

Assessing peer grading in a poster session

Hugo Lopez¹, Louis Vasselin², Karine Le Barch³, Melanie Etheve-Quellejeu³, Samuela Pasquali⁴, Antoine Taly⁵✉;

¹Center for Research and Interdisciplinarity; ²Licence Frontieres du vivant, Center for Research and Interdisciplinarity, Paris, France; ³Facult des Sciences Fondamentales et Biomedicales, Universite de Paris, Paris, France; ⁴Laboratoire de Cristallographie et RMN Biologiques, CNRS UMR 8015, Universite de Paris, Paris, 75006, France; ⁵Laboratoire de Biochimie Theorique, CNRS, Universite de Paris, Paris, France. Correspondence to: antoine.taly@ibpc.fr

Poster DOI: 10.6084/m9.figshare.8481932 — <https://github.com/CyberCRI/PeersGraderApp> — [@AntTaly](https://twitter.com/AntTaly)



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. Students are also evaluated for their ability to evaluate their peers, comparing their evaluations to those performed by others (teachers and peers). Data from three years allow us to analyze the quality of the evaluation performed by the students and the possibility to grade it: i) the dynamic range of grades given by students tends to be smaller than in the case of teachers; ii) the evaluation conducted by students appears not to be biased in terms of gender but could be detrimental to visible minorities.

Evaluation grid

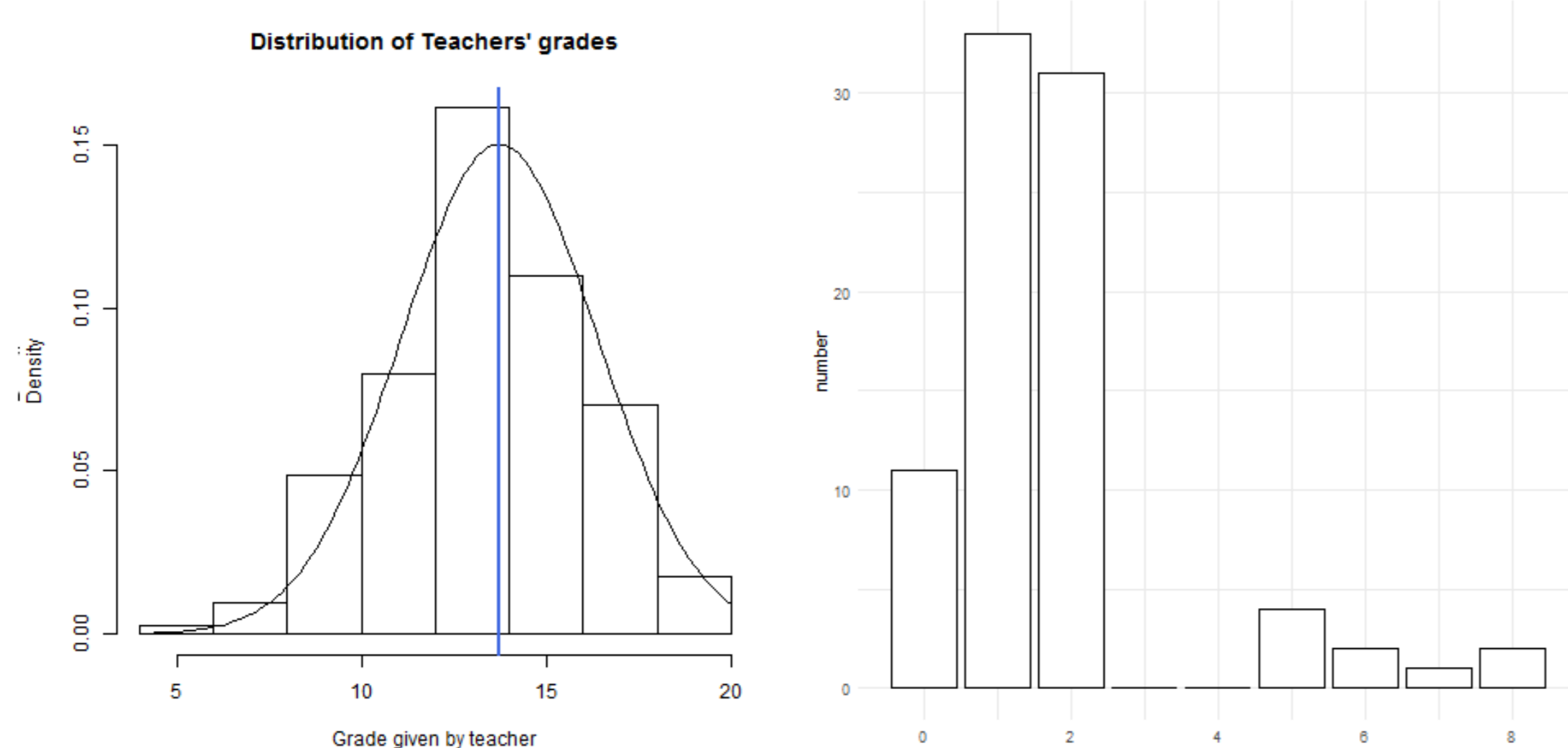
To facilitate and favor reproducible grading we use an evaluation grid.

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 contains 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

'Poster session rubric' by Céline De Flori, Cristina Claver Sicilia, Oussama Benkhacheche and Antoine Taly est mis à disposition selon les termes de la [licence Creative Commons Attribution - Partage dans les Mêmes Conditions 3.0 France](https://creativecommons.org/licenses/by/4.0/).

Teachers Evaluations

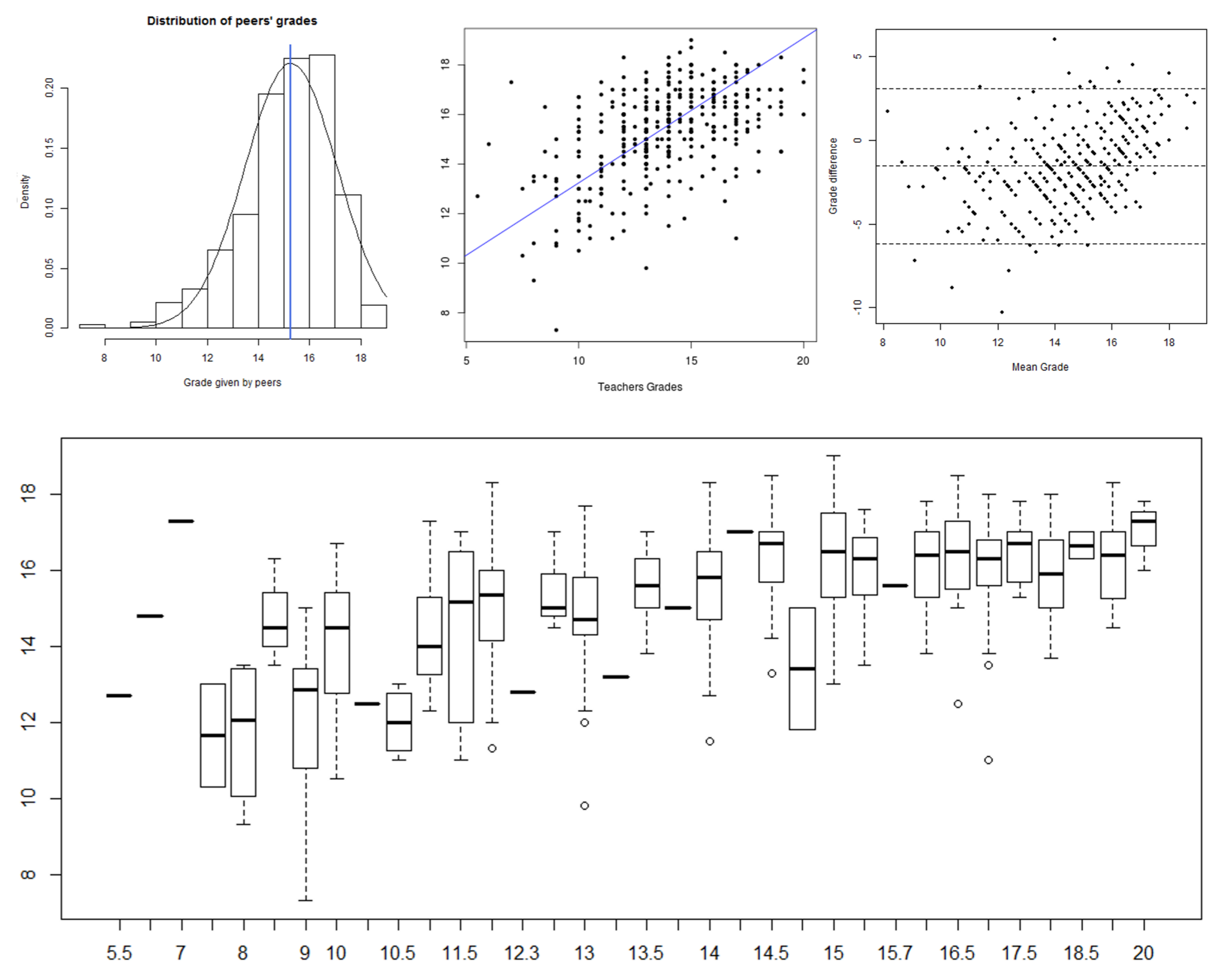
The dynamic range of grades given by teachers is large ($s_T = 2.7$), including failing grades (< 10). When two teachers are evaluating a student their grades are separated by less than 2 points in 90% of the cases.



Quantitative analysis of grades given by teachers in poster sessions. Left distribution of grades ($m_T = 13.7$, $s_T = 2.7$). right variability of grades when two teachers grade a single student.

Peers evaluations

The dynamic range of grades given by peers is smaller ($s_P = 1.8$ vs $s_T = 2.7$), in addition the grades are higher on average ($m_P = 15.3$ vs $m_T = 13.7$).



Quantitative analysis of grades given by peers in poster sessions. Left distribution of grades ($m_P = 15.3$ and $s_P = 1.8$); middle correlation of mean grades given by peers and teachers. A regression line is shown ($R^2 = 0.25$); Right Altman and Bland representation of grades given by peers versus teachers; Bottom average grades given by peers as a function of teachers' grades for a single group.

Detectable Biases?

We tested whether the poster session suffered from gender bias. The results suggest that there is no gender bias.

	Females	Males	P - value
Peer's grades	15.40	15.06	0.108
Teacher's grades	13.92	13.36	0.050
Written exam	8.51	8.61	0.8576

Test of gender bias Test of gender bias in Teacher's and Peer's grades with the (anonymous) written exam as a control.

We tested whether the poster session suffered from bias towards minorities. The results do not show effect on teachers' grades but suggest that there could be a bias towards minorities in Peer's grades.

individ	averageProfs	averagePeers	normalness	finalGrade	ethnicity
G01a	15.7	16.7	15.5	15.9	FRENCH
G01b	17.3	17.7	16.5	17.2	FRENCH
G01c	14.8	17.7	15.5	15.7	ARAB
G02a	12	12.5	16.5	13.3	INDIAN
G02b	11.5	13.3	16.5	13.2	HISPANIC
G02c	12	15.8	17	14.2	INDIAN

	Teachers grades		Peer grades	
		P- values		P-values
French names	14.50		15.41	
Foreigner names	14.47	0.915	15.07	0.043
Arab names	14.47	0.944	15.10	0.191
African names	14.66	0.758	15.26	0.684

Bias towards minorities. Comparison of mean grades as a function of origins (inferred from names)

We note however that it is difficult to conclude given the relatively small number of individuals tested here and the grade distributions of teachers and peers.

References

- [1] Hugo Lopez, Karine Le Barch, Mélanie Ethève-Quellejeu, Samuela Pasquali, and Antoine Taly. Poster session with peer grading. 2018.
- [2] Antoine Taly, Ccile de Flori, Christina Claver Sicilia, and Oussama Benkhacheche. poster session evaluation grid. 8 2018.
- [3] Antoine Taly, Francesco Nitti, Marc Baaden, and S Pasquali. Molecular modelling as the spark for active learning approaches for interdisciplinary biology teaching. *Interface focus*, 9(3):20180065, 2019.

Funded by

