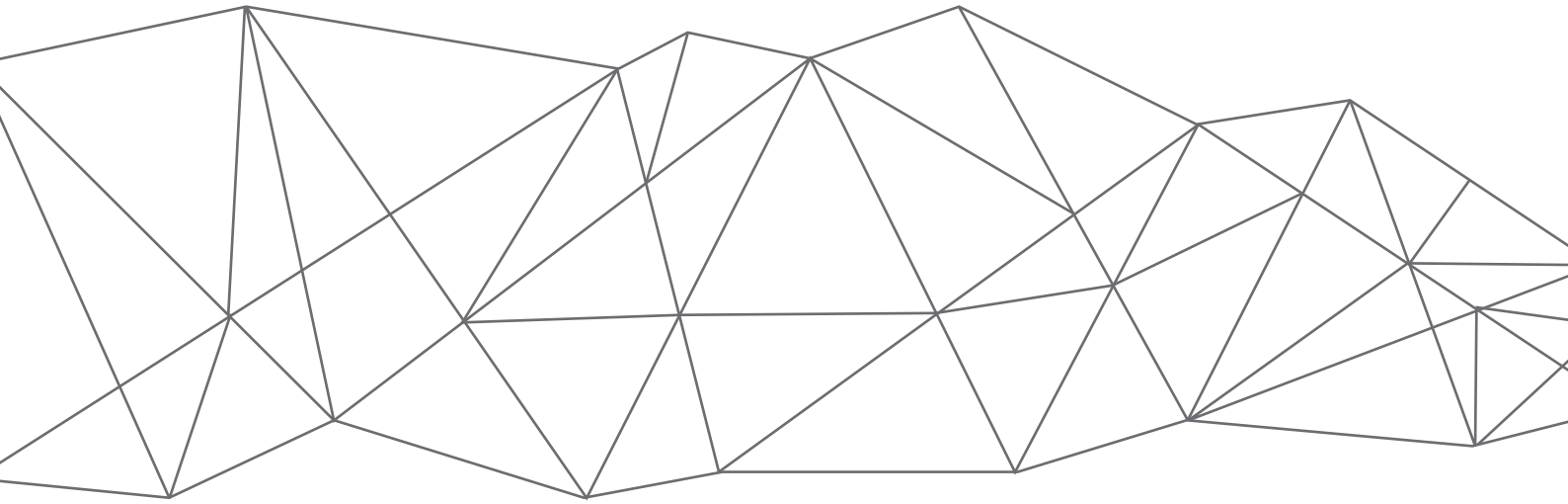
Anja Friedrich and Boas Pucker

Peer-review as a teaching method



Working Paper der AG Forschendes Lernen
in der dghd

Diese Working Paper Reihe ist ein Produkt der AG Forschendes Lernen in der Deutschen Gesellschaft für Hochschuldidaktik (dghd). Sie erscheint als Online-Publikation an der Carl von Ossietzky Universität Oldenburg und richtet sich an alle, die an Forschendem Lernen interessiert sind. Veröffentlicht werden wissenschaftliche und praxisnahe Beiträge zum Forschenden Lernen. Die Autor_innen müssen nicht Mitglied der AG Forschendes Lernen sein. Veröffentlichungen sind in deutscher und englischer Sprache möglich. Alle eingereichten Beiträge durchlaufen einen Begutachtungsprozess.

Redaktion und Kontakt: Dr. Janina Thiem (fl-workingpaper@uol.de).

Das Herausgeber_innen-Team bedankt sich beim Vorstand der dghd für dessen Unterstützung der Reihe.

Die Working Paper sind abrufbar unter https://www.uni-oldenburg.de/fl-workingpaper/

Herausgeber: Carl von Ossietzky Universität Oldenburg

Druck: BIS-Druckzentrum

Die Veröffentlichung der Working Paper Reihe wird aus Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen FKZ 01PL16056 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autor_innen.

FLiF+FoL
Forschungsbasiertes Lernen im Fokus + Forschungsorientierte Lehre

forschen@studium

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

Anja Friedrich[1,*] and Boas Pucker[2,3]

Peer-review as a teaching method

Working Paper Nr. 2, 2018

* shared first authorship

1 Faculty of Psychology and Sports Science, Bielefeld University, Bielefeld Germany
2 Faculty of Biology, Bielefeld University, Bielefeld, Germany
3 Center for Biotechnology, Bielefeld University, Bielefeld, Germany

AF: anja.friedrich@uni-bielefeld.de
BP: bpucker@cebitec.uni-bielefeld.de

Carl von Ossietzky Universität Oldenburg

Abstract

Peer-reviews are a common and valued teaching tool at Anglo-American and Asian universities. Previous studies recommended a scaffolded peer-review process and pre-specified criteria. The current study investigated the feasibility and acceptance of scaffolded peer-reviews as a teaching method in German college students. Participants were 7 psychology students and 13 life science students. The students had to write a project report about a psychological experiment or genome research projects. All reports underwent a scaffolded peer-review process according to pre-specified criteria. The students' feasibility and acceptance ratings were evaluated using customized questionnaires. The results indicated a good feasibility and acceptance in both courses, although the small samples and the different measurements impair comparability and restrict generalization. Descriptive data and qualitative comments indicated similarities between psychology and life science students. In line with evidence from other countries, this subsample of German college students provided first empirical evidence that the scaffolded peer-review might be a feasible and well-accepted teaching method. Future studies should include methodological improvements (e.g. control condition).

Abstract (deutsch)

Peer-reviews sind eine beliebte Lehrmethode in anglo-amerikanischen und asiatischen Universitäten. Die bisherige Literatur empfiehlt einen strukturierten Peer-review-Prozess nach vorher festgelegten Kriterien. Diese Studie untersucht die Machbarkeit und Akzeptanz eines angeleiteten Peer-reviews als Lehrmethode bei deutschen Studierenden. Teilnehmende waren 7 Psychologie- und 13 Studierende der Lebenswissenschaften. Die Studierenden verfassten einen Projektbericht über ein psychologisches Experiment oder Genomuntersuchungen. Alle Berichte durchliefen einen angeleiteten Peer-review Prozess nach vorher festgelegten Kriterien. Die Machbarkeits- und Akzeptanzratings wurden mit neu entwickelten Fragebögen erfasst. Die Ergebnisse deuteten auf eine gute Machbarkeit und Akzeptanz hin. Die deskriptiven Daten und qualitativen Kommentare implizierten Ähnlichkeiten zwischen Studierenden aus Psychologie und Lebenswissenschaften. In Übereinstimmung mit der bisherigen wissenschaftlichen Evidenz lieferte diese Studie erste Hinweise darauf, dass diese kleine Stichprobe von deutschen Studierenden den angeleiteten Peer-review-Prozess als machbare und akzeptable Lehrmethode bewerten. Zukünftige Studien sollten methodische Verbesserungen berücksichtigen (z.B. Kontrollgruppe).

1 Introduction

1.1 Research-based learning and the advantages of peer assessment

Research-based learning is a teaching principle that aims to imbed research principles and processes (e.g. discovering and solving scientific problems) in academic learning contexts (Huber, 1970). It often focusses on the scientific process itself and rarely on other aspects of research related to this process, such as peer assessments. However, peer assessments are an integral part of today's research process, as double-blind peer-reviews are the gold standard for publishing articles in academic journals (e.g. Ware, 2008). Thus, students should learn to conduct peer-reviews during their academic career. Moreover, peer assessment is a common tool in higher education: Students give and receive feedback on presentations, they review their scientific writing in group projects, and they even grade each other's papers. Although the definitions of peer assessment vary, most include the rating or evaluation of a student's work by his or her peers (e.g. Falchikov, 1995).

As early as 1998, Topping conducted a narrative review on the typology and scientific evidence of peer assessments in higher education. The review included 109 original studies. The typology encompassed 17 aspects of peer assessments, including the subject area, objectives, output, privacy, ability, place, time, official weight, as well the constellations of assessors and assesses, respectively.

In addition to (meta-)cognitive benefits, such as greater self-awareness and earlier reflection of errors, Topping found that peer assessments increased the students' personal responsibility and motivation, teamwork, active learning, tolerance for criticism as well as a greater insight into the assessment process. Potential disadvantages included a low acceptance of peer feedback from poorer students, a lack of responsibility for the peer assessment and the abuse of peer power.

The results from 31 studies provided enough evidence for an acceptable reliability and validity of peer assessments. While reliability was measured by correlating several peer assessed scores or grades, validity was obtained by correlating peer assessments with self- and tutor assessments. Although Topping does not provide quantitative reliability data for most of the included studies, exemplary data from some studies report correlations between different peer assessments of $r = .85 - .88$, indicating a high reliability. The validity assessments revealed more modest correlations for the comparison of peer with self assessments ($r = .61 - .64$) and tutor assessments ($r = .34 - .55$).

However, the students' acceptance of peer assessments varied and was not related to the actual reliability, thereby indicating a low face validity. Finally, Topping provided step-by-step instructions to ensure a higher quality of peer assessments:

1. “Clarifying expectations, objectives and acceptability,
2. matching participants and arranging contact,
3. developing and clarifying assessment criteria,
4. providing quality training,
5. specifying activities,
6. monitoring the process and coaching,
7. moderating reliability and validity,
8. evaluating and providing feedback” (Topping, 1998, 265–267)

The following year, Dochy, Segers and Sluijsmans (1999) wrote a narrative review on 63 studies examining the use of self-, peer and co-assessment in higher education. When compared to teacher assessments, the peer assessments showed a good validity and reliability. Furthermore, the students enjoyed the peer assessment process, appreciated the individual’s performance, took more responsibility, and reported better interpersonal relationships with their peers. Risks in peer assessments included friendship markings, collusive markings (= a lack of differentiation between different group members), decibel markings (= the most dominant student receives the best marks) and parasite markings (= social loafing). Summarizing their findings, the authors formulated seven guidelines for peer assessments that match points described by Topping e.g. training of students and definition of criteria prior to the peer assessment. They highlight the fact that peer assessment takes time and that it could be used as a learning tool. As a preparation for the application of peer assessment for educational purposes, they recommend the development of suitable stuff development programs to prepare teachers for this task.

Correspondingly, Cho, Schunn and Wilson (2006) showed that not only peer assessment, but also peer grading of students’ writing is a valid and reliable approach. Participants were 708 students from 16 different courses across ten different subjects. The authors examined the results from a scaffolded peer-review process, during which three to six students reviewed another students’ writing assignment using three pre-specified criteria (flow, logic, insight). Results indicated a good validity and an acceptable reliability from the instructors’ perspectives across all subjects. Reliability increased with a larger number of peer-reviewers. Oppositely, the students estimated that the peer grades had a low validity and reliability, especially in average students. A possible explanation for this discrepancy was that the instructor could compare his own assessment with the peer assessments, while the students only had their own assessments of other writing tasks as comparisons. Despite the poor face validity, the peer assessments and peer grades provided good validity and reliability, indicating that students are good assessors. Furthermore, being an assessor improved the students’ own writing, which highlights the benefits of being assessor as well as assessee.

The importance of the last finding was further stressed in a study with 43 undergraduate students (Li, Liu & Steckelberg, 2010). Using a web-based tool, the students evaluated two projects from their peers and in turn received feedback from two other students on their own projects. Then, the quality of the reviews the students' provided for their peers' projects (assessors) and the quality of the reviews they received (assessees) were evaluated by the instructors. Surprisingly, only the quality of reviews the students provided for their peers' projects (assessors) correlated with the quality of the student's final project with a large effect size. The quality of the reviews that the students received (assessees) did not correlate with the quality of the final project. Thus, the role of assessor might have been more important than previously thought.

Two studies with large sample sizes examined the effects of peer assessments and possible moderators. Sun, Harris, Walther and Baiocchi (2015) conducted a matched randomized crossover experiment with 378 STEM students who alternately participated in peer assessments during the semester. The students who gave and received peer assessment on their homework scored better on the short-term unit quizzes and the long-term final exam than those who did not. Wen and Tsai (2006) investigated the students' perceptions and attitudes towards peer assessment. 280 Taiwanese students participated in the study. The results indicated a generally positive view of peer assessment, with a more positive attitude in male students and less negative attitude in students with previous experience in peer assessment. These factors (short/long term, gender and experience) should be considered when interpreting evidence of peer assessment studies.

1.3 Peer assessment in writing projects: The peer-review

While peer assessments in general are more common research topics, fewer studies concern themselves with peer assessments in writing projects. A common term to describe the assessment of a written product (e.g. report, article, thesis) is the term peer-review. Thus, the peer-review process constitutes a subgroup of the more general term peer assessment.

Pelaez (2002) compared the efficacy of problem-based writing combined with peer-review to traditional lectures. 35 students from a Human Physiology course participated in the study. Despite the fact that students in the peer-review group had problems relating concepts, they received better marks in the multiple choice questions during midterm and the final exam ($p < .001$), leading to the conclusion that problem based writing with peer-review is superior to traditional lectures.

Another application for peer assessment was investigated by Reynolds and Thompson (2011). They compared Biology students who received a writing course with peer assessment ($N = 47$) to those who did not ($N = 143$). All students received the regular one-on-one consultations with mentors from the faculty. Results showed that significantly more students in the writing course received degrees with honors. Furthermore, they showed a

higher quality of critical thinking and writing skills as assessed by the instructors. These findings support the implementation of peer-review processes as a preparation for thesis writing.

In a recent study, Meihami and Razmjoo (2016) asked 11 Iranian teachers and 56 Iranian undergraduate students about the challenges of implementing peer assessment in writing courses. The qualitative analysis of the semi-structured interviews (teachers) and the open-ended questions (students) revealed that the lack of assessment literacy in teachers and learners are the greatest obstacles when implementing peer assessment. Similar concerns were raised in a study with computer science students (Venables & Summit, 2003). The effect of assessment literacy was confirmed in an earlier study by Min (2006), who examined the effects of assessment training on the quality of peer-review processes in a Taiwanese EFL class. After five hours of training, the students incorporated more reviewer comments into their drafts and the quality of the revised drafts improved significantly. This indicates that moderate training significantly improves the peer-review process in writing tasks.

Aside from learning and grading, the students' acceptance and feasibility ratings of the peer-review process have to be considered. Most of the aforementioned studies only allude to the students' acceptance in their discussion, based on qualitative commentary. Venables and Summit (2003) pose a notable exception. They examined the students' acceptance and feasibility ratings of a peer-review process about literature review essays using three items:

- "When you were first given the essay assessment and description, what did you think about the idea of having to mark other students' work?"
- "What was the most positive thing you can say about the essay assessment?"
- "What was the least positive thing you can say about the essay assessment?"

In their sample of 61 computer science students, most students (n = 38) had initial reservations about peer assessment. 20 students responded to the second question that they "learned a lot". The most common concern was the "time required for task" (n = 9). The results are limited by the low response rate, as 16 students did not answer the second and 32 students did not answer the third question.

One study reported contrary findings. Yankulov and Couto (2012) compared 393 peer-reviews on research proposals to instructor ratings. While the peer-reviews were able to differentiate between pass and fail, the marks were highly variable. Furthermore, the students had trouble distinguishing a good from an excellent performance. Despite previous peer-review training and pre-specified criteria, this study did not reveal positive effects of peer-reviews in scientific writing. The lack of experience in grading proposals could be one explanation for this observation. Students could face problems when assessing performances that are superior to their own. Since the number of excellent students is usually

very small, it could be assumed that good performances cannot be distinguished from excellent ones. Additionally, the variation between different instructor grading was not investigated. A low correlation between grades from different persons could be a general issue that is not particularly related to peer-review. Furthermore, Yankulov and Couto used a very open-ended writing task for their peer-review process, which might contribute to the higher variance in the students' ratings.

1.4 The current study

1.4.1 Peer-review and research-based learning

The scientific evidence suggests that students profit from peer-reviews in various writing projects across different subjects – as long as they are trained properly and receive specific criteria. By using peer-reviews as a method of research-based learning, the students gain the opportunity to learn from the positive aspects as well as the mistakes of their peers' work. Furthermore, they need to present their findings in a coherent manner, so that the assesse understands what was good about their text and what needs to be changed. Hence, peer-reviews do not only improve the assessee's work on a content-level, but also the assessors' reflection of what constitutes a good scientific text on a meta-level. In turn, this could lead to an improvement of the assessors' own work.

Considering that peer-reviews are a form of preliminary evaluation, the necessity for some kind of product to be evaluated is obvious. Evidently, peer-reviews are a teaching method best applied in the later stages of research-based learning, when the research questions and methods have already been defined. Thus, peer-reviews are especially suitable when checking and presenting the students' research findings in a scientific thesis, report or article.

1.4.2 Topping's typology and the current study

According to Topping's typology, the current study focuses on psychology and life science students. The objective is to increase cognitive and metacognitive abilities using qualitative and quantitative reviews. The output will be research papers and the reviews will be supplementary to the staff assessment. The reviews will not contribute to the final grade and the directionality is one-way. The reviews will be conducted anonymously (life sciences) and publicly (psychology) without face-to-face contact. The individual assessors and the assessee groups will be in the same year, thus having similar abilities. The peer-review process will take place outside of class. It will be compulsory for both, assessors and assessees. Finally, the reward consists of course credit.

1.4.3 Objectives

Considering that peer assessments are widely used and evaluated in Anglo-American and Asian universities, the lack of scientific evidence examining their feasibility and acceptance in other countries is disconcerting. More studies are needed to test whether this approach is useful in other countries with different educational backgrounds and different subjects. In addition, the scientific evidence regarding the application of peer assessment in writing projects is lacking and should be expanded. We set out to develop and evaluate the application of peer-review methods in highly structured courses in Germany.

2 Peer-review as a teaching method in different courses

The peer-review process was applied in two courses: one course about experiments in psychology and one course about applications of next generation sequencing technologies in genome research. Results described in this report were deduced from observations in these courses, which took place within one year.

2.1 Example 1: Psychology course

The course was offered at Bielefeld University, Germany. All participants were recruited in an elective two semester course about planning, conducting and presenting scientific experiments in clinical psychology. The experiments were conducted in project groups of three to four students. Over the course of the first semester (10/2016–03/2017), the participants planned and partially conducted the experiments. The second semester (04/2017–09/2017) additionally included the interpretation of the experiments' results. In order to receive course credits, each project group had to write a scientific article about their experiment. The drafts of these articles were then forwarded to three or four students from different project groups who were not involved with the project. They reviewed the article according to the peer-review instructions (*appendix 1*) and graded it (*appendix 2*) within eight weekdays. The peer-review instructions were discussed beforehand in class and students received the opportunity to ask questions.

2.1.1 Participants

Fifteen graduate psychology students were enrolled in the course, seven of them participated in the current study. All participants had a Bachelor degree and studied in their first semester for a Master of Science. All of the participants were female. Other demographic characteristics were not recorded. The study was conducted from October 2016 to September 2017. The participants had to be graduate psychology students and participants of the project course. No other in- or exclusion criteria were applied.

2.1.2 Peer review instructions

The peer-review instructions *(appendix 1: Peer-review instructions for psychology students)* were developed for the current study and presented in a rubric. The content was based on the authors' experience with scientific writing and common mistakes among students. Every important section of the article (title, abstract, introduction, current study/ practical relevance, method, results, discussion, and literature) was explained in detail using the three categories aim/function of the section, the required content and aspects of the academic writing style. The peer-review instructions were utilized two times: First for the preparation of the article by the project group and then for the evaluation of the article by the peer-reviewers.

2.1.3 Peer-review grading sheet

The peer-review grading sheet (*appendix 2: Peer-review grading sheet*) was developed for the current study and presented in a table. The content was based on the peer-review instructions. Seven sections of the article (title, abstract, introduction/practical relevance, method, results, discussion, and literature) were evaluated via free text and a rating scale. The free text ratings consisted of aspects that the reviewer appreciated ("Things I liked") and aspects that the reviewer felt needed improvement ("Things you could improve"). The rating scale was a ten point Likert scale ranging from one ("very bad") to ten ("very good").

2.1.4 Measurement: Feasibility and acceptance questionnaire

After the course, the participants received the respective acceptance and feasibility questionnaires via e-mail. If the participants failed to respond, a reminder was sent to them via e-mail. The participants were not reimbursed for their participation, neither did they receive course credit for the acceptance and feasibility questionnaire. The study adhered to the ethical guidelines of the German Psychological Society (DGP) and the Association of German Professional Psychologists (BDP). All participants provided written informed consent.

The feasibility and acceptance questionnaire (FAQ) was developed for the current study (**Table 1**). It consisted of 14 items and two scales, with seven items belonging to the feasibility scale and seven items belonging to the acceptance scale. Thus, each sum scale had a range of 7 to 70. Higher values indicated a higher acceptance and a better feasibility of the peer-review process. Additionally, two items pertained to the assessor perspective (F2: "I had enough time to read and review the other project article.", A3: "I think that my suggestions helped the other authors to improve their article.") and two items to the assessee perspective (F1: "I had enough time to incorporate the reviewers' comments.", A2: "The reviewers comments improved my own article."). All items were rated on a ten point Likert scale ranging from one ("does not apply at all") to ten ("applies completely"). As the FAQ was developed for the current study, cut-off scores and information about psychometrics were not available. However, the short duration of the FAQ (approx. 5 min) indicated a good economic quality.

Table 1: Feasibility and acceptance questionnaire (FAQ). Higher values indicate a higher acceptance and better feasibility

	trifft überhaupt nicht zu (1) *does not apply at all*									trifft vollständig zu (10) *applies completely*
Machbarkeit *feasibility*										
1. Ich hatte genug Zeit, um die Kommentare der Reviewer einzuarbeiten. *1. I had enough time to incorporate the reviewer comments.*	1	2	3	4	5	6	7	8	9	10
2. Ich hatte genug Zeit, um den anderen Projektbericht zu lesen und zu bewerten. *2. I had enough time to read and review the other project article.*	1	2	3	4	5	6	7	8	9	10
3. Der Arbeitsaufwand für den Peer Review Prozess war angemessen. *3. The effort fort the peer review process was appropriate.*	1	2	3	4	5	6	7	8	9	10
4. Der Erwartungshorizont war verständlich. *4. The peer review instructions were comprehensible.*	1	2	3	4	5	6	7	8	9	10
5. Der Erwartungshorizont war hilfreich. *5. The peer review instructions were helpful.*	1	2	3	4	5	6	7	8	9	10
6. Die Bewertungskriterien waren verständlich. *6. The peer review grading sheet was comprehensible.*	1	2	3	4	5	6	7	8	9	10
7. Die Bewertungskriterien waren hilfreich. *7. The peer review grading sheet was helpful.*	1	2	3	4	5	6	7	8	9	10
Akzeptanz / Vorteile *Acceptance / benefits*										
1. Der Peer Review Prozess hat mir gefallen. *1. I liked the peer review process.*	1	2	3	4	5	6	7	8	9	10
2. Die Kommentare der Reviewer haben mir dabei geholfen, meinen Projektbericht zu verbessern. *2. The reviewers' comments improved my own project article.*	1	2	3	4	5	6	7	8	9	10
3. Ich glaube, dass meine Bewertung und meine Kommentare der anderen Projektberichte den Autoren geholfen haben, ihren Projektbericht zu verbessern. *3. I think that my suggestions helped the other authors to improve their article.*	1	2	3	4	5	6	7	8	9	10
4. Das Lesen und Bewerten der anderen Projektberichte hat mir dabei geholfen, meinen eigenen Projektbericht zu verbessern. *4. The peer review method helped me with improving my own project article.*	1	2	3	4	5	6	7	8	9	10

5. Ich würde den Peer Review Prozess in einer anderen Veranstaltung gerne nochmal machen. *5. I would like to repeat the peer review process in other courses.*	1	2	3	4	5	6	7	8	9	10
6. Ich würde die Methode für künftige Projektseminare weiterempfehlen. *6. I would recommend the peer review method for future project courses.*	1	2	3	4	5	6	7	8	9	10
7. Ich glaube, dass mich diese Übung gut auf eine wissenschaftliche Karriere vorbereitet hat. *7. I think that this exercise prepared me for a scientific career.*	1	2	3	4	5	6	7	8	9	10
Kommentar: comments:										

2.1.5 Results

The feasibility and acceptance of the peer-review method were investigated with the FAQ. Due to the small sample (n = 7), all results should be interpreted with caution. The sum score of the feasibility scale ranged from 52 to 59 (M = 57.14, SD = 2.91). The sum score for the acceptance scale ranged from 40 to 54 (M = 47.71, SD = 5.62). All items on both scales were rated greater than 5 on a scale from one to ten. As higher values indicated a better feasibility and acceptance of the peer-review method, the students evaluated the method favorably. While the item about a potential repetition of the peer-review process (item 5) received the lowest acceptance rating (M = 5.71, SD = 1.38), the time and effort it took received the lowest feasibility rating (M = 6.71, SD = 1.79). On the other hand, students reported that their own suggestions helped their peers (M = 8.14, SD = 2.41) and that they received helpful instructions (M = 9.43, SD = .54). The acceptance of this method seemed to vary stronger than the feasibility among the participants.

Psychology students commented that the peer-review method taught them important skills and that it was a useful method. A negative aspect was the time it took to complete the peer-review method. One student suggested using the peer-review method on individual reports instead of group reports, as group reports already included the feedback of several students.

2.3 Example 2: Peer-review in 'Applied Genome Research'

In parallel, the teaching method peer-review was tested in two repetitions of the seminar 'Applied Genome Research'. This course provides a general understanding of the concept of sequencing and processing of next generation sequencing data. Participation in this course is not obligatory for any of the study programs at Bielefeld University. Students perform their own analysis on recently published and simulated data sets. Starting with raw sequencing reads they perform multiple data processing steps by utilizing various bioinformatic tools. Finally, results are interpreted in the biological context based on scientific literature. Presenting current research topics in oral presentations, writing about own results, and reviewing the work of peers add to this research experience.

2.3.1 Participants

Thirteen life science students from different study programs including biology, genome-based systems biology, biochemistry, molecular cell biology, molecular biotechnology, bioinformatics and genome research participated in the evaluated courses. They were in the last semesters of their Bachelor of Science (N = 10) or just started their Master of Science (N = 3). Six students participated in the first course and seven students in the second one. Almost all participants were between 20 and 25 years old. Most of the participants were female (70%) representing roughly the general gender ratio in the involved study programs. The courses described here were conducted in March 2017 (winter semester) and August 2017 (summer semester), respectively. All participants were verbally informed about the content and the aim of this report. Moreover, they had the opportunity to ask questions, before they decided to participate. Correspondingly, the participants gave verbally informed consent.

2.3.2 Measurement: Online survey

The students were instructed to conduct the peer-review process according to the peer-review instructions for life science students (*appendix 3*). The feasibility and acceptance of the peer-review in relation to other teaching methods applied in a seminar were investigated via an online survey (**Table 2**).

Table 2: Acceptance and feasibility questionnaire in the course about Next Generation Sequencing.

item	Question	Yes	Maybe	No
1	Würdest du den Kurs weiterempfehlen / soll der Kurs wieder angeboten werden? Would you recommend this course / should it be offered again?			
2	Waren die Vorlesungsteile hilfreich? Were the lecture components helpful?			
3	Waren die Übungen hilfreich? Were the exercises helpful?			
4	Waren die Fragen auf den Folien hilfreich? Were the questions on the lecture slides helpful?			
5	Waren die Wiederholungen mit Möglichkeiten zum Stellen von Fragen hilfreich? Were summaries as well as the chance to ask questions helpful?			
6	Waren die studentischen Vorträge hilfreich? Were the talks by students helpful?			
7	War das Schreiben des Protokolls hilfreich? Was it helpful to write a protocol?			
8	War der peer-review-Prozess zur Kontrolle der Protokolle hilfreich? Was the peer-review process for evaluation and correction of the protocols helpful?			
9	Waren die Anleitungen zum peer-review-Prozess ausreichend? Were there sufficient instructions to perform the peer-review process?			

Table 2 displays the items about the acceptance (item 8) and feasibility (item 9) of the peer-review method in the course about Next Generation Sequencing (NGS). The items 1 to 7 examined the acceptance of other components of the NGS course (e.g. presentations, exercises, quizzes). The entire questionnaire was presented anonymously through a Doodle poll. Adding comments was also possible to enable improvement suggestions. Participants were not able to see the answers of the other participants.

2.3.3 Results

In general, the feedback from life science students was positive and constructive. Since individual reports were written, no feedback was given to the students during the writing process. Most students agreed that the amount of work is high compared to the classical system, in which the reports are controlled by teachers. Nevertheless, some students explicitly stated that this method should be kept for repetitions of the courses. Despite the increased workload several students evaluated this method as very useful. Quantitative interpretation of the results was limited by the relatively small sample size.

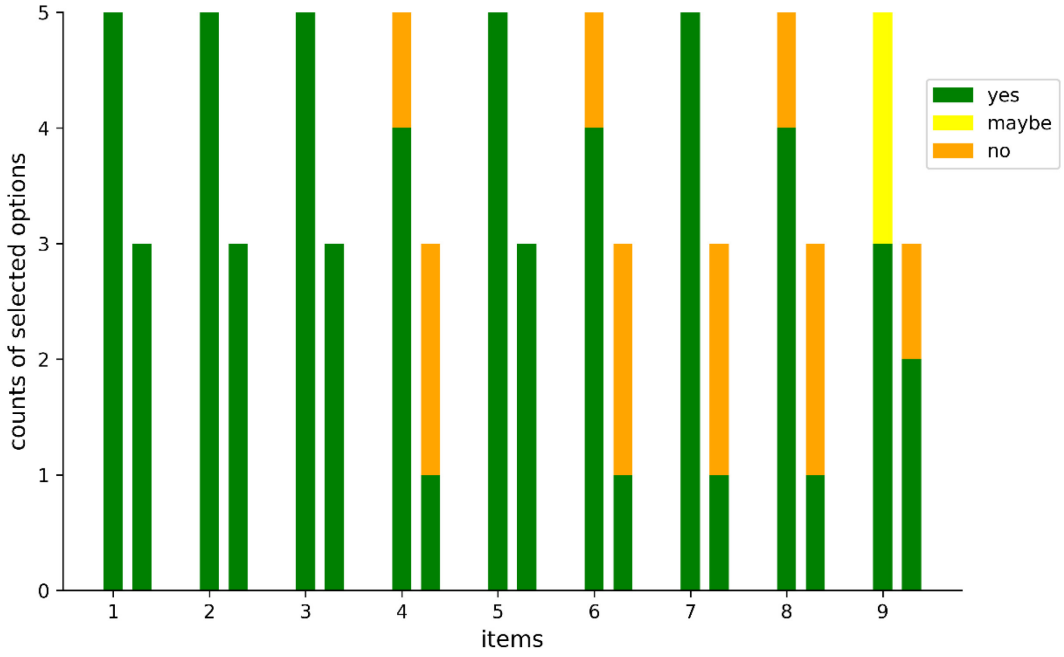


Figure 1: Survey results for the comparison of peer-review relevance for teaching in relation to other established methods. All participants of two repetitions of a seminar were kindly asked to fill out a short poll. Sample sizes (at least partly completed surveys) were n = 5 (NGS1) and n = 3 (NGS2), respectively. Peer-review was evaluated by item 8 (acceptance) and 9 (feasibility).

3 Discussion of collected experiences

3.1 Summary of observations and alternative explanations

3.1.1 Summary

The current report evaluated the feasibility and acceptance of a scaffolded peer-review process in German college students from different subjects. Students from different subjects rated the scaffolded peer-review process as a feasible teaching method.

The psychology students reported varied, but overall positive opinions regarding the peer-reviews' acceptance. The possible repetition and the required time received lower acceptance ratings, the evaluation of their own suggestions and the peer-review instructions received higher acceptance ratings.

While the psychology students provided heterogeneous but positive opinions regarding the process' acceptance, fewer life science students accepted the scaffolded peer-review. However, two thirds of the life science students responded favorably.

A direct comparison of the courses is not possible, as different measurements were used. Overall, the results provide first empirical evidence that the scaffolded peer-review process is a well-accepted and feasible teaching method in this sample of German psychology and life science students.

3.1.2 Additional Considerations

The results of the current study should be interpreted with caution due to the small sample size and the fact that not all participants from the psychology course who conducted the peer-review process completed the feasibility and acceptance questionnaires. It is possible that only those students who liked and profited from the peer-review process filled in the questionnaires and those who did not refused. This could bias the results leading to a more positive evaluation of the peer-review method. The same bias might be present in the NGS course, as the completion of the peer-review evaluation was voluntary.

Furthermore, the instruments were not validated in a pilot study and the feasibility and acceptance ratings in the NGS course were based on only one item, which possibly does not cover the whole construct of feasibility or acceptance. Thus, the positive results of the current study regarding the peer-review process might partially be caused by this methodological shortcoming and not by the students' positive evaluations.

3.2 Scientific integration

This work reports similar results for the use of the peer-review process in two different courses. Furthermore, this report does not describe an application in a specific scientific writing course, as was the case in the study from Meihami and Razmjoo (2016), or conducted on unit quizzes (Sun, Harris, Walther & Baiocchi, 2015). The fact that the psychology and life science students still liked the scaffolded peer-review process hints at the possibility that the use of this method might also be feasible in other contexts.

Important factors that influence the feasibility and efficacy of peer-reviews in teaching are gender, with a more positive attitude in male students, and previous experience with peer-reviews (Wen & Tsai, 2006). The application of this method in two courses revealed that female students also profited from the experience of peer-reviewing their fellow students' writing, although gender comparisons could not be conducted due to the small sample size. None of the students in this study had previous experience with peer-reviews, thus supporting the idea that inexperienced students also benefit from peer-reviews. Lin, Liu and Yuan (2001) examined the acceptance, feasibility and influential variables on the success of web-based peer assessment in 58 computer science students. They reported that students with high executive thinking style who received specific feedback benefitted the most, whereas students with low executive thinking style who received holistic feedback did not profit as much. Thus, a high executive thinking style and specific feedback provided the best results. Despite the fact that the current report did not contain information about the students' thinking styles, the positive results and the qualitative commentary indicate that the specific feedback was very helpful, lending support to the findings of Lin, Liu and Yuan (2001).

In general, the scientific studies on peer-reviews in teaching are very heterogeneous. Hence, the current report can be understood as a pilot study and did not consider all possible factors and outcome measures, resulting in a limited comparability. This pertained to the analysis of the quality of the papers (e.g. Cho, Schunn & Wilson, 2016), the marks on (final) exams (Pelaez, 2002) and the degree (Reynolds & Thompson, 2011). Despite the limited comparability, the small sample size and the different measurements, it was possible to replicate the positive findings on peer-reviews as a teaching method and provide first pilot data for a future application in a larger sample of German students.

3.3 Peer-reviews in research-based learning

3.3.1 Underlying mechanisms of peer-reviews

The results of the current study imply that peer-reviews are a viable method of research-based learning. Consequently, the underlying mechanisms that determine the success of research-based learning might also be present in the specific peer-review method.

Hänze and Moegling (2004) proposed that research-based learning is characterized by a high degree of self-determination coupled with competency. This degree of self-determination is also present in peer-reviews: The students assume the role of assessors, which provides them with the opportunity to assess their peers' work according to the pre-specified criteria. In addition to their own understanding of what constitutes a good scientific report, the pre-specified criteria aid the review process, as they provide guidelines to assist the students' pre-existing competency. Self-determination as well as competency support the development of intrinsic motivation, a powerful predictor of learning.

However, the presence (or absence) of these mechanisms remain speculative, as they were not assessed (e.g. via questionnaires) in the current study.

Other teaching elements like predefined exercises achieved better results during the evaluation than peer-review. Possible explanations are the high time effort associated with the peer-review process and the novelty of this method. Since peer-review is rarely used as a teaching method at German universities, long term test with repeated application should be performed. However, there are other teaching methods like writing a protocol or giving a oral presentation which were not favored over the peer-review process. Since peer-review requires writing a text first, the low acceptance of the writing process could influence the evaluation result.

3.3.2 Practical implications

The current study provides several practical implications for the implementation of peer-reviews as teaching method in academic settings. These practical implications pertain to three phases: Preparation, realization and evaluation.

In the preparation phase, the academic staff has to prepare for the peer-review process by defining the typology of the peer-review process. Then, they have to specify clear and concise criteria that will be used to write and review the scientific report, e.g. in a rubric. These two steps should take place before involving the students.

In the realization phase, the students have to be prepared for the peer-review process by informing them about the process and advantages of peer-reviews. This can be done in written form, e.g. in a syllabus. Students should be trained in the use of the pre-specified criteria and they should have the opportunity to ask questions (scaffolded process). The instructor has to decide beforehand whether or not they will assist the students during the peer-review phase and inform the students accordingly. It is also important to consider whether or not the assessors have to send their reviews to the instructor. This step compromises the assessors' anonymity while probably increasing the students' efforts, due to the fact that the students are aware that the instructors will read their reviews.

Finally, the acceptance and feasibility of the peer-review process should be evaluated using quantitative and qualitative assessment methods. The results may be presented to the students in a feedback session. The latter was not done in the current study, but may be helpful in order to increase the students' acceptance of the method.

A variety of different scientific skills can be improved through the use of peer-review as a teaching method. First, writing a scientific text for a lab protocol based on performed experiments and appropriate literature is the perfect training for writing theses, grant proposals, or scientific papers. Critical reading of peers' protocols leads to additional reflection about the topic. Since written communication is a fundamental skill in sciences, students should get plenty offers to practice. This is especially important for German students in subjects which require English texts.

3.4 Limitations and strengths

Since teaching was the main focus while collecting the presented data, the current report suffered from some limitations regarding the sample, the instruments and the scaffolded peer-review process. Despite relative high ratios of completed evaluations, the sample size was very small, impairing quantitative analyses. Furthermore, all courses mostly comprised of women. The latter aspect was not surprising, as both subjects, psychology and life sciences, are predominantly female subjects. Thus, the gender distribution in the sample reflected the actual distribution in the involved subjects. Furthermore, the acceptance results of the psychology students showed heterogeneous findings. Nevertheless, this heterogeneity occurred within the positive range. This means that even the items that received lower ratings (e.g. repetition of the peer-review process) still received positive ratings (> 5). Finally, the scaffolded peer-review process was not conducted anonymously and without a control condition, possibly facilitating environmental biases.

Irrespective of possible limitations, the current report contained several strengths regarding the sample, instruments and the design of the scaffolded peer-review process. The sample consisted of different disciplines, including psychology, biology, genome-based systems biology, biochemistry, molecular cell biology, molecular biotechnology, bioinformatics and genome research. This wide range of study subjects allowed for a better generalizability. Lastly, the design of the peer-review process was based on literature recommendations (e.g. scaffolded process, pre-specified criteria), increasing the validity.

3.5 Conclusion and future prospects

Overall, the current report provided preliminary empirical evidence for the feasibility and acceptance of scaffolded peer-reviews as a teaching method in a small sample of German psychology and life science students, replicating results from Anglo-American and Asian countries. Future studies on scaffolded peer-reviews should include a control condition, a larger sample size, more men and more subjects. Studies should extend the range of analyzed subjects especially into the humanities, where this method is not frequently used (The British Academy, 2007). Since the benefit of this method could depend on the subject, a comparison between subjects with frequent application of this method in the scientific practice to subjects without frequent use could be beneficial.

Practical implications are that peer-reviews are a valuable teaching tool in college students. Additionally, peer feedback could save the time of lecturers in larger courses by filtering out basic and frequently occurring mistakes in protocols. It is important to note that the peer-review-process should be scaffolded and follow pre-specified criteria. Moreover, constant supervision is necessary to ensure sufficient quality.

Literature

Cho, K., Schunn, C. D., & Wilson, W. (2006). Validity and reliability of scaffolded peer assessment of writing from instructor and student perspectives. Journal of Educational Psychology, 98, 891-901.

Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: a review. Studies in higher education, 24, 331-350.

Falchikov, N. (1995). Peer feedback marking: Developing peer assessment. Innovations in Education and Training International, 32, 175-187.

Hänze, M., & Moegling, K. (2004). Forschendes Lernen als selbstständigkeitsorientierte Unterrichtsform. Persönliche Voraussetzungen und motivationale Wirkmechanismen. Psychologie in Erziehung und Unterricht, 51(2), 113-125.

Huber, L. (1970). Forschendes Lernen: Bericht und Diskussion über ein hochschuldidaktisches Prinzip. Neue Sammlung 10(3).

Li, L., Liu, X., & Steckelberg, A. L. (2010). Assessor or assessee: how student learning improves by giving and receiving peer feedback. British Journal of Educational Technology, 41, 525-536.

Lin, S. S. J., Lui, E. Z. F., & Yuan, S. M. (2001). Web-based peer assessment: feedback for students with various thinking styles. Journal of computer assisted learning, 17, 420-432.

Meihami, H., & Razmjoo, S. A. (2016). An emic perspective toward challenges and solutions of self- and peer-assessments in writing courses. Asian-Pacific Journal of Second and Foreign Language Education, 1.

Min, H.-T. (2006). The effects of trained peer review on EFL students' revision types and writing quality. Journal of Second Language Writing, 15, 118-141.

Pelaez, N. J. (2002). Problem-based writing with peer review improves academic performance in physiology. Advances in Physiology Education, 26, 174-184.

Reynolds, J. A., & Thompson, R. J. jr. (2011). Want to improve undergraduate thesis writing? Engage students and their faculty readers in scientific peer review. CBE – Life Sciences and Education, 10, 209-215.

Sun, D. L., Harris, N., Walther, G., & Baiocchi, M. (2015). Peer assessment enhances student learning: The results of a matched randomized crossover experiment in a college statistics class. PLOS One.

The British Academy. (2007). Peer Review: the challenges for the humanities and the social sciences. URL: https://www.britac.ac.uk/sites/default/files/Peer_Review-report.pdf [06.03.2018].

Topping, K. (1998). Peer assessment between students in colleges and universities. Review of Educational Research, 68, 249-276.

Venables, A., & Summit, R. (2003). Enhancing scientific essay writing using peer assessment. Innovations in Education and Teaching International, 40, 281-290.

Ware, M. (2008). Peer review: benefits, perceptions and alternatives. Publishing Research Consortium, 4, 1-20.

Wen, M. L., & Tsai, C. C. (2006). University students' perceptions of and attitudes towards (online) peer assessment. Higher Education, 51, 27-44.

Yankulov, K., & Couto, R. (2012). Peer review in class: metrics and variations in a senior course. Biochemistry and Molecular Biology Education, 40, 161-168.

Appendix 1: Peer-review instructions for psychology students

Section	Aim / Function	Content	Style
Title	• inform • search • main results? • awaken interest	• central variables • central method • type of article (e.g. meta-analysis) • consistent with structure of the article	• provocative question • (may be lurid) • two parts: catcher + information
Abstract	• overview • inform	• short description of every part	• precise • no abbreviations, no complex terms • max. 300 words
Introduction	• introduce the reader • Why is this study important? • overview of the current research • leads to own research question	• summary of empirical evidence • structure corresponds to the order of the hypotheses • emphasize the relation to the research question • appropriate literature: – sample, quality, new results, amount – presentation of literature is precise and neutral	• one paragraph = one thought • summaries • use and define (technical) terms • short sentences
Current Study / Relevance	• present academic void	• research question • scientific and practical relevance • hypotheses – appropriate amount – statistical verifiability – have to be derived from the text – do not forget to include the control condition in your hypotheses!	• avoid „we“ • active phrasing
Method	• transparency • replicability	• clear structure (1. Procedure, 2. instruments, 3 data analysis) • procedure – experimental / quasiexperimental design – at the beginning of the methods section – variables (dependent, independent) – exact timeline and instructions • sample – age – study timeframe – description of the G*power analysis – voluntariness, anonymity and compensation – inclusion and exclusion criteria • operationalization – selection of instruments according to research question – instruments are listed according to the introduction and the hypotheses – answer formats, amount of items and cut-off values – meaning of high / low values • information about psychometric criteria – economic validity, reliability, validity /	• check the order! • examples of specific items might be helpful

		constructed according to a valid instrument • consider potential confounding variables • analyses – in line with hypotheses – missing data, requirements and effect estimates	
Results	• information	• type of analysis • tables and figures according to APA • descriptive results • short repetition of hypotheses • results structured according to hypotheses	• APA • text and tables should complement and not repeat each other • report statistic values • "dry", no interpretation • statistical symbols / abbreviations in italics • round decimals to the second decimal
Discussion	• meaning of the results • (critical) interpretation	• summary of the results • Which hypotheses can be confirmed and which ones have to be rejected? • scientific integration • alternative explanations • limitations • future prospects – theoretical / practical implications • conclusion	• same order as hypotheses • do not stop with the limitations • parallel with introduction • analytic style; creative, but always in line with the results
Literature			• APA (no abbreviated journal titles, "et al." when more than 6 authors, always proved volume and issue numbers) • complete citations – e.g. software (SPSS, G-Power) • consistent citation style • check the publication dates and the AASM-source

Formal criteria

- font: Times New Roman, size 12
- line spacing: 1.5
- max. 10.000 words (excl. literature)

content

- there has to be a source / citation for every scientific statement
- please use a consistent punctuation (e.g. a dot or a comma as a separator)
- please avoid grammar or spelling mistakes
- use scientific language (e.g. no "we"; neutral, descriptive, consistent tense (in each section), consistent language (German/English))
- abbreviations are introduced, stay consistent and are used consequently
- use paragraphs to structure your article

Appendix 2: Peer-review grading sheet

Please record your assessment for each section in the table below. Use the peer-review instructions as a basis for your evaluation. In addition to the quantitative scale you have to provide a short description for each section: What did you like, and what could be improved?

section	things I liked	things you could improve	rating (0=very bad, 10=very good)
title			0 1 2 3 4 5 6 7 8 9 10
abstract			0 1 2 3 4 5 6 7 8 9 10
introduction ➔ practical relevance			0 1 2 3 4 5 6 7 8 9 10
method			0 1 2 3 4 5 6 7 8 9 10
results			0 1 2 3 4 5 6 7 8 9 10
discussion			0 1 2 3 4 5 6 7 8 9 10
literature			0 1 2 3 4 5 6 7 8 9 10

Appendix 3: Peer-review instructions for life science students

Peer-review instructions for 'Applied Genome Research'

The following instructions can be applied to each of the following parts of the course:

1. Generation and processing of sequencing reads
2. Assembly
3. Annotation
4. Resequencing and variant analysis
5. Gene expression analysis

Section	Aim/Function	Content (Example for Assembly)	Style
Abstract	• overview	• short description of complete content	• precise • no abbreviations • no special expressions • max. 300 words
Introduction	• introduce reader to topic • demonstrate relevance • Current research in field	• Development of assembly strategies over time • Relevance of assemblies for biology • current research articles or at least review articles	• One point per paragraph • Summary of literature • define specific expressions • short sentences • figures to illustrate process
Method	• enable reproduction	• clear description of assembly process • all arguments with explanations • description of assembly evaluation	• Clear description in right order • Detailed description • Examples
Results	• Information	• Assembly statistics e.g. N50, assembly size	• Description in text • Tables for extended data • Figures to illustrate results
Discussion	• Interpretation • Critical analysis • Put in context • Comparison to literature	• Summary of results • Comparison to other NGS assemblies • Improvement suggestions • Conclusion	• ‚limitations' should not be the last point • Parallel to introduction • Based on own results
Literature			• Consistent (select style of one journal) • Use tool for reference management (e.g. zotero)

Style

- Font size 11pt, e.g. Arial (CourierNew for sequences)
- Line spacing 1.5x
- Full justification

Line and page numbering (important for review!)Content

- Reference for all scientific facts
- Check grammar and spelling
- Scientific writing (no ‚we'/ ‚I'; neutral, descriptive, consistent time, consistent language – German/English)
- Abbreviations are explained once and are consistently used in the following text
- Use paragraphs to structure text

Informationen zu den Autor_innen / information on the authors

Anja Friedrich

Anja Friedrich is a clinical psychologist working on her PhD at Bielefeld University. Her research interests encompass sleep disorder diagnosis and therapy, dreams and trauma research. Her fascination with higher education began in 2016 when she started her teaching duties and the Bielefeld Certificate for Higher Education. In her spare time, she practices her didactic skills by competing in science slams.

Boas Pucker

Boas Pucker obtained a Master's degree from Bielefeld University. His academic background is molecular biology and bioinformatics with a focus on plants. He is currently researching at the University of Cambridge, while pursuing a PhD at Bielefeld University.