Is grading in a poster session influenced by biases?

Hugo Lopez, Louis Vasselin, Karine Le Barch, Melanie Etheve-Quelquejeu, Samuela Pasquali, Antoine Taly

Is grading in a poster session influenced by biases?

Hugo Lopez*, Louis Vasselin**, Karine Le Barch***, Melanie Etheve-Quelquejeu***, Samuela Pasquali**** & Antoine Taly*****

Abstract

Scientific article reading is an important competence for undergraduate students in the sciences. To help acquire this skill we chose to propose students a poster session with peer evaluation. Students are grouped in teams to which subjects are assigned. Each team prepares a poster during the semester and prints it before the exam. Students then enter a rotation system in which they alternatively present their poster to their peers and one teacher, or evaluate the posters of other groups. Each session is composed of a presentation, questions and evaluation. The evaluation is performed both by the teachers and peers. The question of biases emerges because this poster session replaces an anonymous exam. Data from four years allows us to analyze the evaluation performed by the students and teachers: i) the dynamic range of grades given by students tends to be smaller than in the case of teachers; ii) the evaluations appear not to be biased in terms of gender but might be detrimental to visible minorities.

INTRODUCTION

Posters are used to present information on a research subject. For most conferences, poster sessions are included, during which the author stands by the poster while other participants can come, view the presentation and interact with its authors. As noted by others, this is often the first form in which students share their results, before writing articles (Kennedy, 1985; Wimpfheimer, 2004). This is probably one of the reasons for the use of poster sessions in-class (Kennedy, 1985; Wimpfheimer, 2004), which has now a large use history (Baird, 1991; Sisak, 1997; Widanski et al., 2016).

In addition to preparing students for the management of posters *per se*, many learning outcomes have been proposed: i) change in the perception of science (Gerczei, 2016), ii) work on difficult concepts (Sisak, 1997), iii) work on communication skills and statistical reasoning (Logan et al., 2015). These interesting learning outcomes have been proposed to be sustained by increased motivation (Altintas et al., 2014). In addition, a poster session can be used for evaluation (Marino et al., 2000; Wimpfheimer, 2004), including to replace an exam (Marino et al., 2000). Often this is done with some sort of scoring/grading sheet (Widanski et al., 2016; Wimpfheimer, 2004). As noted by others, the use of a poster session for evaluation, as for other alternative evaluation methods, could facilitate the inclusion of underrepresented students (Gerczei, 2016; Marino et al., 2000).

A literature search shows that many parameters of the poster sessions can be adapted to various situations and learning outcomes, in particular: i) the evaluation can be done by teachers or by peers (Kadnikova, 2013; Wimpfheimer, 2004), ii) the posters can be presented or not (Gerczei, 2016; Wimpfheimer, 2004), iii) the work can be done alone or in group (Gerczei, 2016; Marino et al., 2000).

As part of an interdisciplinary science course, we decided to use this "Poster Session" concept as a format for the final exam. Interdisciplinarity is crucial for students' future development in academia or industry alike. However, it’s sometimes difficult to have homogeneous knowledge for all of the students. We designed an interdisciplinary course (3 credits) at the level of undergraduate students in the third year. We study the DNA biomolecule from the viewpoint of chemistry, physics, biology and cognitive sciences. This course is called ”from the double helix to the clinics" and is attended by 120 to 150 students each year.

One of the issues in these courses is the final evaluation. We have therefore built a setup that relies on peer

evaluation where students prepare their posters, participate in discussions with their peers and teachers, like in other cases (Adkins & Lyon, 2012), but also evaluate each other's work. Classical conception of a poster allows students to focus on a precise subject, to read scientific articles, to clearly design the poster, to present it in a controlled time, and to be able to answer the questions.

Among the distinctive properties of our approach two should be kept in mind:

- students prepare their posters in groups and present them individually thanks to a rotation system that allows them to alternatively present their poster and evaluate that of others.
- While dealing with relatively large cohorts we chose to maintain teacher's evaluation and constructed a system that allows combining teacher and peer evaluation. The whole system has been implemented in a script to be more easily managed by us and others.

The poster session was introduced to replace a written exam. Anonymity is known to have effects on evaluation (Panadero & Alqassab, 2019). We therefore wondered whether the loss of anonymity might have been associated with the (re)emergence of potential biases. Among potential biases we focussed on gender and ethnicity. Indeed, ethnicity-related stereotypes are known to have effect on group work and there exist stereotypes on communications skills (Bandyopadhyay et al., 2022). Being committed to increase inclusivity, it is important to document putative discriminatory situations, in particular in the french context in which information is lacking.

Methods

We present here the poster session itself in detail to ensure that no parameter that could play a role is omitted. The analysis in terms of ethnicity is also presented.

Sizing : 2 months prior

In order to provide our students a collaborative experience and the opportunity for each to present twice within 2 hours, we opted for groups of 3 students and 6 sessions of 15 minutes (adding up to 2 hours). Depending on the total number of students, some groups were reduced to 2 students, each one then presenting 3 times. When a

group was reduced to one student, because of one or two students being absent, on spot adaptation were performed (see #Managing changes). To ensure all students, presenting twice or more, were reviewed by knowledgeable experts (Ph. D. students, post-docs or professors), the number of experts was set to be at least half the number of groups; each expert monitoring 2 groups.

Groups creation and distribution of assignments : 2 months prior

The students chose the group they join in. Groups and students were assigned code names such as “G02” and “G02a”, respectively. To ensure fairness, the subjects or articles are assigned randomly (Table 1).

Table 1. Table of groups, individuals and assignments for one class

Please select your team by adding your name in front of a letter (a, b, c).		
Group G01	a: John Snow	Probing the salt dependence of the torsional stiffness of DNA by multiplexed magnetic torque tweezers, Kriegel F. & al., Nucleic Acids Research, 2017, Vol. 45, No. 10, 5920-5929.
	b: Jane Doe	
	c: Jade Smith	
...		
Group G12	a: Stella Green	Modified RNAs in CRISPR/Cas9: An Old Trick Works Again Alfonso Latorre, Ana Latorre, and çlvaro Somoza, Angew. Chem. Int. Ed. 2015, 54, 4710 – 4712.
	b: Sylvia Seal	
	c: ~~Brian Breeze~~	
Struck students (ex: *Brian Breeze*) when absent on D-Day.		

Training : +1.5 month prior

One or two month before the D-day, a lecture “How to design a Poster” was provided to the students, for practical methodology and to clarify the grading criteria. The evaluation grid (Taly et al., 2018) was then shared with the students.

A small lecture on document search and plagiarism was also given. The preparation guidelines were as follow :

Box 1. Guidelines submitted to all teams two months before D-day

Following the table of assignments, please find your academic article online, read and understand its concepts. Search online for more information to help your understanding.
Your team will have to create an accurate and elegant A0 or A1 poster presenting the article content.
On D-day, each student will present their team’s poster to a few students and teachers.

Presentation will last 5 mins, Q&A session 5 mins, evaluation 5 mins.

The evaluation grid clarifies what is expected from you.

Printing : 7 days prior

Early bird groups had access to free high quality printing, in A1-A0 format, via the University's press service. A pdf had to be sent at least one week prior to the activity. Late bird groups had to handle and pay for professional printing by themselves.

Activity guidelines and unfolding

As a peers grading activity started, we orally briefed our students with the following overall guidelines :

Box 2. Guidelines submitted to students on D-day

Today's activity will let you display your team's work and evaluate other students' works.
Today's activity will be divided into 6 sessions of 15 minutes, from #1 to #6.
You will have 5 min for presentations, 5 min for Q&A, 5 min for assessment.

You will evaluate the other team's work for 3 or 4 sessions.
You will present your work for 2 or 3 sessions.
Your personal planning is provided by the teaching team. Check your schedule and write it down on your sheet.

At each session's end, take notes on a paper, so you collect a list of grades such as :
- Session 1 : presentation G01, grade given 16/20.
- Session 2 : presentation G03, grade given 14/20.
- Session 3 : I presented, no grade.
- ...
- Session 6 : presentation G08, grade given 17/20.

N.B. you need to be graded by at least one teacher. At your last presentation, if you haven't been graded by a teacher, please contact us.

After the last session, report all these evaluations on this form :
URL : https://tinyurl.com/posterday170607

The teachers were also assigned an identifier at that time (Profxx).

Planning of rotations

In a class of 30 students, 10 groups of 3 presenting their work, the students dispatch on session n°1 out of 6 would look like the scheme in Figure 1.:

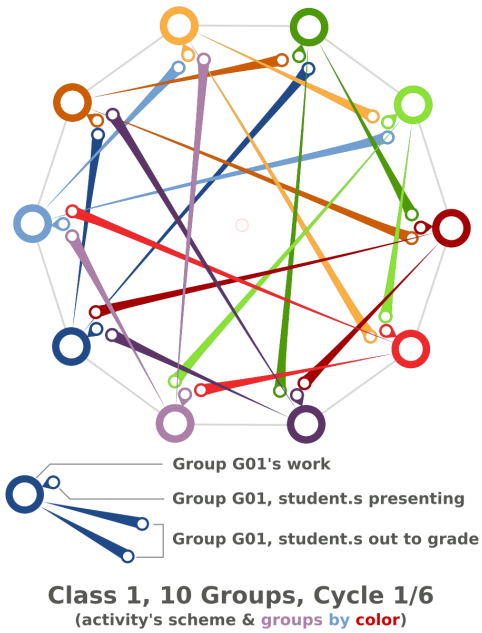

Figure 1. Illustration of the rotation system.

Rotation tables were used to inform the students of where they are supposed to be for each of the *n* sessions, where in our case *n=6* (see an example as supplementary material). A shift between session *x* and session *x+1* allows everyone to move to another work. Along the activity, each professor is assigned to 2 groups, alternating her or himself from one group to the other.

Evaluation

The teaching team provided each student with a printed evaluation grid. After a few trials, and a co-working session with student's representatives, we came up with the table shown in Figure 2.

Poster session's rubric

CATEGORY	Poster and/or Oral	Exemplary	Very Good	Competent	Partially Competent	Unsatisfactory	Points
		4 points	3 points	2 points	1 point	0 points	
A: **Scientific content.** N.B. B, C and D depend on A (if A=x then B, C and D <= x)	P+O	The subject is largely covered and in depth (beyond the original article), including from the angle of various disciplines.	The subject is largely covered and in depth (beyond the original article)	The subject is partially covered (only the original article)	Only a fraction of the subject covered or the poster contain many unrelated content.	Nothing or off topic.	/4
B: **Accuracy**	P+O	All points are accurate.	Anecdotal errors.	Limited errors.	Significant errors.	Major errors	/4
C: **Sources**. N.B.: if C = 0 → A=0 and B=0.	P+O	Work on multiple sources, including scientific articles. Critical and convincing analysis of the sources used.	Work on multiple sources, including scientific articles.	Use of more than one source.	Work limited to given article and course material (e.g. reference book).	Non justified claims, plagiarism, non respect of authors rights.	/4
D: **Poster and pedagogy**	P	The poster is structured, homogenous and conscientious. When necessary, explanations are supported by images or schemes that are clear and appealing.	The poster is structured but with non homogenous elements (color, languages, etc). No images/schemes.	The poster is poorly structured.	The poster is poorly structured and contains major errors.	No poster.	/4
E: **Answer to questions**	O	Answer to all questions, capability to go beyond the initial paper and to interact. The student is conscious of the limits of the responses given. When needed, different analyses are confronted.	Answer to most questions, minor errors.	Answer to part of the questions, minor errors.	Limited answer to part of the questions. Significant errors.	No answer.	/4
TOTAL							/20

Figure 2. Evaluation grid.

Online system and data gathering

PeersGraderApp is the open source online application we developed to support our peer grading activities (Lopez & Vasselin, n.d.). Participants open the link, follow their assignments, then submit evaluations as needed. The form first collects information relative to the person who fills the form: full name, email address for quick contacts, team and user's identifier. A second section is made of triads "Session name / Group presenting / Grade given if any", repeated for as many sessions as needed.

Algorithm and processing

Based on the answers provided, 4 elements are calculated. These elements are defined as follows :

- **Professors grade (Prg)** [range:0-20]: average of grades a student received from her/his professor(s).
- **Peers grade (Peg)** [range:0-20]: average of grades a student received from her/his peers.
- **Grading assessment (Grad)** [range:0-20]: average of grading assessment grades of a student's evaluation on peers presenting to her/him. The more the grade given to a presenter is distant from this presenter's overall averaged Peg + Prg, the less the judgment was accurate, the lower the grading assessment score gets. A difference of ≤1pts results in grading assessment of 20/20 ; ≤2pts : 18 ; ≤4pts : 16; ≤8pts : 10 ; >8 points Grad=0/20. For example, given student Alice received a mean grade of 16/20, student Beau gave her 15/20, Beau's Grad=20/20. Celestine, who gave Alice 7/20 has a judgment of 16-7=9 points from mean grade and so, gets a grading assessment Grad=0/20 on this single evaluation.
- **Final grade (F):** the final grade of the student for this workshop, with default as :

$$F = Peg \times 0.25 + Prg \times 0.50 + Grad \times 0.25.$$

Peg, Prg and Grad being averages of multiples Peg, Prg, Grad values associated with this student. The weights displayed have been chosen so that half of the overall grade still comes from teachers. After the event, the final results are displayed in the form of a final HTML table with associated grades (Figure 3).

Figure 3. Example of PeersGraderApp final results.

indivId	averageProfs	averagePeers	normalness	finalGrade	indivFamily	indivEmail
G01a	13	15.6	16.5	14.5	Antic Ant	antic.ant@student.org
G01b	14	17.3	11.7	14.3	Big Bunny	big.bunny@student.org
G01c	14	16.1	16.2	15.1	Candid Camel	candid.camel@student.org
G02a	16	16.5	12.3	15.2	Digging Dog	digging.dog@student.org

Teacher's checklist

On D-day, our pre-activity checklist was as follows :

Box 3. Organizer's pre-activity checklist

Human resources

- 1 teacher for 2 groups of students so each student is graded by one teacher
- **12** groups of 3 students presenting **2** times each = **6** teachers needed
- One teacher also manages timing : 5mn for presentation, Q&A, grading.

Material

- *n* classrooms, ideally allowing to have no more than 3~4 groups per classroom so to reduce noise
- student's badges labeled with *name* and group-based *student ID* (ex:G02a). Use right wording on badge so to avoid confusion.
- 1 × tape roll per classroom
- 1 × large erasable marker per classroom, write group above presentation spots
- 1 × start/stop whistle

Managing change : the last minute missing students

Example of last minute changes were as follows :

- When a group falled from 3 to 2 students: the two teammates replaced their missing teammate equally for her or his presentations without further consequence.
- When a group was represented by a single student: the poster was presented only part of the time.
- When a whole group was missing : the poster was not presented.

When a poster was not presented, as per above, the visiting students were redirected to review other groups, i.e. to the spots which the missing group was expected to review, so as to fill the gap.

Origins determination

No systematic stats in France.

Controversial subject such that it is possible only for a prospective study after obtaining a specific authorization.

However our study is retrospective.

Two systems were used, ethnea (Torvik & Agarwal, 2016) and namsor (*Namsor: Name Checker for Gender, Origin and Ethnicity Classification*, n.d.), in order to construct a consensus classification. Four groups were used:

Africa, Arab, Asia and Europe. Jewish names that are classified as western asia in one case and french in the other were classified as european. Other ambiguities were resolved by looking at online information (linkedin profiles, CV) either with pictures or information on the CV (e.g. spoken languages). Eleven cases could not be disambiguated with this method and were therefore tagged as not determined. The final classification was incorporated in an anonymous file to comply with the French law.

Ethenea results were also used directly to construct two groups French/Foreigner

Results and Discussion

Preparation of the poster session

As shown in methods the poster preparation as well as presentation was largely framed by a series of instructions and tools, that can be compared to other protocols:

- We rapidly identified that the poster preparation could be facilitated with some help given to students. Similarly to Widansky and colleagues (Widanski et al., 2016), although independently, we felt the need for input from a librarian. In our case, given that the initial article is provided by the teacher, the support is focussed on plagiarism rather than literature search. A more complete support would probably be helpful to students.
- To frame the poster preparation and its evaluation we used an evaluation grid (Figure 2). The grid had several purposes. First, it was meant to guide students so their evaluation of a peer at each session is based on a clear system of judgment. The grid was also expected to add legitimacy to the peer grading activity, showing explicitly teachers' expectations. This is similar to the use of scoring/grading sheets by others (Widanski et al., 2016; Wimpfheimer, 2004). At variance to other examples our grid uses 5 levels of accomplishment. The grid was improved with students because many of them were uncomfortable with the previous 4 levels grid. Adding an additional level made the situation easier to manage as shown by the absence of complaints

after the modification of the grid.

- We did not ask for an intermediate assignment. Such an intermediate assignment has been found by others to allow students to receive feedback and ensures that the work is not done at the last moment (Gerczei, 2016; Widanski et al., 2016). However we printed posters for the students, in order to avoid printing issues for low income students (Gerczei, 2016). The printing was performed provided that students had sent their poster in advance, which ensured that the work was not done at the very last moment.
- The questions that were asked during the poster session by teachers and peers were completely open and were not based on a question bank (Marino et al., 2000). Although a question bank has advantages, it also limits the freedom of teachers which we found interesting, as it allows teachers to give feedback to students.

An important issue with a poster session used as an exam is to have enough faculty to actually grade all students. This is not a problem with small groups of students, but, in the case of larger cohorts of students, pushed others to use peer review exclusively (Gerczei, 2016; Kadnikova, 2013; Marino et al., 2000), or an online poster session (Menke, 2014). We instead chose to keep the teacher's evaluation even if that means involving a large number of instructors. We note that the numbers given above are similar to those given by Marino and others, i.e. 10 faculty for a class of 80 (Kadnikova, 2013). As discussed below it was relatively easy to have more teachers involved.

Teachers' Evaluations

The first observation was that it was easy for teachers to join the workshop, even for those with no experience of this type of teaching. The only uncommon element in this otherwise familiar context is the evaluation system. First, the evaluation grid needed to be presented in detail to the teachers. The second uncommon aspect was that the exam involved peer evaluation. One significant parameter of

our system was therefore the choice of weights used to make the final grade (equation 1). These weights had been chosen to ensure that most of the grade comed from teachers: here it represented the majority of the grade given that 50% of the grade was coming directly from them and that they also had a strong effect on the evaluation of students ability to grade others. This situation had been conceived with the idea that it would not deviate too much from the usual situation in which the teachers are making all of the grade, which in turn might facilitate the adoption of the system by teachers. The effect of this parameter on adoption has not been formally studied but we did not get any negative feedback from teachers on this aspect.

The second satisfactory observation is that the dynamic range of grades is similar to what is observed in other evaluations. Dynamic range of grades given by teachers is large (Figure 4), including failing grades (<10), at variance to what was observed by others (Kadnikova, 2013). In some cases students were graded by two teachers. The difference between the two evaluations are shown in Figure 4b. Most of the cases (75) are associated with small differences (<= 2), but a few cases (9) were associated with large differences (5-8). When these cases were identified immediately, the teachers were interrogated for details. The gathered information, although anecdotal, suggest that students improved from first to second presentation, avoiding initial errors during the second presentation. Altogether this data suggest that the grades given by teachers are rather reproducible.

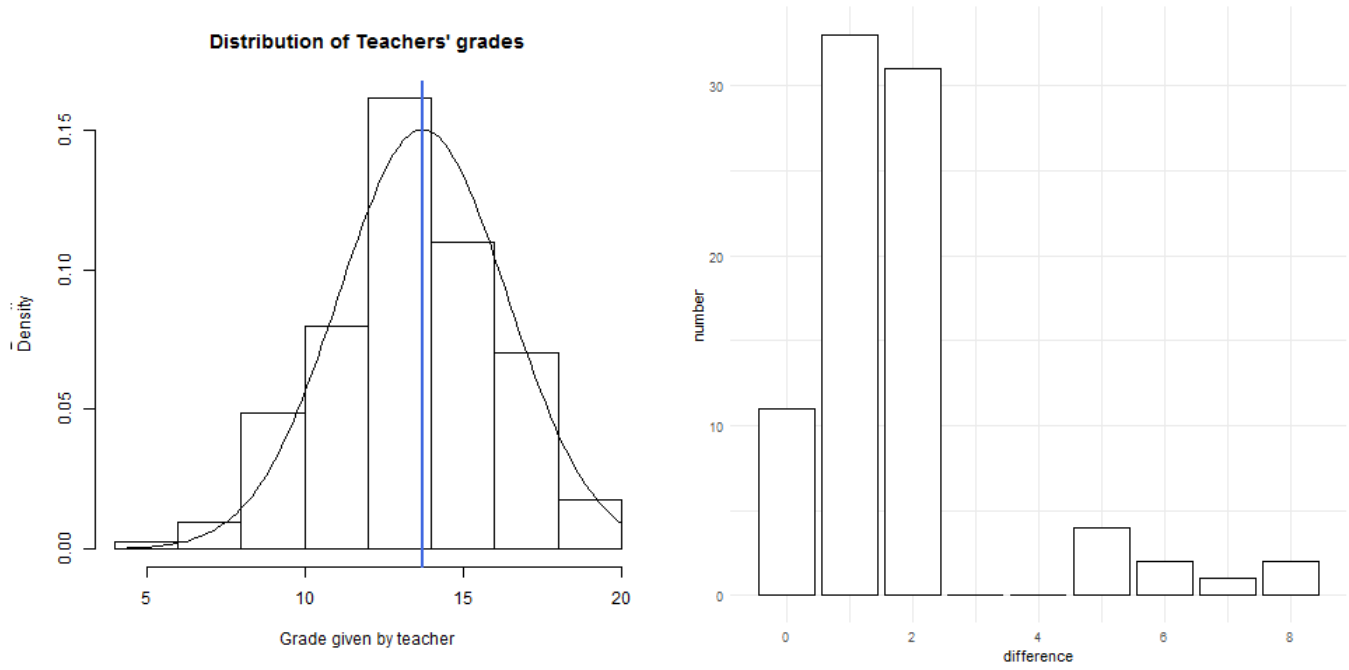

Figure 4. Quantitative analysis of grades given by teachers in poster sessions. Left: distribution of grades. Mean (14.2) is shown as a blue line; right: variability of grades when two teachers grade a single student.

Peer's grades

Interestingly, peers' grades are relatively well correlated to teachers' grades (Figure 5b) with a value of R^2 (0.16) that is similar to that found by Gerczei (Gerczei, 2016). Noteworthy, the grades given by students are slightly higher than those given by teachers (compare Figure 5a and Figure 4a). This observation is however not true for the whole grading scale. A bland-Altman representation shows that, if the grades given by teachers are used as reference, the students tend to overgrade weak students and undergrade strong students (Figure 5c). To explain this observation two explanations can be put forward: i) students might not be willing to give poor grades to each other, therefore all students get grades ranging 10 and higher from their peers. ii) another hypothesis is that students lack the ability to detect all the mistakes made by their classmates and grade them without a perfect understanding of the content.

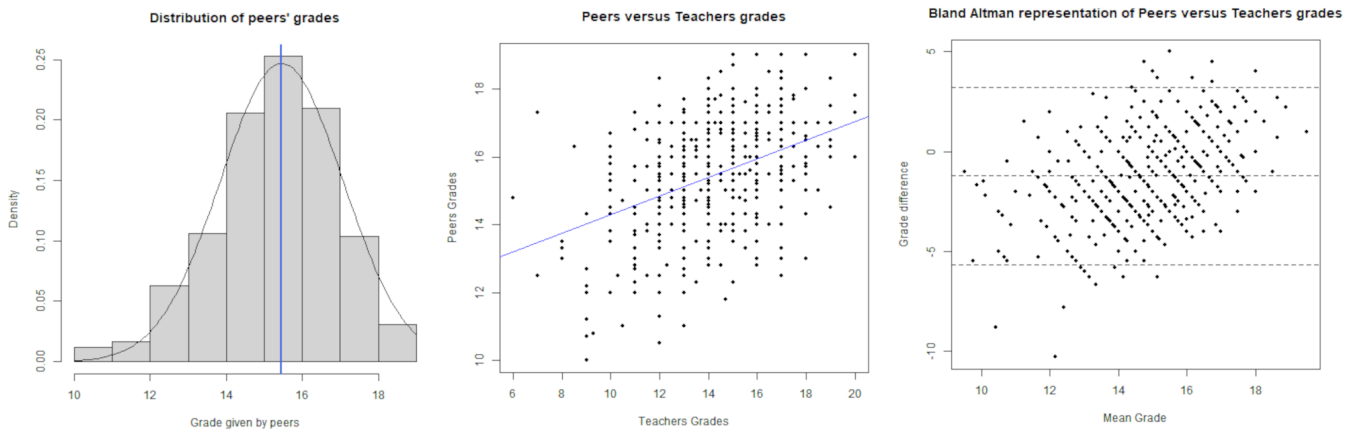

Figure 5. Quantitative analysis of grades given by peers in poster sessions. Left: distribution of grades. Mean (15.5) is shown as a blue line; middle: correlation of mean grades given by peers and teachers' grades. A regression line is shown (R^2=0.16). Right: Altman and Bland representation of grades given by peers versus teachers.

Biases

The poster session was introduced to replace a written exam. Anonymity is known to have effects on evaluation (Panadero & Alqassab, 2019). We therefore wondered whether the loss of anonymity might have been associated with the (re)emergence of potential biases. Among potential biases we focussed on gender and ethnicity. The original data has been transformed into anonymous data with origin and is available online (Antoine et al., 2021).

We first compared grades obtained by males and females. The mean grades of female students are higher than that of males, in the case of grades given by both other students and teachers (table 2). The difference in average is however not significant for both peers' and teachers' grades. Therefore there does not seem to be a significant effect against females. We keep in mind however that our population of students is composed of a majority of females (69%) such that some biases might be damped and further studies on larger cohorts might be worth undertaking.

The analysis presented above relies on a binary analysis that implies that individuals define themselves as either male or female, essentially neglecting the difference between sex and gender. For

this reason, during the third and fourth year an additional option was added where students could choose the option ‘prefer not to say’. Three students chose to answer the question on gender with this option. Although it could be interesting to analyse the situation of those students in further detail, the small number of students involved as well as the fact that there could be more than one reason to use this option prevented further analysis.

	Females	Males	P - value
Peer’s grades	15.5 (1.6)	15.3 (1.7)	0.25
Teacher’s grades	14.3 (2.4)	14.0 (2.2)	0.23

Table 2. Grades recieved as a function of students’ sex. The standard deviation is shown in parentheses.

We then explored the potential biases towards visible minorities. The creation of databases involving people's ethnicity is restricted in France. Data on the ethnicity of students is therefore not available at variance to what is seen elsewhere (Cramer, 2021). The relevance of collecting ethnicity is even subject to controversies (Beaud & Noiriel, n.d.; Marichalar, n.d.). Because our study has been conducted *a posteriori* the groups need to be inferred from names. Interestingly, inferences, despite being imperfect, have been shown to be usefull to measure discrimination (Rieke et al., 2022). We therefore opted for an automatic treatment in which the origin of people is inferred from their names (Mazières & Roth, 2018). As detailed in methods the groups were defined using two methods (ethnea and nomsor). A consensus was constructed with groups that would be large enough to allow significant sampling; four groups were used: Africa, Northern-Africa(Arab), Asia and Europe. These groups have sizes of 50, 69, 52 and 309 respectively. Noteworthy, a large number of disagreements appeared between both methods (102 out of 491). This observation is not necessarily surprising given the differences in methods and populations used to develop them, and could be due to changes in the usage of first names among immigrants documented recently (Coulmont & Simon, 2019). It appeared necessary to reduce these ambiguities because the size of our sample is too small to accommodate

them (Mazières & Roth, 2018). One ambiguity seemed to arise in the case of Jewish names that are classified as western asia with one method and french with the other. Given that these cases most probably correspond mainly to french students, they were classified as european here. Other ambiguities appeared less systematic. In general they could be resolved by looking at online information (linkedin profiles, CV) either with pictures or information on the CV (e.g. spoken languages). This classification is of course itself subject to biases, but those biases are arguably coherent with the biases studied here. Eleven cases could not be disambiguated with this method and were therefore tagged as not determined.

The grades received by students were analyzed as a function of those groups (Table 3). Anova analysis shows no difference for grades received from peers but a significant difference for the grades received from teachers. In that later case the grades received by students in the Africa and Asia groups are lower.

	Africa	Arab	Asia	Europe	P - value
Peer's grades	15.3 (1.7)	15.4 (1.8)	15.1 (1.8)	15.6 (1.5)	0.31
Teacher's grades	13.8 (2.8)	14.4 (2.1)	13.7 (2.1)	14.4 (2.4)	0.006
Difference	1.4 (2.3)	1.1 (2)	1.4 (1.9)	1.1 (2.4)	0.07

Table 3. Grades received as a function of students' origin defined with four categories. The standard deviation is shown in parentheses.

The grades were also analyzed when a binary classification was used (French versus Foreigner) (Table 4). In that case no significant difference in the grades as a function of origin are observed in the grades given by teachers, but the grades received from peers were lower for students with 'foreign' names.

	Foreigner	French	P - value
Peer's grades	15.3 (1.7)	15.7 (1.5)	0.003
Teacher's grades	14.1 (2.4)	14.4 (2.4)	0.12

Difference	1.2 (2.2)	1.3 (2.4)	0.6

Table 4. Grades received as a function of students' origin defined as a binary. The standard deviation is shown in parentheses.

The grades can also be compared by looking at the distribution of grades in the various groups (Figure 6). This representation shows that the distributions are not necessarily normal making any further analysis difficult at best.

Altogether the analysis of the impact of the origin on grades suggest that there might be an effect. However, this effect is not robust in that it is not identical in teachers and peers and depends on the way the origin groups are made. It is tempting to speculate that the cohort we are working on is large enough to allow these effects to appear but too small for them to be robust. This is especially true because we rely on an idirect classification of visible minorities because of legal limitations.

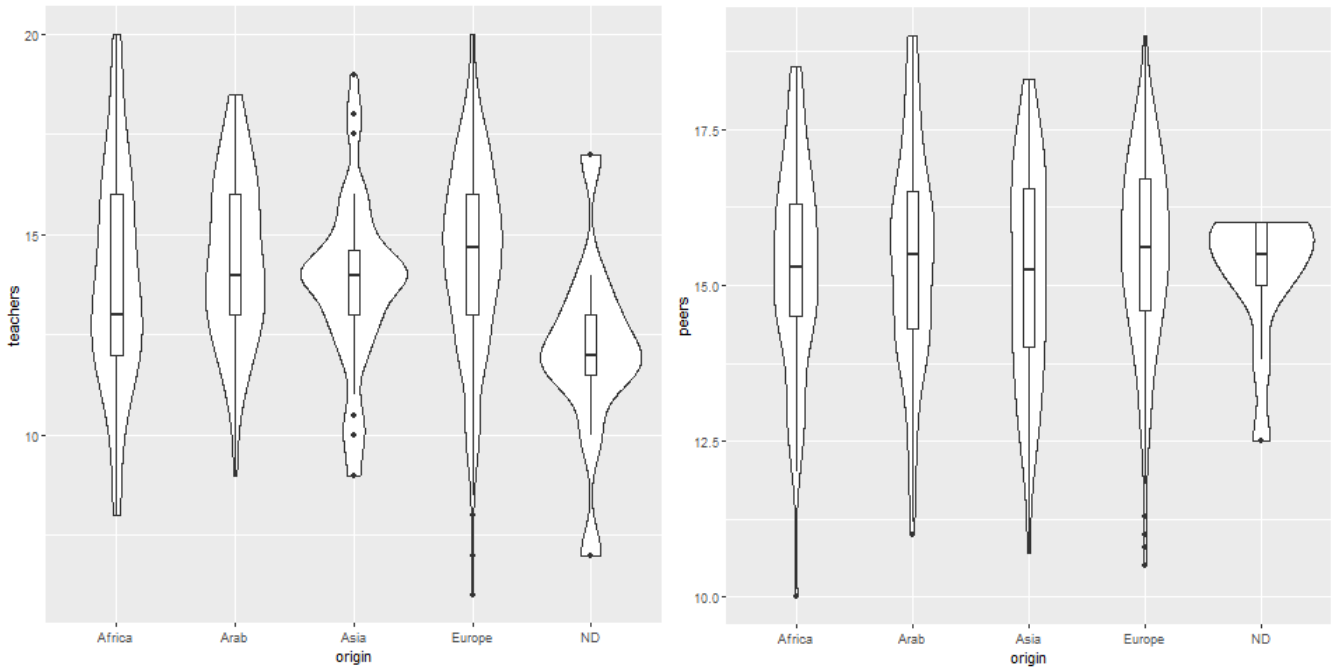

Figure 6. Violin plots showing the distribution of the grades as a function of origins, defined as four groups. Left, teachers' grades; right, peers' grades.

Conclusion

The evaluations appear not to be biased in terms of gender but might be detrimental to visible minorities. This bias however does not appear to be robust, i.e. it is seen in either teachers or peers depending on the clustering used. It is somewhat reassuring that no bias could be formally demonstrated, but we would nevertheless also consider that they are not definitely ruled out. This issue should probably be kept in mind to be studied in different contexts and on even larger cohorts of students to characterize it further.

AUTHOR CONTRIBUTIONS

Conceptualization, Antoine Taly; Data curation, Hugo Lopez, Louis Vasselin and Antoine Taly; Funding acquisition, Antoine Taly; Investigation, Karine Le Barch, Mélanie Etheve-Quelquejeu, Samuela Pasquali and Antoine Taly; Project administration, Antoine Taly; Software, Hugo Lopez; Writing – original draft, Antoine Taly; Writing – review & editing, Hugo Lopez, Louis Vasselin and Karine Le Barch.

Corresponding Author

antoine.taly@ibpc.fr

Acknowledgments

Céline De Flori, Cristina Claver Sicilia, Oussama Benkhabcheche are acknowledged for their help with the evaluation grid.

REFERENCES

Adkins, D. R., & Lyon, J. S. (2012). Promoting Research to the Masses: Assessing the Impact of a Poster Walk. *International Journal for the Scholarship of Teaching and Learning*, *6*(2). https://eric.ed.gov/?id=EJ1135550

Altintas, N. N., Suer, A. Z., Sari, E. S., & Ulker, M. S. (2014). The Use of Poster Projects as a

Motivational and Learning Tool in Managerial Accounting Courses. *Journal of Education for Business*, *89*(4), 196–201. https://doi.org/10.1080/08832323.2013.840553

Antoine, taly, Vasselin, L., & Hugo, L. (2021). *Poster session grading* [Data set]. Zenodo. https://doi.org/10.5281/zenodo.5176789

Baird, B. N. (1991). In-Class Poster Sessions. *Teaching of Psychology*, *18*(1), 27–29. https://doi.org/10.1207/s15328023top1801_7

Bandyopadhyay, S., Boylan, C. T., Baho, Y. G., Casey, A., Asif, A., Khalil, H., Badwi, N., & Patel, R. (2022). Ethnicity-related stereotypes and their impacts on medical students: A critical narrative review of health professions education literature. *Medical Teacher*, *0*(0), 1–11. https://doi.org/10.1080/0142159X.2022.2051464

Beaud, S., & Noiriel, G. (n.d.). Race et sciences sociales. In *Http://journals.openedition.org/lectures*. Agone. Retrieved September 12, 2022, from https://journals.openedition.org/lectures/48106

Coulmont, B., & Simon, P. (2019). How do immigrants name their children in France? *Population Societies*, *565*(4), 1–4.

Cramer, L. (2021). Equity, Diversity and Inclusion: Alternative strategies for closing the award gap between white and minority ethnic students. *ELife*, *10*, e58971. https://doi.org/10.7554/eLife.58971

Gerczei, T. (2016). Impact of an In-Class Biochemistry Mini-conference on Students' Perception of Science. *Journal of Chemical Education*, *93*(9), 1521–1527. https://doi.org/10.1021/acs.jchemed.5b00612

Kadnikova, E. N. (2013). "Molecules-in-Medicine": Peer-Evaluated Presentations in a Fast-Paced Organic Chemistry Course for Medical Students. *Journal of Chemical Education*, *90*(7), 883–888. https://doi.org/10.1021/ed2003737

Kennedy, J. H. (1985). Poster presentations for evaluating laboratory coursework. *Journal of Chemical Education*, *62*(12), 1104. https://doi.org/10.1021/ed062p1104

Logan, J. L., Quiñones, R., & Sunderland, D. P. (2015). Poster Presentations: Turning a Lab of the Week into a Culminating Experience. *Journal of Chemical Education*, *92*(1), 96–101.

https://doi.org/10.1021/ed400695x

Lopez, H., & Vasselin, L. (n.d.). *GitHub - CyberCRI/PeersGraderApp: PeersGrader is an opensource webapp to ease the organisation of peer grading activities by teachers for classes.* GitHub. Retrieved February 8, 2022, from https://github.com/CyberCRI/PeersGraderApp

Marichalar, P. (n.d.). *Penser sans, penser contre. Au sujet de Race et sciences sociales de Stéphane Beaud et Gérard Noiriel.* 36.

Marino, R., Clarkson, S., Mills, P. A., Sweeney, W. V., & DeMeo, S. (2000). Using Poster Sessions as an Alternative to Written Examination—The Poster Exam. *Journal of Chemical Education*, *77*(9), 1158. https://doi.org/10.1021/ed077p1158

Mazières, A., & Roth, C. (2018). Large-Scale Diversity Estimation Through Surname Origin Inference. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, *139*(1), 59–73. https://doi.org/10.1177/0759106318778828

Menke, J. L. (2014). Implementation of Online Poster Sessions in Online and Face-to-Face Classrooms as a Unique Assessment Tool. *Journal of Chemical Education*, *91*(3), 414–416. https://doi.org/10.1021/ed400665n

Namsor: Name checker for gender, origin and ethnicity classification. (n.d.). Retrieved February 8, 2022, from https://namsor.app/

Panadero, E., & Alqassab, M. (2019). An empirical review of anonymity effects in peer assessment, peer feedback, peer review, peer evaluation and peer grading. *Assessment & Evaluation in Higher Education*, *44*(8), 1253–1278. https://doi.org/10.1080/02602938.2019.1600186

Rieke, A., Southerland, V., Svirsky, D., & Hsu, M. (2022). Imperfect Inferences: A Practical Assessment | 2022 ACM Conference on Fairness, Accountability, and Transparency. *FAccT'22*, 767–777. https://doi.org/10.1145/3531146.3533140

Sisak, M. E. (1997). Poster Sessions as a Learning Technique. *Journal of Chemical Education*, *74*(9), 1065. https://doi.org/10.1021/ed074p1065.2

Taly, A., Benkhabcheche, O., De Flori, C., & Sicilia, C. C. (2018, August 20). *Poster session evaluation grid* [Figure]. Figshare; figshare. https://doi.org/10.6084/m9.figshare.6984068.v1

Torvik, V. I., & Agarwal, S. (2016, March). *Ethnea -- an instance-based ethnicity classifier based on geo-coded author names in a large-scale bibliographic database: International Symposium on Science of Science*. http://hdl.handle.net/2142/88927

Widanski, B., Thompson, J. A., Foran-Mulcahy, K., & Abafo, A. (2016). Providing Students with Interdisciplinary Support To Improve Their Organic Chemistry Posters. *Journal of Chemical Education*, *93*(1), 93–97. https://doi.org/10.1021/acs.jchemed.5b00020

Wimpfheimer, T. (2004). Peer-Evaluated Poster Sessions: An Alternative Method to Grading General Chemistry Laboratory Work. *Journal of Chemical Education*, *81*(12), 1775. https://doi.org/10.1021/ed081p1775