaching a course, grading is often the least enjoyable task. Fortunately, there are many guidelines to help lecturers design exams, both as international publications (e.g., Biggs 1996; Downing and Haladyna 2011; Race et al. 2005) and as institutional guidelines (e.g., ETH Zurich 2013 in case of the ETH Zurich). According to the "*Guidelines on Grading Written Examinations*" of the ETH Zurich (ETH Zurich 2013), an exam is a good exam if the three principles objectivity, reliability and validity are respected. Each of the three principles is outlined below with a question and a statement:

- Objectivity:
  - Is the measurement independent of the person who measures and the circumstances of the measurement?
  - Different people reach the same conclusion, independent from each other!
- Reliability:
  - Is the measurement reliable? Does it deliver reproducible results?
  - Different versions of the measurement yield the same result!
- Validity:
  - Do I measure what I intend to measure, or something else?
  - The measurement measures what it pretends to measure!

Thereby, the last principle (validity) can only be achieved if the two others are assured. Hence, validity is the prime aim of designing an exam. While these three principles are specifically defined for written exams, they should also hold for any other type of examination at the university level.

The problem with these three principles is that it is impossible to ever achieve them. For example, every exam always also tests a certain level of fluency of a language. The students not only have to know something, but first they have to be able to understand the question and then to explain/express what they know. This is true for written exams, but even more for oral exams. The latter in addition also tests to a certain level the self-confidence of a student and hence his/her personality. If student A does not know very much during an oral exam (i.e., little factual knowledge), but presents everything very elegantly and self-confidently (good soft-skills), he/she will unavoidably get a better final mark than student B, who knows as little as student A (i.e., equal factual knowledge) but is less self-confident in her/his presentation (poor soft-skills). The same problem occurs if reports are evaluated and graded, for example reports of end-of-semester projects. In this case, the language proficiency plays an important role, but also the visual impression (document layout, figures). Hence, not only the project work is evaluated, but also several other soft-skills, which are not intended to be tested and marked. For example, a figure that is very well made and clear, but does not contain much information, still impresses the reader and may

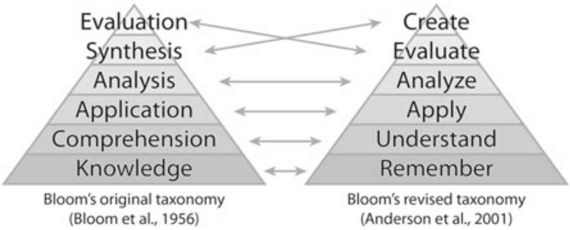

Fig. 1 Graphical representation of Bloom's original and revised taxonomy of educational objectives after (Bloom et al. 1956) and (Anderson et al. 2001), respectively. The taxonomy describes the increasing levels and complexity of learning and thinking, which should also be addressed by educational developments and examinations

positively influence the evaluation and grading. With all these examples, it should be clear that a true validity cannot be reached when evaluating and grading students.

The best way to at least come close to fulfilling the three principles described above is to use multiple-choice (MC) exams. MC-exams are influenced very little by language issues or any other confounding factors, and they allow to really test the pure knowledge of students. However, MC-exams are very limited in testing the higher levels of learning and thinking according to Bloom's taxonomy (Fig. 1; Anderson et al. 2001; Bloom et al. 1956; Krathwohl 2002), such as the analysis and evaluation of results. In Bloom's taxonomy (Fig. 1), each level builds on all lower levels. For example, students can only apply a concept to a specific problem (third level) if they know it (first level) and understand it (second level); and only when they are able to apply the concept, they can analyze the corresponding results or analyze results from a different application. For university-level teaching (in particular on the Master's level), it is desirable to reach the highest levels in Bloom's taxonomy. Of course, the students need to know and understand the necessary concepts and be able to apply them to specific problems. However, at the end of the Master's program, the students should also be able to analyze and evaluate results (e.g., not just understand, but have an opinion when reading a scientific article) and eventually also be able to create new applications (i.e., have their own ideas). Fortunately, modern curriculum reforms respect this; university courses transform more and more from knowledge-oriented to skill-oriented. However, the way students are evaluated and graded has not changed considerably and does often not keep up with the educational developments.

The highest levels of Bloom's taxonomy are also required in the student's future career (e.g., in industry, as Ph.D. student). There, the three principles for evaluations are not guaranteed either. An evaluation (e.g., by a customer; in a job interview) will always depend on the person who evaluates (i.e., no objectivity), it will not be reliable because in many cases there is no option for a repetition and hence no indication for reliability, and evaluations deliberately consider many different aspects of a piece of work or of a person (i.e., no validity). In other words, traditional exams at universities do not represent typical evaluations in real life; hence, they do not prepare the students very well for their future career. However, universities should do exactly this: prepare the students for their future career, not only in terms of knowledge, but also in terms of how they and their work are evaluated. Therefore, I believe that, at least on the Master's level, university teachers should be allowed to ignore the three principles above for a "good" exam and design evaluation procedures that also consider crucial soft-skills, such as teamwork, communication and presentations skills, technical writing skills, or self-confidence, which count as much as or even more than the technical/scientific knowledge.

In this article, I present an evaluation procedure of end-of-semester project work that applies self- and peer-evaluation by the students themselves in a Master's level course. First, I describe the course, the end-of-semester projects, and the evaluation procedure in detail and provide all the documents handed out to the students in the Appendix. Second, I report on the outcome of the evaluations and compare them to my own evaluation. I implemented the evaluation procedure in 2014 and therefore can present data for three consecutive years. Last, I discuss the pros and cons of the presented evaluation procedure and give some advice about critical points that can render the evaluation procedure a success.

2 Course Description and Project Work

Here I provide an overview of the course "Numerical Modeling of Rock Deformation" (NMRD), in which I implemented the innovative evaluation procedure. The NMRD-course is a regular semester course at the Department of Earth Sciences, ETH Zurich. It comprises two contact hours per week during each fall semester (14 weeks; Table 1) and is worth four European Credit Transfer System (ECTS) points. For a 2-h/week course, three ECTS points are more common. However, the NMRD-course requires a significant amount of extra work compared to other courses because of the semester-end project work and its special evaluation procedure described in this article. Therefore, the extra ECTS point is justified. The course is offered on the Master's level in the Structural Geology Module (Master in Earth Sciences → Major in Geology → Structural Geology Module). However, the course is open to all Master students of the ETH Zurich. In particular, it is recommended for Earth Sciences Master students with Major in Engineering Geology (Master in Earth Sciences → Major in Engineering Geology) and it is regularly taken by students from civil engineering, material sciences, or environmental sciences. Also, Ph.D. students from all these disciplines regularly take the NMRD course. Usually, the course is taken by 15–20 students each fall semester, out of which 13–18 students seek the final mark.

Table 1 Schedule for the course "numerical modeling of rock deformation" at the ETH Zurich. The last two weeks are devoted to the individual exercises the students work on in groups of two

Block	Week	Topic
1	1	Introduction to the course and introduction to MATLAB
	2	Continue with introduction to MATLAB and exercises
	3	Kinematic models—strain ellipses (MATLAB exercise)
2	4	Basics on continuum mechanics
	5	Basics on rheology
	6	Introduction to the finite-element (FE) method
3	7	FE—Exercises in 1D and isoparametric elements
	8	FE—Going to 2D
	9	FE—2D elastic deformation
	10	FE—Stress calculation and visualization
	11	FE—2D viscous flow
	12	FE—Heterogeneous media and wrap-up
4	13	Time for exercises
	14	Time for exercises

2.1 Course Content and Schedule

The primary aim of the course is to enable students to solve simple mechanical/geological problems using the finite-element (FE) method. Applications are either elastic structural problems (e.g., stress analysis) or viscous flow problems (e.g., buckle folding, shear zones). Thereby, they do not use existing commercial or open-source codes, but develop the necessary FE-codes themselves during the course. Hence, further principal aims of the course are to understand all relevant mathematical and mechanical concepts to formulate the mechanical problem (i.e., continuum mechanics, rheology), understand the basics of the FE-method, and learn how to program an FE-code (Table 1).

To teach all of this, I divided the NMRD-course into four blocks (Tables 1 and 2), which are distinctively different in their content and didactical approach. These blocks are:

1. ***Introduction to MATLAB***: Some students take the NMRD-lecture without any experience in MATLAB or programming; others have quite some experience. Hence, the principal aim for this block is to bring the students to the same (or similar) level of prior programing skills. To achieve this, the students work on MATLAB/programing exercises by themselves or in small groups of two or three. I and one teaching assistant help/teach the students one-to-one at the computer workplaces.
2. ***Theoretical background***: During this block, the students learn the basics of continuum mechanics and rheology, as well as the FE-method. These are traditional lectures using a frontal teaching style. I derive the relevant equations on the

blackboard, and the students copy everything by hand to their personal lecture notes. At the same time, I provide PowerPoint and pdf-slides online with all the material that I explain on the blackboard for the students to prepare and repeat the lectures.

3. ***FE-theory and -exercises***: At the end of block 2, I teach the basics of the FE-method. Now in this block, students use, apply, and extend the FE-code going to more and more advanced applications (1D → 2D elastic problem → 2D viscous problem). For this, I provide short theoretical inputs in the form of PowerPoint presentations or self-study tutorials, but the majority of the time the students spend programing their FE-codes. As a starting point for each major step and because of time restrictions, I provide FE-codes "with holes", where the students program the critical gaps. I and one teaching assistant help/teach the students one-to-one at the computer workplaces.
4. ***Application to a particular problem (see Sect. 2.2)***: In the end-of-semester project work, the students finally apply their FE-code to an individual problem, which they solve in groups of two. There is no teaching anymore. However, I and one teaching assistant help/teach the students one-to-one at the computer workplaces.

Table 2 Description of the teaching styles during the four distinct blocks of the course "numerical modeling of rock deformation" at the ETH Zurich

Block	Teaching style
1	• Mainly computer/programing exercises using MATLAB • Students get short handouts and problem descriptions • Students may work alone or in small groups of 2–3
2	• Theoretical background lectures • Traditional frontal teaching using the blackboard • Students copy everything by hand to their personal lecture notes; however, PowerPoint and pdf-slides are provided online with all the material from the blackboard
3	• Theoretical input (~1/4) and MATLAB/programing exercises (~3/4) are mixed • Theoretical input is provided using short PowerPoint presentations or self-study tutorials; the latter mostly mixed between theory and exercises • FE-codes "with holes" serve as the basis for the exercises. Hence, students do not program everything themselves, but fill in critical gaps
4	• Individual exercises the students work on in groups of two • No lecturing anymore • The students can discuss with the teacher and/or the assistant

2.2 *End-of-Semester Project Work*

The first three blocks of the NMRD-course are rather theoretical and not really applied to any specific problem. The aim is to understand the basics of the FE-method and learn how to program an FE-code. However, the overarching aim of the NMRD-course is to enable the students to solve a specific problem. Therefore, block 4 is the culmination of the course, where the students apply their prior knowledge (and their FE-code) to a specific problem of their choice. I ask the students to form groups of two, although I also allow them to work alone or in groups of three in exceptional cases.

I want the students to profit from this last part of the lecture as much as possible. Therefore, each group can choose the problem/application they want to work on themselves. They can choose between an elastic and a viscous problem, for both of which they have developed a running FE-code during the course. Hence, for these types of applications, the main challenge is to set up the geometry and the numerical grid, as well as using the correct boundary conditions. Students may also choose a more theoretical/mathematical/methodological problem, which typically involves deriving new equations and implementing new FE-terms in their code. However, students may also define their own project (in discussion with me) based on their personal interests. Such projects may be defined in relation to a student's Master's thesis, another lecture, or any other personal interests.

To help students decide on the project they want to work on, I provide them with a short explanatory note some two weeks before they have to make the choice. This explanatory note can be found in Appendix A. Three examples of end-of-semester project descriptions, exactly the way the students receive them from me, are provided in Appendix B, Appendix C, and Appendix D. These examples comprise two viscous flow problems (Appendix B: development of parasitic buckle folds in a multilayer stack; Appendix C: horizontal sill emplacement below a sharp lithological boundary) and one elastic problem (Appendix D: stress distribution around two tunnels).

3 Course Assessment

For the final grading, it was important to me to assess the student's achievement of the major aim of the course, i.e., being able to solve a specific problem. Because this is rather a skill than just plain knowledge, it was clear to me that I will not base the final mark on a written or oral exam, in which I could only assess knowledge, but not the desired skill. Therefore, the final mark is based on the student's performance in their end-of-semester project work. To do that, the students not only have to solve their individual problem, but also write a four-page report explaining their solution strategy, show and interpret the results, and draw some conclusions. This report is also written in groups of two and should be structured very much like a scientific article with its typical sections. To help the students structuring their report, I provide them with some general remarks regarding this report, which can be found in Appendix E. Ultimately, each group has to hand in their report together with all the necessary FE-codes to solve their specific problem.

3.1 Self- and Peer-evaluation

After the students handed in their report and FE-codes, the work is evaluated by three parties:

- By the teacher (me)
- By the student her/himself (i.e., self-evaluation)
- By two fellow students from the same course (i.e., peer-evaluation)

The students have to take responsibility and actually set a mark for themselves and for the report they peer-evaluate. The marks from the self- and peer-evaluation are not altered in any way by me but count one-to-one for the final mark. To calculate the final mark, all three marks are averaged with equal weight (i.e., 1/3). However, students working alone on their end-of-semester project only receive one peer-evaluation while students working in groups of two receive two peer-evaluations. Therefore, in the latter case, the two marks from the peer-evaluations are averaged first and only this average counts 1/3 for the final mark. The key how to calculate the final mark is shown Tables 3 and 4 for the two cases.

The redistribution of the reports among all students for the peer-evaluation is done by me based on a pre-defined key. To define this key, I follow two simple rules:

- One report cannot be peer-evaluated by two students of the same group of two
- Two students (and also two groups) cannot peer-evaluate each other. That means no swapping of the reports.

However, to redistribute the reports for the peer-evaluation, I also try to consider the similarity of the different projects and the student's interests. The resulting redistribution key resembles an exchange of the reports in some circular way (see Figs. 2, 3 and 4).

Table 3 Marking scheme in case a student works alone on his/her end-of-semester project (exceptional case). In this case, the peer-evaluation is done by only one fellow student

Mark from	Counting	Mark
Teacher	1/3	A
Self-evaluation	1/3	B
Peer-evaluation	1/3	C
Average mark		$(A + B + C)/3$
Final mark		Average mark rounded to 0.25

Table 4 Marking scheme in case two students work together on their end-of-semester project (regular case). In this case, the peer-evaluation is done by two fellow students

Mark from	Counting	Mark
Teacher	1/3	*A*
Self-evaluation	1/3	*B*
Peer-evaluation Student 1 Student 2	 1/6 1/6	 C_1 C_2
Average mark		$\frac{A+B+(C_1+C_2)/2}{3}$
Final mark		Average mark rounded to 0.25

The students cannot just set a mark, but they have to justify their decision and argue about the positive and negative aspects of the report. To do that, they are asked to write an evaluation report for both their own work and the work they peer-evaluate. These evaluation reports should be about one-page long and conclude with the mark they set. To help the students during the self- and peer-evaluation, I provide them with some evaluation guidelines, which can be found in Appendix F. However, I do not provide a standardized evaluation form, because I really want that the students find their own way of doing these evaluations. For the peer-evaluation, they can also choose to be anonymous or reveal their identity, very much the same as in a scientific review process.

During my evaluation of the end-of-semester projects, I also check the numerical FE-codes in detail. Hence, my evaluation reports always contain three sections: (1) general impression/general comments concerning the entire work, (2) specific comments/questions to the report, and (3) specific comments/questions to the FE-code. For the self-and peer-evaluations, I ask the students to primarily focus on the report, and not so much on the FE-code. However, it is each student's choice to which degree he or she wants/has to read the FE-code to be able to do the evaluation.

After all the evaluation is done, each student receives his/her personal evaluation bundle containing the following documents/information:

- Final mark with an explanation how it is calculated based on the marks by me, the self-, and peer-evaluation (see Tables 3 and 4)
- Evaluation report from me with my mark
- Evaluation report from themselves (self-evaluation) with their mark
- One or two peer-evaluation reports with the corresponding marks, depending on whether they worked alone or in a group of two
- Annotated project report from me with mostly editorial comments and suggestions.

4 Assessment Outcome and Interpretation

Figures 2, 3 and 4 show the marks of the end-of-semester projects of the years 2014 (15 data points), 2015 (13 data points), and 2016 (16 data points), respectively, yielding a total of 44 data points for the three years. Shown are the individual marks from me (the teacher), the self-, and peer-evaluation, the average mark calculated using the formulas in Tables 3 and 4, and the final mark, which is the average mark rounded to 0.25. Red and green colors highlight marks that are below and above my given mark, respectively. Also, the redistribution key for the peer-evaluation is shown in the third column of each figure. Figures 5 and 6 show the distribution of these marks for the three years individually and combined, respectively.

Generally, it is striking that all marks lie above 4.00, which is the critical mark to pass the course (see Swiss marking scale in Appendix F). Hence, nobody in three years failed the NMRD-course. In addition, the average mark lies between 5.36 and 5.45, which seems relatively high on a scale from 1.00 to 6.00. This mark distribution is relatively common for Geology courses on the Master's level at the ETH Zurich (Master in Earth Sciences → Major in Geology), which may have several reasons:

Course grades 2014

Exercise	Student	Peer-review from Student	Mark					
			Teacher	Self-evaluation	Peer-evaluation	Average	**Final**
1	1a	5a	5.50	5.50	5.75	5.54	**5.50**
	1b	6a		5.50	5.50	5.54	**5.50**
2	2a	4a	5.25	5.50	5.50	5.42	**5.50**
	2b	9b		5.50	5.50	5.42	**5.50**
3	3a	7a	5.50	5.50	6.00	5.67	**5.75**
4	4a	5b	5.75	5.25	5.25	5.42	**5.50**
5	5a	3a	5.50	5.25	5.75	5.50	**5.50**
	5b	2a		5.50	5.75	5.58	**5.50**
6	6a	8a	5.25	5.50	5.25	5.33	**5.25**
	6b	9a		5.50	5.25	5.33	**5.25**
7	7a	6b	4.75	5.50	5.25	5.17	**5.25**
	7b	2b		5.00	5.25	5.00	**5.00**
8	8a	1b	6.00	5.75	5.50	5.75	**5.75**
9	9a	7b	5.50	5.00	5.75	5.33	**5.25**
	9b	1a		4.50	5.25	5.17	**5.25**

1.50	above teacher
1.25	
1.00	
0.75	
0.50	
0.25	
0.00	
0.25	below teacher
0.50	
0.75	
1.00	
1.25	
1.50	

Fig. 2 Individual marks for the students of the NMRD-course of fall 2014. Listed are the marks from me (teacher; fourth column), from the students themselves (self-evaluation; fifth column), from the fellow students (peer-evaluation; sixth column), the average mark calculated using the formulas in Tables 3 and 4 (seventh column), and the final mark (average mark rounded to 0.25; eighth column). The third column shows who did the peer-evaluation for whom. The red and green color scales indicate how much the self- and peer-evaluation is below or above the teacher's evaluation, respectively

Course grades 2015

Exercise	Student	Peer-review from Student	Mark: Teacher	Mark: Self-evaluation	Mark: Peer-evaluation	Average	**Final**
1	1a	9a	5.25	5.75	5.75	5.58	**5.50**
2	2a	9b	5.25	5.50	5.75	5.38	**5.50**
	2b	4a		5.50	5.00	5.38	**5.50**
3	3a	8b	5.25	4.75	5.00	5.00	**5.00**
4	4a	6a	5.50	4.00	5.25	4.92	**5.00**
5	5a	1a	5.25	5.25	5.75	5.29	**5.25**
	5b	2b		4.50	5.00	5.04	**5.00**
6	6a	2a	5.75	6.00	5.50	5.75	**5.75**
7	7a	5b	5.50	5.75	5.00	5.42	**5.50**
8	8a	7a	5.75	5.25	5.00	5.29	**5.25**
	8b	5a		5.25	4.75	5.29	**5.25**
9	9a	3a	5.75	5.75	5.25	5.58	**5.50**
	9b	8a		5.25	5.25	5.42	**5.50**

1.50	above teacher
1.25	above teacher
1.00	above teacher
0.75	above teacher
0.50	above teacher
0.25	above teacher
0.00	
0.25	below teacher
0.50	below teacher
0.75	below teacher
1.00	below teacher
1.25	below teacher
1.50	below teacher

Fig. 3 Same as Fig. 2, but for fall 2015

Course grades 2016

Exercise	Student	Peer-review from Student	Mark: Teacher	Mark: Self-evaluation	Mark: Peer-evaluation	Average	**Final**
1	1a	3b	5.25	5.50	5.25	5.42	**5.50**
	1b	5a		5.50	5.75	5.42	**5.50**
2	2a	6b	6.00	5.50	6.00	5.83	**5.75**
	2b	7b		5.00	6.00	5.67	**5.75**
3	3a	8a	4.25	6.00	5.75	5.29	**5.25**
	3b	4b		5.00	5.50	4.96	**5.00**
4	4a	6a	6.00	5.50	5.00	5.58	**5.50**
	4b	2a		5.50	5.50	5.58	**5.50**
5	5a	4a	5.75	5.25	5.25	5.54	**5.50**
	5b	7a		5.25	6.00	5.54	**5.50**
6	6a	3a	5.25	5.00	5.75	5.29	**5.25**
	6b	1b		5.50	5.50	5.46	**5.50**
7	7a	1a	5.50	5.75	5.75	5.63	**5.75**
	7b	8b		5.75	5.50	5.63	**5.75**
8	8a	2b	5.00	5.50	5.00	5.25	**5.25**
	8b	5b		5.50	5.50	5.25	**5.25**

1.75	above teacher
1.50	above teacher
1.25	above teacher
1.00	above teacher
0.75	above teacher
0.50	above teacher
0.25	above teacher
0.00	
0.25	below teacher
0.50	below teacher
0.75	below teacher
1.00	below teacher
1.25	below teacher
1.50	below teacher
1.75	below teacher

Fig. 4 Same as Fig. 2, but for fall 2016

- The Bachelor's program at the ETH Zurich is very competitive. Only the top students actually make it through to the Master's level.
- For incoming international Master students, who did their Bachelor's degree abroad, the ETH Zurich is very restrictive. In many cases, the prior knowledge is assessed before acceptance of a student to the ETH Master's program.

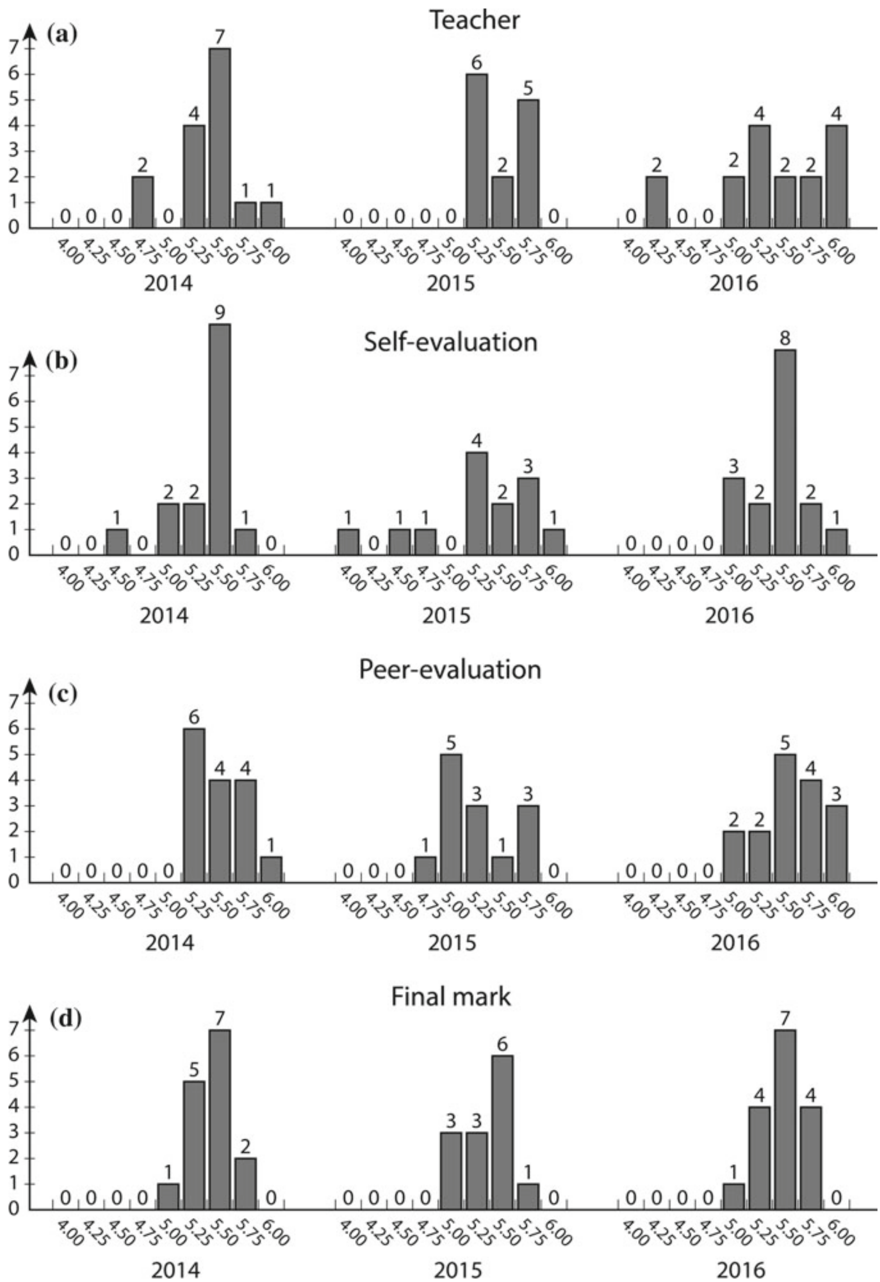

Fig. 5 Distribution histograms for all three years (2014–2016; left to right) of the marks given by me (teacher; **a**), from the students themselves (self-evaluation; **b**), from the fellow students (peer-evaluation; **c**), and the final mark (**d**)

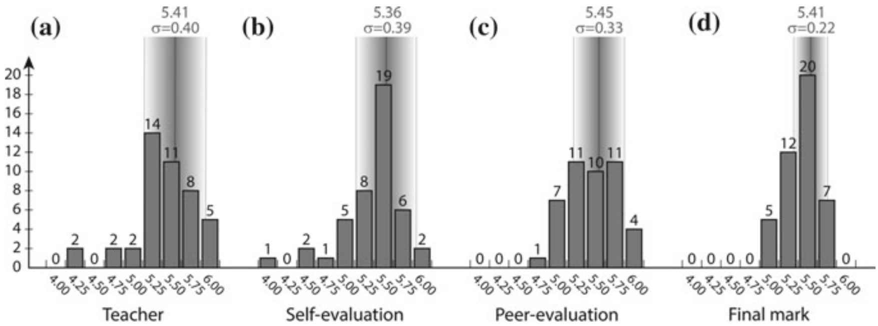

Fig. 6 Distribution histograms for all three years combined of the marks given by me (teacher; **a**), from the students themselves (self-evaluation; **b**), from the fellow students (peer-evaluation; **c**), and the final mark (**d**). Vertical gray areas indicate the average mark (bold gray line) and the standard deviation (shaded area, σ), both of which are also given as gray numbers

- Compared to other courses, the NMRD-course is a relatively technical/mathematical course for Geology Master students. Hence, the NMRD-course is usually only taken by students that enjoy being exposed to this branch of geological work.

For these reasons, students in the NMRD-course are typically very motivated to learn the FE-method and enjoy working with MATLAB and solve mathematical problems. From the feedback that I get from the students, I know that they are very motivated to work on their end-of-semester projects and apply their gained knowledge to a specific problem. As a result, the individual marks are all very good in this course.

The distribution histograms (Fig. 6) show that the average marks are extremely similar independent of who is marking (teacher, self-, or peer-evaluation, final mark). In particular, the average final mark is exactly the same as the average teacher's mark. However, the distribution of the final marks is narrower (standard deviation $\sigma = 0.22$) than each individual contribution (teacher $\sigma = 0.40$, self-evaluation $\sigma = 0.39$, and peer-evaluation $\sigma = 0.33$). With this narrower distribution of the final marks, nobody neither reaches the highest possible mark (6.00) nor a mark smaller than 5.00, even though these marks have individually been given during the teacher's evaluation or during self- and peer-evaluation. This is clearly an effect of averaging the different contributions when calculating the final mark (Tables 3 and 4). During this averaging, the more extreme marks (6.00 or <5.00) are averaged out.

It could have been expected that the self-evaluation is overly positive because students may hope to get a better final mark by systematically overestimating themselves. However, the distribution histograms (Fig. 6) clearly show that this is not the case, at least not on average. In fact, the contrary is the case. The average mark from the self-evaluations is the lowest average mark. This demonstrates that the students are relatively critical with themselves and evaluate their own work objectively.

Concerning the peer-evaluation, I had no clear expectation. It could have been both overly positive (students support their colleagues) or negative (students compete with each other). However, the distribution histograms (Fig. 6) again show that neither is the case, at least on average. If anything, the peer-evaluation is slightly more positive than the teacher's evaluation and the self-evaluation. However, the distribution of marks from the peer-evaluation is somewhat narrower ($\sigma = 0.33$) than the other two contributions (teacher $\sigma = 0.40$, self-evaluation $\sigma = 0.39$), which is due to the lack of relatively lower marks (<4.75). This shows that during peer-evaluation, students do not like or do not dare to give lower marks.

4.1 Marks Compared to Teacher's Marks

Figure 7 shows cross-plots comparing the marks from the self-evaluation, peer-evaluation, and the final mark with my (the teacher's) mark (a–c, respectively), as well as the comparison between peer- and self-evaluation (d). The last is provided as a comparison, but is not discussed here in detail. All cross-plots also show the linear regression line with its slope (s). It is remarkable that the slope of the linear regression is very small for both self- ($s = 0.008$) and peer-evaluation ($s = 0.024$). This means that on average, the marks from the self- and peer-evaluation are almost uncorrelated with the teacher's marks. In other words, when I give a relatively lower mark (lower than about 5.00), the students (both self- and peer-evaluation) tend to give a higher mark than me; when I give a relatively higher mark (higher than about 5.50), the students tend to give a lower mark than me.

However, the linear regression may be biased by the low mark of 4.25 that I gave to one group in 2016 (Figs. 4 and 5a). I had to reduce the mark for this particular group primarily because of some confusions and mistakes in the FE-code, while the report was reasonably good. Because during the peer-evaluation students focus on the report, they have not realized the problems in the FE-code. Therefore, the two peer-evaluations are much more positive than my evaluation (Figs. 4 and 7b). It goes without saying that the group of two themselves did not realize the mistakes in the FE-code; otherwise they would not have made them. Therefore, also the two self-evaluations are much more positive than my evaluation (Figs. 4 and 7a). If the relatively low mark of 4.25 is removed from the datasets, the slope of the linear regression in Fig. 7 (s_l) increases significantly for both the self- ($s_l = 0.096$) and peer-evaluation ($s_l = 0.143$).

For the final mark (Fig. 7c), the linear regression is much steeper ($s = 0.317$) than the self- and peer-evaluation and the relatively low mark of 4.25 does hardly affect the linear regression ($s_l = 0.362$). The significantly larger slope can be explained because the final mark consists to one-third of the teacher's mark. Hence, the linear regression partly reflects an auto-correlation. However, the final mark is still more narrowly distributed than my mark (see also Fig. 6). In other words, when I give a relatively lower mark (lower than about 5.00), the final mark tends to be higher than

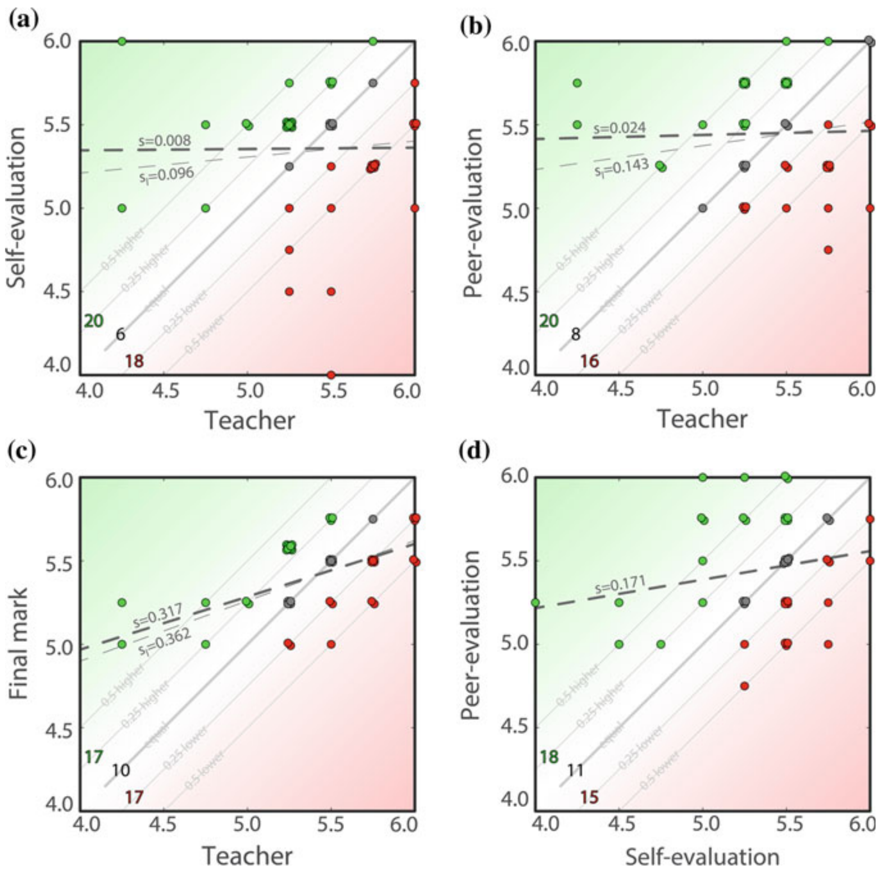

Fig. 7 Cross-plots showing the relationship between **a** the self-evaluation and the teacher's evaluation, **b** the peer-evaluation and the teacher's evaluation, **c** the final mark and the teacher's evaluation, and **d** the peer- and the self-evaluation. Bold light gray lines indicate where the two marks are equal; thin light gray lines indicate where the marks on the vertical axis are 0.25 or 0.5 units above or below the marks on the horizontal axis. Red, green, and black numbers indicate how many marks on the vertical axis are below (plotting within the red field), above (plotting within the green field), or equal to marks on the horizontal axis (plotting on the bold light gray line), respectively. Dashed dark gray lines correspond to the linear regression for each entire dataset (bold line; slope s) and each dataset neglecting the teacher's mark of 4.25 (thin line; slope s_l). Note that some data points are slightly shifted to make them all visible

mine; when I give a relatively higher mark (higher than about 5.50), the final mark tends to be lower than mine.

In principle, in Figs. 7a–c, I assumed my mark as the reference for all other marks (mark from self- and peer-evaluation and final mark). Red and green colors highlight the marks that are below and above this reference, respectively. As an additional information, Figs. 8 and 9 show the distribution of how far the marks are below and above the reference for the three years individually and combined, respectively. Hence, these are similar plots as in Figs. 5 and 6, but plotted relative to my (the teacher's) mark. Therefore, very similar observations can be made. On average, the marks from the self-evaluation are slightly below my marks (average deviation -0.051), demonstrating that the students mark themselves slightly more critical than me. The marks from the peer-evaluation are slightly above my marks (average deviation +0.040), showing that the students are slightly more positive with their colleagues than me. However, these average deviations are almost negligible. This is even truer for the final mark, which on average deviates only +0.006 from my mark. Hence, on average, the students (almost) get the identical final mark I would have given them alone. However, the distribution of the deviation for the final mark is significantly narrower (standard deviation $\sigma = 0.33$) than the self- ($\sigma = 0.56$) and peer-evaluation ($\sigma = 0.51$). Again, this is an effect of the averaging scheme (Tables 3 and 4), which averages out larger deviations from the individual contributions. In the end, 36 out of 44 students (81.8%) have a final mark within ± 0.25, and 42 students (95.5%) have a final mark within ± 0.50 of my (reference) mark.

5 Discussion

Generally, the evaluation procedure works extremely well, which was a positive surprise at the beginning when I introduced it. Both the self- and peer-evaluation is done in a very serious and rigorous way and leads to marks very similar to mine. At the same time, I always believed that students can deal with responsibility and use it in a positive way if they get the opportunity. Hence, I also expected that the students do a great job during the self- and peer-evaluation. Already before I introduced this evaluation procedure, but now even more, I think it is safe to give responsibility into the students' hands and it even has very positive effects on their performance and motivation.

5.1 Training of Soft-Skills

The results of the evaluation procedure over the course of three years (Figs. 6 and 9) show that on average the students (almost) get the identical final mark I would have given them alone. Therefore, the question arises if this entire evaluation procedure is obsolete. In my opinion, this is clearly not the case because during the evaluation procedure, the students learn and train some crucial soft-skills, such as

- Having an opinion! During the evaluation, the students have to build an opinion about the work they evaluate. They cannot be indifferent toward the evaluation, but they have to seriously deal with the work because of the next point.

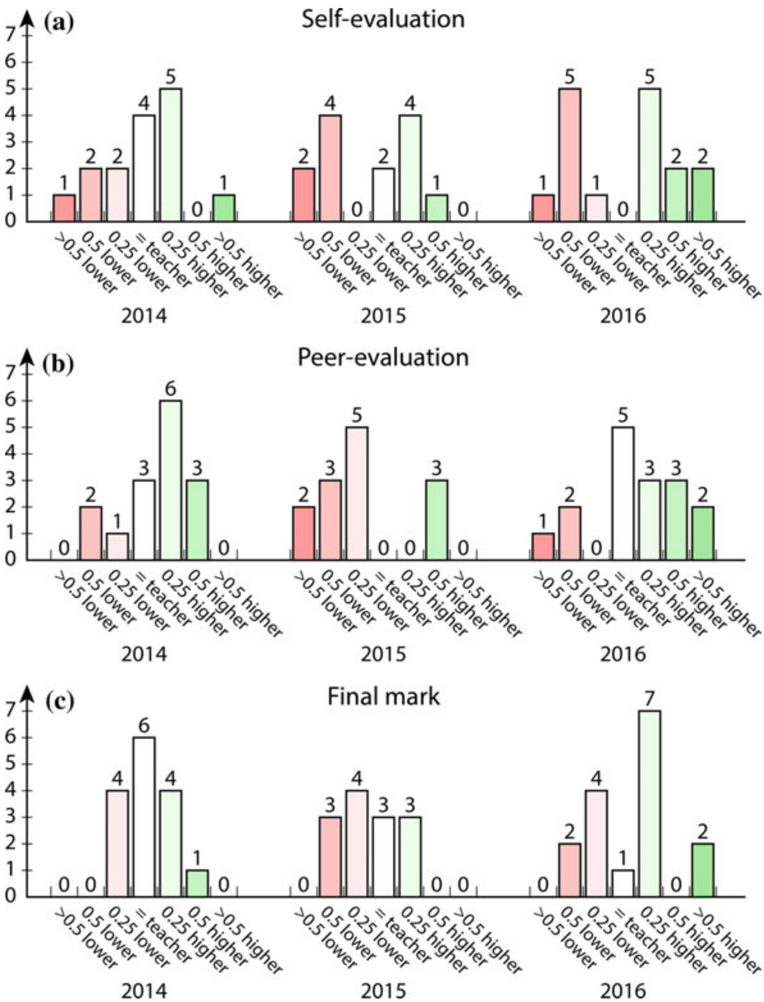

Fig. 8 Distribution histograms of the number of marks below (red) and above (green) the teacher's mark for all three years (2014–2016; left to right) for the self- (**a**) and peer-evaluation (**b**), as well as for the final mark (**c**)

- Taking responsibility! The students have to conclude their self- and peer-evaluation with a mark, which is not altered in any way by me and hence counts one-to-one for the final mark. This is an important point for me, because I strongly believe that students can deal with responsibility and use it in a positive way if they get the opportunity.
- Being objective and honest with yourself and your own work! This is a very crucial soft-skill both as a scientist and in industry jobs. Very often in a job, one has to decide if a piece of work is good enough (i.e., finished) to be presented to the

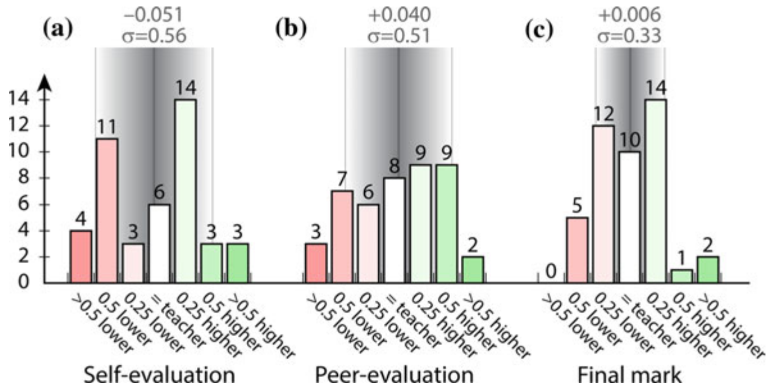

Fig. 9 Distribution histograms of the number of marks below (red) and above (green) the teacher's mark for all three years combined for the self- (**a**) and peer-evaluation (**b**), as well as for the final mark (**c**). Vertical gray areas indicate the average deviation from the teacher's mark (bold gray line) and the standard deviation (shaded area, σ), both of which are also given as gray numbers

supervisor or a customer or to be submitted to a journal. This requires an objective and honest self-evaluation.

- Being objective and honest with a colleague! This is a very crucial soft-skill when working in teams. It is important to be able to provide critical feedback to somebody, who is a close collaborator or friend. In particular, it is very important to be able to provide criticism in a way that the colleague can accept it, which is only possible when it is objective and fair.
- Evaluating work, for which you are not the exact expert! During peer-evaluation, the students have to judge if the given results and interpretations are feasible without having worked on the specific problem themselves. As a scientist or in industry jobs, this situation occurs all the time when doing a peer-review for a journal article or when evaluating the work of a contractor.

All of these soft-skills are very important for the student's future career. The self- and peer-evaluation mimics a scientific review process or an industry workflow of reporting. Unfortunately, such soft-skills are not typically trained at universities, where the primary focus is on knowledge and its applications. However, modern education should be more skill-oriented than knowledge-oriented, in particular at the university level. Using the presented evaluation procedure, I try to cover both aspects, technical knowledge and its application, but also the required soft-skills to perform the necessary evaluation of the work.

Going back to Bloom's taxonomy of educational objectives (Fig. 1), it is satisfying to see that the NMRD-course reaches the highest levels of learning and thinking. Just with the end-of-semester project alone (which would only be evaluated by me), students would only reach the level of "Application/Apply" (i.e., apply the FE-code to a specific problem) and "Analysis/Analyze" (i.e., analyze and interpret the results), which are in the middle of the taxonomy. However, with the presented self- and peer-evaluation procedure, the students reach the level of "Evaluation/Evaluate", which is the highest and second highest level in the original and revised taxonomy, respectively. Those students who also chose to define their own end-of-semester project even reach the level of "Create", which is the highest level in the revised taxonomy. However, all students generate their own new knowledge by working on and solving their individual end-of-semester project. In that sense, all students reach the level of "Synthesis/Create".

5.2 Advantages of the Self- and Peer-evaluation Procedure

Besides the training of soft-skills described above, the presented evaluation procedure of the NMRD-course has several additional advantages. In particular, I feel that the students are extremely motivated to work on their end-of-semester project because they can choose or even define their own project. This allows them to work on something they are really interested in, which can be different if such projects were distributed randomly by the teacher. This very high motivation expresses itself by how much time the students spend on their project and in particular on some small details that they would like to understand. It happened quite regularly that some students even come up with their own questions related to the project, which go beyond the original set of questions in the project posed by me.

The final mark of any course at the ETH Zurich (and probably at most universities) has to be individual. In other words, marking group work is officially not allowed. Nevertheless, many courses are marked based on group work, in particular on the Master's level. The reason is that many courses become more and more skills-oriented rather than knowledge-oriented, which is relatively difficult to assess in individual written or oral exams. Group work solving a particular problem is better suited to assess the desired problem solving-skills. The presented evaluation procedure offers a solution to this problem, because thanks to the self-evaluation, it assesses group work with individual marks. In this procedure, only two-thirds of the final mark is the same for the group (peer-evaluation and mark from the teacher; Tables 3 and 4), while one-third is individual (self-evaluation). It does not happen too often that the two students of a group have different final marks; however, it happened to 4 out of 18 groups (22.2%) in the course of three years.

5.3 Disadvantages and Criticism of the Self- and Peer-evaluation Procedure

Of course, the presented evaluation procedure also has a few disadvantages. In particular, it is a relatively large effort in terms of time and work for everybody involved. On one hand, I (the teacher) have to define and describe all the individual end-of-semester projects and in fact solve all of them myself to make sure they are suitable for the students in terms of difficulty and time requirements. If the students come up with their own ideas for these projects (which I encourage a lot), I even cannot copy and modify the projects from the previous years but have to design them every year from scratch, which is very time-consuming. Also, I have to define the redistribution key for the peer-evaluation and do an evaluation myself for every report and FE-code. On the other hand, the students not only have to solve their individual project, but also read one report from another group in detail and deliver two evaluation reports (self- and peer-evaluation). Despite all of this additional work compared to a course using a more traditional evaluation procedure, the average final mark for the students is (almost) identical to the teacher's mark. However, the extra work (at least of the students) is rewarded with one extra ECTS point compared to other 2-hour/week lectures.

One criticism of the presented evaluation procedure might be that the final mark is not objective, because every student receives the evaluation from different people. Therefore, the marking standard might be different for different students. This is a justified criticism. However, one could argue that this is just the way it is in real life. The acceptance of a scientific paper in a peer-review process also depends on the choice of reviewers (and maybe even their daily mood); or whether or not a service can be sold to a customer also depends on the person representing that customer. Because such differences in the evaluation of the same piece of work exist in real life, teachers should expose our students to this challenge and train them accordingly. Traditional individual exams, which are all evaluated and marked by the same teacher, pretend that there is one unique evaluation and marking standard. On the Master's level, I believe that this is not what an teacher should teach the students anymore. Also, there is no reason to believe that my (the teacher's) evaluation is more objective than the student's own evaluation.

5.4 Communication: One Key Point for the Self- and Peer-evaluation Procedure

One very important point for the presented evaluation procedure is a good communication. For almost all students, it is the very first time they do such a self- and peer-evaluation, and in particular to give marks that count. Therefore, it is crucial to provide sufficient explanation, documentation, and guidance prior and during the entire procedure. In the case of the NMRD-course, I explain the evaluation procedure at the very beginning of the course, so the students know very early what they are supposed to do at the end of the semester. Also, I provide a short information in the middle of the semester (Appendix A) and some general remarks (Appendix E) and the evaluation guidelines (Appendix F) when the students actually start with the end-of-semester projects. In addition, the students have access to some four reports from previous years to get an idea how the report may look like. All of this is crucial to make the self- and peer-evaluation an enjoyable and successful experience for the students.

5.5 *Further Ideas/Outlook*

During the three years I implemented the presented evaluation procedure, a number of extension and modification possibilities came to my mind. However, I never implemented any of them, and therefore, I cannot report on any experience. These are just ideas, including

- The evaluation of the report and the FE-code may be separated more strictly in the sense that I only evaluate the FE-code (not even looking at the report) and the peer-evaluation only considers the report (the students would not even get the FE-code). This would put even more responsibility into the students' hands because they had the sole responsibility to evaluate the report. At the same time, it would reduce my workload significantly. In this scenario, the self-evaluation would still consider both the report and the FE-code.
- Instead of written evaluation reports, it would be very interesting to do oral examinations consisting of two parts:
  - First, a true oral examination of/discussion with the student receiving the mark (about 20 min). The examination panel would consist of myself and the student(s), who do the peer-evaluation in the current evaluation procedure (i.e., they have read the report in detail). Hence, one or two fellow students would take an active part in examining their colleague together with me. This situation would be extremely interesting to see how the students interact with each other. The students, who were exposed to this situation could learn and train soft-skills related to oral argumentation that go beyond the written evaluation.
  - Second, finding the final mark in a discussion (about 10 min). Everybody together (I, the fellow student(s) doing the peer-evaluation and the student him/herself) should agree on the final mark in a discussion. So, the idea is not to average the different contributions, but to have a discussion and argue about the final mark. If there is no consensus, the established averaging scheme (Tables 3 and 4) may still be adopted.

    Such an oral examination incorporating fellow students into the examination panel and finding an agreement on the final mark would be extremely interesting for everybody involved. It involves a lot more interpersonal exchange than the current evaluation procedure, which is just based on written reports. In terms of workload, the students do not have to write two evaluation reports anymore (self- and peer-evaluation); instead they have to take part in 2–3 oral examinations (their own and 1–2 in the examination panel). So, the workload may be about the same as in the current evaluation procedure. However, organizing all the oral examinations with all the students would be very tedious and time-consuming.

- The oral evaluation procedure just described above may also be combined with written evaluation reports. Of course, the workload would increase significantly. At the same time, the amount and quality of feedback each student would receive also increases significantly because the advantages of both written and oral evaluations were combined.

6 Conclusions

The students really enjoy working on their end-of-semester project, because they can apply their gained knowledge to a specific problem. The fact that they can choose, or even design themselves their own project contributes significantly to their motivation. Concerning the self- and peer-evaluation, my main conclusion is that they both work very well. After three years of experience, it is clear that the students can handle the given responsibility and take the self- and peer-evaluation very seriously. They are objective during the evaluation and do not evaluate themselves or their colleagues overly positive or overly negative. If anything, the self-evaluation tends to be slightly more negative and the peer-evaluation slightly more positive than my (the teacher's) evaluation. On average, the final mark is exactly the same as my mark.

In addition, the evaluation procedure allows the students to learn and train several crucial soft-skills allowing them to conduct an objective and fair evaluation of their own and their colleague's work. The presented evaluation procedure allows teaching and training such soft-skills in addition to the technical/scientific knowledge, which is otherwise not trained enough at universities at the Master's level. Based on the student feedback I received over a period of three years, I can conclude that the evaluation procedure itself significantly enhances student motivation. The main reason for this is an opportunity they do not usually get as students, that is taking responsibility for their own evaluation and marks. However, giving responsibility into the students' hands as a teacher acknowledges their competence and capabilities, which is something very motivating.

Acknowledgements I thank all the past students that took the NMRD-lecture at the ETH Zurich. Originally, the presented grading procedure was indeed an experiment, which the students pursued on their free will. Only thanks to their enthusiasm, I initiated this experiment and through very constructive and helpful feedback from my students, I could refine and improve the presented grading system over the years. I also thank Adrian Gilli, our Department's educational support and teaching expert, for encouraging me to initiate this grading procedure at the very beginning. Reviewed by Soumyajit Mukherjee (IIT Bombay).

Appendix A

In the following, a short document is provided that helps the students decide what kind of end-of-semester project they want to work on.

Table 5 List of possible end-of-semester projects to give the student ideas what they could/want to work on

Project	Elastic problem	Viscous problem	Mathematical/theoretical problem
Examples	• Stress distribution inside a bridge • Pressure distribution around a borehole • Pressure distribution at tunnel level below a topography • Stress distribution in a landslide situation • Any other engineering-type application (i.e., stresses in a piece of something) • Lithospheric bending	• Geological buckle folding, e.g., neutral line in folds • Growth of diapirs and salt domes • Sigma-/delta-clast formation • Deformation in shear zones, e.g., drag fold formation • Glacier rock glacier flow	• Visco-elastic rheology, or non-linear rheology • Seismic wave propagation • Surface processes/erosion • 3D problem (viscous or elastic) • Problems in cylindrical or spherical coordinates
Suited for	• Engineering geologists • Geologists interested in elastic and brittle deformation	• Geologists interested in viscous/ductile deformation	Everyone who wants to • learn more technical details about the FEM • use the FEM in the MSc thesis

A1. *Some notes on the examination*

For getting a mark for this course, you are expected to work on an individual project in groups of two (if possible) and hand in a short report about this project. Further details about this report, I will tell you later in the semester. But it is time to speak about the type of project you should work on. Because I want that you can profit from this project as much as possible, I let you decide what the broad application of your project should be. Possible choices are provided in Table 5.

If you have your own ideas, just talk to me. I am sure we find a good project for you.

Important (until end of next week): Please form groups of two (if possible) and tell me or write an e-mail about your choice of project.

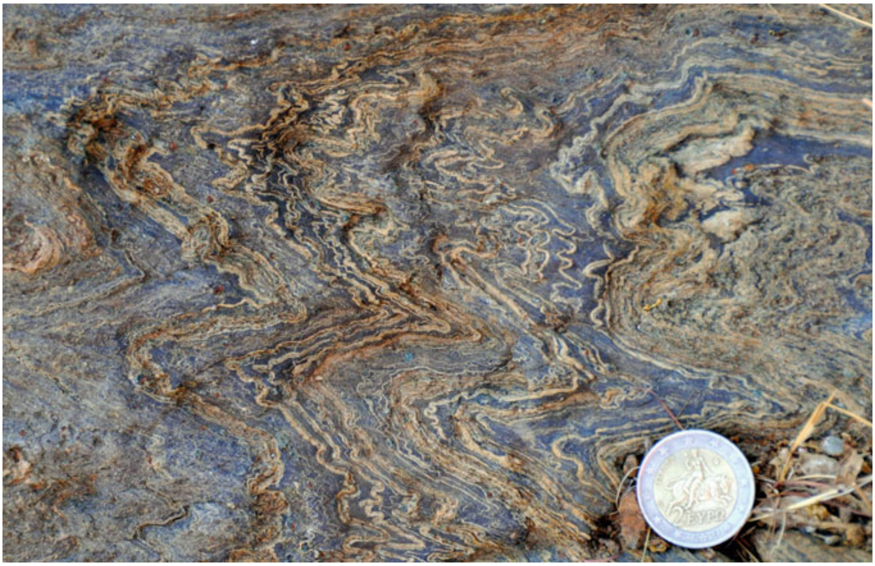

Fig. 10 Parasitic folds in the Cycladic Blueschist Unit. The individual layers develop an order of magnitude smaller wavelength compared to the entire multi-layer stack, which is referred to as parasitic folds or second-order folds. Location: Northern Syros Island, Greece (37° 29′ 28.86″N/24° 54′ 26.14″E)

Appendix B

In the following, an example of an individual semester-end project work is provided, which would be solved by a group of two students. The following description corresponds exactly to the problem set the students receive.

B1. *Exercise: Parasitic folds*

Folds are one of the most spectacular geological structures that form in compressional tectonic regimes (Fig. 10). In many cases, folds can be modeled as a linear viscous (Newtonian) fluid in two dimensions. Very often, folds occur on multiple scales, whereas the smaller-scale folds develop so-called parasitic folds (S-, Z-, and M-folds; Fig. 11) in relation to the larger-scale fold.

In this exercise, you should model multilayer folding and the development of parasitic folds, very similar to my Master's Thesis (Frehner and Schmalholz 2006). Consider the multilayer geometry in Fig. 12, which consists of three stronger ($\eta_{\mathrm{Layer}} = 100$) and five weaker layers ($\eta_{\mathrm{Matrix}} = 1$). All lengths are normalized; hence, in the following, all lengths are given as units, and not meters. The layers are 20.2 units long. This corresponds to the dominant wavelength, λ_d, of the two-layer system; i.e., neglecting the thin central layer:

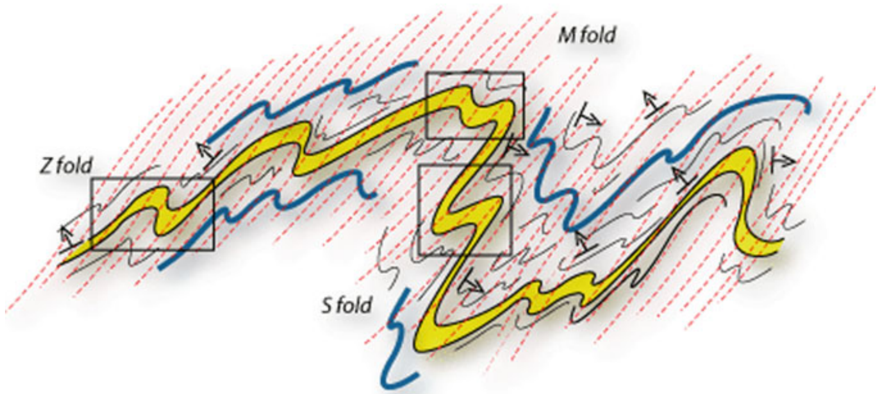

Fig. 11 Sketch of asymmetric parasitic folds (second-order folds) in association with a larger-scale fold (first-order fold). Also shown are the third-order folds, which are parasitic folds to the second-order folds. Picture taken from the lecture notes of Jean-Pierre Burg, ETH Zurich (http://www.files.ethz.ch/structuralgeology/JPB/files/English/8folds.pdf)

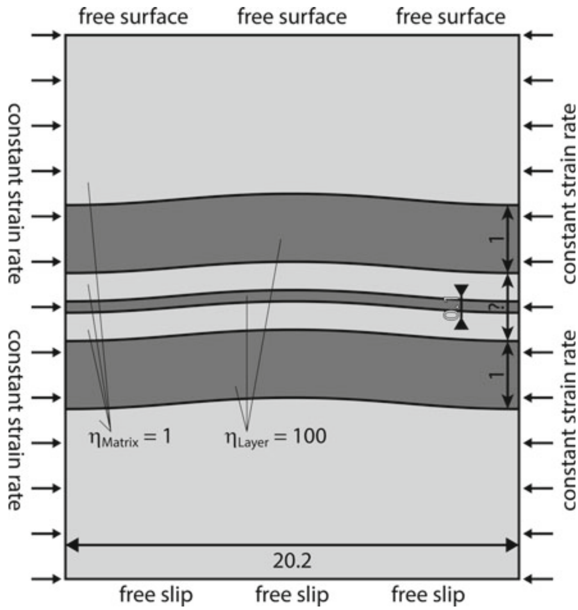

Fig. 12 Sketch of the model setup to model the development of asymmetric parasitic folds in a multilayer stack. All material parameters, model dimensions, and boundary conditions are provided here

$$\lambda_d = 2\pi H \sqrt[3]{n_{Layer} \frac{\eta_{Layer}}{6\eta_{Matrix}}}, \tag{B.1}$$

where H is the thickness of the individual strong layers and η_{Layer} is the number of strong layers (i.e., 2). The upper and lower strong layers have a thickness of 1 unit, and the central strong layer has a thickness of 0.1 units. The central strong layer should be placed symmetrically in the center between the two outer strong layers. However, the distance to them is a free parameter.

All layer interfaces need to have a sinusoidal initial geometry to initialize the folding process (indicated in Fig. 12). The amplitude of this sinusoidal initial geometry should be small (use 1×10^{-4}). In addition to the sinusoidal initial geometry, the central strong layer should have a random initial geometry to allow it developing its own dominant wavelength. This random initial geometry, you may program as

```
rand_array = Arand*h0*detrend(cumsum(rand(1,nx)));
```

where `Arand` is the amplitude of the random geometry (use `Arand` $= 2 \times 10^{-1}$), `h0` is the thickness of the thin layer (i.e., `h0` $= 0.1$), and `nx` is the number of nodes of the numerical grid in x-direction.

The ***boundary conditions*** for this multilayer folding model are:
Left and right boundary: Constant strain rate
Bottom boundary: Free slip
Top boundary: Free surface.

The free surface boundary condition corresponds to a zero-stress Neumann boundary condition and can be applied in the FE-code by not defining any Dirichlet boundary conditions for the corresponding degrees of freedom. Free slip corresponds to the following two conditions:

- Zero velocity perpendicular to the boundary
- Traction-free movement parallel to the boundary. This can be applied in the FE-code by not defining any Dirichlet boundary conditions for the boundary-parallel velocity.

The constant strain rate boundary condition corresponds to the following two conditions:

- Traction-free movement parallel to the boundary.
- Velocity perpendicular to the boundary is a function of the coordinate value (the x-coordinate in this case), i.e., $v_x = \dot{\varepsilon}_{xx} x$. Use $\dot{\varepsilon}_{xx} = -1$ in your model.

B2. *Questions*

- Plot the distribution of the pressure inside and around the multilayer folds as they progressively deform.
- Find a good value for the distance between the two outer strong layers to develop nice parasitic folds.
- Describe (and possibly quantify) the development of the small- and large-scale folds and how they are related to each other. How is the temporal relationship between the small- and large-scale folds? When do the small-scale folds become asymmetric?
- Where do S-, Z-, and M-folds develop in relation to their position on the larger-scale fold?

B3. *Notes/Tips*

- Use a time increment of $\mathrm{d}t = 5 \times 10^{-3}$.
- Simulate approximately 120 time steps.
- Put your lower and upper boundaries "far away" from the layer, i.e., approximately 15x the layer thickness.
- You do not need to include gravity in this exercise, because it is irrelevant on this scale.
- Make some color snapshots during the simulation to visualize the model evolution.
- Maybe, you have to run the exact same model setup several times. These alleged equal simulations are still different because you use a random initial geometry for the central strong layer. Use a simulation that gives "nice" results.
- You realize that you model is mirror symmetric with a vertical symmetry axis. Maybe you can think of only modeling half the model to save some numerical costs. When plotting the results, you can "double" the model again.

What kind of boundary conditions do you have to use on the symmetry axis?

Appendix C

In the following, a second example of an individual semester-end project work is provided, which would be solved by a group of two students. The following description corresponds exactly to the problem set the students receive.

C1. *Exercise: Sill emplacement*

Diapirs are important geological structures that occur when relatively lighter material rises through relatively denser material. This occurs, for example, at the slab interface of a subducting plate, where rock partially melts and rises through the upper mantle toward the crust (Fig. 13). If a diapir encounters a sharp lithological contrast, it may not be able to ascend further and starts to move sideways along the lithological boundary forming a horizontal sill. Hence, diapirism is not only an important process for vertical magma emplacement, but also for underplating, and the formation of large intrusions.

In many cases, diapirs can be modeled as a linear viscous (Newtonian) fluid in two dimensions. In this exercise, you should investigate the interplay between the vertical and horizontal movement of a diapir/sill encountering a sharp lithological boundary. To do that, an initial model geometry is assumed (Fig. 14) that resembles a diapir in its relatively late stage just after encountering the lithological interface. Consider the initial model geometry in Fig. 14. It consists of three materials: the diapir, the surrounding mantle, and a layer above the diapir. All values for the dimension of the model, density, and viscosity are given in the figure.

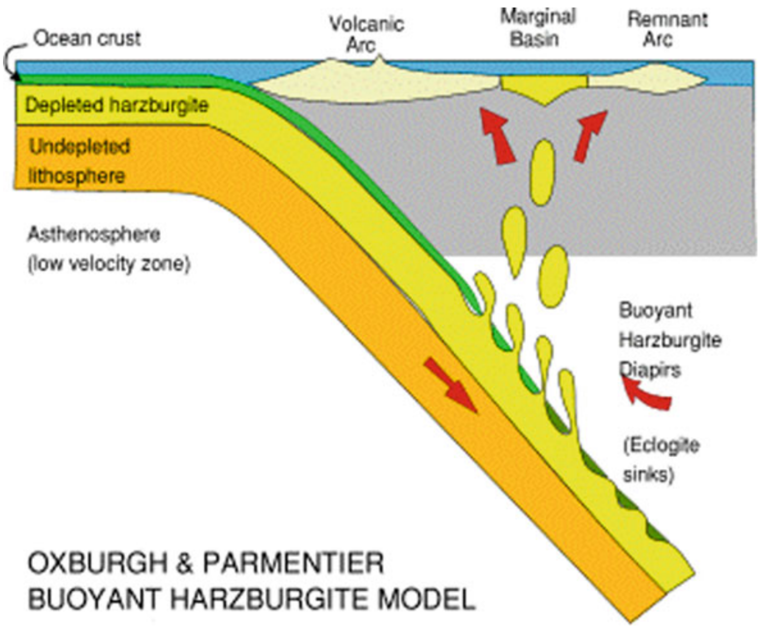

Fig. 13 Sketch of diapirs rising from a subducting slab below an intraoceanic island arc. This buoyant harzburgite model originates from (Oxburgh and Parmentier 1978)

The ***boundary conditions*** for this diapir model are:

Left, right, and bottom boundary: Free slip boundary conditions

Top boundary: Free surface boundary conditions

The free surface boundary condition corresponds to a zero-stress Neumann boundary condition and can be applied in the FE-code by not defining any Dirichlet boundary conditions for the corresponding degrees of freedom. Free slip corresponds to the following two conditions:

- Zero velocity perpendicular to the boundary
- Traction-free movement parallel to the boundary. This can be applied in the FE-code by not defining any Dirichlet boundary conditions for the boundary-parallel velocity.

C2. *Questions/Tasks*

- Plot the pressure distribution in and around the diapir as it progressively amplifies. Where is the pressure greatest/smallest? Are there pressure concentrations?
- Track both the amplitude of the topography above the diapir and the horizontal extent of the diapir's head. Compare and quantify the two values (maybe take the ratio or make a cross-plot) for different values of the viscosity of the upper layer. Which viscosity should the upper layer have for an efficient horizontal movement of the diapir's head?

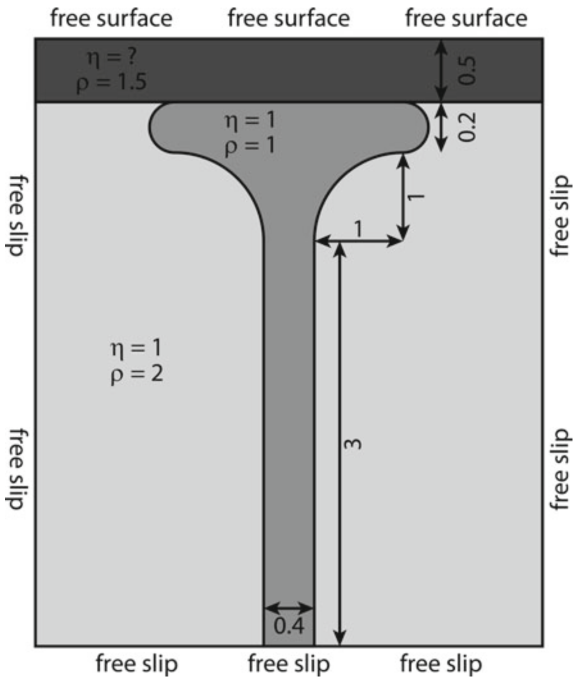

Fig. 14 Sketch of the model setup to model sill emplacement along a sharp lithological boundary when a rising diapir head encounters the latter. All material parameters, model dimensions, and boundary conditions are provided here

C3. *Tips and tricks for building the model grid*

Creating the model is somewhat tricky because the model contains some curved interfaces. However, if you use a constant number of elements in both x- and y-directions, you can keep the structure of the two arrays `x2d` and `y2d`, only that you have to program the actual coordinate values differently (not using `ndgrid`). The final grid looks something like Fig. 15 (this is a low-resolution numerical grid for visualization reasons). You recognize that the y-coordinates are constant in x-direction. So, layer-by-layer, you can create a y-coordinate array.

Once you have this, it is probably best to loop through this y-coordinate array. For every y-coordinate, you can define the x-coordinate of the bold interface in the middle of the model and from this, you can calculate the grid spacing in x-direction (Δx) on the left and right side of this interface.

C4. *Notes/Tips*

- The horizontal extent of the model is not provided here. Choose a "large enough" model to avoid boundary effects.
- Gravity-driven self-consistent models are actually quite tricky because they can numerically become unstable (the solution can explode). Therefore, this model is somewhat scaled and does not really contain physical parameters. In other words, the lengths are not in meters but in arbitrary units; the densities are not in kg/m^3,

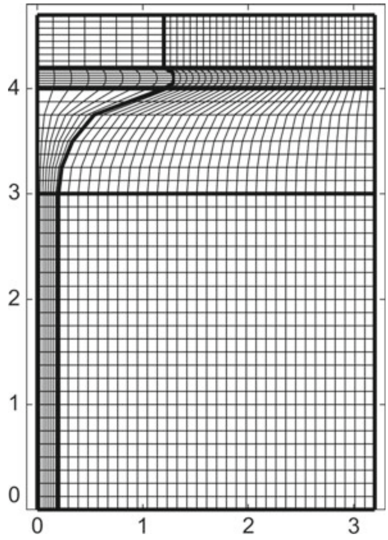

Fig. 15 Low-resolution version of the numerical FE-grid that discretized the model shown in Fig. 14

but also in arbitrary units; and so on. It is important to stick to the given values for this exercise not to have numerical problems.

- Use a time increment of $\Delta t = 5 \times 10^{-2}$.
- Simulate approximately 120 time steps.
- Make some color snapshots during the simulation to visualize the model evolution.
- You realize that you model is mirror symmetric with a vertical symmetry axis. Maybe you can think of only modeling half the model to save some numerical costs. When plotting the results, you can "double" the model again.

What kind of boundary conditions do you have to use on the symmetry axis?

Appendix D

In the following, a third example of an individual semester-end project work is provided, which would be solved by a group of two students. The following description corresponds exactly to the problem set the students receive.

D1. *Exercise: Second Gotthard tunnel*

In 2016, the Swiss public voted (57% yes against 43% no; https://www.bk.admin.ch/ch/d/pore/va/20160228/det599.html) for the construction of a second Gotthard road tunnel. The main pro-argument was the increased safety, because with two tunnels the traffic in the two directions can be separated from each other. The main contra-argument was the protection of the alpine environment. Now after the vote, the planning for this construction has started.

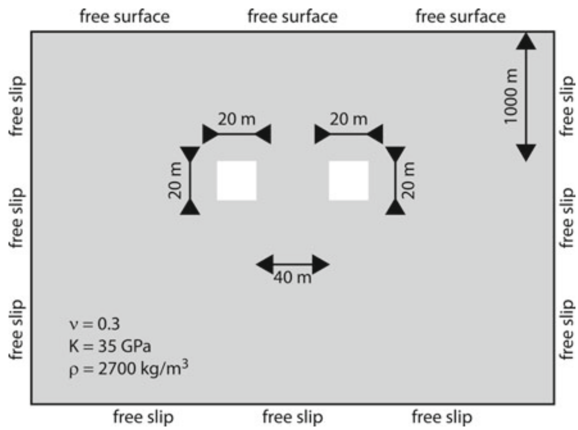

Fig. 16 Sketch of the model setup to model the second Gotthard tunnel. All material parameters, model dimensions, and boundary conditions are provided here

In this exercise, you should use the FE-method to investigate how the second tunnel affects the stress distribution around the first tunnel. Consider the two-dimensional elastic model shown in Fig. 16 with the material parameters given in Table 6.

The ***boundary conditions*** in this model should be the following:

Left, bottom, and right boundary: Free slip

Top boundary: Free surface

The free surface boundary condition corresponds to a zero-stress Neumann boundary condition and can be applied in the FE-code by not defining any Dirichlet boundary conditions for the corresponding degrees of freedom. Free slip corresponds to the following two conditions:

- Zero displacement perpendicular to the boundary
- Traction-free movement parallel to the boundary. This can be applied in the FE-code by not defining any Dirichlet boundary conditions for the boundary-parallel displacement.

To avoid boundary effects, the left, bottom, and right boundaries should be "far away" from the tunnel.

The goal of this exercise is to study the stress distribution in the subsurface in the case of one (initial state) and two tunnels under the gravitational load that acts on the rock mass and compare the two cases.

Table 6 Elastic material properties to model the second Gotthard tunnel

Parameter	Surrounding rock (granite)	Tunnel (air)
Poisson's ratio (ν)	0.3	0.1
Elastic bulk modulus (K)	35 GPa	0.1 MPa
Density (ρ)	2700 kg/m^3	1 kg/m^3

D2. *Questions*

1. Plot the distribution of the second invariant of the stress tensor, which is given as $\sigma_{\mathrm{II}} = \sqrt{\left(\sigma_{xx} - \sigma_{yy}\right)^2 / 4 + \sigma_{xy}^2}$ in both cases (1 and 2 tunnels), and plot the difference between these two cases.
2. Describe the stress distribution around the tunnel in both cases. How does it look like? Is the stress distribution symmetric around the tunnel? Are there areas of stress concentrations?
3. How does the stress distribution around the first tunnel change when the second tunnel is added? Does it increase or decrease, and where does it change?
4. How would you adjust the safety installations in the first tunnel before drilling the second tunnel?

D3. *Tips and Tricks*

- You are supposed to run two different models, one with one tunnel, and one with two tunnels. I suggest that the numerical grid for both simulations is exactly the same. Then you can compare your results one-to-one, because the numerical points of the two simulations are at the same position.
- So, I suggest to use three phases in your model (i.e., the `Phase`-array contains values 1, 2, and 3), 1 for the granitic rock, 2 for the left tunnel, and 3 for the right tunnel. Then you also have to define three sets of material parameters (i.e., 2 elastic parameters and density), one for each phase. When you do that, you can run the first simulation where the material parameters for phase 3 are equal to the rock parameters and the second simulation where the material parameters for phase 3 are equal to the air parameters.
- To compare these two simulations, you can either
  - save the stresses after the first simulation, i.e.,

```
save filename STRESS
```

At the end of the code for your second simulation, before you plot the stresses, you can load the stresses from the first simulation again:

```
T2 = STRESS;
load('filename')
T1 = STRESS;
```

`T2` are the stresses in the case of two tunnels and `T1` are the stresses in the case of one tunnel. And then you can plot the difference between the two.

- or write a loop around your entire code, which is executed twice, e.g.,

```
for iTunnel = 1:2
```

and in each iteration choose the material parameters according to one or two tunnels. After executing this loop, you have both stress fields.

- To avoid an overly dense resolution away from the areas of interest, it would be helpful to use a non-uniform resolution (non-constant Δx and Δy). You may try using a smoothly increasing spatial resolution from the left boundary toward the tunnels, and a smoothly decreasing spatial resolution from the tunnels away toward the right boundary (and similar in the y-direction). This way, you can save some numerical points away from the tunnels, where you do not need a high resolution.

Appendix E

In the following, some general remarks that are handed out to the students are provided. These general remarks specify some details concerning the project work, the report, and the evaluation procedure.

E1. *General remarks*

Your mark will be based on your performance in a small project carried out in groups of two (if possible). I tried to design projects of similar difficulty. The idea of the project is for you to demonstrate that you understand the FE-method sufficiently to solve a practical problem. Of course, the primary goal is to get the correct answer. However, it is not a disaster if your code contains a bug or gives strange results. Therefore, it is important that you provide sufficient documentation, such that even if your program does not work correctly, your solution strategy is clear.

For getting a mark, you have to hand in the following things ***until xxx***:

- ***A report, not longer than 4 pages (!)***, describing how you solved your problem and summarizing your results. See below.
- ***All MATLAB codes*** necessary for solving your problem. The codes should automatically produce the figures in your report. Use comments in your code (!); otherwise, I will not be able understand it (and cannot find the errors ☺).

I leave the organization of your four-page report up to you; the following may act as a guideline:

- Your report should only contain solution strategies and results/interpretations/discussions to your problem. So, do not waste time and space for
  - repeating the questions given in your project. I know them already.
  - introduction and motivation. This report is not for a general audience.
- Shortly describe the governing equations describing your problem and how they are discretized with the FE-method (e.g., $\mathbf{Ku}=\mathbf{F}$; what is $\mathbf{K}$, $\mathbf{u}$, and $\mathbf{F}$?).
- Describe the boundary and initial conditions and how they are implemented.
- Maybe draw a diagram or flowchart of your code, or outline it in keywords. This also helps you writing the actual code. Focus on the things that you implemented, and not on the things **done** together in the course.
- Include figures, flowcharts, sketches, etc. ***(!)***. Do not forget figure labels. Figures are the most important step for explaining results and to draw conclusions. Figures in the report should be reproducible by the code that you hand in.
- Describe the results, interpret and discuss them, and draw some conclusions.

The last two weeks of the semester, you have time to work on your project during the lecture hours. You are free to discuss with me and the course assistant. During the week and semester break, I am still available for discussions and questions (also per e-mail). ***Your questioning will not have any influence on your mark.*** Consider this exercise as an integral part of the lecture. I want you to profit from this exercise as much as possible…!

You should email your report and all MATLAB files to me or bring them to my office ***until xxx***.

E2. *Getting a mark*

After handing in everything, your report and code will be marked by three parties:

- I mark your report and your code.
- You mark your report ***yourself!*** (i.e., self-evaluation)
- Two ***fellow students*** from the same course will mark your report. (i.e., peer-evaluation)

Your final mark will be the average value from the three marks.

Because the reports will be rotated among all students, reports from groups of two will be evaluated by two other students, while reports from single students will be evaluated only by one other student. In the first case, the two evaluations from the two fellow students will be averaged to result in on mark counting one-third of the total mark. Which student will mark whose other student's report will be determined by me based on the similarity between the different projects.

This somewhat revolutionary marking procedure allows you to not only learn some technical details about the FE-method and how to apply the method to your problem, but also to gain some crucial "soft skills", which may be useful for your further career, such as

- Honest self-evaluation
- Evaluating reports, where you are not the exact expert
- Be objective when dealing with colleagues
- Be objective with yourself
- Accept criticism from colleagues without taking it personal

Of course, the mark that you give to yourself and to a colleague of yours has to be justified. Therefore, not only the mark is required, but a written statement that justifies your decision. To help you during the evaluation process (both self-evaluation and peer-evaluation), I will provide some evaluation guidelines and some reports from the same course from previous years as a comparison. So, everybody's task will be threefold:

- Write a report and hand in all the codes.
- After handing in everything, evaluate your own report yourself based on the evaluation guidelines.
- Evaluate one other report from a colleague.

The self- and peer-evaluation should happen as soon as possible after the deadline for handing in the report, i.e., soon after ***xxx***, but latest until the new semester starts, that is until ***xxx***.

Now have fun…

E3. *General tips and tricks*

During this course, you wrote two working codes, one for 2D elastic media and one for 2D incompressible linear viscous (Newtonian) media. Now it is time to apply one of these codes. So, most of you do not need to include any new physics or any new equations, but building the model (i.e., the geometrical setup and setting up the numerical grid) is the most challenging part. Therefore, I suggest that you first try to draw the numerical setup on a sheet of paper and try to understand how you can define the external and internal boundaries and interfaces mathematically, so that you can then program them.

[In the original document, I list some technical tips and tricks in MATLAB, which are not of relevance here.]

Appendix F

In the following, the evaluation guidelines that are handed out to the students are provided. These guidelines specify some details concerning the evaluation procedure and explain how to evaluate the reports for the self- and peer-evaluation.

F1. *Guideline for report evaluation*

Your report and code will be marked by three parties: me, ***yourself*** (i.e., self-evaluation), and ***one or two fellow students*** (i.e., peer-evaluation) (depending whether you are in a group of two or alone). So, everybody will have to evaluate his/her own report plus one other report from a colleague. Of course, the mark that you give to yourself and to a colleague of yours has to be ***justified***. Therefore, not only the mark is required, but a ***written statement*** that justifies your decision. This written statement should be roughly ***one-page long***.

To help you evaluating (both self- and peer-evaluation), I provide some questions that you may/should answer during the evaluation, very similar to a review process of a scientific manuscript. I deliberately do not provide a strict evaluation form, because you should keep some flexibility and independence. As additional help, you will also have access to some reports from the same course but from previous years as a comparison.

F2. *Possible questions for evaluation*

Because your reports are reasonably short, you are not supposed to explain the purpose of the study and how it fits into a bigger picture. Therefore, you should also not evaluate these points. During both self- and peer-evaluation, you may consider the following questions:

Table 7 Swiss marking system

Mark	Meaning	Comment
6.0	Excellent	Best possible mark
5.5	Very good	
5.0	Good	
4.5	Satisfactory	
4.0	Sufficient	Passed the course
3.0	Insufficient	Failed the course
2.0	Poor	
1.0	Very poor	

- What is your general impression of the report?
- Is the structure of the report logical? Is there a good flow in reading the report?
- Are all assumptions clearly declared and justified?
- Is the applied method and solution strategy clear?
- Are the results clear and well presented?
- Is there a clear distinction between results and interpretations?
- Are the conclusions supported by data/results?
- Are the figures clear (labels, captions, colors, legend, etc.)?
- Do the figures contribute to a better understanding of the text?
- Are all figures necessary?
- Are all citations (if any) correctly listed at the end of the report?
- Did anything strike you as particularly clever/novel/elegant/interesting?

F3. *Marking the report*

In particular for self-evaluation for students in groups of two:

- How was the work distributed?
- Did someone contribute considerably more than the other?

After you evaluated the report, you should mark the report (both yours and your colleague's). For marking the report, you should follow the standard Swiss marks, which have the meaning given in Table 7. You can give marks with an accuracy of 0.25 (i.e., 4.75, 5.25).

F4. *Other remarks*

- I am not asking you to check your ***colleague's code*** in detail (I will take care of this). However, the code will be provided and it is your decision to which detail you want/have to look at the code for your evaluation.
- Even though I will not check on this, I ask you to ***do the self-evaluation alone***, even if you are in a group of two.
- This somewhat experimental evaluation procedure requires and trains some "soft skills", which I want you to keep in mind when performing the evaluation:

- Be as ***objective*** as possible. Forget the person behind the report (both yourself for self-evaluation and your colleague for peer-evaluation).
- Be ***honest with your colleague***. If you consider (parts of) your colleague's report bad, tell so in your written statement, but justify and explain in a way that your colleague can accept your criticism.
- Be ***honest with yourself*** during self-evaluation.

References

Anderson LW, Krathwohl DR, Airasian PW, Cruikshank KA, Mayer RE, Pintrich PR, Raths J, Wittrock MC (2001) A taxonomy for learning, teaching, and assessing: a revision of bloom's taxonomy of educational objectives. Longman, New York. ISBN 0-415-34279-1

Biggs J (1996) Enhancing teaching through constructive alignment. High Educ 32:347–364. https://doi.org/10.1007/BF00138871

Bloom BS, Engelhart MD, Furst EJ, Hill WH, Krathwohl DR (1956) Taxonomy of educational objectives: the classification of educational goals, handbook 1: cognitive domain. Longmans, Green and Co. Ltd., London. ISBN 0-582-28010-9

Downing SM, Haladyna TM (2011) Handbook of test development. Lawrence Erlbaum Associates Inc, Publishers, Mahwah. ISBN 0-8058-5265-4

ETH Zurich, Department for Educational Development and Technology (LET) (2013) Guidelines on grading written examinations. ETH Zurich, Zurich

Frehner M, Schmalholz SM (2006) Numerical simulations of parasitic folding in multilayers. J Struct Geol 28:1647–1657. https://doi.org/10.1016/j.jsg.2006.05.008

Krathwohl DR (2002) A revision of Bloom's taxonomy: an overview. Theory Pract 41:212–218. https://doi.org/10.1207/s15430421tip4104_2

Oxburgh ER, Parmentier EM (1978) Thermal processes in the formation of continental lithosphere. Philos Trans Royal Soc Math Phys Eng Sci 288:415–429. https://doi.org/10.1098/rsta.1978.0025

Race P, Brown S, Smith B (2005) 500 Tips on assessment. Routledge Falmer, Abingdon. ISBN 0-415-34279-1