CHAPTER 8. ENHANCING FORMATIVE ASSESSMENT AS THE WAY OF BOOSTING STUDENTS' PERFORMANCE AND ACHIEVING LEARNING OUTCOMES

Nikita Minin, Masaryk University

Coach: Eszter Simon, Masaryk University

Keywords: audio feedback, Bloom's taxonomy, constructive alignment, formative assessment, international relations, quasi-experiment, sequential assessment

Introduction

After assisting for the Master's degree course on international relations theories, I noticed that despite studying hard and spending considerable time preparing for classes, students found it difficult to grasp the idea of writing a critical position paper, shaping their opinion into a concise and coherent argument. While some managed to make considerable progress throughout the semester, the majority did not noticeably improve their writing and argumentative skills, and thus, failed to achieve the learning outcomes of the course. I assumed that such performance resulted from flaws in the practice of formative assessment and decided to improve that component of the course.

Accordingly, I made three alterations in this year's iteration of the course. First, the introductory lecture devoted more time to organizational instructions, providing students with the assessment rubric for position papers as well as examples of good papers. Second, the quality of the feedback was improved through addressing both strong and weak sides of the papers. In addition, instead of traditional brief textual remarks in the university's Information System (IS), students were provided with audio feedback on their position papers ahead of classes. Thirdly, I made use of sequential assessment. Namely, original audio feedback was provided without the grade; then the grade was inserted into the IS soon after the class.

The research herein demonstrates that the teaching innovation was successful in part and has managed to address the teaching challenge to an extent. Quantitative analysis in the form of t-tests showed that the innovation had positive impact on the final performance of the treatment group. Qualitative analysis revealed that improved instructions as well as the re-iteration of certain points during in-class discussions were useful and helped students to better understand the task at hand. Moreover, most students appreciated verbal feedback and preferred it over the written one. At the same time, students did not clearly see the benefits of separating the feedback from the grade.

Context of the innovation

The innovated course International Relations Theory and Energy Security was taught at the Department of International Relations and European Studies of Masaryk University (Brno), where teachers are fairly free to prepare their syllabi. However, the teaching staff of the Energy Security Studies Master's degree program, of which this course was a part, coordinate the activities they expect students to complete across various courses in order to make those as diverse as possible. This course has weekly ninety-minute sessions, each of which is divided into a forty-five-minute lecture and a forty-five-minute seminar. Seminars are focused on discussing the required readings for each class and students' weekly position papers that they prepare before the class takes place. Since this particular Master's degree is taught in English, the class was comprised of five international students from various countries. My main responsibilities for the course included facilitating seminar discussions and dealing with students' position papers. In addition, I gave two lectures.

Theoretical background

Biggs (2018) came up with the idea of constructive alignment, suggesting that activities should correspond to learning outcomes, helping students to acquire necessary competences. However, my previous experience with the course showed that position papers were not contributing to student learning the way they were expected to. In theory, this activity was aimed at helping students to be able to apply theoretical approaches to the analysis of particular issues of today's international politics and assess the viability of those approaches. These outcomes belong to the third and sixth levels of Bloom's taxonomy (Krathwohl 2002), which are adequate for the Master's level.

Nevertheless, students found it difficult to grasp the idea of writing a critical position paper, shaping their opinion into a concise and coherent argument related to the theoretical aspects of international relations. The majority did not noticeably improve their writing and argumentative skills, having difficulties achieving learning outcomes. Thus, the problem seemed to reside in the practice of formative assessment, which had hitherto offered only brief textual remarks in the IS, explaining the grade.

In order to make it work for students, feedback should be constructive and formative, stimulating their reflective learning and allowing for improvement throughout the course (Gibbs 2015; Juwah et al. 2004). Hence, I decided to change how feedback is provided to students, and thus, address the teaching challenge in a theory-driven way. First, the introductory lecture of this year's course provided explanations on what position papers were, their structure and how they contributed to the learning outcomes. Moreover, in addition to a thorough clarification of the assessment criteria, students were provided with the assessment rubric for position papers and examples of good papers from the previous years.

Second, another measure addressed the quality of feedback by taking into account both strong and weak sides of the papers and presented them in the 'sandwich' format. Furthermore, instead of traditional written feedback, the students were provided with audio comments on their position papers before classes. Listening to an audio feedback is more personal and comprehensible for students, since it makes them focus on what they are listening to: 'There are reports that students are altogether more likely to listen to feedback than to read feedback, and to return to the same piece of feedback more frequently when it is audio feedback' (The Higher Education Academy 2012: 3).

The last part of the innovation was sequential assessment. Namely, the podcasts were provided without the grade; then the grade was inserted into the IS soon after the class. This way, I avoided using feedback for simply justifying the grade. There was also a strong probability that it would encourage students to read the feedback as the only way to get a sense of how well they had performed (The Higher Education Academy 2012: 4).

Research design

I opted for a two-fold evaluation of the impact of the innovation, combining both quantitative and qualitative methods.

Quantitative component

The quantitative component used a quasi-experimental design. The treatment group included students enrolled in the innovated course. I had two control groups: the first one (control group 2016) was comprised of the students from the previous iteration of the very same course; the second one (control group 2017) was comprised of the students enrolled into the Czech version of the course, which was not subject to the innovation. For all 3 groups I was the instructor using the same assessment criteria, assuring the comparability of data. Quantitative analyses were based on descriptive statistics as well as on one-tailed independent t-tests with the standard α=0.05 cut-off point for evaluating the statistical significance of the innovation's impact, admitting α=0.1 cut-off point for marginal significance[1], using the R software[2]. One-tailed t-test was used because the hypotheses below are unidirectional expecting a positive impact.

Specifically, my first hypothesis expected that the treatment group would obtain better final grades compared to the control groups. Accordingly, I compared the average final grade of students in the treatment group to those of the members of the control groups. Since the grades at the university have the letter denomination, I transformed those into numbers, using the official Masaryk University grade classification (see table 1).

1 Such an approach has been gaining wider acceptance recently, see Pritschet et al. 2016.

2 I am indebted to Petr Ocelík, during whose course I could implement the innovation and who has also helped with the quantitative analyses in the R software.

Table 1. Grade classification for academic courses

Level	ECTS letter grade	Value
Excellent	A	1
Very good	B	1.5
Good	C	2
Satisfactory	D	2.5
Sufficient	E	3
Failed	F	4

Source: Masaryk University 2012.

In hypothesis 2, students of the treatment group were expected to perform better over time than students in either of the control groups. Position papers selected from the beginning, middle and the end of the course allowed me to assess students' progress towards learning outcomes over time. The chosen papers for all groups were devoted to the topics of Realism, the Copenhagen School and the Welsh School of Security Studies, assuring their comparability.

Operationalizing this second hypothesis led me to specify it into three sub-hypotheses. In short, I expected that students in the treatment group would show greater improvement from one assignment to another than the students in either of the control groups. That is, the change in group means from position paper 1 to position paper 2 (H2a), from 2 to 3 (H2b); and from 1 to 3 (H2c) should be higher for the group exposed to the innovation.

Qualitative component

The qualitative component, which aimed at strengthening the findings of the research and focusing on students' perceptions of their own progress and the usefulness of the innovation, was comprised of the textual analysis of the data collected in the forms of minute papers and a final questionnaire.

Minute papers were collected for all relevant sessions (n=10) and were designed to encourage students to listen to the feedback they were given, providing information regarding students' satisfaction with the formative assessment. They also allowed me to trace the trend (or lack thereof) towards the convergence of students' self-assessment and the actual grading. In addition, they helped me with adjusting the feedback to the particular needs of each student.

The final questionnaire was filled in at the end of the semester. It was used to grasp students' overall experience with the innovation and to understand whether or not the improvement in

their performance could be attributed to the innovation. It was also instrumental for me in deciding which aspects of the innovation should be preserved, abandoned or adjusted in the future.

Results

Quantitative component

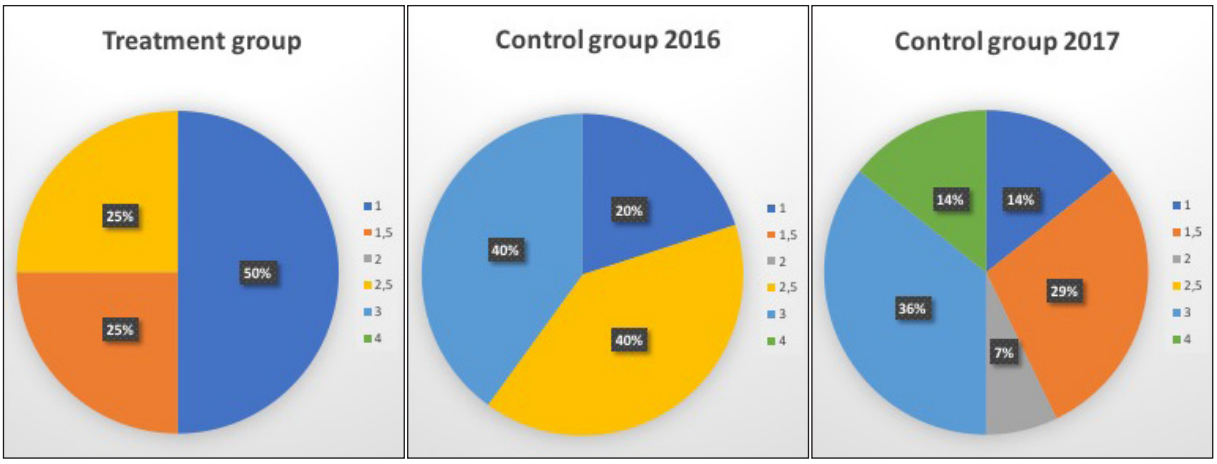
First, I tested hypothesis 1, using descriptive statistics. Comparing the average performance of groups over their final grades revealed that the treatment group demonstrated a tangibly better score of 1.5 compared to the control groups' scores of 2.4 and 2.36[3] (table 2). In other words, based on the grade classification presented in table 1, the average grade of the treatment group was B (very good), while the averages of both control groups were closer to D (satisfactory). In addition, figure 1 shows that the best grades corresponding to the values of 1 and 1.5 comprised 75 per cent of all the grades obtained, which equalled only 20 per cent and 43 per cent for control groups 2016 and 2017 respectively. At the same time, the share of the 2 worst grades in the treatment group was null, while it reached 40 per cent for control group 2016 and 50 per cent for control group 2017. Thus, the treatment group performed better.

Table 2. Comparing the change of student performance on grades (hypothesis 1) and position papers (hypotheses 2a, 2b, 2c).

	Control Group	N		Mean		t-value	Adj. df	p-value[3]
		Treatment Group	Control Group	Treatment Group	Control Group			
H1	2016	4	5	1.5	2.4	-1.765	6.923	0.061†
	2017	4	14	1.5	2.36	-2.001	7.942	0.04*
H2a	2016	5	5	0.7	0	2.333	5.539	0.969
	2017	5	14	0.7	0.18	1.634	16.299	0.939
H2b	2016	5	5	-0.1	0.3	-0.843	7.549	0.213
	2017	5	14	-0.1	0.36	-1.071	13.114	0.152
H2c	2016	5	5	0.6	0.3	0.572	7.997	0.709
	2017	5	14	0.6	0.6	0.134	10.056	0.552
Test: independent t-tests, one-sided * p≤.05; † p≤.1								

3 One of the five students from the treatment group had health-related issues throughout the semester, which has affected his overall performance. Hence, while there were data collected from his position papers, there was no final grade in the system at the time of writing, which was treated as 'not available'.

Figure 1. Shares of final grades values across groups



The t-tests support this conclusion as well (table 2). There was a significant difference in the scores for treatment group and control group 2017 (p=0.04). There was also a marginally significant difference in the scores of treatment group and control group 2016 (p=0.061).

Figure 2: Shares of grades across groups at various stages of assessment

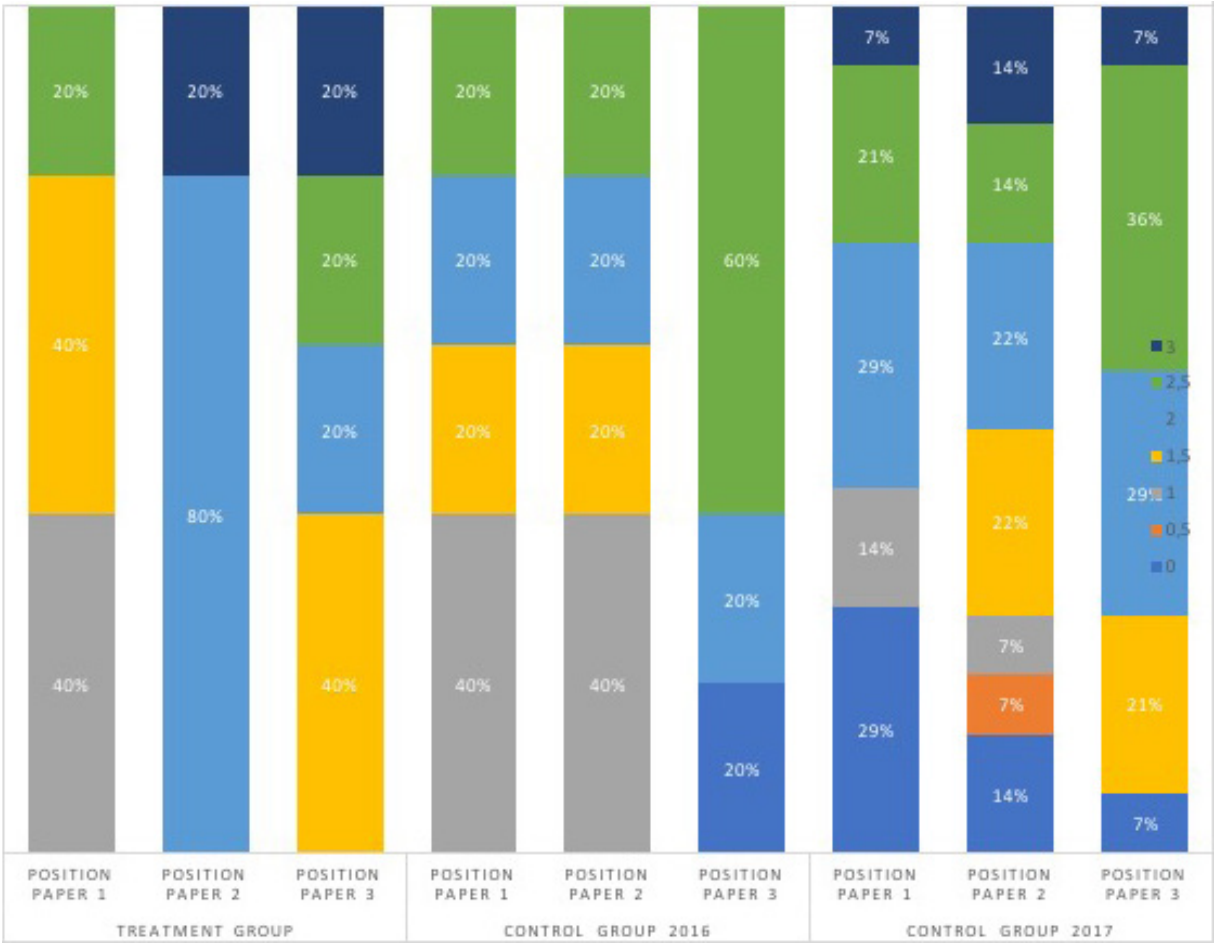
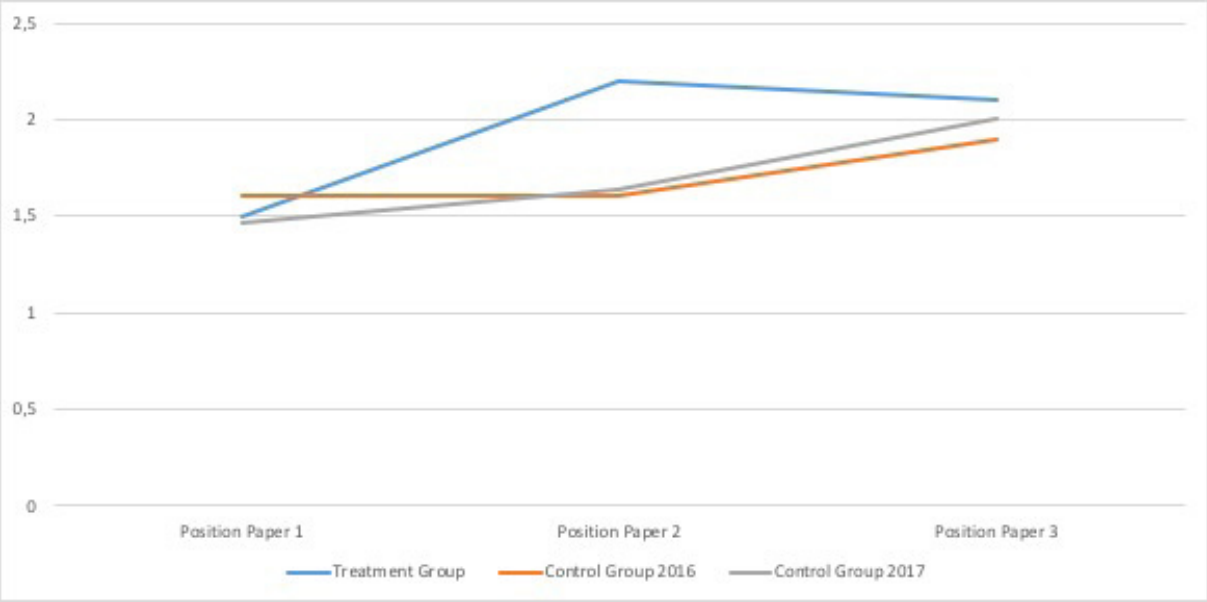


Figure 2 helps to make a preliminary assessment of hypothesis 2 and its derivatives that expect that students in the treatment group performed better over time than students in either of the control groups. Position papers were graded from 0 to 3 points, with 3 being the highest grade. The treatment group had no cases of two lowest grades, while in both control groups some students received 0 or 0.5 points for some of the assignments. Also, in the case of the treatment group the percentage of the highest grade surpassed those of the control groups. There is also a clear trend of improvement in the former case (i.e. increasing percentage of higher grades), while in the latter cases the higher grades stagnate.

Figure 3 provides intriguing details about these trends. First, on average, the treatment group performed better than the control groups. Second, neither of the control groups made higher progress throughout the semester than the treatment group. Similarly, treatment group made a drastic improvement from the first selected position paper to the second one, while control group 2017 made a very modest improvement and control group 2016 showed no improvement at all. At the end, all groups grasped the idea of writing critical papers, but the treatment group did it a lot earlier and, thus, it partially explains why it outperformed both control groups in grade averages.

Figure 3. Group averages at various stages of assessment



Nevertheless, t-tests showed that (H2a) in terms of the progress from assignment 1 to 2 there was no statistically significant difference in the scores between the treatment group and either of the control groups. The same is true for hypotheses 2b and 2c. Therefore, it is not possible to quantitatively confirm hypotheses 2a, 2b and 2c and to attributing differences in performances at various stages of assessment to the innovation. The reasons for this lack of finding are two-

fold. Firstly, the sample size of the treatment group was fairly small, which tends to distort statistical analysis. Secondly, the treatment group and control group 2017 had greatly differing number of students (n=5 vs n=14), which could also negatively affect the results.

Qualitative component

Qualitative data shed lights on student perception of the innovation and their own performance. When it comes to the minute papers, the four regularly attending students of the treatment group[4] were overwhelmingly satisfied with the quality and clarity of the feedback they were provided, agreeing with the points I raised therein. After a scale of satisfaction was added into the minute papers[5], out of twenty minute papers, eighteen indicated full satisfaction, while the other two expressed satisfaction.

However, it is instrumental to look into those where students disagreed with some parts of the feedback. Out of total thirty-eight papers collected, ten of them included some critical remarks[6]. In four cases, students disagreed that their papers were overly descriptive and lacked critical thinking. Another four critical remarks were made by the same student about the same issue: he could not grasp the idea of using empirical evidence instrumentally, that is, to substantiate his position. Although I tried to adjust my feedback and be more explicit after the first occurrence, it did not have much effect. The reason most certainly was that the student was coming from another discipline and had difficulties with adopting the norms and requirements of political science.

Notably, almost all students in their minute papers for the first two sessions of the course mentioned that they needed more guidance with regard to the structuring of position papers. One student even asked for more examples (in addition to two example papers already provided). While the fact that such requests disappeared from the minute papers as the course progressed signalled that students obtained the skill at stake throughout the semester, it also means that I should further improve the initial instructions.

As for the final questionnaire, all students believed that the initial instructions and in-class comments regarding the position papers were useful (all of them graded it as 9 or 10)[7]. Three out of four students also indicated their satisfaction with the form and quality of the feedback provided. Similarly, three out of four recommended to preserve audio feedback in the future iterations of the course. Therefore, based on the opinion of the majority, these aspects of the innovation were successful: initial instructions as well as the re-iteration of certain points during the in-class discussions were satisfactory and helped students to understand the task at hand better. More-

4 This was a 5-point Likert scale, where 1 indicated 'not satisfied at all' and 5 'very satisfied'.
5 Minute papers underwent a few changes to improve data collection.
6 Two of those remarks referred to technical issues (language, deadline).
7 To measure student satisfaction and dissatisfaction, I used a 10-point Likert scale that ranged from dissatisfaction (1) to satisfaction (10).

over, verbal feedback, although not uniformly, was preferred over the written one.
At the same time, only one student strongly agreed with the idea that separating feedback from grading helped in concentrating more on the feedback itself. Two other students could neither agree nor disagree, while the remaining student opposed the aforementioned practice. Hence, students have not clearly seen the benefits of this innovation, which refutes the theoretical reasoning that conditioned the implementation of this measure. Therefore, I will omit it in the future.

Conclusion

My analysis demonstrates that the teaching innovation was quite successful and have managed to achieve most of its expected outcomes. On the one hand, comparing final grades of the treatment group to those of the control groups showed that the average grade of the former was B (very good), while the average of the latter was closer to D (satisfactory). Moreover, the t-tests showed that the difference was statistically significant in one case and marginally significant in the other, confirming that the innovation had a positive impact on the final performance of the treatment group. At the same time, the analysis revealed that, on average, the treatment group performed better throughout the semester and grasped the idea of writing critical papers a lot earlier than either of the control groups even though this difference has not reached statistical significance.
Furthermore, the qualitative analysis confirmed that instructions as well as re-iteration of certain points during the in-class discussions helped students to better understand the task at hand. Moreover, verbal feedback was preferred over the written one. However, students did not clearly see the benefits of separating the feedback from the grading. Hence, I plan to keep the innovation in future iterations of the course, albeit not the two-stage feedback. Collecting students' feedback via minute papers proved to be very useful, and, therefore, I plan to use it again in order to reassess and, if necessary, to refine my innovation further.

References

Biggs, J. (2018) *Constructive alignment*, available at http://www.johnbiggs.com.au/academic/constructive-alignment/, accessed 7 August 2018.
Gibbs, G. (2015) 'Making feedback on assignments effective: principles and guidance for tutors', *Transforming the experience of students through assessment*, available at https://www.testa.ac.uk/index.php/resources/best-practice-guides?download=3:feedback-guide-for-lecturers, accessed 7 August 2018.
Juwah, C., Macfarlane-Dick, D., Matthew, B., Nicol, D., Ross, D. and Smith, B. (2004) *Enhancing student learning through effective formative feedback*, York: The Higher Education Academy.
Krathwohl, D.R. (2002) 'A revision of Bloom's taxonomy: an overview', *Theory into Practice* 41:4, pp. 212-218.

Masaryk University (2012) *ECTS grading system at Masaryk University,* available at http://web.aiu.ac.jp/_old_international/wp-content/uploads/2012/03/Grading-system-at-MU-2.pdf, accessed 7 August 2018.
Pritschet, L., Powell, D. and Horne, Z. (2016) 'Marginally significant effects as evidence of hypotheses: changing attitudes over four decades', *Psychological Science* 27:7, pp. 1036-1042.
The Higher Education Academy (2012) *10 Strategies to engage students with feedback*, available at http://teche.ltc.mq.edu.au/wp-content/uploads/2016/08/10_strategies_to_engage_students_with_feedback.pdf, accessed 7 August 2018.

Nikita Minin is a PhD student at the Department of International Relations and European Studies at the Faculty of Social Sciences, Masaryk University. His current research focuses on the European nuclear market and regional energy cooperation. He has been taking part in various teaching activities at the department since 2016, most notably as a teaching assistant for the Theories of International Relations and Energy Security course.
nikita.minin@mail.muni.cz