

## Enhancing Scientific Communication Skills: a Real-World Simulation in a Tertiary-Level Life Science Class Using E-Learning Technology in Biomedical Literature Perception, Reflective Review Writing on a Clinical Issue, and Self and Peer Assessments

Elisabeth Eppler<sup>1,2,3</sup>  • Jan Meyer<sup>2</sup> • Steffen Serowy<sup>3</sup> • Karl Link<sup>4</sup> • Barbara Pauk<sup>5</sup> • Luis Filgueira<sup>2,4</sup>

### Abstract

This educational study aimed to explore the feasibility and acceptance of a literacy exercise adopted from the realworld of scientific publishing in a cell and tissue biology course. For that purpose, a tertiary-level multimodality science course, which integrated a blended learning faculty and student lectures, journal club, and wet laboratory sessions including a research project as well as examinations, was complemented by essaywriting of a review and peerreviewing of five manuscripts. All tasks contributed to the final course mark. Special emphasis was laid on the usability of different E-Learning applications for scientific writing and teacher- and peerassessment procedures. Further, potential influences of student characteristics on their peer- and self-assessments as well as their acceptance of the feedback from their peers were evaluated. Seventy-five undergraduate students from different Bachelor programs were included in the study. Plagiarism check and double-blind assessments of the essays were performed using "Turnitin.com." Students self-assessed their paper and received feedback from five peers and the teacher. Peer assessment was more severe than the teacher- or self-assessment, and the peer mark correlated best with the final course mark. Students with better marks assessed more generously, and there had moderate tendencies for influences of gender and background on peer feedback behavior. The students perceived the writing and assessment exercises, especially being peer-assessed, as demanding, but rewarding and a great learning experience. The additional tasks were feasible using E-Learning technology, which should foster future biomedical courses to train writing skills and the ability to cope with different roles in the scientific community.

**Keywords** Communication skills · Scientific writing · Peer assessment · Self-assessment · E-learning · Blended learning · Plagiarism · Peer review

---

### Electronic supplementary material

✉ Elisabeth Eppler  
elisabeth.eppler@unibas.ch

Science literacy has been recognized as a key element of science education (Osborne 2002; Holbrook and Rannikmae 2007; Crotwell-Timmerman et al., 2010; Hand et al. 2010; Hewson 2010). In tertiary biomedical education, in addition to learning life science content, research protocols and methods, and laboratory and manual skills, students are required to improve their writing skills, become familiar with scientific literature, and be able to read and understand research articles. Furthermore, essay writing is often used in exams to support learning and understanding scientific concepts (Ellis et al. 2007).

Increasing numbers of educational studies have positively evaluated the value of peer assessment in science teaching and learning (Gielen et al., 2011; Nulty, 2010; Poon et al. 2009; Topping 2009; Venables and Summit 2003; Vu and Dall’Alba 2007; Xu et al. 2016), which could even be enhanced by combination with peer feedback (Cartney 2010; Liu and Carless 2006; Tseng and Tsai 2007; van den Berg et al. 2006a, b) and self-assessment (Eppler et al. 2018; Wenzel 2007). In short, peer assessment is well-established in science teaching, but coping with different roles in academic writing is still challenging for authors, reviewers, and editors (Xu et al. 2016), which applies even more to learners (Sidalak et al. 2017), also in health-related training (Galipeau et al. 2015). Thus, limited data exist on teaching how to write and peer-assess a scientific paper in a biomedical course, at the undergraduate (Weaver et al. 2016) and graduate (Guildford 2001; Holstein et al. 2015) level, and even less on potential influences of students’ characteristics on peer assessments and their own perception of being assessed.

While the workload of peer review was high for the instructors at the beginning of the millennium (Guildford 2001), modern E-Technology applications are in the march to reform the established processes of academic writing, peer review, and publishing (Sidalak et al. 2017). Nevertheless, their usability in biomedical curricula has not been systematically explored to date. On the one hand, E-Learning is gaining increasing attention for “gamification” (McCoy et al. 2016), and on the other hand, using E-Learning to align teaching with “authentic science” has been proposed (Olitsky et al. 2018). In both applications, real-world simulations intend to evoke emotional responses, which reportedly stimulate students’ learning experience and engagement in science (King et al. 2015).

As outlined above, various authors have identified peer- and self-assessment and peer feedback as valuable science literacy exercises at the secondary and tertiary level, particularly in biomedical education (Eppler et al. 2018; Holstein et al. 2015; Perera et al. 2010). Nevertheless, few courses have been conducted under “real world” conditions with individual students acting as authors undergoing an anonymous peerreview procedure or as reviewers (Guildford 2001; Holstein et al. 2015; Kwon et al. 2018; Weaver et al. 2016). Gender-related differences have been recently emphasized in peer-assisted study sessions (Geerlings et al. 2016); however, as far as known to the authors, such real world writing courses have not addressed to date individual students’ characteristics such as gender and cultural origin when acting as authors, peer- and self-assessors, nor did they systematically integrate E-Learning elements including plagiarism check.

This educational study aimed to explore the feasibility and learning value of a literacy exercise adopted from the realworld of scientific publishing in a cell and tissue biology course with established tools of blended learning. Special emphasis was laid on guiding all processes of scientific writing; awareness of plagiarism issues; timely submission of a manuscript and the process of peerreviewing, both in the roles as an author and a reviewer; and the usability of different E-Learning applications for these processes. The study investigated the outcome of peer-, self-, and teacher assessments and the potential influence of student characteristics on

the assessments as well as the students' acceptance of the feedback from their peers. Finally, this teaching approach, which has meanwhile been applied by the authors in diverse biomedical science courses in Australia and Switzerland, is discussed regarding its feasibility for medical curricula, since these are proposed to have an increasing demand for science literacy training, but limited faculty and time resources (see below).

## Methods

### Study Group

Seventy-five science students from different Bachelor programs, enrolled in a cell and tissue biology Bachelor unit (ANHB3313) with the senior author (L.F.) as unit coordinator and principal instructor, participated in this study. An established third-year undergraduate theoretical and practical course was modified. The only essay-like work included in the previous course was a report on their laboratory-based project. In the new course, the participants had the opportunity to additionally train their skills in literature search, reading and writing of scientific papers, predominantly by blended learning using diverse E-learning applications for teaching and a real-world simulation of a peerreview process. They were informed about the study at the beginning of the semester and all students agreed to participate. Students were thoroughly advised (Supplementary material 1) in the course manual (Filgueira 2010a) that in the real world of publishing, scientific manuscripts undergo a stringent peer review before acceptance for publication and that the current exercise mimics that process. They were informed that essays would be evaluated through double-blind assessment and that coping with their peers' anonymous comments may be a challenge, but prepare them for the real academic world. It was emphasized that receiving comments from six reviewers (five peers and the teacher) would form a solid basis for diversity in feedback. The study was conducted in accordance with the ethics rules and regulations at the University of Western Australia. All students participating in the course gave their written informed consent to participate in this educational research project.

### General Unit Organization, Course Contents, and Teaching Methods

For the present study, the course consisted of faculty lectures ( $2 \times 1$  h/week, over 12 to 13 weeks), tutorials ( $1 \times 3$  h/week, over 8 weeks), and laboratory sessions ( $1 \times 3$  h/week, over 6 weeks), as well as 25 h allocated in the time table for essay writing, peer review, and editing. Six UWA points were accredited (Filgueira 2010a) corresponding to 25% of a semester workload of 24 points. This corresponds to 30 ECTS (European Credit Transfer and Accumulation System), i.e., about 6 contact hours and 8–12 h total workload per week during the 13-week semester (UWA 2014).

To progress well, essential prerequisites on course entry were basic knowledge of cell biology, histology, and tissue structure, and additional reading was recommended to students with no or deficient background (Filgueira 2010a). No specific knowledge and understanding about high-quality scientific communication were required. However, based on the UWA's education principles, certain attitudes were stressed such as "(...) awareness of the great overlap between many disciplines (both biological and others) and the need to communicate openly and share skills and knowledge (...), the need for an open and imaginative approach

to problem solving, (...), a rigorous and critical approach to experimental design and interpretation. (...). Active participation in all class activities and particularly in group activities. It relies on vigorous and rigorous discussion and careful examination of all points of view” (Filgueira 2010a). Further aims were stimulating field-specific and generic skills such as managing tasks, solving problems, working with others, and communication (Filgueira 2010a).

Faculty lectures focused on scientific topics such as imaging of structure and function in cells and tissues, tissue formation, growth factors, stem cells, disease, tissue engineering and repair, developmental biology including molecular mechanisms and epigenetics, angiogenesis, genetically modified animals, host genetics, viral infections, ethics, and methodology (Filgueira 2010a). Lectures were recorded through Lectopia ([www.lectopia.com.au](http://www.lectopia.com.au), n.d.) and made available online via the teaching and learning administration program WebCT/Blackboard ([www.blackboard.com](http://www.blackboard.com), n.d.), which was familiar to the students from preceding courses.

During tutorials in the first weeks of semester, students were introduced to scientific skills including biomedical literature search using PubMed and Web of Science, to peerassessment and scientific writing. Complementary literature on presentation and communication skills was recommended, e.g., Exley and Dennick (2009) and Harmon and Gross (2010). Students were instructed by university library staff in the reference administration program Endnote (<https://endnote.com/> n.d.). During the laboratory sessions, students were introduced to safety measures and cell biological methods in a rotation system covering “cutting,” “staining,” “immunostaining,” and “cell culture,” alternating with journal club sessions. In a group project, they acquired firsthand laboratory experience in tissue preparation and analysis combined with organizational and presentation skills (Filgueira 2010a).

### Overview on Course Tasks and Assessments

A course manual (Filgueira 2010a) with the topics (Supplementary material 2), tasks, procedures, timeline, and assessments (Table 1) was presented during the first week of semester. The tasks included student group mini-lectures on assigned topics (Table 1), which were peer- and teacher-assessed using a structured questionnaire as previously described (Eppler et al. 2018). In brief, clear and concise language, appropriate and logical flow of the lecture were reviewed as well as clarity and structure of the PowerPoint slides and figures. Special emphasis was laid on interest of the lecture, appropriate coverage of the content, and referencing.

During the weekly journal club sessions, groups of 2–3 students chose one out of three provided original articles, read, summarized, and critiqued it according to criteria taught during the tutorials, and presented it to a small group of peers. Also, this task was self-, peer-, and teacher-assessed (Filgueira 2010a). In the laboratory class, teams of 2–3 students had to deliver a written report and an oral presentation on a practical research project. In total, the team work tasks contributed 30% to the final mark.

Individual tasks were a multiple choice mid-semester examination test and a final theory examination at the end of semester (Table 1). All students had to write an essay in the form of a review article, which was assessed by the teacher using a scale (Table 2) ranging from 15 points (accepted for publication) to 0–5 points (insufficient and rejected) and by five peers according to similar criteria (Supplementary material 1). The final essay mark, which contributed 15% to the final course mark (Table 1), was calculated as the average of the teacher mark on the one hand and the mean peerassessment marks on the other hand (Table 2). This resulted

**Table 1** Assessment weighting and items

Assessment weighting and time schedule		
Component	Weight (%)	Due date
Journal club group presentation	5	1st part of semester
Mini-lecture <sup>a</sup>	5	1st part of semester
Essay (review journal article)	15	Mid semester
Essay evaluation (peer review)	5	Mid semester
Mid semester test	10	Mid semester
Project report and presentation	20	End of semester
2-h theory examination	40	End of semester
Assessment items		
Item title	Description	
Journal club (weeks3–5)	Twenty minutes critical presentation of an original research article	
Mini-lecture (weeks2–6)	A lecture of 20 min will be delivered about a given topic <sup>a</sup> in a group of 3 students. The content and delivery will be assessed by peers and teacher	
Mid semester test (week 7)	Multiple Choice Question (MCQ)-based test about the material from lectures and group presentations	
Essay (review journal article) (weeks1–8) and evaluation (peer review) (weeks9–10)	Based on a given topic and at least five recent original research papers in high ranking journals, a review manuscript will be submitted in the format of either Cell, Nature, or Science. The manuscript will be reviewed and assessed by five peers doing ANHB3313, but writing on a different topic	
Project report and presentation (week 13)	Based on a given project, the students will present their results, similar to a conference paper, and write a report, similar to an original journal article in the format of a peer review international journal	
2-h theory examination (examination period)	Short answers and essay-like examination assessing material from lectures, tutorials, and laboratories	

<sup>a</sup> Mini-lecture topics: 1. Stem cell—cancer cell link: are cancer cells just stem cells out of control? 2. Caloric restriction—the solution for a longer life span? 3. Circadian rhythm—influence on liver function. 4. Inflammation—the main culprit of aging? 5. Stem cells for the regeneration of the brain—do they exist and where are they located? 6. Viagra and cialis—what are these drugs and what effects do they have on different cell types? 7. Effects of chronic HIV infection on bone and cardiovascular system. 8. Self-repair of heart damage—how does it work and how could it be improved through biomedical interventions? 9. Parathyroid hormone—does it cause or does it help with osteoporosis? 10. The human embryonic trophoblast invades the maternal uterine tissue like cancer if not controlled—what mechanisms regulate its controlled invasion?

in the teacher mark having the highest individual weighting. All students had to review five essays from their peers, for which they could receive a maximum of  $5 \times 1$  point/per completed assessment, which contributed 5% to the final mark ([Supplementary material 1](#)).

### Essay Writing, Plagiarism Awareness, and Peer Review Process—Task, Instructions, and Use of E-Learning Tool

Students were provided a list of 10 essay topics related to the course content (1. Wound healing and tissue repair; 2. Hair—to grow or not to grow; 3. Why do children grow—but stop growing when reaching a certain age? 4. Muscle—we like it as food, we need it to move—what else? 5. Brain—infection and inflammation; 6. We are what we eat—influence of food on the human body; 7. Our senses and how we perceive the world; 8. Symbiotic coexistence of humans and other organisms; 9. Ethnicity and disease; 10. Animals in biomedical research). At first, students had to choose a topic and look for at least five related original articles published

**Table 2** Calculation of the essay assessment marks

Assessment	Individual points to achieve	Criteria	Assessment mark
From teacher	15 points	Accepted for publication	Maximum 15 points
	10–14 points	Minor revision required	
	6–9 points	Major revisions required	
	0–5 points	Insufficient and rejected	
From 5 peers	3 points	Essay covered all criteria, excellent and highly recommendable for publication without changes	Average from the individual peerpoints × 5 = maximum 15 points
	2 points	Good essay recommended for publication with minor changes	
	1 point	Essay not recommendable for publication without major revision	
	0 point	Article of very poor quality, difficult to read, does not make sense	
Self	15 points	Same scale as the teacher	0 points

in the preceding 10 years in the journals *Aging Cell* (© John Wiley & Sons Ltd. and The Anatomical Society); *Cell* and *Cell Stem Cell* (© Elsevier Inc.); *Journal of Cell Biology* and *Journal of Experimental Medicine* (© Rockefeller University Press); *Journal of Immunology* (© The American Association of Immunologists, Inc.); *Nature*, *Nature Cell Biology*, *Nature Medicine*, *Nature Neuroscience*, and *PNAS* (©Macmillan Publishers Ltd.); and *Science* (© 2018 American Association for the Advancement of Science). These five articles had to be submitted in semester week 4 (Table 1) through WebCT/Blackboard for the first line of feedback from the teacher regarding their appropriateness as a basis for the essays. Further, as a template, students were provided with related review articles published in *Nature*, *Science*, or *Cell*. The students were permitted to use as many other appropriate sources of scientific information as required. Some sources were recommended, such as Alberts et al. (2009), Gilbert (2010), Kiernan (2009), Kierszenbaum and Tres (2011), Lodish et al. (2007), Mescher (2009), Pollack (2001), Ross et al. (2006), Sunderland (2000), and the in house manual and examination paper “Cell and Tissue Organisation ANHB3313” (Filgueira 2010a, b). Complementary to the tutorials and advice in the course manual (Supplementary material 3), literature on scientific writing in biomedicine was recommended such as Truss (2003), Glasman-Deal (2009), Hall (2011), Hewson (2010), Hofmann (2010), Lindsay (2011), Matthews and Matthews (2007), and McMillan (2011) as well as additional websites, e.g., <http://www.lab.anhb.uwa.edu.au/mb140/> n.d. and <https://www.nih.gov/health-information> n.d.. Students were also familiar with the recommended IRIS online search tool (Supplementary material 3) provided by the University library. Essays with a given length and structure (Supplementary material 3) had to be submitted electronically through the text originality analysis and rating program Turnitin (<https://turnitin.com> n.d.) until semester week 8, however, could be resubmitted and rechecked for plagiarism as many times as required, up to the deadline. The Faculty of Science has implemented Turnitin into E-Learning for students for many years from the first year on for all essaywriting activities. Each submitted anonymous essay was distributed by the teacher through Turnitin to five peers who covered a different topic. Students were advised to evaluate, give feedback, and rank the five manuscripts (first =

best, fifth = worst) by Turnitin, which allows pdf-format highlighting and inserting comments, corrections, and suggestions, according to given guidelines (Supplementary material 1), in brief: 1. Five or more original research articles used? 2. Do the original research articles and other supporting information address the topic? 3. Was the required format applied? 4. Is spelling and language of high scientific quality? 5. Do introduction, main body text and conclusion flow well and does the author tell a nice and coherent story? and 6. Does the abstract summarize the main body text? Students were asked to assign points to each manuscript by a scale with a maximum of 3 points (Supplementary material 1). As each essay was assessed by five peers, it could achieve a maximum of 15 points from the peers (Table 2).

### Qualitative and Quantitative Study Data

The overall course outcome with regard to student marks was compared with experience from preceding courses. Student data, including gender, status of enrolment (either a local student from Australia or New Zealand or an international student) were collected. Student marks for the different tasks and summative final mark for the course (high distinction 80–100%, distinction 70–79%, less than distinction < 70%) were recorded. Data analysis was focused on the written essays and their assessments. After the peerreview exercise, but before receiving their marks, the students had the opportunity to self-assess their essay using the same scale as the teacher (Table 2). This self-assessment did not contribute to the essay and final course marks, but was evaluated in the study. The peer-, self-, and teacher assessment marks of the essays were evaluated for correlations with gender, enrolment status, and individual final marks of the students in their roles as authors and reviewers.

E-Learning applications such as Turnitin were explored for their feasibility for efficient submission and distribution of the essays. Plagiarism ratio was measured on Turnitin whereby a ratio > 25% was considered suspicious and students had to discuss this with the teacher and revise their essay before their final submission. Student feedback on the peer assessment experience was evaluated by an online questionnaire (Table 3), which was created based on experiences from previous evaluations (authors' unpublished data). To assure independence of the students' feedback from the teaching, the questionnaire could be voluntarily submitted after the final examination only after cessation of the course. During the ongoing course, students could put questions, answers, and comments on a blog established on WebCT/Blackboard, and via E-mail to the teacher. Interactions were recorded by the teacher (L.F.), and if necessary, answered. After cessation of the course, qualitative selection, shortenings, and grammar corrections (for better readability) without change of the message and interpretation of the statements from all students who either answered in the open comment field of the online questionnaire or in the course blog and by E-mail were performed by a co-author who did not act as supervisor in this course (E.E.).

### Statistics

All data were analyzed using GenStat (VSN International, Hemel Hempstead, UK). Quantitative variables were expressed as percentage or marks/points, respectively, and used for calculation of average marks  $\pm$  standard deviation. The Shapiro-Wilk test was used to check for normality and Pearson's correlation coefficient ( $r$ ) was calculated. A correlation  $r \leq 0.39$  was considered as weak,  $0.4 \leq r < 0.6$  as moderate, and  $\geq 0.6$  as strong. A two-tailed  $p$  value < 0.05 was considered to be statistically significant.

**Table 3** Online questionnaire

I felt comfortable to assess	- a mini-lecture of a peer.	<input type="checkbox"/> Yes	<input type="checkbox"/> No
	- a paper presentation of peers.	<input type="checkbox"/> Yes	<input type="checkbox"/> No
	- an essay written by a peer.	<input type="checkbox"/> Yes	<input type="checkbox"/> No
Was the assessment of your	- mini-lecture fair?	<input type="checkbox"/> Yes	<input type="checkbox"/> No
	- paperpresentation	<input type="checkbox"/> Yes	<input type="checkbox"/> No
	- essay	<input type="checkbox"/> Yes	<input type="checkbox"/> No
The criteria for the assessment of the essays were helpful and adequate.		<input type="checkbox"/> Yes	<input type="checkbox"/> No
The reading and assessing of other essays improved my capacity to write an essay.		<input type="checkbox"/> Yes	<input type="checkbox"/> No
The listening and assessing of other mini-lectures improved my capacity to deliver a lecture.		<input type="checkbox"/> Yes	<input type="checkbox"/> No
The listening and assessing of paper presentations improved my capacity to present a paper?		<input type="checkbox"/> Yes	<input type="checkbox"/> No
The process of assessing the work of my peers as well as my own work has induced me to reflect on strategies in learning and presenting concepts?		<input type="checkbox"/> Yes	<input type="checkbox"/> No
The fact that my essay was peer reviewed has motivated me to put more effort into it.		<input type="checkbox"/> Yes	<input type="checkbox"/> No
What are the main advantages of peer assessment:			
What are the main disadvantages of peer assessment:			

## Results

### Students' Characteristics and Final Unit Marks

Seventy-five students were included in this study, 53 (71%) females and 22 (29%) males. The majority (85%) were local/standard students from Australia or New Zealand ( $n = 64$ ), and a smaller group of international students ( $n = 11$ ). Altogether, the final marks were in a similar order of magnitude (average  $76.9 \pm 9.0$ , range 43–90 points) as in preceding courses (data not shown). As final unit marks, out of 74 students (see below), 49% (25 female = 70%, 11 male = 30%) achieved high distinction, 39% (20 female = 69%, 9 male = 30%) distinction, and 12% (7 female = 78%, 2 male = 22%) less than distinction. The subgroup of international students scored on average significantly lower ( $p = 0.019$ ), which is similar to previous cohorts and evaluations in a similar study subject area at our Faculty of Science (authors' unpublished results).

### Course Assessment Marks

In the mini-lectures, on average, 4.6 (out of 5.0, 92%)  $\pm 0.3$  points were accredited and in the journal club presentations 4.5 (out of 5.0, 90%)  $\pm 0.3$  points. In the laboratory group project presentations, students achieved on average  $6.9 \pm 0.7$  points and for the research project report  $9.1 \pm 1.7$  points out of 20.0 points (80%) collectively (Table 1). In the mid-semester test, on average, 6.1 (out of 10.0, 61%)  $\pm 1.1$  points were achieved, and in the final examination 29.6 (out of 40.0, 74%)  $\pm 6.2$  points. All students collected 5 points for the peer review process, and for the essays, on average 11.5 (out of 15.0, 77%)  $\pm 2.3$  points were achieved from the peers' and teacher's assessments collectively (Table 2).

## **Analysis of the Submission and PeerReview Process and Plagiarism CheckUsing E-Learning**

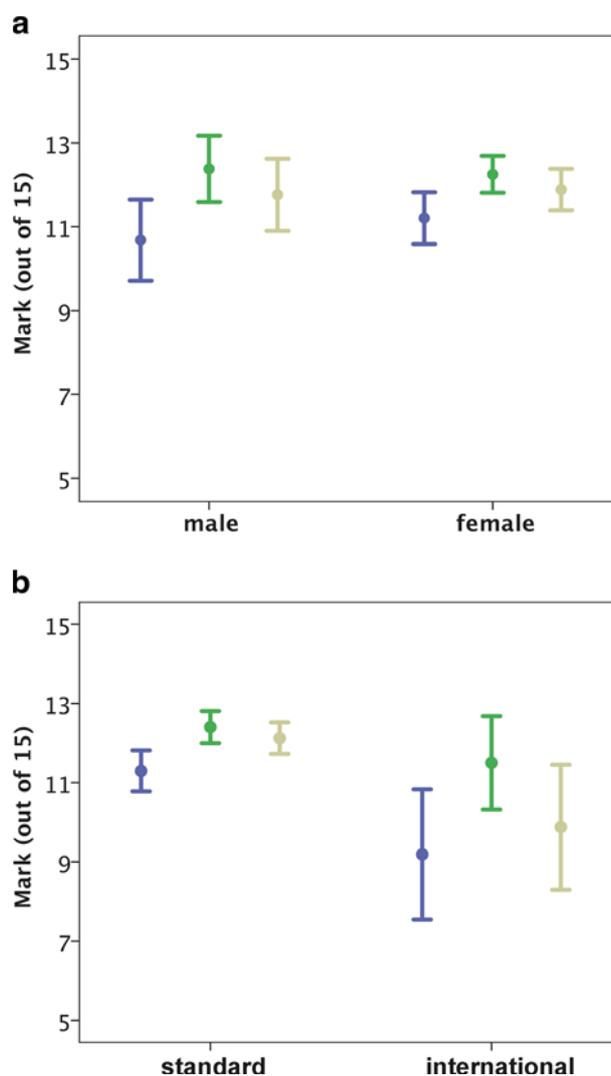
The writing and peerreviewing tasks were easy to implement within the established framework of blended learning. Modern E-Learning tools such as WebCT/Blackboard and Turnitin were feasible for the administration of the peerreview process, including distribution of assignments and tasks from the teacher to the students, online submission, and anonymous distribution of students' written work to peers as perceived by the teacher. All students reviewed five essays of their peers and achieved 5 points for this task. One student was not able to submit the essay, but participated in the peerassessment process. One student's essay was not peer-assessed due to technical problems in the submission process. These two students were therefore excluded from some analysis, which was performed with  $n = 74$  or  $n = 73$ , respectively. Sixty-six students also used the opportunity to voluntarily self-assess their essay.

Turnitin was efficient in identifying students with a tendency towards plagiarism. In their first drafts, 20% of the students (15/74) had plagiarism ratios  $>25\%$ , which was considered significant. Students were required to revise passages that were too close to published texts used as references in their essay text. They were also made aware of the seriousness of this issue by the teacher who used Turnitin results as a basis for discussion. The final submitted text was checked again for plagiarism, which was insignificant for all essays.

### **Analysis of Peer-, Self-, and TeacherAssessments**

On average, in the essays, 11.5 (out of 15.0) points were achieved from the peers and the teacher collectively, whereby the teacher mark was weighted equally with the averaged marks from the peers (Table 2). The self-assessments resulted in the highest average marks (Fig. 1a,b), and the teacher's marks were slightly, but not significantly, lower, whereas peer-assessed marks were significantly lower than the teacher's and self-assessed marks. Gender had no significant influence on the essay marks. The subgroup of international students graded on average significantly lower (Fig. 1b) in the teacher's ( $8.4 \pm 3.7$  vs.  $12.0 \pm 1.6$  points) and peers' assessments ( $8.8 \pm 2.1$  vs.  $11.1 \pm 2.0$  points) of the essays than the standard students. While they marked their own essay lower than did the standard students ( $11.7 \pm 1.4$  vs.  $12.4 \pm 1.6$  points), their self-assessed marks were significantly higher than the teacher's ( $p = 0.013$ ) and peers' ( $p = 0.012$ ) marks. Nevertheless, the peer-, self-, and teacher-assessed marks significantly correlated with the final course marks, and among these, the mean peer marks (Fig. 2a) on average correlated best ( $r = 0.47$ ) with the final course marks (Fig. 2b). This is similar for the peermarks of the international and standard-enrolled students, however at different levels (Fig. 2c). When individually analyzed (Fig. 3a), the peermarks for the written essays were moderately lower than the teacher mark (both maximum of 15 points) with the majority ranging from  $-1$  (peer mark  $<$  teacher mark) to  $+0.5$  (peer mark  $>$  teacher mark).

The essay marks distributed by the students in the role of assessors were compared with their final course marks, gender, and enrolment status. Females tended to assess more generously, but this trend was not significant (data not shown). However, there was a significant positive linear relationship between final course marks of assessors and average essay marks, which they provided to their peers; for instance, students with high distinction provided on average 2.33 points  $\pm 0.02$  (out of 3 possible points, 78%) as compared to students with distinction (2.08 points  $\pm 0.03$ , 69%) and students with less than distinction (2.01 points  $\pm 0.07$ , 67%). International students tended to be significantly more severe in

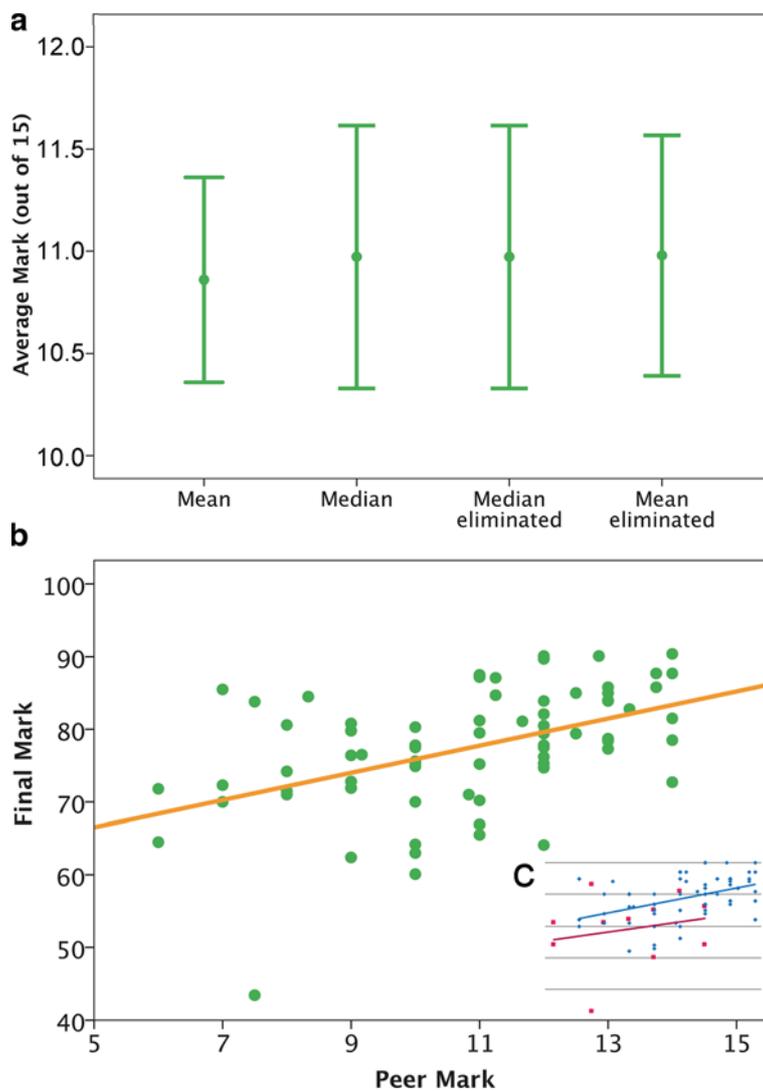


**Fig.1** Average marks out of a maximum of 15 points (Y-axis) achieved for the written essay by **a** male and female authors (X-axis) and **b** local/standard students and international students (X-axis), respectively. Blue line: peer mark averaged from the individual peer marks (maximum 3 points)  $\times$  5 peers = maximum 15 points. Green line: self-assessed mark, which did not contribute to the essay marks. Yellow line: teacher-assessed mark, which weighted as high as the peer-assessed marks collectively. Bars show 95% confidence interval

their assessment (1.80 points  $\pm$  0.03, 60%) as compared to their peers with standard enrolment status (2.26 points  $\pm$  0.02, 75%), even when corrected for having on average lower marks as a subgroup.

### Students' Perception of the PeerReview Process

After cessation of the course, the anonymous online standard-questionnaire (Table 3) was returned by all students, mostly with very short positive comments. Only five students added

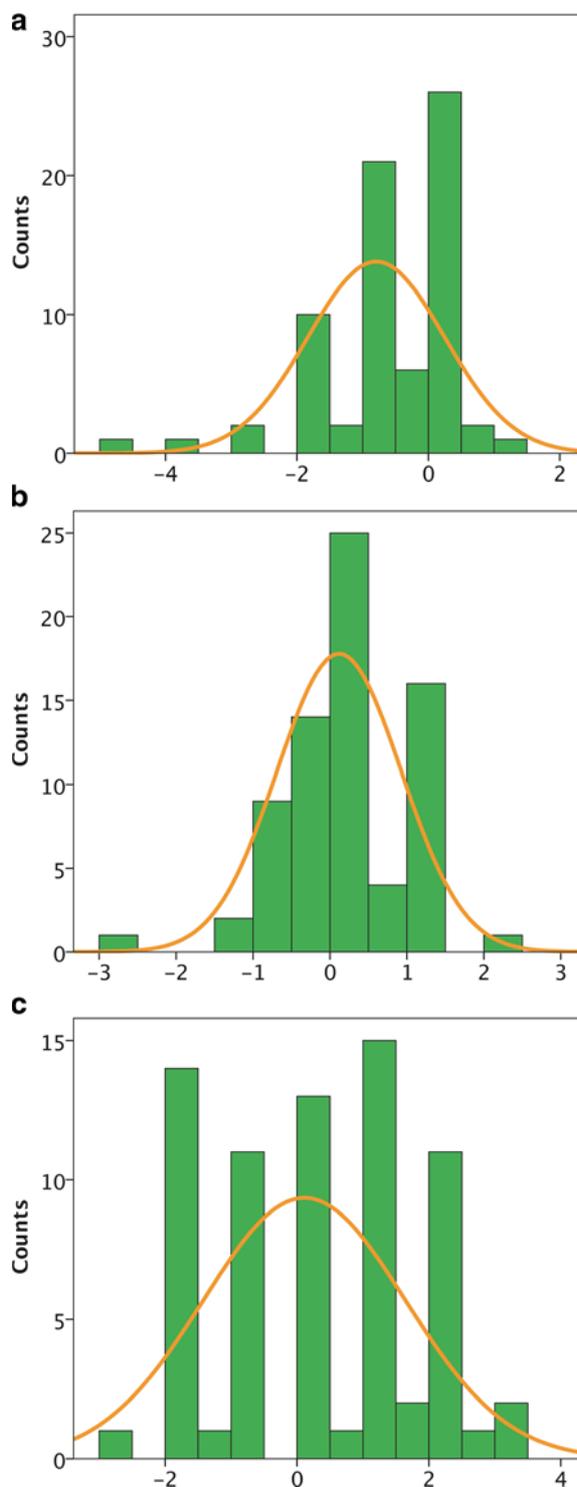


**Fig.2** Analysis of the peer assessment marks achieved for the written essays. **a** Average marks out of maximum of 15 points (Y-axis) calculated as mean of the individual marking (maximum 3 points)  $\times$  5 peers = maximum 15 points as applied in the course (X-axis). Alternative calculations revealed no significant difference as compared with the median, the median after elimination of the maxima, and the mean after elimination of the maxima. **b** Correlation of the mean marks achieved by the students from the peer evaluations (X-axis, maximum 15 points) with the final unit mark (Y-axis, maximum 100%). **c** Insert depicts similar correlations separated for local/standard (blue circles) and international students (red squares), respectively

more critical detailed comments to the question “What are the main advantages/disadvantages of peer assessment?” (Table 3):

“But it doesn't matter what mark I get (...) I would love to get a great score but I did really enjoy writing the essay. I learnt a lot about my topic, which was really interesting. I just hope that I portrayed that in my essay. Thank you”

**Fig.3** Frequency analysis (*Y*-axis) of the differences (*X*-axis) in marking for the written essays. **a** Calculation of individual peer mark (maximum 3 points  $\times$  5 peers = maximum 15 points) minus the teacher mark (maximum 15 points, see Table 2). **b** Calculation of the mean peer mark (maximum 3 points  $\times$  5 peers = maximum 15 points) after elimination of the maxima minus the original mean peer marks. **c** Calculation of the median peer mark (maximum 3 points  $\times$  5 peers = maximum 15 points) minus the mean peer marks



“I found this exercise very helpful. It was nice to be able to read fellow students' work and to look at a large range of topics.”

“Many students were too heavily influenced by (...) incorrect presumptions about what to expect from an article of this nature and it showed when they reviewed (...). At the same time it was terrific to have to accept feedback from your peers because this is obviously very, very relevant to everyday life once we enter the real world of science. Even with its flaws, I found this to be one of the best assignments I have done at uni because of how it has prepared me for post grad life.”

“(…). I think that it was a good exposure for us to review (...), but the feedback (...) were too different to be conclusive. Sometimes they even conflict with each other. I think many of us got marked down just because our essays were not written to the preferred style of the reviewer, which i do not think is objective marking at all. (...) even though the peer-review was a good exercise, (...) our marks should be determined by an experienced professional, instead of it being affected by individuals with little or no experience (...) Although you did mention that you moderated the articles, I do not know how much of an impact the comments of the previous reviewers had on your revision of the articles. (...) to have our grades partially determined by our peers with varying opinions is a little discomforting. Maybe it would be better to have the same exercise, but without having the scores obtained from peers as part of our final marks. I think most of us would be more honest with our markings as well.”

“This peer assessment was indeed a fun experience and something new that i've never done before in my life.”

Further, using E-mail and predominantly the course blog, 23 individual statements were collected during the ongoing course. These were generally positive and students agreed that this exercise was a great learning experience:

“I really think self evaluation is such a good idea”

“(…) now that I have been in the reviewers shoes (...), its easier to pick out the positives and negatives of my essay.”

Generally, students accepted the received peermarks:

“I don't fully support some of the comments (...), however I believe that I do still have some aspects I can improve on, thus I'm willing to generally accept the marks (...).”

“I think the marks I got were okay in the peer review activity.”

Only some of the excellent performing and some international students were disappointed with the feedback and marks they received or skeptical towards the peerreviewing process and its influence on the course mark as they expressed during the course in the blog:

““I don't wish to whine about my marks but some of the comments I received were ridiculous - some markers seem to consider themselves more knowledgeable about my 5 articles than I myself.”

“My comments are totally contradictory (...)”

“In one of my reviews, I had someone write 2 comments.... One to say I get a grade of 2 for not having a good format (...) and my title(...) set out wrong.”

Some students were even frustrated about the harsh feedback and reacted very emotionally as was further reflected by these (shortened) vivid discussions on the course blog:

“It is incredibly frustrating to be criticized by an imbecile, (...) having your sentences marked as "too long" or "difficult to read" simply because large sentences overload the brain of your marker.”

“Oh I know exactly what you mean (...). It really frustrates me when (...) people have a read of my article, then say it was “difficult to understand” (...) They expect to understand my review after reading it once, when I spent more than a week trying to understand all the articles myself??”

“HA, exactly! (...) Oh and apparently my abstract wasn't very eye-catching. I mean, come on, how many exciting abstracts have you read?? The purpose of the abstract is to convey the main results, which I did!”

“I was criticized for (...) referencing (...) That's no reason to nearly fail someone.”

Nevertheless, some revealed a strong sense of humor or weakened their critical opinions during the blog discussions:

“Ha ha (...), there are no "sour grapes" here with the marks i received. :)”

“my own way of (...) criticism towards overconfident reviewers who seem to feel a rush of power when protected by the anonymity of this review process.”

“(...) hypocrisy of criticizing someone's grammar with a sentence that has many spelling and grammatical mistakes itself (...) was amusing (...)”

“Politically incorrect yes, but I guess I'm just a trouble-maker.”

“Haha I know what you mean. How can you possibly get 3/3 from 4 (...) and then have 1 person give you a 1/3 for not writing grammar (...)”

“i know i got some pretty odd and weird comments (..) i only LOL. literally.”

Nevertheless, in this vivid blog discussion, some students took an opposite position:

“the length of a sentence does not just "overload the brain of your marker", it also disrupts the flow (..) the marking scheme was based on whether the article was ready for publishing. do you think an article (...) with little flow, is ready for publishing? i certainly don't.”

“(...) it's refreshing to have 5 new pairs of eyes looking at my article. it gives you a new perspective on your writing. okay, so maybe the marking is inconsistent, but that's just a difference of opinion between people.”

“People need to relax. It is only a review process...”

Repeatedly, the role of the teacher as an experienced and objective reader was emphasized, which mostly consolidated the concerns of the students regarding their marks:

“(...) a deep breath is needed here...given that it was probably everybody's first time (...) i think we should all cut each other a little slack. Let's not take anything to heart, at least there is an academic, who marks and evaluates (...) professionally, (...) who will be balancing this all out and giving our marks to us.”

“(...) People didn't know how to mark correctly, but at the end of the day i hope (...) the final mark will be based on assessment by an academic.”

“Nonetheless I'm sure Luis will look at large disparities and fix things as they pop up/are complained about. My rant is now over (...)”

“(...) that's why we got Luis to make sure everything is marked fairly in the end, right?”

## Discussion

Overall, the course was very efficient and students highly valued the course elements. The newly introduced tasks of writing and peer review did not significantly affect the final marks as compared to previous courses (authors' unpublished data), probably because the total workload was kept at a comparable level. The students perceived the peer assessment process as a great learning experience. Particularly, reading texts written by peers was greatly valued as well as critical analysis of peer texts and providing written feedback. Some students emphasized that reviewing five essays was very time-consuming and demanding. Thus, in the subsequent year, the contribution of this task to the final mark was increased from 5 to 10%, and of the essay writing even to 30%. Interestingly, in a previous study (Guildford 2001), students did not report substantially improved writing skills, so that the author concluded that it was more the effect of the manuscripts undergoing reviewing than the process of writing itself, which enhanced their quality. This finding supports the necessity to stimulate writing skills additionally, e.g., in a multimodal course by defined guidelines. However, Guildford emphasized, as most striking effects, enhanced understanding of the scientific publishing process and writing in a professional style and format of a "major" journal (Guildford 2001).

Implementation of more E-Learning tools into the blended learning structure of the course concept was easy and feasible, probably because the students were already familiar with them. While the workload of peer review for the instructors was high at the beginning of the millennium (Guildford 2001), use of E-Learning tools facilitated electronic submission, peer review, and plagiarism checks. Turnitin was suitable for the administration of the peerassessment process, including individual distribution of assignments and tasks by the teacher to the students, which is similar to previous findings (Chen and Tsai 2009; Tseng and Tsai, 2007), particularly for double-blind anonymous distribution of students' written work to peers as well as providing feedback and marks. Confidentiality and anonymity are crucial (Guildford 2001), and our established concept that all students review their peers' work (Eppler et al. 2018) avoids effects such as advantageous personal scores reported for those students who assessed others (Schönrock-Adema et al. 2007).

According to studies investigating plagiarism at the tertiary level in Australia and other countries, one would expect that a certain number of students would not comply with the well-defined rules of academic honesty (Brimble and Stevenson-Clarke 2008; Maxwell et al. 2008; Wagner et al. 2009). Those studies have therefore proposed that checking and control for plagiarized paragraphs and sentences are the best measure to prevent plagiarism and educate students. We consider as important didactic advantage of this E-Learning setting that Turnitin not only simplified control of the submitted essays, but also trained students to become progressively aware of this issue. By the possibility of checking and re-checking submitted essay drafts, a dynamic learning process was initiated. Students not only realized when there was too much similarity to published texts, but also benefited from every improvement in correct referencing as was perceived by the teacher. Thus, self-check and control, and allowing students to revise their texts, made them more aware of their attitude towards academic honesty. Furthermore, discussions with the teacher about their wrongdoing based on Turnitin results may be even more efficient than check for plagiarism alone, which is well in agreement with the reported importance of personal feedback by the teacher in addition to peer feedback (Guildford 2001; Holstein et al., 2015).

While the teacher's marks were slightly, but not significantly, lower, the peer marks were significantly lower than both, teacher and self-assessed marks. The latter were the highest average marks, which is similar to other reports (Inayah et al., 2017; Leach 2012; Strong et al. 2004). Our observation that peers were more severe in their judgments than the teachers was different to some studies (Inayah et al., 2017; Kommalage and Gunawardena 2011; Leach 2012; Strong et al. 2004), but similar to other reports (Albanese et al. 1991; Eppler et al. 2018; Lurie et al. 2006); in particular, medical students with low levels of peer-assessed interpersonal attributes were proposed to be more negative in their judgments of classmates (Lurie et al. 2006). In the present study, some students were skeptical with regard to the peer marks and felt affected from the harsh criticism by their peers, which is reflected by the emotional course blog discussion. Also, in another study from Australia, students disliked when peers marked their work and these marks were included in their final mark (Moore and Teather 2013). Nevertheless, in the present study, the peer marks correlated best with the final course marks, even though this educational exercise only contributed 15% to the final mark.

Remarkably, the best marks were achieved for teamwork activities. Team-based learning improved students' overall performance (Poon et al., 2009; Topping, 2009; Van den Berg et al., 2006a,b; Vu and Dall'Alba 2007; Wenzel, 2007), which was particularly noted for low-scoring students and heterogeneous classes (Nieder et al. 2005; Pizzimenti et al., 2016; Spandorfer et al., 2014; Vu and Dall'Alba, 2007). This effect may result from collaborative efforts and a range of skills required for the complex task (Gibbs, 1995; Plymouth University Guidelines for Group Work and its Assessment, 2013). Similarly, in the present course, students from different Bachelor studies and background achieved higher scores as collective marks. As team-based work may not fully reflect the individual contributions (Plymouth University Guidelines for Group Work and its Assessment, 2013), it is recommended not to allocate more than 40% of the mark to teamwork (Gibbs, 1995). Accordingly, in this course, 30% of the mark was allocated to team-based activities. Thus, the relatively lower marks achieved by individual MC-exams and theoretical final examinations may better reflect individual achievements similar to the marks given by the teacher and peers for individual essays. Reliability and validity of peerassessment are well established (Topping 2009); however, alternative calculations of the peer mark could be considered for future studies. For instance, elimination of the maxima lead to a more even distribution of the differences between peer- and teacher-assessed marks ranging predominantly between - 1 and + 1.5 points (out of a maximum of 15 points), which would have led to benefits for some, but disadvantages for other students (Fig. 3b) while the median peer mark would have led to a broader distribution pattern of the differences when compared to the mean peer marks (Fig. 3c). Finally, in the present study, the students expressed their trust in the teacher's ability to consolidate the assessments, which underlines the above reported role of teacher feedback in addition to peer feedback. Thus, current data might encourage increased investigation on peerassessment of essays in the future. More studies should strengthen students' confidence in the objectivity of this assessment system. For instance, Guildford reported elimination of petitions for better grades after invention of peerreviewing (2001). Indeed, some students emphasized that clear criteria served as a basis for the essay mark, such as suitability of the essay for publication. Similarly, this supports the experience that transparent criteria help students understand their marks. A next exercise step for future studies could be to give feedback to the students on their peer review or to give them the

opportunity to see the teacher's feedback on the same essays that they have assessed for comparison.

Further, there was a significant positive linear relationship between final course marks of the students and average essay marks provided by them in the role of assessors to their peers, which indicates that stronger students tended to be more generous in their marking. Even when corrected for having on average lower marks as a subgroup, international students tended to be more severe in their peer assessment. The data regarding peer- and self-assessments indicate a discrepancy between the cohort of local students and the subgroup of international students, even in careful light of the different cohort sizes. Also, the lower scoring of the subgroup of international students was similar to previous cohorts. This phenomenon has been previously evaluated in a similar study subject area at the UWA (authors' unpublished data) and was mainly attributed to the different educational background of international students as compared to Australian and New Zealand students. Thus, the difference between the two cohorts has to be seen as culturally driven. To our knowledge, the present study is the first time that such a correlation has been shown. While there exist some data on impact of cultural background and parental expectations on primary and secondary education (e.g., Li et al. 2017; Cooper et al. 2018), little research has been done to date on the influence of cultural background on assessment and evaluation at the tertiary level. According to publications addressing this issue in the context of experience in the educational system of the USA, one would expect that cultural background and parental expectations may have a substantial influence on how academic tasks such as peerreview are done and perceived, as well as how students assess themselves (Zurcher 1998; Skiba et al. 2002). In a course for nursing students in Spain, self-assessment revealed a pronounced "influence of culture on people's behaviour and how difficult it is to step outside of one's own culture and understand others" and made students reflect "on the beliefs and values of my own culture and of other cultures and (...) analyse(d) their positive and negative aspects", even "changed" students (Siles-González and Solano-Ruiz 2016). In addition to gender differences, also ancestry and background were recently identified to influence participation in science for Australia's students (Cooper et al. 2018). As outlined there, migrating students are equipped with "a set of overarching pattern of values, belief systems, lifestyles, opportunities, customs and resources"(Hackling et al. 2016), which may shape their perceptions (Cooper et al. 2018). Interestingly, first-generation and foreign-background students over 16 years of age at Australian schools were significantly more likely to participate in a science project as compared to Australian-background students (Cooper et al. 2018); however, more data are required on this issue. As societies become culturally more diverse and student exchange programs are highly promoted, future studies should address the question of cultural diversity and educational assessment procedures also at the tertiary level.

Most importantly, this exercise enabled the students to engage in a peerreview process that they may encounter later in their career, in the real-world of scientific publishing. As expressed by many students, this exercise was one of the most valuable that they had experienced during their entire undergraduate education. They thought that it prepared them well for life after graduation including how to cope with harsh feedback. Recently, the reactions of learners to the harsh criticism of peerreviews in the real-world of publishing initiated the invention of coached peerreview exercises using modern E-Technology such as blogs (Sidalak et al. 2017). Also, in the present study, blogs were obviously a good platform for some students to express their feelings about the review process.

In the subsequent year, the course concept was repeated with a new cohort of students. This time, the real-world simulation was changed to stimulate the willingness for careful revision of the manuscript according to the received feedback, which can be low in a classroom scenario (Van den Berg et al. 2006a). Further, students dislike peerreview only for marking (Moore and Teather2013). In a previous study, some students wished for the publication of their writing, which may be difficult to realize (Guildford 2001). To this end, the final versions of all students' essays were edited by the unit coordinator as "Collection of Essays written by 3<sup>rd</sup> Year Undergraduate Students in 2011 at The University of Western Australia in ANHB3313: Cell and Tissue Organisation", and published online by a commercial publisher: "Plasticity of the Human Body" (Bennett et al. 2011), "Biological Emotions of the Heart" (Gomes et al. 2011), "Biology of Vascular Diseases" (Harris et al. 2011), "Biology of Aging" (May et al. 2011), "The Mystery of the Human Breast" (Paplia et al. 2011), "Diabetic Vascular Disease" (Price et al. 2011), "Blood-Brain Barrier-Matter of Life and Death" (Ritschka et al. 2011), and "Bone Biology" (Yap et al. 2011). This real publication procedure was perceived as a very rewarding experience by the students (authors' unpublished data).

With regard to the clinical topics and journals used, the authors consider that the concept may be feasible also for medical students at a similar educational level, which may well support them in different roles of their professional career such as preparing a scientific paper or a conference abstract. Reflective writing on clinical issues has further gained increasing attention to aid in coping with challenging situations and enhancing empathy (Borgstrom et al. 2016; Cowen et al. 2016), but its role needs further investigation (Lin et al. 2016; Ottenberg et al. 2016; Wald and Reis, 2010). Recent findings support the positive impact of mandatory science courses on medical students' research productivity and career preferences (Möller and Shoshan2017). However, there are only limited time and personal resources to establish additional exercises in a medical curriculum. For instance, high workload and lack of academic supervision are among the factors most listed as obstacles for medical students to engage in research (Chang and Ramnanan2015; Jones et al. 2017; Memarpour et al. 2015). Thus, in particular for medical students, there is a demand to implement practical exercises in writing and academic publishing in curricular laboratory or clinical courses. Even short-time interventions improved active scientific writing skills in global health (Deonandan et al. 2017). Thus, offering such a real-world writing skills course in the framework of a regular semester unit may be supportive for the majority of students not enrolled in other specifically research-focused programs. This educational exercise, which simulates the process of scientific writing and publishing, may be implemented in science teaching units at the tertiary level.

### **Limitations of the Study**

First of all, the study in its presented form was conducted only once. Already in the subsequent year, course modifications were implemented based on these experiences, such as a different weighting of the tasks adapted to workload. Second, real-world publication of the essays was achieved only in the subsequent year, which was outside of the present study. More data are needed as a basis to foster peer assessment as a well-accepted assessment method also for final marks. As emphasized, the data regarding the subgroup of international students have to be looked at with caution due to the difference in sample sizes, as well as in light of the particular educational background in Australia. However, the preliminary data may stimulate future

studies to address students' characteristics and their potential influences on their roles as authors and peerreviewers.

## Conclusions

Science literacy is a core issue in science teaching and learning. It includes literature research, reading, critical thinking, active writing of an original academic work, as well as peer and self-evaluation. The described educational exercise contained all these parameters and offered an appropriate assessment of the students' performance in science literacy related to the course topic. The authors would like to encourage such additional writing exercises, particularly with the support of modern E-learning tools.

## Implications for Practice

Writing is demanding for students and needs intensive support and feedback by academic teachers. A real-world scenario including active coached writing, and anonymous peer assessment using modern E-Learning tools including plagiarism check, is feasible for implementation into a pre-existing semester course. Peer assessment may serve for assessment of essays in tertiary education, but more data are needed to strengthen reliability and acceptance by the students. There exists a moderate trend towards cultural differences in self- and peer assessments, which requires further study. This real-world scenario may be considered for advanced life science and medical students, but more studies are needed.

**Acknowledgments** The study was supported by the School of Anatomy and Human Biology and an ISL Grant, UWA, Perth, Australia.

**Compliance with Ethical Standards** The study was conducted in accordance with the ethics rules and regulations at the University of Western Australia.

**Conflict of Interest** The authors declare that they have no conflict of interest.

## References

- Albanese, M. A., Schuldt, S. S., Case, D. E., & Brown, D. (1991). The validity of the lecturer ratings by students and trained observers. *Academic Medicine*, *66*, 26–28.
- Alberts, B., Bray, D., Hopkin, K., Johnson, A. D., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2009). *Essential cellbiology* (Ed. 3 ed.). UK: Garland Science, Taylor and Francis.
- Bennett, C., Barsden, R., Cooke, D., Browne, A., Bradley, C., Bester, A., O'Kelly, H., Metzger, P., Lewis, J., & Filgueira, L. (2011). *Plasticity of the human body* (Kindle ed.). Seattle: Amazon.
- Borgstrom, E., Morris, R., Wood, D., Cohn, S., & Barclay, S. (2016). Learning to care: medical students' reported value and evaluation of palliative care teaching involving meeting patients and reflective writing. *BMC Medical Education*, *16*, 306.
- Brimble, M., & Stevenson-Clarke, P. (2008). Perceptions of the prevalence and seriousness of academic dishonesty in Australian universities. *The Australian Education Researcher*, *32*(3), 19–44.

- Cartney, P. (2010). Exploring the use of peer assessment as a vehicle for closing the gap between feedback given and feedback used. *Assessment and Evaluation in Higher Education*, 35(5), 51–564.
- Chang, Y., & Ramnanan, C. J. (2015). A review of literature on medical students' band scholarly research: experiences, attitudes, and outcomes. *Academic Medicine*, 90, 1162–1173.
- Chen, Y. C., & Tsai, C. C. (2009). An educational research course facilitated by online peer assessment. *Innovations in Education and Teaching International*, 46, 105–117.
- Cooper, G., Berry, A., & Baglin, J. (2018). Demographic predictors of students' science participation over the age of 16: an Australian case study. *Journal of Research in Science Education*. <https://doi.org/10.1007/s11165-018-9692-0>.
- Cowen, V. S., Kaufman, D., & Schoenherr, L. A. (2016). A review of creative and expressive writing as a pedagogical tool in medical education. *Medical Education*, 50, 311–319.
- Crotwell-Timmerman, B. E., Strickland, D. C., Johnson, R. L., & Payne, J. R. (2010). Development of a "universal" rubric for assessing undergraduates' scientific reasoning skills using scientific writing. *Assessment and Evaluation in Higher Education*, 35, 1–39.
- Deonandan, R., Sangwa, N., Kanters, S., & Nsanzimana, S. (2017). Writing skills enhancement for public health professionals in Rwanda. *Advances in Medical Education and Practice*, 8, 253–256.
- Ellis, R. A., Taylor, C. E., & Dury, H. (2007). Learning science through writing: associations with prior conceptions of writing and perceptions of a writing program. *Higher Education Research and Development*, 26, 297–311.
- Eppler, E., Serowy, S., Link, K., & Filgueira, L. (2018). Experience from an optional dissection course in a clinically-orientated concept to complement system-based anatomy in a reformed curriculum. *Anatomical Sciences Education*, 11, 32–43.
- Exley, K., & Demick, R. (2009). *Giving a lecture: from presenting to teaching* (2nd ed.). London: Routledge.
- Filgueira, L. (2010a). *Cell, tissue and development: ANHB 3323*. Perth: School of Anatomy, Physiology and Human Biology, The University of Western Australia.
- Filgueira, L. (2010b). *Cell and tissue organisation ANHB3313 examination paper*. Perth: School of Anatomy and Human Biology, The University of Western Australia.
- Galipeau, J., Moher, D., Campbell, C., Hendry, P., Cameron, D. W., Palepu, A., & Hébert, P. C. (2015). A systematic review highlights a knowledge gap regarding the effectiveness of health-related training programs in journalology. *Journal of Clinical Epidemiology*, 68, 257–265.
- Geerlings, P., Cole, H., Batt, S., & Martin-Lynch, P. (2016). Peer assisted study session (PASS): does gender matter? *Journal of Peer Learning*, 9, 10–25.
- Gibbs, G. (1995). *Learning in teams. A tutor guide* (1st ed.). Oxford: Oxford Centre for Staff Development.
- Gielen, S., Dochy, F., & Onghena, P. (2011). An inventory of peer assessment diversity. *Assessment and Evaluation in Higher Education*, 36, 137–155.
- Gilbert, S. F. (2010). *Developmental biology* (9th ed.). Sunderland: Sinauer Associates, Inc.
- Glasman-Deal, H. (2009). *Science research writing for non-native speakers of English*. London: Imperial College Press.
- Gomes, S., Lee, V., Kagan, G., Pal, S., Iswan, N., Stepan, A., Mortimer, C., & Filgueira, L. (2011). *Biological emotions of the heart* (Kindle ed.). Seattle: Amazon.
- Guildford, W. H. (2001). Teaching peer review and the process of scientific writing. *Advances in Physiology Education*, 25, 167–175.
- Hackling, M., Ramseger, J., & Chen, H. (2016). *Quality teaching in primary science education*. Springer Nature, Berlin: Switzerland.
- Hall, G. M. (Ed.). (2011). *How to write a paper* (4th ed.). Hoboken: BMJ Books.
- Hand, B., Yore, L. D., Jagger, S., & Prain, V. (2010). Connecting research in science literacy and classroom of science teaching journals in Australia, de UK and the United States, 1998–2008. *Studies in Science Education*, 46(1), 45–68.
- Harmon, J. E., & Gross, A. (2010). *The craft of scientific communication*. Chicago: University of Chicago Press.
- Harris, S., Hardy, L., Yousef, D., Gee, S., Jevadi, N., Tang, S., & Filgueira, L. (2011). *Biology of vascular diseases* (Kindle ed.). Seattle: Amazon.
- Hewson, P. W. (2010). Literacy and scientific literacy: a response to Fensham. *Canadian Journal of Science*, 2(2), 207–213.
- Hofmann, A. H. (2010). *Scientific writing and communication: papers, proposals, and presentations* (1st ed.). Cary: Oxford University Press.
- Holbrook, J., & Rannikmae, M. (2007). The nature of science education of enhancing scientific literacy. *International Journal of Science Education*, 29(11), 1347–1362.
- Holstein, S. E., Mickley Steinmetz, K. R., & Miles, J. D. (2015). Teaching science writing in an introductory lab course. *Journal of Undergraduate Neuroscience Education*, 13, A101–A109.

- <https://endnote.com/> (n.d.) (last access: 07.08.2018)  
<https://turnitin.com> (n.d.) (last access: 07.08.2018)  
<http://www.lab.anhb.uwa.edu.au/mb140/> (n.d.) (last access: 07.08.2018)  
<https://www.nih.gov/health-information> (n.d.) (last access: 07.08.2018)
- Inayah, A. T., Anwer, L. A., Shareef, M. A., Nurhussen, A., Alkabbani, H. M., Alzahrani, A. A., Obad, A. S., Zafar, M., & Afsar, N. A. (2017). Objectivity in subjectivity: do students' self and peer assessments correlate with examiners' subjective and objective assessment in clinical skills? A prospective study. *BMJ Open*, 7, e012289.
- Jones, M., Hutt, P., Eastwood, S., & Singh, S. (2017). Impact of intercalated BSc on medical student performance and careers: a BEME systematic review: BEME guide No. 28. *Medical Teacher*, 35, 10.
- Kiernan, J. A. (2009). *Histological and histochemical methods: theory and practice* (4th ed.). Banbury: Scion Publishing Ltd..
- Kierszenbaum, A., & Tres, L. (2011). *Histology and cell biology: an introduction to pathology* (3rd ed.). USA: Mosby.
- King, D., Ritchie, S., Sandhu, M., & Henderson, S. (2015). Emotionally intense science activities. *International Journal of Science Education*, 37, 1886–1914.
- Kommalage, M., & Gunawardena. (2011). Evaluation of physiology lectures conducted by students: Comparison between evaluation by staff and students. *Advances in Physiology Education* 35, 48–52.
- Kwon, J. Y., Bulk, L. Y., Giannone, Z., Liva, S., Chakraborty, B., & Brown, H. (2018). Collaborative peer review process as an informal interprofessional learning tool: findings from an exploratory study. *Journal of Interprofessional Care*, 32, 101–103.
- Leach, L. (2012). Optional self-assessment: some tensions and dilemmas. *Assessment and Evaluation in Higher Education*, 37, 137–147.
- Li, M., Xue, H., Wang, W., & Wang, Y. (2017). Parental expectations and child screen and academic sedentary behaviors in China. *American Journal of Preventive Medicine*, 52, 680–689.
- Lin, C.-W., Lin, M.-J., Wen, C.-C., & Chu, S.-Y. (2016). A word-count approach to analyse linguistic patterns in the reflective writings of medical students. *Medical Education Online*, 21, 29522.
- Lindsay, D. (2011). *Scientific writing = thinking in words*. Clayton, South Victoria, Australia: CSIRO Publishing.
- Liu, N. F., & Carless, D. (2006). Peer feedback: the learning element of peer assessment. *Teaching in Higher Education*, 11(3), 279–290.
- Lodish, H., Berk, A., & Kaiser, C. A. (2007). *Molecular cell biology*. London: W.H. Freeman & Co, Macmillan Publishers.
- Lurie, S. J., Nofziger, A. C., Meldrum, S., Mooney, C., & Epstein, R. M. (2006). Effects of rater selection on peer assessment among medical students. *Medical Education*, 4, 1088–1097.
- Matthews, J. R., & Matthews, R. W. (2007). *Successful scientific writing: a step-by-step guide for the biological and medical sciences* (1st ed.). Cambridge: Cambridge University Press.
- Maxwell, A., Curtis, G. J., & Verdanege, L. (2008). Does culture influence understanding and perceived seriousness of plagiarism? *The International Journal for Educational Integrity*, 4(2), 25–40.
- May, L., Monoharan, S., Wingfield, S.-L., McMahan, A., Rule, G., Melvin, Z., Clark, M., Clark, A., Clancy-Love, K., & Filgueira, L. (2011). *Biology of aging* (Kindle ed.). Seattle: Amazon.
- McCoy, L., Lewis, J. H., & Dalton, D. (2016). Gamification and multimedia for medical education: a landscape review. *The Journal of the American Osteopathic Association*, 116, 22–34.
- McMillan, V. (2011). *Writingpapers in thebiological sciences* (5th ed.). Boston: Bedford/St. Martin's.
- Memarpour, M., Poostforoush, F. A., & Ghasemi, R. (2015). Evaluation of attitude to, knowledge of and barriers towards research among medical science students. *Asia Pacific Family Medicine*, 14(1), 1.
- Mescher, A. (2009). *Junqueira's basic histology: text and atlas* (12th ed.). USA: McGraw-Hill Medical.
- Möller, R., & Shoshan, M. (2017). Medical students' research productivity and career preferences; a 2-year prospective follow-up study. *BMC Medical Education*, 17, 51.
- Moore, C., & Teather, S. (2013). Engaging students in peer review: feedback as learning. *Issues in Educational Research*, 23(2, Special issue), 196–211.
- Nieder, G. L., Parmelee, D. X., Stolfi, A., & Hudes, P. D. (2005). Team-based learning in a medical gross anatomy and embryology course. *Clinical Anatomy*, 18, 56–63.
- Nulty, D. D. (2010). Peer and self-assessment in the first year of university. *Assessment & Evaluation in Higher Education*, 36, 493–507.
- Olitsky, S., Becker, E. A., Jayo, I., Vinogradov, P., & Montcalmo, J. (2018). Constructing “authentic” science: results from a university/high school collaboration integrating digital storytelling and social networking. *Research in Science Education*. <https://doi.org/10.1007/s11165-018-9699-6>.
- Osborne, J. (2002). Science without literacy: a ship without a sail? *Cambridge Journal of Education*, 32, 203–218.

- Ottenberg, A. L., Pasalic, D., Bui, G. T., & Pawlina, W. (2016). An analysis of reflective writing in the medical curriculum: the relationship between reflective capacity and academic achievement. *Medical Teacher, 38*, 724–729.
- Paplia, P., Osman, A., Prempeh, L., Plint, G., Butchard, L., Brooks, S., Malaga, G., Koh, S. L., Tan, J., & Filgueira, L. (2011). *The mystery of the human breast* (Kindle ed.). Seattle: Amazon.
- Perera, J., Mohamadou, G., & Kaur, S. (2010). The use of objective structured self-assessment and peer-feedback (OSSP) for learning communication skills: evaluation using a controlled trial. *Advances in Health Science Education: Theory and Practice, 15*, 185–193.
- Pizzimenti, M. A., Pantazis, N., Sandra, A., Hoffmann, D. S., Lenocho, S., & Ferguson, K. J. (2016). Dissection and dissection-associated required experiences improve student performance in gross anatomy: differences among quartiles. *Anatomical Sciences Education, 9*, 238–246. <https://doi.org/10.1002/ase.1574>.
- Plymouth University. (2013). *Educational development. Guidelines for group work and its assessment—March 2013* (1st ed.). Plymouth: Plymouth University 5 p. URL: [https://www.plymouth.ac.uk/uploads/production/document/path/2/2427/Guidelines\\_for\\_Assessing\\_Group\\_work\\_Dec\\_2012.pdf](https://www.plymouth.ac.uk/uploads/production/document/path/2/2427/Guidelines_for_Assessing_Group_work_Dec_2012.pdf) [accessed 17 February 2017].
- Pollack, G. H. (2001). *Cells, gels and the engines of life: a new, unifying approach to cell function*. Seattle: Ebner and Sons Publishers.
- Poon, W. Y., McNaught, C., Lam, P., & Kwan, H. S. (2009). Improving assessment methods in university science education with negotiated self- and peer-assessment. *Assessment in Education: Principles, Policy & Practice, 16*(3), 331–346.
- Price, I., Smith, A., Pantula, R., Wilson, T., & Filgueira, L. (2011). *Diabetic vascular disease* (Kindle ed.). Seattle: Amazon.
- Ritschka, B., Stackpoole, E., Tedja, A., Brown, T., Luitingh, T., Symons, Y., Foster, N., & Filgueira, L. (2011). *Blood-brain barrier-matter of life and death* (Kindle ed.). Seattle: Amazon.
- Ross, M. H., Romrell, L. J., & Pawlina, W. (2006). *Histology: a text and atlas* (6th ed.). Philadelphia: Lippincott Williams & Wilkins.
- Schönrock-Adema, J., Heijne-Penninga, M., van Duijn, M. A., Geertsma, J., & Cohen-Schotanus, J. (2007). Assessment of professional behaviour in undergraduate medical education: peer assessment enhances performance. *Medical Education, 41*, 836–842.
- Sidalak, D., Purdy, E., LUCKETT-Gatopoulos, S., Murray, H., Thoma, B., & Chan, T. M. (2017). Coached peer review: developing the next generation of authors. *Academic Medicine, 92*, 201–204.
- Siles-González, J., & Solano-Ruiz, C. (2016). Self-assessment, reflection on practice and critical thinking in nurse students. *Nurse Education Today, 45*, 132–137.
- Skiba, R. J., Knesting, K., & Bush, L. D. (2002). Culturally competent assessment: more than nonbiased tests. *Journal of Child and Family Studies, 11*(1), 61–78.
- Spandorfer, J., Puklus, T., Rose, V., Vahedi, M., Collins, L., Giordano, C., Schmidt, R., & Braster, C. (2014). Peer assessment among first year medical students in anatomy. *Anatomical Sciences Education, 7*, 144–152.
- Strong, B., Davis, M., & Hawks, V. (2004). Self-grading in large general education classes: a case study. *College Teaching, 52*, 52–57.
- Sunderland, G. S. F. (2000). *Developmental biology. Part 1: principles of development in biology*. Cary: Sinauer Associates, Oxford University Press.
- The University of Western Australia (2014) [http://www.international.uwa.edu.au/data/assets/pdf\\_file/0004/2633719/Accepted-Full-Load-Equivalents-updated-Nov-2014.pdf](http://www.international.uwa.edu.au/data/assets/pdf_file/0004/2633719/Accepted-Full-Load-Equivalents-updated-Nov-2014.pdf). Perth, WA, Australia.
- Topping, K. J. (2009). Peer assessment. *Theory Into Practice, 48*, 20–27.
- Truss, L. (2003). *Eats, shoots and leaves (the zero tolerance approach to punctuation)*. London: Profile Books.
- Tseng, S. C., & Tsai, C. C. (2007). On-line peer assessment and role of the pper feedback: a study of high school computer course. *Computers & Education, 49*, 1161–1174.
- Van den Berg, I., Admiraal, W., & Pilot, A. (2006a). Peer assessment in university teaching: evaluating seven course designs. *Assessment and Evaluation in Higher Education, 31*(1), 19–36.
- Van den Berg, I., Admiraal, W., & Pilot, A. (2006b). Design principles and outcomes of peer assessment in higher education. *Studies in Higher Education, 31*(3), 341–356.
- Venables, A., & Summit, R. (2003). Enhancing scientific essay writing using peer assessment. *Innovations in Education and Teaching International, 40*, 281–290.
- Vu, T. T., & Dall’Alba, G. (2007). Students’ experience of peer assessment in professional course. *Assessment and Evaluation in Higher Education, 32*(5), 541–556.
- Wagner, E., Fiack, S., Graf, C., & Rowlands, I. (2009). Science journal editors’ views on publication ethics: results of an international survey. *Journal of Medical Ethics, 35*, 348–353.

- Wald, H. S., & Reis, S. P. (2010). Beyond the margins: reflective writing and development of reflective capacity in medical education. *Journal of General Internal Medicine*, 27, 746–749.
- WeaverKF, MoralesV, NelsonM, WeaverPF, ToledoA, GoddeK (2016) The benefits of peer review and a multise­mester capstone writing series on inquiry and analysis skills in an undergraduate thesis. *CBE Life Science Education* 15
- Wenzel, T. J. (2007). Evaluation tools to guide students' peer-assessment and self-assessment in group activities for the lab and classroom. *Journal of Chemical Education*, 84(1), 182–186.
- www.blackboard.com (n.d.) (last access: 04.07.2018).
- www.lectopia.com.au (n.d.) (last access: 04.07.2018).
- Xu, J., Kim, K., Kurtz, M., & Nolan, M. T. (2016). Mentored peer reviewing for PhD faculty and students. *Nurse Education Today*, 37, 1–2.
- Yap, C., Ma, J., Gow, S., Wilson, L., Toro, A., Amirudin, S., Pleydell-Bouvarie, M., Visser, C., & Filgueira, L. (2011). *Bone biology* (Kindle ed.). Seattle: Amazon.
- Zurcher, R. (1998). Issues and trends in culture-fair assessment. *Intervention in School and Clinic*, 34, 103–106.

### Affiliations

Elisabeth Eppler<sup>1,2,3</sup> • Jan Meyer<sup>2</sup> • Steffen Serowy<sup>3</sup> • Karl Link<sup>4</sup> • Barbara Pauk<sup>5</sup> • Luis Filgueira<sup>2,4</sup>

<sup>1</sup> Department of Biomedicine, University of Basel, Pestalozzistrasse 20, CH-4056 Basel, Switzerland

<sup>2</sup> School of Human Sciences, The University of Western Australia, Perth, WA, Australia

<sup>3</sup> Department of Neuroradiology, University Hospital of Magdeburg, Magdeburg, Germany

<sup>4</sup> Anatomy, Department of Medicine, University of Fribourg, Fribourg, Switzerland

<sup>5</sup> European Languages and Studies, School of Humanities, The University of Western Australia, Perth, WA, Australia