UCL
Université catholique de Louvain

LEARNING BY ASSESSING: USING ONLINE STUDENT PEER-ASSESSMENT TO IMPROVE STUDENTS' LEARNING IN POLITICAL SCIENCE

BAUDEWYNS Pierre [1], DOCQ Françoise [2], LEGRAND Vincent [3], MICHEL Marianne [4], PELLON Gaëlle [5], REUCHAMPS Min [6], VANDAMME Dorothée [7]

ISPOLE Institut de Sciences Politiques Louvain Europe

Working Papers

N° 10 - 2015

1 Institut de Sciences Politiques Louvain-Europe
2 Institut de Pédagogie universitaire et des Multimédias
3 Institut de Sciences Politiques Louvain-Europe
4 Entité fonctionnelle en charge des services généraux du Système d'Information
5 Institut de Sciences Politiques Louvain-Europe
6 Institut de Sciences Politiques Louvain-Europe
7 Institut de Sciences Politiques Louvain-Europe

La Série « Working Papers ISPOLE » a pour objectif de promouvoir le développement et la diffusion des travaux de recherche sur « le, la et les politiques » tant au niveau belge qu'européen et international. Les thématiques traitées touchent le pouvoir, les identités politiques et les phénomènes de mémoire collective, les institutions, les acteurs et les comportements politiques (vote, participation …), les conflits internes et internationaux et les politiques publiques (administration et management public). Une attention particulière est portée aux différentes méthodologies qualitatives et quantitatives d'analyse des phénomènes politiques tant en termes comparatifs que d'études de cas.

Il s'agit d'offrir un espace de valorisation des recherches en cours menées au sein de notre Institut et au-delà. Les « Working Papers ISPOLE » permettent la diffusion des travaux des membres de l'Institut ou chercheur-e-s et professeur-e-s participants ou étant associés à ses activités. Leur ambition est également de faire connaître certains des meilleurs travaux des doctorant-e-s de l'Institut.

Dans le but de mettre à la disposition de la communauté scientifique des recherches originales, cette série de Working Papers est conçue comme une catégorie intermédiaire de publication permettant aux chercheur-e-s de rendre accessibles des travaux d'intérêt et de qualité, qui pourront par la suite faire l'objet d'une publication conventionnelle (ouvrage, article dans une revue scientifique).

Les textes publiés sont signés et référencés (DIAL UCL). Ils sont publiés et archivés sur le site internet de l'ISPOLE. Les auteur-e-s conservent leurs droits sur les textes.

Responsable éditoriale : Dr. Virginie Van Ingelgom

Gestionnaire de la collection : Mme Karine Verstraeten

Les working papers ISPOLE sont disponibles en ligne sur le site : *www.uclouvain.be/402389*

Notre site : *www.uclouvain.be/ispole*

Récemment parus :

- Follebouckt, X. (2013). « "Nous devons parler d'une seule voix". Le Parlement européen face à la Russie : délibérations et positionnement (2004-2009) », *ISPOLE Working Paper n°5*, Louvain-la Neuve : Institut de Sciences Politiques Louvain-Europe, UCL.
- Duchesne, S (2013). « L'entretien collectif comme méthode expérimentale d'objectivation de l'identité nationale », *ISPOLE Working Paper n°6*, Louvain-la Neuve : Institut de Sciences Politiques Louvain-Europe, UCL.
- Laborderie V. (2014). « La relation entre une scission de BHV et d'hypothétiques frontières d'Etats post-Belgique », *ISPOLE Working Paper n°7*, Louvain-la Neuve : Institut de Sciences Politiques Louvain-Europe, UCL.

Si vous désirez de plus amples informations sur la Série des Working Papers ISPOLE, nous vous invitons à contacter la gestionnaire de la collection, Mme Karine Verstraeten (Karine.Verstraeten@uclouvain.be)

LEARNING BY ASSESSING: USING ONLINE STUDENT PEER-ASSESSMENT TO IMPROVE STUDENTS' LEARNING IN POLITICAL SCIENCE

Abstract:

While the value of student peer-assessment (SPA) is now largely acknowledged, it can be problematic for teachers to implement such a method in large groups and to ensure that students accept and benefit from it. This *Working Paper* focuses on this issue learning from the experience of implementing SPA with first year bachelor students following an introductory course in political science at the *Université catholique de Louvain* (UCL). To this end, this article is twofold: on the one hand, it presents the project that was implemented both in its objectives and its design and, on the other hand, it discusses its impact on the students, as measured by original data collected along the implementation of SPA. The findings show that online SPA has a positive impact for the students in terms of increasing their (perceived) abilities. The research also delivers interesting results concerning students' post-SPA perceptions of their ability to assess their peers' works and of their peers' ability to assess their own works.

Résumé:

Alors que la valeur de l'évaluation des étudiants par leurs pairs est à présent largement reconnue, les enseignants peuvent rencontrer des difficultés à appliquer une telle méthode dans de grands groupes, à garantir l'adhésion des étudiants à celle-ci et s'assurer qu'ils en bénéficient. Ce *Working Paper* adresse cette problématique en tirant profit d'une expérience d'évaluation par les pairs mise en œuvre dans un cours d'introduction à la science politique s'adressant à des étudiants de première année à l'Université catholique de Louvain (UCL). À cette fin, l'article a un double volet : d'une part, il expose le projet en question tant dans ses objectifs que dans son architecture et, d'autre part, il débat de son impact sur les étudiants, tel qu'il ressort des données originales collectées durant sa mise en œuvre. Les résultats démontrent que l'évaluation des étudiants par les pairs a un impact positif sur ceux-ci en termes d'accroissement de leurs compétences (perçues). La recherche a également produit des résultats riches d'enseignement au sujet des perceptions des étudiants quant à leurs capacités à évaluer leurs pairs et vice versa après l'expérience.

Citation :

Baudewyns, P., Docq, F., Legrand, V., Michel, M., Pellon, G., Reuchamps, M., Vandamme, D. (2015), « Learning by assessing: Using online student peer-assessment to improve students' learning in political science », ISPOLE Working Papers, n° 10, Louvain-La-Neuve : Institut de Sciences Politiques Louvain-Europe, UCL

1. Introduction

Student peer-assessment (SPA) is known as an innovative collaborative method of learning. The classical definition was provided by the seminal work of Topping (1998). He defined student peer-assessment as "an arrangement in which individuals consider the amount, level, value, worth, quality, or success of the products or outcomes of learning of peers of similar status". Since then, student peer-assessment has been extensively used to improve students' performance, reflective writing skills, oral presentations, peer-group evaluation, to increase students' attendance and engagement, to develop cognitive skills and to stimulate critical thinking. While the value of SPA is now largely acknowledged (Topping, 1998; Cho & al., 2006), it can be problematic for teachers to implement such a method in large groups and to ensure that students accept and benefit from it.

In the Belgian context, implementing SPA at the Bachelor's level is challenging in multiple ways. First, while this practice has now received some attention, Belgian universities still lag a bit behind. Second, the use of SPA as a learning and assessment method means to challenge students' learning habits and to ask them to replace individual and competitive learning by collaborative learning. Both challenges raise several questions about the institutional, personal or technological capacity to adapt. Among all the issues raised by SPA, the question of how students benefit from this practice still needs to be studied. This contribution will therefore focus on this specific question and will discuss more specifically the implications for students of implementing SPA with first year bachelor students following an introductory course in political science at the *Université catholique de Louvain* (UCL). Relying on a pre- and post-questionnaire data and on students' results, our main argument is that online SPA can significantly help students to improve their learning abilities. To this end, this article is twofold: on the one hand, we present the project that was implemented both in its objectives and its design and, on the other hand, we discuss its impact on the students, as measured by original data that was collected along the implementation of SPA.

2. The promises and pitfalls of student peer-assessment

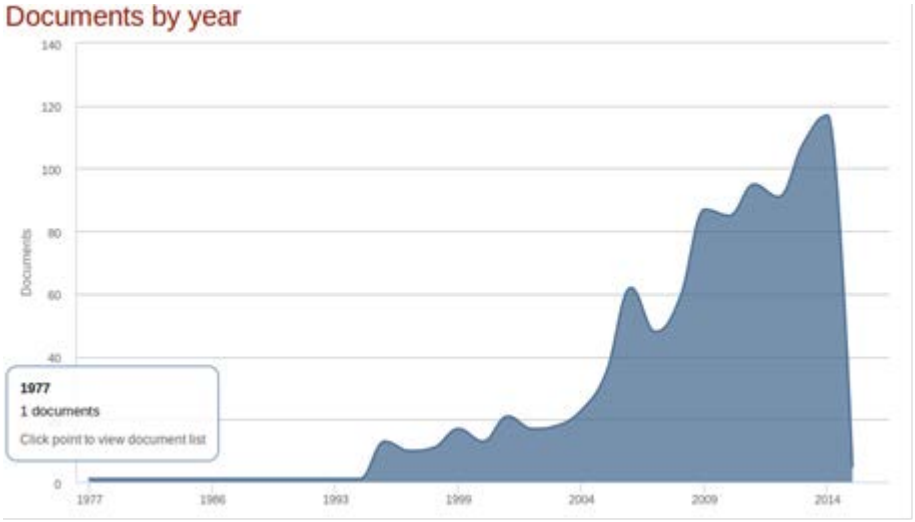
Student peer-assessment has become a popular object of research in all sectors of higher education. As displayed in Figure 1, the number of scholarly articles keeps growing since the 1990s. Since the publication of Topping in 1998 on 'Peer assessment between students in colleges and universities', publications on SPA have tripled. In 2014, more than one hundred articles were published on this subject.

The literature on the subject is, however, very unequally distributed (see Figure 2 2). Most of the articles come from Anglo-Saxon countries. The top five countries in terms of number of published articles with keywords student peer-reviewing or peer-assessment are the United States, the United Kingdom, Australia, Taiwan and Spain. Belgium lags behind with less than 20 published articles as listed by the Scopus database in 2014.

Among all the fields using and researching on student peer-reviewing, social sciences is the leading field of research with more than 63% of the total number of publications. The 5 main social sciences journals publishing on SPA include education journals *Assessment and Evaluation in Higher Education* (72 articles published in 2014), *American Journal of Pharmaceutical Education* (17), *Medical Education* (16), *Computers and Education* (15) and *Innovations in Education and Teaching International* (14). As SPA is most of the time a

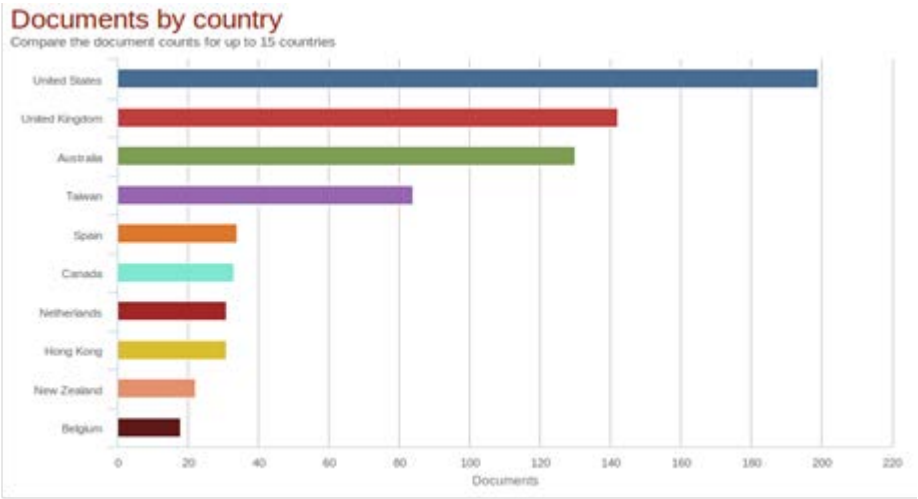
computer-assisted learning tool, the field of computer sciences is the second most important field of research with 26,6% of the total number of published articles.

Figure 1. Publications on student peer-assessment from 1977 to 2014



Source: Scopus Database

Figure 2. Number of publications on student peer-assessment per country



Source: Scopus Database

In social sciences SPA has triggered off an increasing number of publications investigating the promises and the pitfalls of this learning tool. In 1998, Topping made the following statement: "Given the many different types of peer assessment, establishing a single overarching theory or model of the process seems likely to be difficult". Sixteen years later, Topping's statement is still valid. Indeed, in social sciences, scholars have explored a great variety of student peer-assessment dimensions. The main research question concerns the impact of SPA on students' performance. Most of the scholars investigating this question have pointed out the beneficial effects of SPA on students' performance. The main argument developed in the literature concerns the impact of the change of role. Evidence (Bloxham & West, 2004; Sitthiworachart, & Joy, 2008; Chang & Tseng, 2009, Gove, 2012; Karandinou,

2012; Birjandi, & Hadidi Tamjid, 2012; Lynch, McNamara & Seery 2012; Li, Liu & Zhou, 2012; Young, 2012) shows that changing roles from being assessed to being an assessor improves students' ability to judge and improve their own work. Although the positive effects of SPA is acknowledged by most of the scholars, some authors have also called attention to the potential pitfalls. For instance, Gielen & al. (2011) investigate the very concept of performance used in research on SPA. According to their study, students' performance should also be assessed on the basis of a number of quality goals such as the installation of social control; the preparation for self-monitoring and self-regulation in lifelong learning; and the active participation of students in the classroom.

The second largest cluster of publications aims at investigating the viability and reliability of SPA compared to teachers' assessment. The bulk of publications focuses on the comparison between the marks or grades awarded between peers compared to the mark a teacher would have given. The comparison is mostly quantitative in spite of the critics addressed against this minimalist approach of grading (Devenny, 1989). Evidence shows that student peer-assessment has a high degree of reliability. In 1998, Topping reported that 72% of the studies comparing teachers' and students' marks reported an acceptable degree of reliability. This result has been confirmed more recently (Ross, 2003 and Cho & al., 2006). In the context of the present contribution, this result is crucial since the design of peer-assessment relies on the idea that SPA is valid and reliable. The second part of the article explains in detail how this principle was applied and how students perceive this.

Students' perception is also the third most important question investigated in SPA research. Bozkurt & Demir (2013) show that students attaining had positive thoughts about SPA. A conclusion similar to Foley's study (2013) conducted at the International School in Lausanne or Pombo & al.'s study conducted at the University of Aveiro (2010). Most of the time, students value the use of peer-assessment in their education. However, Patton's study (2010) tempered this optimistic conclusion. He shows that while students support SPA as a formative assessment, they are more critical when SPA is used as a summative practice. A conclusion that rejoined our own conclusions, as will be set out below.

Apart from these questions, scholars have also focused in their studies on other questions such as the conditions under which SPA may be successful. Takeda & Homberg (2014) examined the effect of gender on cooperative learning systems such as SPA. They concluded that group gender composition has an impact on the success of students performance. Sasmaz's study (2012) demonstrates that female students received significantly higher mean scores than male students. A conclusion that puts into perspective the conclusion of Pope's own survey (2005) showing that females are more stressed by self-assessment than males. According to Van den Berg, Admiraa, & Pilot's study (2006), an optimal design of peer-assessment includes a combination of written and oral peer feedback. The quality of feedbacks given by students has also started off some very stimulating articles that liaise with our study. For Snowball & Mostert (2013), students give more useful feedback on technical aspects, such as presentation and referencing than on content. A conclusion close to what our project demonstrates. On the same subject, Cartney (2010) and Sadler (2010) examined the conditions under which the use of feedback by students can be effective, acknowledging that most of the time they do not make use of feedback they receive.

This short and non-exclusive review of the literature on SPA aimed at illustrating the benefits and the challenges of using SPA as a teaching tool. Clearly, the positive effects of SPA is widely acknowledged even if the conditions under which SPA design is optimal for students' performance still have to be precisely determined. More attention is also needed on how to define and to assess students' performance and learning. With our study, we aim at contributing to this debate. More particularly our study builds on the literature on students' perception on learning and on the reliability and validity of SPA. The present working paper also contributes to the debate on the use of feedbacks by peers and provides some conclusive reflections on Moodle's workshop peer-reviewing.

3. General description of the SPA project

In the beginning of the academic year 2013-2014 we introduced SPA in the introductory course of political science for first year bachelor students. With almost 1200 students, the course of political science is offered to most students in communication, economics, management, political science and social sciences. The course is credited with 5 ECTS and is organized with 2 hours per week of lectures given by three teachers during 15 weeks. Due to the number of students attending the course, the final evaluation was a multiple-choice questions exam only.

The SPA project was implemented to the benefit of around 500 students (on a total amount of 1200, i.e. students in communication, political science and social sciences[8]) in order to achieve three objectives: (1) students' performance, (2) students' motivation and (3) coherence of the curriculum. The last point refers to the 2010 final AEQES report assessing the UCL political science curriculum[9]. This report pointed at the absence of practical courses in political science. It further underlined the need for more innovative teaching methods and for an increase in practice-oriented courses: *"L'équilibre entre cours magistraux et travaux en petits groupes n'est sans doute pas complètement satisfaisant, au moins en Bachelier, mais parfois aussi en Master. S'il y avait davantage de travail en groupes restreints, il serait plus facile d'introduire une plus grande préoccupation du concret et une pédagogie plus active : études de cas pratiques et interventions orales ou exposés d'étudiants dès le Bachelier; en Master, techniques pédagogiques plus innovantes comme des simulations de négociations ou des jeux de rôles. Ce point renvoie à l'existence d'un regret souvent entendu selon lequel il y aurait trop de cours purement théoriques"*[10].

To achieve these goals, the SPA module was implemented on the e-learning MoodleUCL platform (moodleucl.uclouvain.be). This platform is based on the e-learning environment software "Moodle" that is a robust open-source platform and that is supported by a global community. Moodle offers several options for student-to-student and student-to-teacher interaction and the choice of the right Moodle module is depending on the level of structure and amount of students involved. The Moodle "Workshop" module offers the most structured activity for student-to-student interaction. The best way to proceed inside Moodle was firstly, to create a "Forum" where the students could meet each other, secondly, to create a "Wiki"

[8] The other 700 students (in economics and management) benefited from another learning innovation: an introductory MOOC on political science on edX (Louv3x: "Découvrir la science politique").

[9] AEQES is an independent public sector agency, practicing formative evaluation. The Agency was responsible for assessing the quality of higher education and working for its continuous improvement. See AEQUES 2010 final report. http://www.aeqes.be/documents/20100610%20RFS%20UCL.pdf

[10] *Op cit.* p 6.

where the students could work together and finally, to create the "Workshop" where the students could appreciate and assess each other's work.

A Moodle Workshop module is defined by several parameters. The following ones are those that were chosen for the present cases:

- *Grading strategy* determines the assessment form used and the method of grading submissions. There are 4 options: The option that was chosen was "Rubric": a level assessment is given regarding specified criteria.
- *Grade for submission* specifies the maximum grade that may be obtained for submitted work.
- *Grade for assessment* specifies the maximum grade that may be obtained for submission assessment.
- *Overall feedback mode*. If enabled, a text field is displayed at the bottom of the assessment form. Reviewers can put the overall assessment of the submission there, or provide additional explanation of their assessment.
- *Example submissions*. If enabled, users can try assessing one or more example submissions and compare their assessment with a reference assessment. The grade is not counted in the grade for assessment.
- *Mode of examples assessment*: voluntary, must be assessed before one's own submission, available after one's own submission and must be assessed before peer assessment.

The calculation of the Grade for submission depends on the Grading strategy that here is the Rubric strategy. In education terminology, scoring rubric means "a standard of performance for a defined population". Developmental rubrics return to the original intent of standardized developmental ratings, which was to support student self-reflection and self-assessment as well as communication between an assessor and those being assessed. In this new sense, a scoring rubric is a set of criteria and standards typically linked to learning objectives. It is used to assess or communicate about product, performance, or process tasks. A scoring rubric is an attempt to communicate expectations of quality around a task. In many cases scoring rubrics are used to delineate consistent criteria for grading. Because the criteria are public, a scoring rubric allows teachers and students alike to evaluate criteria, which can be complex and subjective. A scoring rubric can also provide a basis for self-evaluation, reflection, and peer review. It is aimed at accurate and fair assessment, fostering understanding, and indicating a way to proceed with subsequent learning/teaching. This integration of performance and feedback is called ongoing assessment or formative assessment.

Scoring rubrics include one or more dimensions on which performance is rated. Dimensions are generally referred to as *criteria*, the rating scale as *levels*, and definitions as *descriptors*.

Table 1. Parameters of the SPA workshops

	Workshop 0 "Training"	**Workshop 1**	**Workshop 2**	**Workshop 3**	**Workshop 4 (partial)**
Grading strategy	Rubric	Rubric	Rubric	Rubric	Rubric
Grade for submission	12	12	30	30	30
Grade for assessment	8	8	20	20	20
Overall feedback mode	Enabled and required				
Example(s) submission	4, after own submission and assessed	4, after own submission and assessed	3, after own submission and assessed	0	0
Submission attachment	1				
Assessment Criteria	6	6	**7**	**7**	**7**
Criteria 1 grading levels	0, 1, 2	0, 1, 2	**0, 1, 2, 3**	**0, 1, 2, 3**	**0, 1, 2, 3**
Criteria 2 grading levels	0, 1, 2	0, 1, 2	**0, 1**	**0, 1**	**0, 1**
Criteria 3 grading levels	0, 1, 2	0, 1, 2	**0, 1, 2**	**0, 1, 2**	**0, 1, 2**
Criteria 4 grading levels	0, 1, 2	0, 1, 2	**0, 1, 2**	**0, 1, 2**	**0, 1, 2**
Criteria 5 grading levels	0, 1, 2	0, 1, 2	**0, 2, 4, 8**	**0, 2, 4, 8**	**0, 2, 4, 8**
Criteria 6 grading levels	0, 1, 2	0, 1, 2	**0, 2, 4, 8**	**0, 2, 4, 8**	**0, 2, 4, 8**
Criteria 7 grading levels	---	---	**0, 3, 6**	**0, 3, 6**	**0, 3, 6**
Submission deadline	Yes, submission period of 10 days				
Assessment deadline	Yes, assessment period of 10 days				
Assessment comparison	Lax f = 1.67	Normal f = 2.50	Strict f = 3.00	Normal f = 2.50	Strict f = 3.00
Averages for submission (502 students)	---	**8.99 (445 sub.)**	**20.88 (447 sub.)**	**22.26 (438 sub.)**	**20.93 (66 sub.)**
Averages for assessments (502 students)	---	**7.00 (428 ass.)**	**15.63 (436 ass.)**	**17.57 (427 ass.)**	**15.47 (63 ass.)**

For example, see workshops 2, 3 and 4 of table 1, the assessment form was divided in seven criteria that were elaborated with the students in class (and reviewed according to students' observations):

1. formal criterion (grammar and spelling), on 4 levels, 3 points (levels 0, 1, 2, 3);
2. instruction compliance, on 2 levels, 1 point (levels 0, 1);
3. language register, on 3 levels, 2 points (levels 0, 1, 2);
4. Text structure, on 3 levels, 2 points (levels 0, 1, 2);
5. relevance of answer, on 4 levels, 8 points (levels 0, 2, 4, 8);
6. use of references, on 4 levels, 8 points (levels 0, 2, 4, 8);
7. use of theoretical knowledge, on 2 levels, 6 points (levels 0, 3, 6).

So, in these last cases (workshops 2, 3 and 4), we have:

- $1 \leq i \leq 7$, where i = each criteria consists of a set of ordered levels that assign a grade;
- N = 7 is the total number of criteria;
- g_i, $1 \leq i \leq 7$ are the levels, so the grade given by students for each criteria;
- max_i, $1 \leq i \leq 7$ are the maximum grade for each criterion.

The final *Grade for submission* is aggregated as:

$$G_s = \frac{\sum_{i=1}^{N} g_i}{\sum_{i=1}^{N} max_i};\ g_i,\ max_i \in \mathbb{R},\ N = 7$$

The calculation of the *Grade for assessment* is depending on the *Grading evaluation method* that here is the *Comparison with the best assessment*. The best assessment is determined by calculating, for all assessment dimensions ($1 \leq i \leq 7$), the arithmetic mean and sample standard deviation across all made assessments in one workshop instance.

The final *Grade for assessment* is aggregated as:

$$G_a(\%) = \left| 1 - \frac{f \times \sum_{i=1}^{N} \partial_i}{N} \right| \times 100\ ;\ N = 7$$

where the f factor determines the level of the assessment comparison (very strict, f = 5.00; strict, f = 3.00; normal, f = 2.50; fair, f = 1.67; very fair, f =1.00)

where the delta for each assessment dimensions ($1 \leq i \leq 7$) are equal to:

$$\partial_i = \left(\frac{\left| Best\_assessment_i - Peer\_assessment_i \right|}{max_i} \right)^2;\ max_i \in \mathbb{R},\ 1 \leq i \leq 7$$

3. Impacts of SPA on students

To assess the impact of SPA on the students, we analyze the student's responses 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 (there are a lot of factors changing from one year/class to another) and does not tell anything about the evolution of students' abilities and representations to which we turn now.

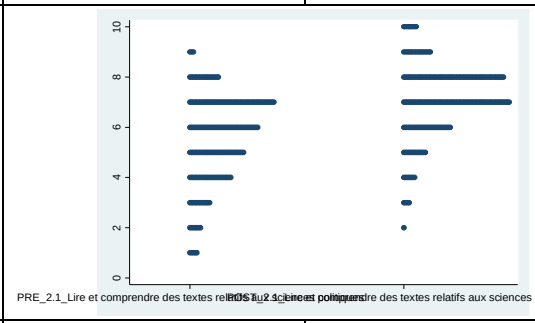
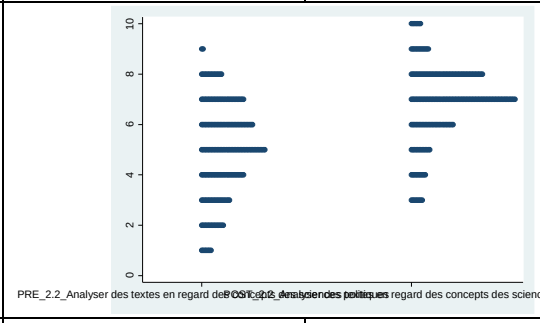
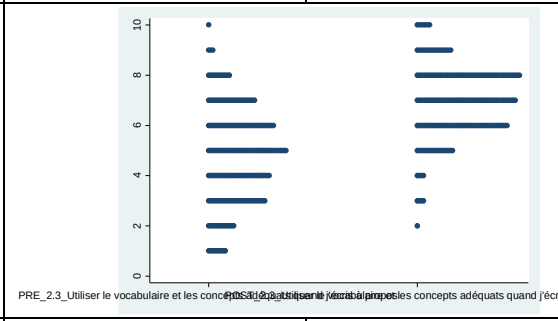
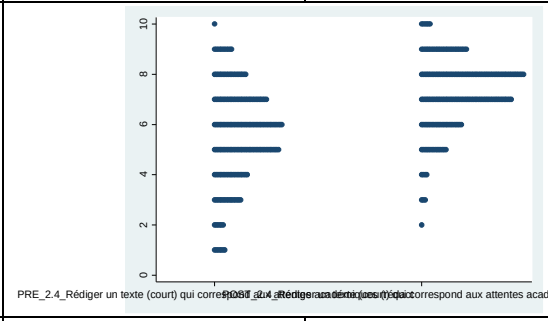
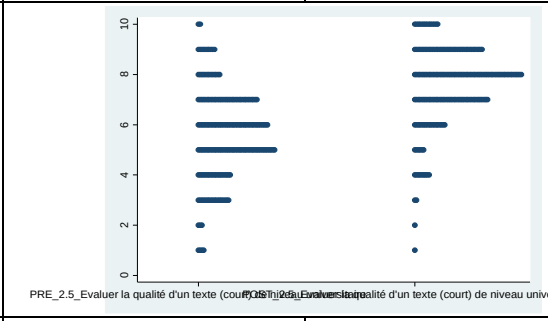
The second step is to explore the evaluation of SPA by the students. To this end, we analyze the results of the pre- and post-questionnaire. Over 500 students did take part to this survey. Nonetheless, only 193 of them (see table below) did both of them. We focus on this sub-sample for the following analyses, as we are interested in assessing the influence of SPA and thus comparing their responses before and after the SPA.

Table 2. Number of students who filled in the questionnaires

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 – feeling of – abilities before and after SPA. We asked them to rate their position regarding six abilities (Figures 3): 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.

Figures 3. Perceptions of the students about their abilities

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

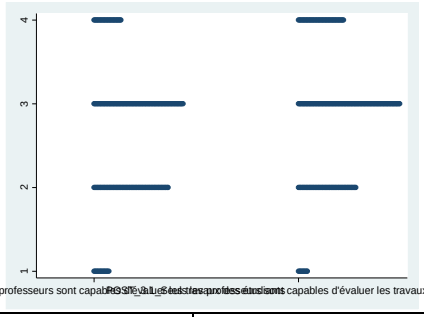
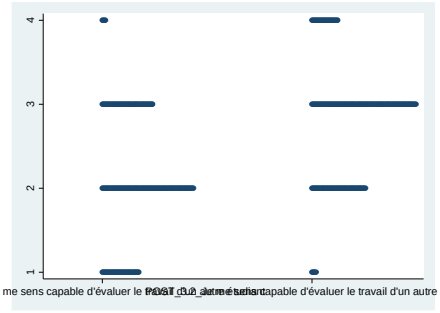
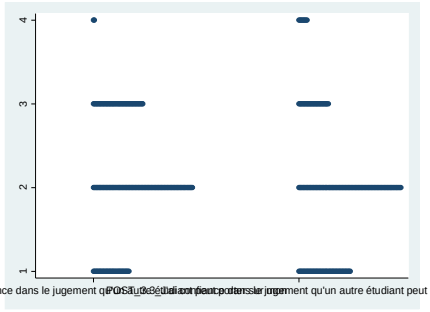
Organizing		
Mean	6,97	7,66
N = 175		

Note: All differences are statistically significant

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 7,66). It can then be argued that SPA had a positive impact for the students in terms of their (perceived) abilities. It is probably unsurprising but these results show the strongest positive effect on (perceived) 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 (Figure 4). More specifically we asked them to 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.

Figures 4. Representations of the students

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		

Note: All differences are statistically significant

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 the 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 nevertheless an important finding and it should be kept in mind when implementing SPA. Before bringing this

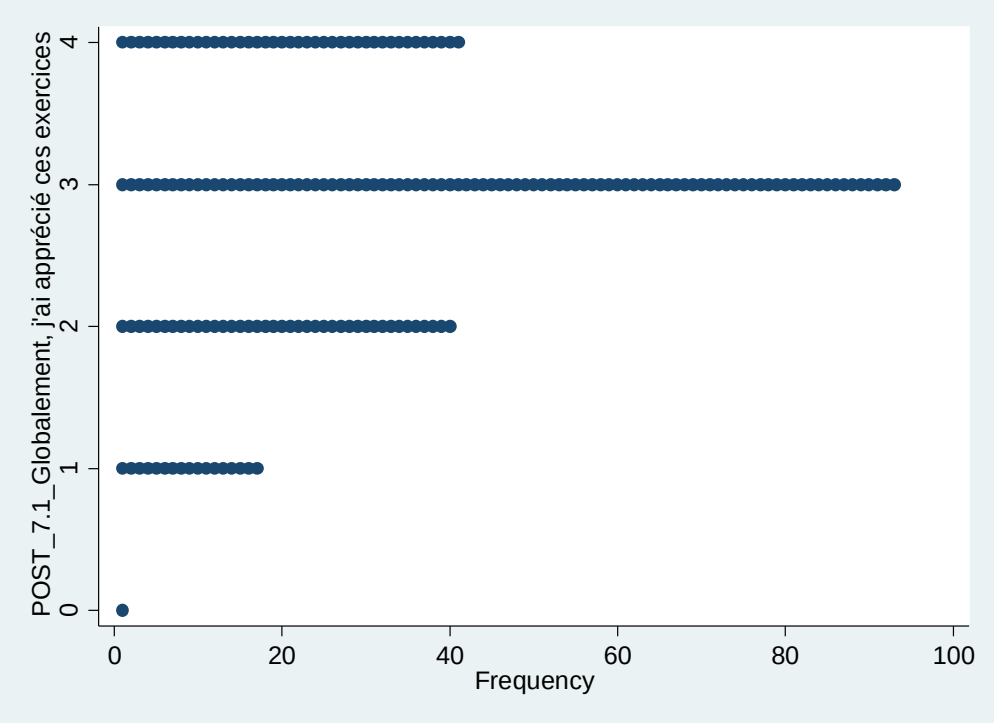
article to a draw, we also look at the evaluation of the exercises and their effects (on a four-points scale).

Table 3. Evaluation of the exercises and their effects

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	

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 mean goes to one's assessment of others' works meaning that students think they have done their best to assess. 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 5).

Figure 5. 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. While they had the opportunity to complete 5 exercises, only four of them were taken into account at the very end, this is why we've merged exercise 4 with exercse 5 . Most students (72,63) are in this category, but we find several students who did not complete four or more exercises.

Table 4. Number of exercises done by the students

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 the amount of completed exercvices might influence their perceptions about their abilities and their representations. It indeed matters. If we compare the mean students 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 is reminding 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 good news is probably that students who did less than four exercises do acknowledge this weakness.

Table 5. Perceptions about abilities by the number of exercises

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 globally 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). The fact of completing all the exercises has therefore only an impact on same items of the abilities and the representations.

Table 6. Representations by the number of exercises

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 performed similar analyses of comparisons of means for the other items discussed in this paper, but the differences were not statistically significant.

4. Conclusion

Student peer-assessment (SPA) is now often praised as an innovative collaborative method of learning around the world. But much is still to learn about how to best design such practice

for large groups of first year students and about its impact on the students themselves. On the basis of a large-scale project of SPA in an introductory course of political science this paper has sought to bring some insights about both questions.

The presentation and the discussion of the online design that was implemented emphasized the importance of a very carefully thought design, especially in a context of several hundreds of first year students. Such design cannot go wrong and transparency is a key factor for the success of SPA. If students cannot trust the design, they will be suspicious of the whole exercise itself. When such transparency can be offered, the results of the pre- and of the post-questionnaires show a statistically significant increase of their (perceived) abilities. This all leads to a positive global evaluation of SPA even if attention should be paid, on the one hand, to the nature of the assessment of their assessment (i.e. retro-evaluation) and, on the other hand, to their perceptions of their peers' abilities to assess their work. These are avenues for future research in student-peer assessment (SPA).

References

Birjandi, P., Hadidi Tamjid, N., "The Role of Self-, Peer and Teacher Assessment in Promoting Iranian EFL Learners' Writing Performance", *Assessment and Evaluation in Higher Education*, 2012, vol. 37(5), pp. 513-533.

Bloxham, S., West, A., "Understanding the Rules of the Game: Marking Peer Assessment as a Medium for Developing Students' Conceptions of Assessment", *Assessment and Evaluation in Higher Education*, 2004, vol. 29(6), pp. 721-733.

Chang, C.-C., Tseng, K.-H., "Use and Performances of Web-Based Portfolio Assessment", *British Journal of Educational Technology*, 2009, vol. 40(2), pp. 358-370.

Cho, K., Schunn, C. D. , Wilson, R. W., "Validity and Reliability of Scaffolded Peer Assessment of Writing from Instructor and Student Perspectives", *Journal of Educational Psychology,* 2006, vol. 98(4), pp. 891.

Devenney, R., "How ESL Teachers and Peers Evaluate and Respond to Student Writing", *RELC Journal: A Journal of Language Teaching and Research in Southeast Asia*, 1989, vol. 20(1), pp. 77-90.

Gielen, S. B., Dochy, F., Onghena, P., Struyven, K., Smeets, S., "Goals of Peer Assessment and their Associated Quality Concepts", *Studies in Higher Education*, 2011, vol. 36(6), pp. 719-735

Gove, S., "Increasing Student Engagement Using Client-Based Peer Assessment in Multi-Role, Whole-Enterprise Simulations", *Cutting-Edge Technologies in Higher Education*, 2012, vol. 6(PARTC), pp. 95-128.

Karandinou, A., "Peer-Assessment as a Process for Enhancing Critical Thinking and Learning in Design Disciplines", *CEBE Transactions*, 2012, vol. 9(1), pp. 53-67

Li, L., Liu, X., Zhou, Y., "Give and Take: A Re-Analysis of Assessor and Assessee's Roles in Technology-Facilitated Peer Assessment", *British Journal of Educational Technology*, 2012, vol. 43(3), pp. 376-384.

Lynch, R., McNamara, P., Seery, N., "Promoting Deep Learning in a Teacher Education Programme through Self- and Peer-Assessment and Feedback", *European Journal of Teacher Education*, 2012, vol. 35(2), pp. 179-197.

Ross, J. A., *The Reliability, Validity, and Utility of Self-Assessment*, University of Maryland, 2006.

Sadler, D., "Beyond Feedback: Developing Student Capability in Complex Appraisal", *Assessment and Evaluation in Higher Education*, **2010**, Vol. 35(5), pp. 535-550

Sitthiworachart, J. B. and Joy, M., "Computer support of effective peer assessment in an undergraduate programming class", *Journal of Computer Assisted Learning*, 2008, vol. 24(3), pp. 217-231.

Snowball, J., Mostert, M., "Dancing with the Devil: Formative Peer Assessment and Academic Performance", *Higher Education Research and Development*, 2013, vol. 32(4), pp. 646-659.

Topping, K., Peer Assessment between Students in Colleges and Universities, *Review of Educational Research*, 1998, vol. 68 (3), pp. 249-276

Young, C., "Initiating self-assessment strategies in novice physiotherapy students: A method case study", *Assessment and Evaluation in Higher Education*, 2013, vol. 38(8), pp. 998-1011.