Title “Learning by assessing: Using online student peer-assessment to improve students’ learning in political sciences”
Authors: Pierre Baudewyns, Françoise Docq, Vincent Legrand, Marianne Michel, Isabelle Motte, Gaëlle Pellon, Min Reuchamps, Dorothée Vandamme.

Introduction

Student peer-assessment (SPA) is known as an innovative collaborative method of learning. While the values of SPA are wildly accepted (Topping, 1998; Cho & al. 2006), it can be problematic for teachers to implement this kind of method in large classrooms and to insure that students accept and benefit of it [add references]. In the Belgian context, implementing student peer-assessment at the Bachelor level is challenging in multiple ways: First, while the number of universities using this kind of method is increasing[1], Belgian universities have still to take the step. The experience presented in this article will, therefore, be of particular interest in terms of sharing of practices. Second, the use of SPA as a learning and assessment method means to challenge student's learning habits and to ask them to replace individual and competitive learning for collaborative learning. Both challenges raise sets of questions about the institutional, personal or technological capacity to adapt [add references]. Among all the issues raised by student peer-assessment, the question of how students benefit of this practice still needs to be studied. This contribution will focus on this specific question and will discuss the implications of implementing SPA in first year bachelor students in political sciences. Using the data of a pre and post questionnaire and of students’ results, our main argument is that online SPA can significantly help students to improve their learning outcomes.

This working-in progress paper will be divided into three parts. The first part briefly presents the project implemented (objectives and design). The second part presents the research that has been carried out. The results of the research will be presented in the third part. The last part is dedicated to a more general discussion on student peer-assessment.

I. General description of the project (Gaëlle) and research design
II. Implementation of the SPA module (Dorothée & Marianne)
III. Questionnaire (Gaëlle Françoise)

À ce stade, peut-être pas nécessaire – dans la section suivante, on revient sur chacune des questions utilisées

IV. Results (Pierre Min Vincent)

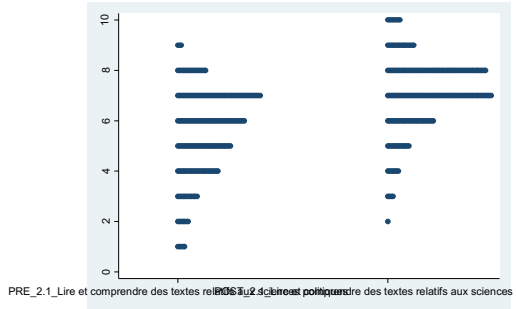
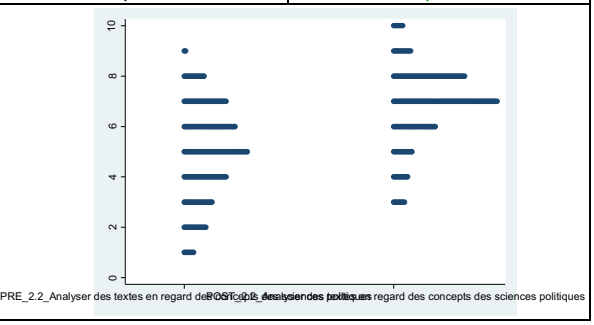
To assess the impact of the SPA on the students, we analyze the responses of the students to the pre- and post-questionnaire. Before proceeding to this analysis, we should first look at the grades obtained by the students via the SPA. The mean score is 14,45 out of a possible maximum of 20, with a standard deviation of 4,05. We should also note that the distribution is skewed towards the right end: final grades of 15, 16 and 17 are indeed the most common grades, accounting respectively for 21%, 32% and 22%. The introduction of SPA in the course assessment brought the

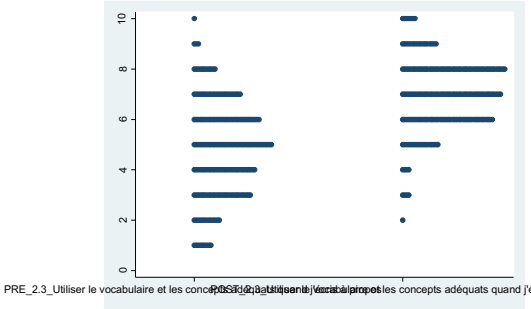
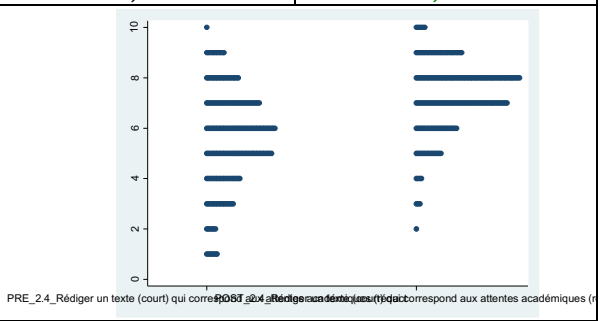
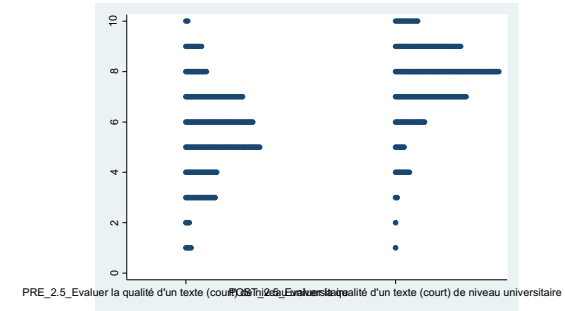
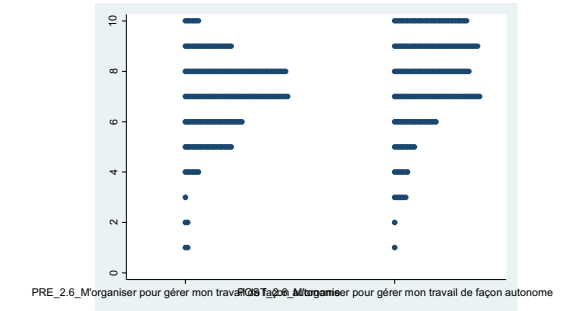
mean up to 11,88 out of 20, compared to 9,08 the year before without SPA – but of course such comparison is always risky and does not tell anything about the evolution of students' abilities and representations to which we turn now.

The second step is to explore the evaluation of SPA by the students. To this end, we analyze the results of the pre- and post-questionnaire. 530 students did take part to this survey. Nonetheless, only 193 of them (see table below) did both of them. We focus on this sub-sample for the following analyses, as we are interested in assessing the influence of SPA and thus comparing their responses before and after the SPA.

Questionnaires	**Frequency**	**Percent**
Only pre-test	222	41,89
Only post-test	115	21,70
Both pre- and post-test	193	36,42
N = 530		

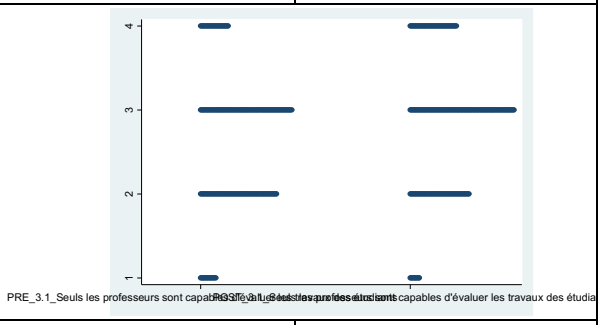
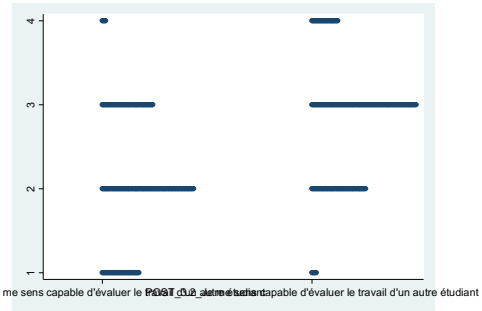
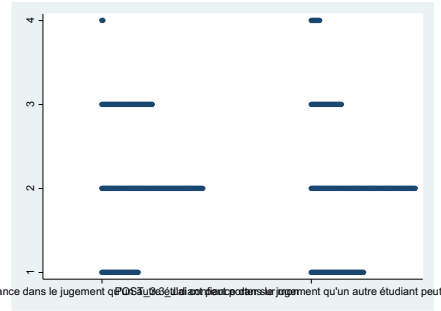
A first way to assess the impact of SPA is to compare the perceptions of the students about their – feelings of – abilities before and after SPA. We asked them to rate their position regarding six abilities: their ability to read and understand political science texts, their ability to analyze texts in light of political science concepts, their ability to use appropriate vocabulary and concepts when writing political science texts, their ability to write (short) texts meeting academic expectations (i.e. academic writing), their ability to assess (short) academic texts, and finally their ability to organize themselves to manage their work with autonomy. We present the distribution for each item on a 0 to 10 scale, both before and after SPA. We also report the mean before and after; the mean after is colored in green when there is a positive difference and in red when there is a negative difference.

Abilities	**Before**	**After**
Reading and understanding		
Mean	5,60	7,10
Analyzing		

Mean	5,05	6,87
Using appropriate vocabulary and concepts		
Mean	4,88	6,93
Writing		
Mean	5,57	7,31
Assessing		
Mean	5,55	7,61
Organizing		
Mean	6,97	7,66
N = 175		

On all six items, there is a positive difference ranging from +0,69 for the ability to organize oneself (which was already before SPA the highest item at 6,97) and +2,06 for the ability to assess academic texts (which is after SPA the highest item at 7,61). It can then be argued that SPA had a positive impact for the students in terms of their abilities. It is probably unsurprising but these results show the strongest positive effect on abilities directly related to SPA: ability to assess (+2,06), ability to use appropriate vocabulary and concepts (+2,05), ability to analyze (+1,82), ability to

write (+1,74). In addition to the perception of their abilities, we also surveyed the students about their representations both before and after SPA. More specifically we asked to them position themselves on a four-points scale about three items: only professors are able to assess students' works, I am able to assess other students' works, I am confident in other students' ability to assess my work.

Representations	**Before**	**After**
Only professors are able to assess		
Mean	2,65	2,85
I am able to assess		
Mean	2,14	2,82
I am confident in others' ability to assess		
Mean	2,09	2,00
N = 175		

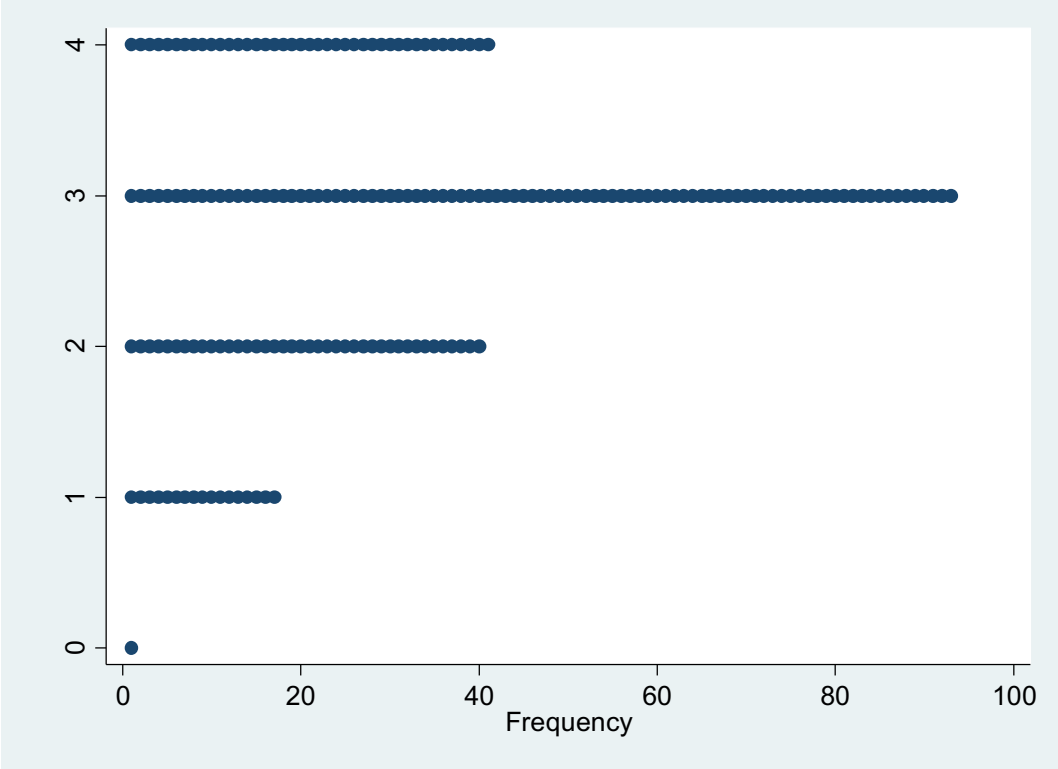
In the pre-questionnaire, the highest mean is 2,65 for the professors and the lowest is 2,09 for the peers' judgment, with a mean of 2,14 for oneself. The latter comes thus closer to the representation of others' ability than to the representation of professors' ability. The analysis of post-questionnaire brings a twofold finding. On the one hand, the gap between the first two items is almost closed as the representation of one's ability goes up to 2,82, whereas professors get 2,85. At the end of the SPA process, participants do therefore feel there are now as able as professors to assess academic works. This is an important value that brings about an important change in their representations of assessment. On the other hand, the representation of their peers' ability went slightly done from 2,09 to 2,00. While this

negative impact should not be too emphasized, as the difference is quite tiny, it is also an important finding and it should be kept in mind when implementing SPA. In the discussion, we will propose ways to prevent this risk. Before turning to the discussion, we should also look at the evaluation of the exercises and their effects (on a four-points scale).

Item	**Mean**
Evaluation of the exercises	
I found the exercises interesting	3,29
It was easy to use Moodle	3,20
I have clearly perceived the link between the exercises and the lectures	2,93
The workload was about 4 hours/week	2,97
I have assessed peers' exercises with fairness	3,44
The evaluations I have received were fair	2,45
Assessing peers' works became easier throughout the four exercises	3,13
Effects of the exercises	
Exercises helped me better understand the course's content	2,87
Exercises helped me better understand how to write a response at university	3,03
Feedbacks from peers helped me improve my works	2,20
Assessing peers' works helped me improve my own works	2,95
N = 190; in green means above 3,00 and in red means below 2,50	

From the evaluation of the exercises, four items stand out: I have assessed peers' exercises with fairness (3,44), the exercises were interesting (3,29), it was easy to use Moodle (3,20) and assessing peers' works became easier throughout the four exercises (3,13). It is interesting to note, in line with the findings from the representations, that the highest point goes to one's assessment of other's works. Students have done their best to assess their peers' works with fairness. By contrast, they feel the evaluations they have received were not as fair (2,45). Similarly, they think the feedbacks from peers did not help them so much to improve their works (2,20). Nonetheless the overall evaluation of the exercises and their effects are quite positive. The final question about their global evaluation goes in the same direction.

Figure 1: Global evaluation (N = 192, Mean = 2,81)



Out of almost 200 students, half of them gave a score of 3 out of 4 as their general evaluation of SPA. Almost the same numbers were very positive (4) and quite negative (2). Less than 10% were very negative about SPA. In total the mean score is 2,81. A further refinement of the analysis can be performed with the number of exercises completed by the students. As explained above, the students had to do three exercises but could also do a training one and an extra one. Therefore in total they could do five, but the target was to do four, this is why we have merged 4 and 5 together. Most students (72,63) are in this category, but we find several students who did not complete four or more exercises.

How many exercises?	**Frequency**	**Percent**
0	1	0,53
1	12	6,32
2	7	3,68
3	32	16,84
4 or 5	138	72,63
N = 190		

We therefore need to investigate whether this might influence their perceptions about their abilities and their representations. It indeed matters. If we compare the mean from the students who did between 0 and 3 exercises and those who did 4 or 5, we find a statistically significant difference on two items – already – before SPA: reading and understanding, on the one hand, and analyzing, on the other hand. This finding brings us outside of the scope of this paper, but it reminds us of the

inequalities among the students. After SPA, one item shows a statistically significant difference: the ability to organize oneself. It comes as no surprise that it has an impact of the number of exercises completed. The goods news is probably that students who did less than four exercises do acknowledge this weakness.

Abilities	**Before**			**After**		
	0-3	4-5	P>|t|	0-3	4-5	P>|t|
Reading and understanding	***5,16***	***5,74***	***0,045***	6,90	7,24	0,669
Analyzing	***4,56***	***5,20***	***0,048***	6,62	6,97	0,638
Using appropriate vocabulary	4,60	4,92	0,383	6,82	7,00	0,594
Writing	5,34	5,69	0,640	7,46	7,24	0,202
Assessing	5,58	5,60	0,792	7,86	7,50	0,070
Organizing	6,82	7,05	0,214	***7,34***	***7,81***	***0,029***
N = 179						

The picture is quite different for the representations where students who will/did not complete all the exercises tend to have a more positive representation on all three items but only one difference is statistically significant: the judgment about others' ability to assess. This item comes again to the front and shows that before SPA students who did 0 to 3 exercises have a more positive view (2,26 vs. 2,02) that decreases a bit more than for the students who did 4 or more exercises (2,02 vs. 1,95).

Representations	**Before**			**After**		
	0-3	**4-5**	**P>|t|**	**0-3**	**4-5**	**P>|t|**
Only professors are able to assess	2,68	2,61	0,804	2,79	2,88	0,134
I am able to assess	2,17	2,13	0,956	2,81	2,84	0,478
I am confident in others' ability to assess	**2,26**	**2,02**	**0,031**	**2,02**	**1,95**	**0,003**
N = 179						

We perform similar analyses of comparisons of means for the other items discussed in this paper, but the differences are not statistically significant. The fact of completing all the exercises has therefore only an impact on same items of the abilities and the representations.

V. Discussion (all together)

Bibliography

Cho, K., Schunn, C. D., & Wilson, R. W. (2006). Validity and reliability of scaffolded peer assessment of writing from instructor and student perspectives. Journal of Educational Psychology, 98(4), 891.

Topping, K. (1998). Peer assessment between students in colleges and universities. Review of Educational Research, 68(3), 249-276.