RESEARCH ARTICLE   **Open Access**

CrossMark

Enhancing students' written production in English through flipped lessons and simulations

M. Laura Angelini[*] and Amparo García-Carbonell

* Correspondence: marialaura.
angelini@ucv.es
School of Education, Universidad
Católica de Valencia "San Vicente
Mártir", Centro de Postgrado
Santísima Trinidad, C/Menéndez y
Pelayo, frente al nº 7, 46100
Burjassot, Valencia, Spain

Abstract

Today, learning is perceived as a challenge that must be faced simultaneously on numerous fronts. Indeed, learning is no longer confined to the classroom. Students have the opportunity to learn inside and outside the classroom walls. Technology plays its part, as does the abundance of information available on social networks and in the mass media. Educators must stay abreast of change as information and potentially useful technological resources leave traditional education behind. Optimising class time through new methods, techniques and resources is paramount in today's education systems. This paper presents the results of a quantitative study of students' written production in English. The English writing skills of engineering students were developed using situational (or class) simulations and a large-scale web-based simulation in real time. Quantitative analysis of students' written production was used to test for differences between experimental and control groups. The goal of this study was to show that simulation-based instruction contributes significantly to students' progress in written production in English. The results showed that students who received simulation-based instruction (experimental group) significantly improved their English writing skills, primarily in terms of organisation and linking of ideas more than students who attended a regular English course (control group).

Keywords: Web-based simulation, Blended learning, Flipped classroom, Active learning

Introduction

Communication has long been a primary goal of foreign language educators. Foreign language students must gain fluency and accuracy to communicate effectively in both written and spoken forms. However, language teachers often teach large classes, and communication can become an ordeal. Blended learning has been gaining ground in language teaching, and certain pedagogical strategies are making headway. Flipped learning is one such strategy. Flipped learning is a specific blended learning model that helps educators optimise class time to encourage communication. In this study, flipped learning was applied, moving lectures outside the classroom and introducing simulation-based lessons to enhance English as a foreign language (EFL) learning, particularly in terms of production skills development. This application of flipped learning inverted the traditional teacher-centred method. Instruction on essay writing, registers and simulation procedures was delivered online outside the classroom, whilst

traditional homework was moved into the classroom environment to identify students' weakness and strengths before students participated in simulations and written tests. The flipped model uses technology to present the theory and background materials. This paradigm shift transforms the roles of teacher and learner (Strayer, 2007, 2012; Tucker, 2012). In this study, instructors became facilitators and guides for learners, who worked in teams during the simulations in class. The learners became the real participants in the classroom (Strayer, 2007, 2012; Tourón, Santiago, & Diez, 2014).

A simulation is an activity in which participants are assigned roles and are given enough information to solve a specific problem. A simulation is based on a representation of a model that imitates a real-world process or system. Key information is provided so that participants can carry out tasks, debate, negotiate from different points of view and solve a specific problem (Klabbers, 2009). It is the participants' responsibility to perform duties and thereby solve the problem without play-acting or inventing key facts (Jones, 2013). Michelson and Dupuy (2014) further discuss simulation and language learning and refer the simulation potential to enact discourse styles associated with social identities.

In the present study, a web-based simulation was used from The International Communication and Negotiation Simulations (ICONS) platform. The ICONS platform, developed at the University of Maryland, combines simulation tools and simulation development dialogue (SDD) methodology to provide clear insights into global socio-political affairs and evaluate alternative courses of action in crisis situations. Simulations performed using the ICONS platform are thus ideal for addressing social issues that relate to education, environmental threats, the sustainable economy and human rights. Scholars have praised simulations as an effective way of instilling ethical responsibilities in students and developing students' global mindset (Crookall, 2010; Crookall & Oxford, 1990). In the debriefing, students reflected on the simulation and the learning component of the whole experience.

The present study, thus, describes the related works comprising simulations in education and flipped learning. The methodological section describes the participants, the quantitative data and studies carried out. Ethical issues and Threats of validity are subsequently addressed together with Results, Conclusions and Future Research.

Related works

Several educational disciplines have embraced simulations. Such disciplines include industry, medicine, nursing, engineering and languages. Despite their relatively short tradition, hundreds of studies have shown the benefits of simulations as they provide immersive experiential learning (Ekker, 2000, 2004; Chang, Peng, & Chao, 2010; Wedig, 2010; Beckem, 2012; Wiggins, 2012; Gegenfurtner, Quesada-Pallarès, & Knogler, 2014; Blyth, 2018). Ekker (2000) studied data on 46 students from four European universities that participated in the Intercultural Dimensions in European Education through On-line Simulation (IDEELS) project. Students' responses to online questionnaires pre- and post-treatment indicated that 90% of students were satisfied with the simulations, reporting a good learning experience. Approximately 73% reported that web-based simulation suited their needs. More than 80% reported that they did not experience difficulties due to cultural differences. Interestingly, all male participants, unlike 22% of female participants, reported that all members of the team contributed to the tasks.

Klabbers (2001) described simulations as learning and instructional resources. According to the author, simulations offer a springboard for interactive learning that develops expertise. Kriz (2003), in turn, contextualised simulation within the educational framework. Simulation is an interactive learning environment that converts problem-oriented learning into purposeful action. According to Kriz, training programmes for systems competence through simulation have shown that simulations favour change processes in educational organisations.

Ekker (2004) conducted an empirical research on simulations applied to education. The author analysed data on 241 subjects who had participated in various editions of IDEELS, examining satisfaction levels and attitudes. The participants had different roles as negotiators, technical consultants, activists or journalists within the "Eutropian Federation Simulation". The three-week simulation consisted of message exchanges, written proposals and "live" conference situations. The software used was a web-based interface driven by a database server. The project resorted to a web-based questionnaire to measure students' satisfaction, personal experiences and attitudes towards the simulation. Findings revealed that students found satisfaction during the simulation, they were activated as the simulation invigorated learning and personal characteristics did not significantly predict or affect users' satisfaction with web-based simulations.

Other studies conducted by Levine (2004) and Halleck and Coll-García (2011) integrated tele-collaborative exchanges and global simulations to turn the foreign language class into its own immersive, simulated environment.

Levine (2004) described a global simulation design as a student-centered, task-based alternative to conventional curricula for second year university students of foreign language courses. The author provided clear guidelines to apply simulations in language courses and identifies strengths such as the use of the content knowledge in the simulation dynamics, target language activation during the simulation phases and collaborative work to carry out the tasks. Furthermore, Halleck and Coll-García (2011) presented the results of a pilot project in which simulation-based learning was used to teach English to engineering students. The study involved 42 undergraduate engineering students at Oklahoma State University, USA, and 56 undergraduate engineering students from Universitat Jaume I, Castellón, Spain. The results of this pilot study shed light on participants' perceptions of how web-based simulations affect the development of language abilities, critical thinking and intercultural awareness. The authors highlighted the importance of a simulated experience in an engineering curriculum. They concluded that a real comprehensive engineering education should provide opportunities to work collaboratively with other professionals in an intercultural setting more than simply solving problems from a textbook.

Burke and Mancuso (2012) in their study of social cognitive theory, metacognition, and simulation learning identified core principles of intentionality, forethought, self-reactiveness and self-reflectiveness in simulation environments. They sustained that debriefing helps build students' self-efficacy and regulation of behaviour. Thus, simulation-based learning combines key elements of cognitive theory and interactive approach to learning. Theory-based facilitation of simulated learning enhances the development of social cognitive processes, metacognition, and autonomy.

Other studies on language teaching and learning have shown that simulations encourage the development and acquisition of language (e.g. Rising, 2009; Andreu-Andrés &

García-Casas, 2011; Author, 2011; Woodhouse, 2011; Michelson & Dupuy, 2014; Blyth, 2018). The scholars coincide that simulations provide greater exposure to the target language, more purposeful interaction, more comprehensible input for learners, a reduced affective filter and lower anxiety in language learning. To mention some, Author (2011) examined perceptions of collaborative work in web-based simulations through evaluations of each student's end-of-course portfolio *[N = 26]*. Students highly valued the collaborative work required in the simulation, which was reflected by the active participation of all team members and by team members' motivation and personal satisfaction. By analysing their own work and that of their teams, the students reported that they had become more resolute and had learnt discourse strategies to persuade others and solve problems. Students also reported that the collaborative work increased their capacity to listen to others' ideas and to learn from others. All this helped increase their intellectual development and knowledge of the world. They also understood specific content faster, improved their language skills and acquired experience in self-assessment. Andreu-Andrés and García-Casas (2011) focused on simulation and gaming as a teaching strategy. Qualitative analysis based on grounded theory was used to study the perceptions of 47 engineering students. These students endorsed experiential learning and reported that learning and having fun reaped rewards. As educators and students became more familiar with the simulations, they developed a greater appreciation of their effectiveness. Students complete simulations with a heightened awareness of what they have learnt and how they can learn more. Another clear example is Woodhouse's (2011) study, in which 33 Thai university students participated in a computer simulation to learn English. Data were collected through personal interviews to learn about students' opinions of the use of simulations to learn a foreign language. The students perceived that the simulation, despite not being face-to-face, did not hinder their learning about sociocultural aspects related to communication in the target language. Students noted that they acquired greater powers of decision, persuasion and assertiveness in communication. Ranchhod, Gurău, Loukis, and Trivedi (2014) make a threefold contribution to the simulation and experiential learning literature. They analyse the representational effectiveness of several learning strategies. Their study builds on Reeve's educationally supportive learning environment through simulations (Reeve, 2013) as the investigation deals with the concrete learning experience generated by the simulation to develop or reinforce theoretical understanding, management experience, and professional skills.

An example of a large-scale simulation was described by Michelson and Dupuy (2014) in which 29 intermediate learners of French at a public university in the Southwest of the United States participated in the study. 12 students of the experimental group participated in the simulation and had specific roles to enact the responsibilities of residents in a commercial area in Paris. 17 students belonged to the control group and did not participate in the simulation. They followed a traditional approach to learn French. Only the experimental students demonstrated abilities to describe how their roles motivated certain linguistic choices and non-linguistic semiotic modes. The study highlights the potential for simulations to boost students' awareness of the target language together with other communication codes.

Blyth (2018) explores the challenges of immersive technologies in foreign language learning and global simulations to enhance language use. The study summarizes the

impact of simulations in language learning and concludes that simulations of language use in authentic contexts boosts real experiential language learning.

A few other studies have examined the effectiveness of technologies and simulations in the language classroom. O'Flaherty and Phillips (2015) provided a broad overview of research on the flipped classroom and links to other pedagogical models such as simulations. They reported considerable indirect evidence of improved academic performance and student and teacher satisfaction with flipped learning. However, further research is required to provide conclusive evidence of how the fusion of these methods enables language and social competence development. Author (2016) investigated combining flipped learning instruction and simulation-based lessons to optimise class time by using and designing simulations with prospective secondary school teachers. Author outlined the benefits of using simulations that are based on literary extracts with a substantial social component.

Methodology

The simulation in this study consisted of three phases: briefing, action and debriefing, all of which required immersion in the English language. During the briefing phase, consistent with the flipped classroom model, students were presented with topics related to the simulation scenario, literature on these topics and videos to be viewed outside the classroom. One benefit of this pedagogical shift is that students have more class time to apply the content knowledge in relevant communication situations than they would if they followed more traditional instruction models. Amongst the communication activities performed in class were minor-scale simulations, debates and forums aligned with problem-based learning. This type of practice helped prepare the students for the larger-scale simulation which covered more topics to analyse and had a different complexity as it was international. This class practice also helped instructors estimate students' understanding of the topic and the type of language that the students used. The instructors provided grammar clarifications and explanations where necessary. Students chose their own teams of four or five members. These teams were the same for the activities and the large-scale simulation. Teamwork was fostered, as was individualised learning. The instructor was able to identify the weaknesses of each student. This initial briefing phase served as preparation for phase 2, during which the web-based simulation took place. This large-scale web-based simulation had several steps: reading and analysing the scenario and assigning individual roles; anticipating other team members' proposals and writing a strategy to persuade other team members to vote for a particular proposal; listening to others and taking notes; and debating, negotiating and, finally, making a decision.

Quantitative data collection

This paper presents the findings of a quantitative study of students' progress in written production in English. The cohort of engineering students who participated in the study had attained the B1 level of English. Moreover, they were enrolled in an intensive optional four-month conversational English course at university. This course corresponded to the B2 level by the Common European Framework of Reference for Languages: Learning, teaching, assessment (CEFR). The CEFR has been designed to provide a coherent and comprehensive basis for the language syllabuses creation,

teaching and learning materials, and the assessment of foreign language proficiency. It is used in Europe but also in other continents. The CEFR is available in 40 languages (Council of Europe, 2001).

There were five subgroups in total. The experimental group had two subgroups (E1 and E2; $N = 50$), which were taught separately in different classrooms. The control group had three subgroups (C1, C2 and C3; $N = 71$), which were taught separately in three classrooms. Smaller groups were more conducive to language learning in both the experimental and the control subgroups. All participants were in the third year of an engineering degree. The experimental subgroups received flipped learning instruction of topics related to the simulation scenario. This means that the students in the experimental subgroups were acquainted with the topics as they had to watch videos and read before the simulation. In class, the simulation guidelines and classroom practice in minor-scale simulations, class debates and forums prepared the students to participate in a large-scale web-based simulation. This latter is conceived as a large virtual exchange with other students from different foreign universities. This web-based simulation was carried out during class-time in the technology lab. Video conferences were held only with groups from other universities in Europe (synchronous simulation). However, there was interaction amongst other groups with different time zone through written messages, recorded voice messages and recorded sessions. Additional file 1 presents a list of materials used. The ICONS web-based simulation consisted of an international summit on current economic, social and security issues. This simulated summit was attended by numerous countries, which were represented by student teams. Attendance was both synchronous and asynchronous. The experimental group worked in teams of four to five members, each with a clear role within the team. These roles were specified in the simulation briefing phase.

The control group, however, was taught under a traditional EFL instruction model, which was based on a B2 course book, 3,5 h-lesson per week in one term. Students sat a final exam at the end of the term. Written production by students in the experimental and control groups was tested pre- and post-treatment. The assessment criteria for the pre- and post-treatment written tests were evaluated on a five-point Likert scale, where 1 indicated 'not accomplished' and 5 indicated 'successfully accomplished' for the three variables *topic development, organising and linking ideas*, and *variety and accuracy in grammar and vocabulary* (University of Cambridge, ESOL Examinations).

Although different skills were worked on during the course, this study focused on written production in English. The experimental group followed simulation-based training, which is illustrated in Fig. 1.

1. **Written pre-test.** Control and experimental groups wrote 250 words about how living in a cosmopolitan city affects their life and lifestyle. Three external examiners assessed the timed essay by applying the adapted writing criteria (University of Cambridge, ESOL Examinations, 2012) of language development, organising and linking ideas, and variety and accuracy in grammar and vocabulary (Additional file 2).

2. **Flipped learning approach in the briefing phase.** Students watched videos, read the news and performed research on several topics related to the web-based simulation scenario. Outside the classroom, they also revised some aspects of grammar

Fig. 1 Procedure workflow
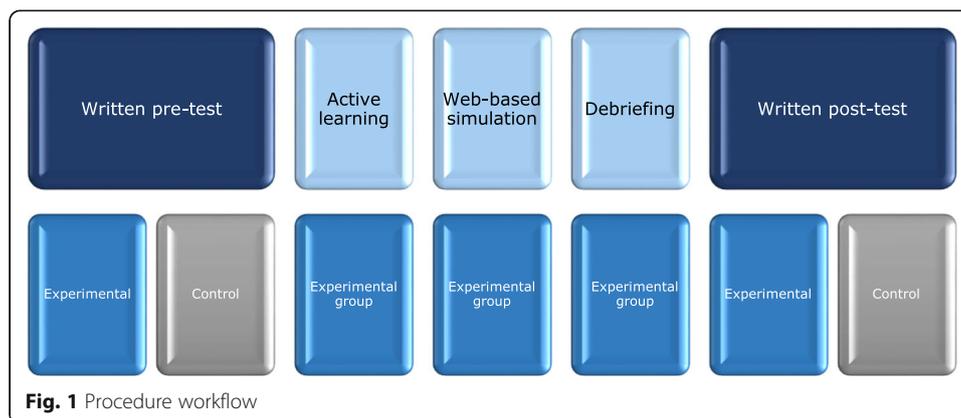
that were occasionally clarified in class. In contrast, class sessions were active learning lessons where students took on responsibilities and participated in minor-scale simulations to debate, negotiate and solve problems. Teamwork was fostered. This phase served as preparation for the action phase, where the web-based simulation took place. Attendance was compulsory, and formative assessment was used to keep a record of students' progress.

3. **Web-based simulation.** Experimental students revised the simulation guidelines and formed teams of four or five members. The students chose their own teams, with no interference from the teacher. Participants became acquainted with the simulation scenario and their roles within the team (simulation scenario can be consulted in Additional file 1). When the action phase took place (synchronously and asynchronously), students analysed the scenario and identified the problems to be solved, planned strategies, participated in debates, set forth and negotiated proposals, and took a final decision. The web-based simulation lasted three weeks. Conversely, control groups followed a more conventional approach to learning English. They had 3,5-h lessons per week and used a general B2 course book to develop the listening, speaking, reading, writing and interaction skills. Lessons attempted to provide them with possibilities of practising these skills and they usually had to do the exercises in the workbook for homework. They sat a final exam at the end of the course. They did think-pair share and group works in the classroom, basically in the speaking exercises.

4. **Debriefing.** A structured debriefing consisted of three phases. The initial phase consisted of reflecting on the simulation experience, discussing it with others and learning and modifying behaviours based on the experience. In this initial phase, facts and concepts were clarified. The second phase dealt with emotions during the simulation, either individually or as a group. The third phase consisted of understanding the different views of each participant and the way each view reflected reality. Thus, the third phase addressed the generalisation and application of the experience to real life (Thatcher & Robinson, 1985).

5. **Written post-test.** This phase was common to both groups (experimental and control). It took place at the end of the course. Participants wrote 250 words on the following topic: 'What do you think about immigration in Spain?' Three

external examiners assessed the timed essay by applying the same criteria (University of Cambridge, ESOL Examinations, 2012) as in step 1.

The goal of the quantitative study was to determine students' progress in written production in English. To achieve this goal, the following tests were conducted:

1. **Pre-treatment homogeneity test.** A Student's t-test was used to compare the means for the experimental and control groups because the distribution of assessments for both groups was normalised (non-significant Kolmogorov-Smirnov test results). Fisher's least significant difference (LSD) method was applied to determine which means were significantly different from others.

2. **Post-treatment comparative analysis of the progress of students in both groups.** Descriptive analysis of the mean scores and standard deviations for the experimental and the control groups was conducted. For the analysis of effect size, Cohen's (1988) procedure was followed. ANOVA was used to identify significant differences between the average progress levels for each group.

3. **Post-treatment analysis of progress for each variable.** A Student's t-test was used to compare mean scores post-treatment. The Kolmogorov-Smirnov test was used to determine the extent to which the distribution of the variables could be considered normal.

4. **Concordance analysis of the three external examiners' assessments.** The concordance of external examiners' assessments was studied to determine whether each examiner exercised independent judgement. An F-test of equality of variances was used to check examiners' variability, variability in variables and students' variability. All analyses were performed in SPSS 25 (under a licence held by the *Universidad Católica de Valencia*).

Ethical issues

Letters of consent were previously signed by members of the five subgroups to comply with the basic principles of research ethics. A sample letter can be found in Additional file 3.

Results

Pre-treatment homogeneity test to compare the mean level of written production in the experimental and control groups

The mean level of written production pre-treatment for the experimental group (5.109) was higher than it was for the control group (4.460). The standard deviation for the control group (1.256) was higher than it was for the experimental group (0.869) (Table 1).

These results indicate considerable variability in the command of the English language displayed by students in the control group. In the experimental group students' command of written English varied to a greater degree.

The Student's t test indicated that the difference between the mean level of written production in the experimental and control groups was significant ($p = 0.001$).

Table 1 Mean level of written production pre-treatment in experimental and control groups

Group	N	Average	Standard deviation	Mean standard error
Experimental group (pre)	50	5.109	0.869	.122
Control group (pre)	71	4.460	1.256	.149

A multiple comparison test (Fisher's LSD method) was applied to determine which means were significantly different from others (Table 2).

Subgroups E1, E2 and C3 had similar levels. C1 and C2 had slightly lower levels. The Student's t-test indicated that the means for the experimental subgroups E1 and E2 were higher and that there was greater variability amongst the control subgroups. This variability in the means for the control group might be associated with the presence of foreign students in subgroup C3. These students had an excellent command of the English language (Table 3).

However, the primary goal of this study was not to identify differences between the means of the experimental and control groups. This study was designed to investigate students' progress post-treatment.

Post-treatment comparative analysis of the progress of students in both groups

ANOVA was used to identify significant differences between the mean level of progress of the experimental and control groups (Table 4). The *p*-value was less than or equal to .05. This result implies that there were significant differences in the mean level of progress of different groups (Table 5).

Analysis of effect size was conducted to determine the magnitude of the change between the mean level of written production pre- and post-treatment.

$$d = \frac{X \text{ exp.imprv} - X \text{ control imprv}}{O' \text{ control imprv}} \quad (1)$$

The effect size was 1.236. This value exceeds the threshold of 0.8, which is the minimum value for the effect size to be considered large (Cohen, 1988). According to

Table 2 Multiple comparison test (Fisher's LSD method) of mean level of written production (pre-treatment) in the 5 subgroups (C1, C2, C3, E1 and E2)

Test	Sig.	Difference	+/− limits
E1 - E2		−0.573	0.587
E1 - C1	*	0.635	0.502
E1 - C2	*	0.679	0.569
E1 - C3		−0.619	0.674
E2 - C1	*	1.208	0.542
E2 - C2	*	1.253	0.690
E2 - C3		−0.046	0.704
C3 - C1		0.044	0.619
C1 - C3	*	1.254	0.636
C2 - C3	*	1.298	0.765

Note: * implies significant difference

Table 3 Homogeneous blocks in terms of written production (pre-treatment)

Group	Cases	Mean	Homogeneous block
C2	15	4.177	**X**
C1	42	4.222	**X**
E1	28	4.857	*
E2	22	5.430	*
C3	14	5.476	*

Cohen, the thresholds for effect size are d = 0.20 (small), d = 0.50 (moderate) and d = 0.80 (large).

Table 6 shows the least significant differences (Fisher's LSD method) in means and the estimated differences between means. Two homogeneous blocks were identified. The first block comprised subgroups E1 and E2. Analysis of the mean level of progress post-treatment did not reveal significant differences. This means both experimental groups were homogeneous.

The second block comprised subgroups C1, C2 and C3. Analysis of the mean level of progress post-treatment did not reveal significant differences. Conversely, when the subgroup E1 was compared with C1, C2 and C3 and when E2 was compared with C1, C2 and C3, significant differences were identified.

To conclude, the initial homogeneity test of both groups pre-treatment indicated that the mean for subgroup C1 was similar to the mean for E1 and E2 and that the mean for subgroup C3 was significantly higher than the mean for C1 and C2. This finding does not invalidate the results of the subsequent comparative analysis of progress, although it is unclear whether the pre-treatment level might have influenced the progress of students in a given subgroup. Nevertheless, the progress of students in subgroup C3 did not differ significantly from the progress of students in subgroups C1 and C2. Students in these subgroups made less progress than did students in the experimental subgroups. This finding shows that the progress of students in the experimental group was significantly greater than the progress of students in all control groups, regardless of students' pre-treatment level (Table 7).

Thus, the simulation-based instruction proved effective at improving students' written production.

Comparative analysis of progress in each variable post-treatment

The independent variables assessed in the comparative study were *topic development*, *organising and linking ideas*, and *variety and accuracy in grammar and vocabulary*.

Table 4 ANOVA of progress in experimental and control groups

Source	Sum sq.	df	Mean square	F	*p*-value
Inter-group	29.34	4	7.336	12.99	0.000
Intra-group	65.50	116	0.564		
Total (Corr.)	94.85	120			

Note: Sum sq. – Sum of squares; df - degrees of freedom

Table 5 Descriptives

Progress variables	N	Mean	Standard deviation	Standard error	95% Interval confidence for mean		min	max
					Lower bound	Upper bound		
TOP DEV. Exp + cont	121	,0970	,14,968	,01361	,0700	,1239	-,23	,53
ORG.IDEAS Exp + cont	121	,1014	,14,884	,01353	,0746	,1282	-,32	,53
GRAM.VOC Exp + cont	121	,1572	,17,218	,01565	,1263	,1882	-,20	,57
total	363	,1185	,15,922	,00836	,1021	,1350	-,32	,57

Topic development

The mean level of progress in topic development for the experimental group was 3,89. For the control group, the mean level of progress was 3,02. Figure 2 shows that in the pre-treatment both control and experimental groups are quite homogeneous. In the post-test, the experimental group perceived greater progress although the control group have also improved (Fig. 3).

The Student's t-test indicated that the mean for the experimental and control groups was non-significant (Table 8).

Dispersion was higher for the control group. This variability amongst students in the control group may owe to the fact that the experimental group was more homogeneous in terms of students' knowledge of English.

Post-treatment progress in topic development was greater for students in the experimental group than for students in the control group. However, the difference was non-significant ($\alpha = .05$).

Organising and linking ideas

The mean level of progress post-treatment was greater for the experimental group 4,76 than it was for the control group 3,48. Figure 4 shows that the mean level was slightly higher for the experimental group and that dispersion was similar for both groups compared to variability observed in the pre-test (Fig. 5).

Table 6 Comparison of the mean level of progress of written production (post-treatment)

Test	Sig.	Difference	+/− limits
E1 − E2		0.126	0.424
E1 − C1	*	1.171	0.363
E1 − C2	*	0.771	0.476
E1 − C3	*	0.726	0.487
E2 − C1	*	1.044	0.392
E2 − C2	*	0.644	0.498
E2 − C3	*	0.600	0.509
C1 − C2		−0.400	0.448
C1 − C3		−0.444	0.459
C2 − C3		−0.044	0.553

Note: * implies significant difference

Table 7 Homogeneous blocks of progress (post-treatment)

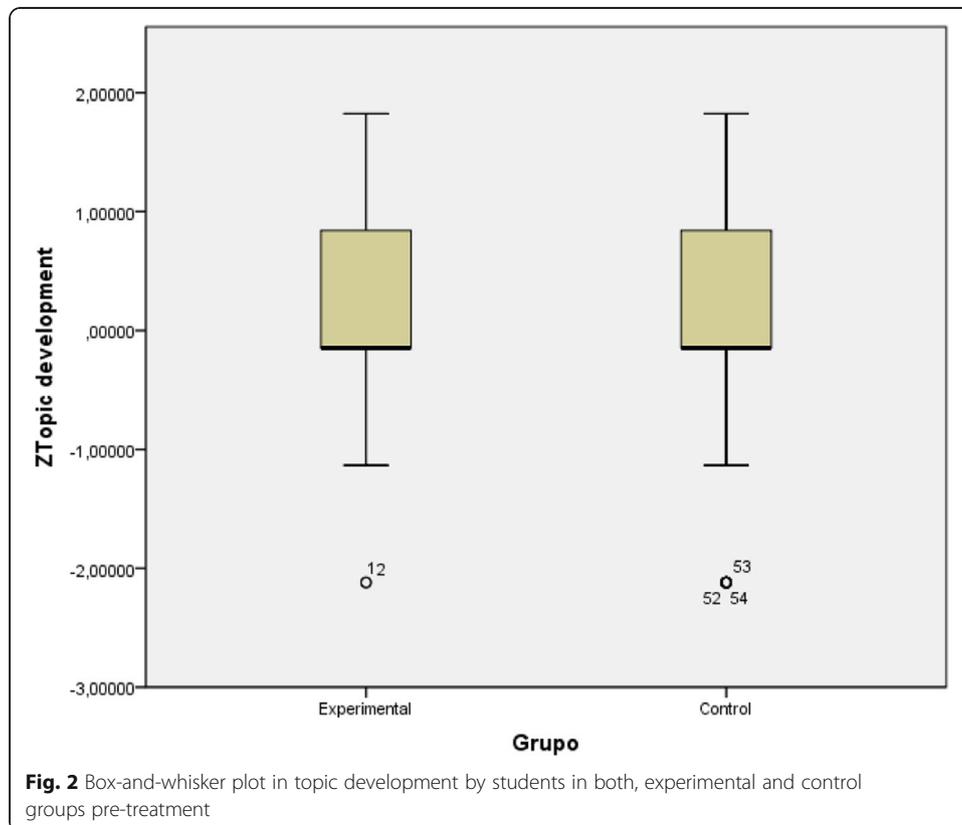| Group	Cases	Mean	Homogenous block
E2	22	1.600	**X**
E1	28	1.726	**X**
C1	42	0.555	*
C2	15	0.955	*
C3	14	1.000	*

The Student's t-test indicated that the *p*-value was less than the level of statistical significance (α = .05). Thus, the difference in progress was significant (Table 9).
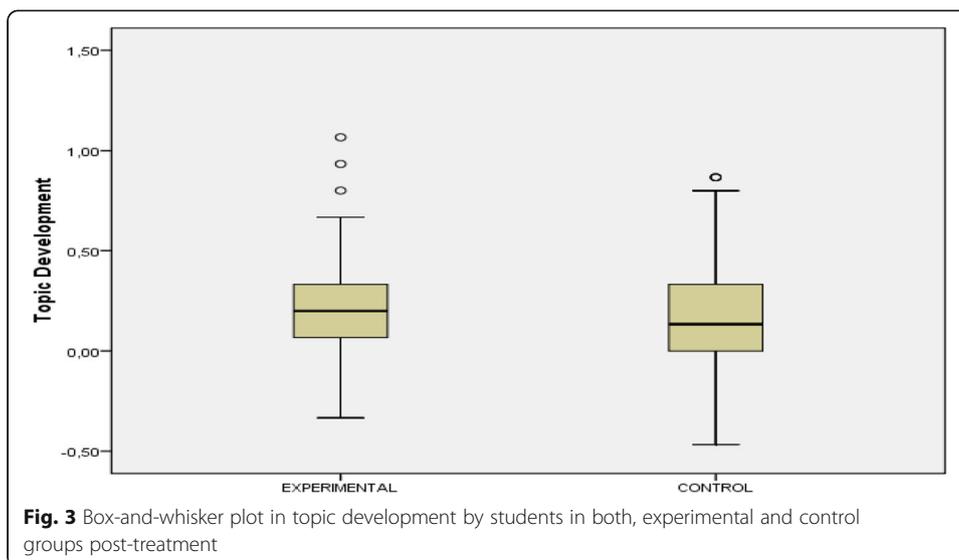
The effect size was 0.876. Because this value was greater than 0.8, the effect can be considered large. This result implies that the experimental group perceived greater progress in organising and linking ideas after the simulation-based lessons.

Variety and accuracy in grammar and vocabulary

The experimental group had a higher mean level (3,78) than the control group (3,60). Figure 6 shows that the mean level was substantially higher for the experimental group compared to the pre-test (Fig. 7).

The Student's t-test indicated that the p-value was substantially less than the level of statistical significance (α = .05). Therefore, the mean level of post-treatment progress in *variety and accuracy in grammar and vocabulary* did not reach significance for the

Fig. 2 Box-and-whisker plot in topic development by students in both, experimental and control groups pre-treatment

Fig. 3 Box-and-whisker plot in topic development by students in both, experimental and control groups post-treatment

experimental group (Table 10). The effect size indicated that the treatment effect was large (effect size of 1.599 > 0.8).

Thus, the results for the three variables of written production indicate post-treatment progress by students in the experimental group. However, this progress was significant (at the 5% level) and had a large effect size for the variable *organising and linking ideas*.

Concordance analysis of external examiners' assessments

In this study, we tested the objectivity and impartiality of the three external examiners' assessments of students' written production pre- and post-treatment.

Pre-treatment

Figure 8 shows the homogeneity of the three external examiners' assessments.

The variability that can be observed in Fig. 8 is not associated with discrepancies in examiners' assessments ($p = 0.674$). Instead, it is due to differences in students' knowledge of English as measured by the three variables that were analysed in this study ($p < 0.00001$). Therefore, the results indicate concordance in the three examiners' assessments pre-treatment.

Post-treatment

The three examiners tended to assess students in the same way in most cases (Fig. 9).

Table 8 Comparative analysis of mean level of progress in topic development post-treatment

Overall score (post-treatment) Topic development	N	Levene's Test		Student's t test					95% CI	
		f	sig.	t	Overall	Sig. (bilateral)	Mean diff.	Std. error of diff.	Lower	Upper
Experimental	50	2.434	.121	1.092	119.000	.277	.030	.027	−.024	.084
Control	71			1.124	114.769	.264	.030	.026	−.023	.083

Fig. 4 Box-and-whisker plot of post-treatment progress in organising and linking ideas in both groups

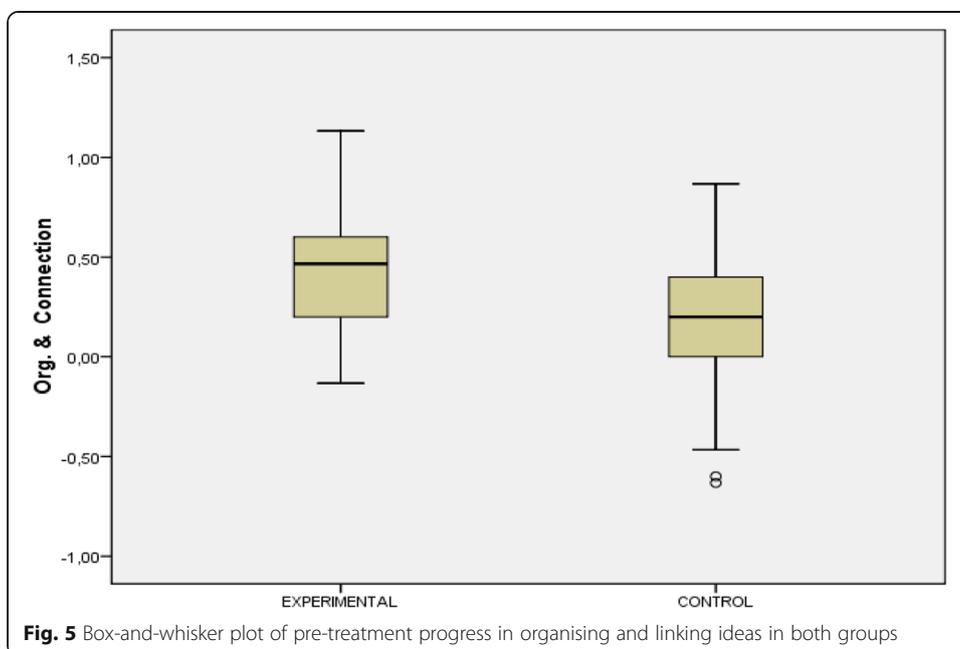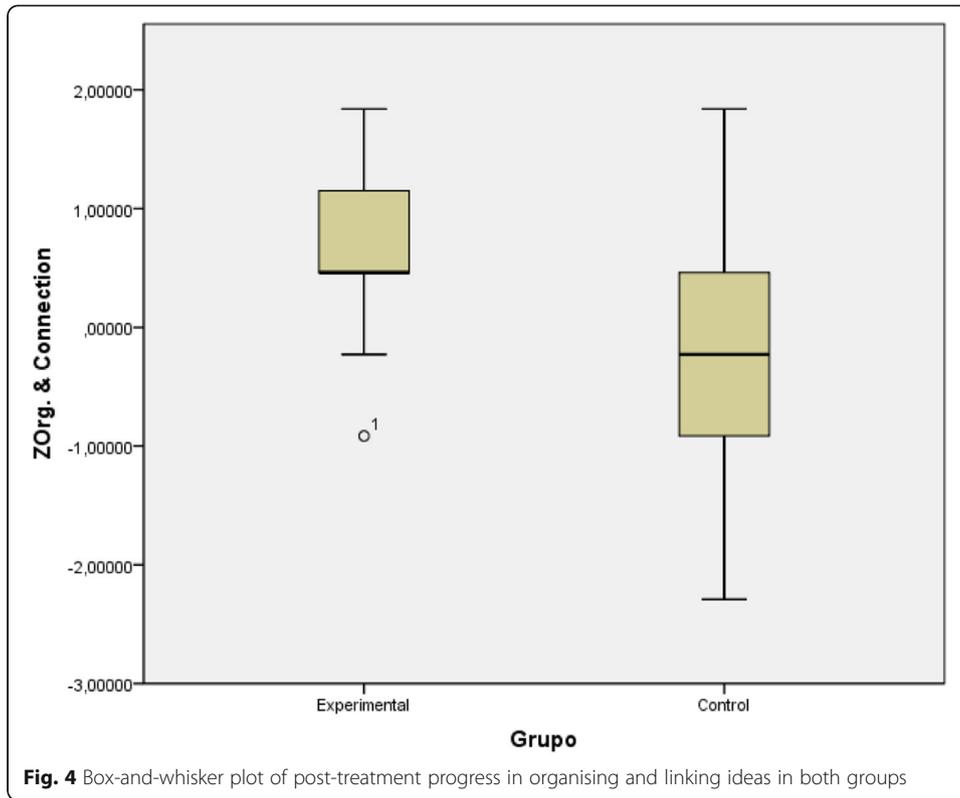**Fig. 5** Box-and-whisker plot of pre-treatment progress in organising and linking ideas in both groups

Table 9 Comparative analysis of the mean level of progress in organising and linking ideas (post-treatment)

Overall score (post-treatment) Organization-Connection	N	Levene's Test		Student's t test					95% CI	
		f	sig.	t	Overall	Sig. (bilateral)	Mean diff.	Std. error of diff.	Lower	Upper
Experimental	50	.005	.005	4.681	119.000	.000	.138	.029	.079	.197
Control	71			4.654	103.429	.000	.138	.029	.079	.197

According to examiners' assessments, students in both groups (i.e. control and experimental) progressed post-treatment. However, the students in the experimental group received higher marks.

Threats to validity
The findings of this study should only be considered in light of its limitations.

Internal validity
As regards the selection bias, the group of participants were not selected from populations with different characteristics. Both, experimental and control groups, were all in third year of an engineering degree. To enrol in the course, students had to prove language proficiency. However, the group was heterogeneous as there were students from ages ranging 21–26 years old, some Erasmus students, and very few students with

Fig. 6 Box-and-whisker plot of post-treatment progress in variety and accuracy in grammar and vocabulary in both groups

Fig. 7 Box-and-whisker plot of pre-treatment in variety and accuracy in grammar and vocabulary in both groups
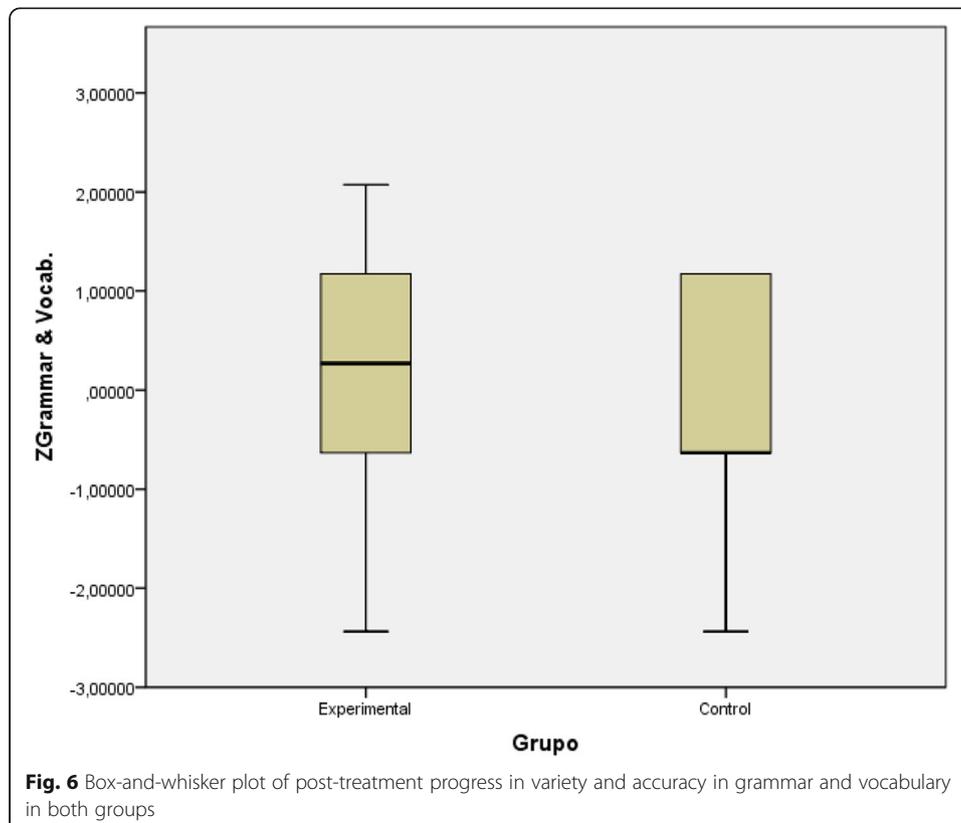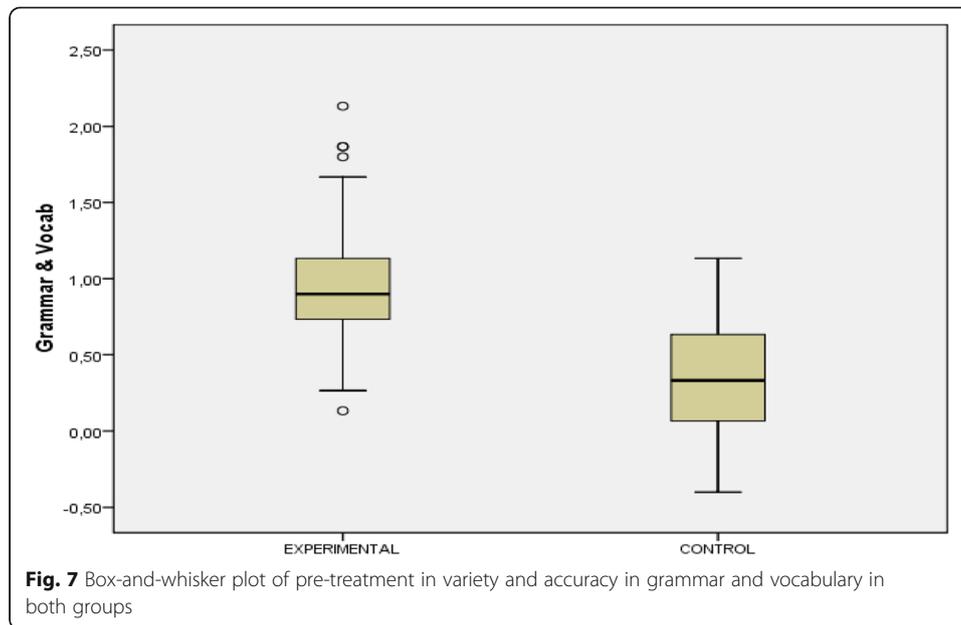
professional experience. Attrition or mortality may have affected the study as data could not be drawn from 7 dropouts in the experimental group and 2 in the control group.
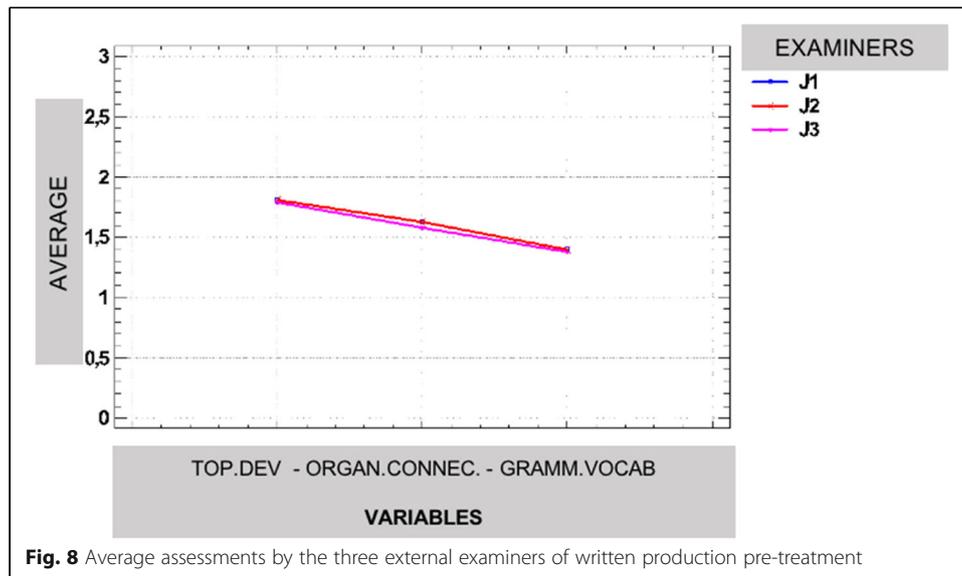
As for the instrumentation, the design of the pre-post test did not vary in any of the groups in spite of different approaches in the development of the lessons. Whereas the control group was more focused on textbook-related activities and developing language skills systematically, the experimental group had autonomous work to do outside of class to learn about specific topics before attending the lessons. However, keeping track of students' activity outside of class was at times difficult to measure. In a few cases, students who did not comply with their homework by reading or watching the videos and they were asked to do so without interfering with the other students, preferably outside of class.

External validity

Situational factors may limit generalizability as the participants of the study were all engineering students who might have found certain limitations to understand the complexities of the problems described in the web-based simulation on social-political issues. However, these types of simulations are often applied to students taking optional

Table 10 Comparative analysis of the mean level of progress in grammar and vocabulary (post-treatment)

Overall score (post-treatment) Grammar and vocabulary	N	Levene's Test		Student's t test					95% CI	
		f	sig.	t	Overall	Sig. (bilateral)	Mean diff.	Std. error of diff.	Lower	Upper
Experimental	50	.044	.835	8.172	119.000	.000	.302	.037	.229	.044
Control	71			7.988	96.633	.000	.302	.037	.227	.197

Fig. 8 Average assessments by the three external examiners of written production pre-treatment

conversational courses as the one presented in the study. It can also be stated that participants' reactions to being studied may have altered their behaviour and therefore the study results. Regarding the experimenter effects, only one of the researchers was in charge of teaching one experimental group. Due to this, we have resorted to three external examiners to bring reliability to the study.

Conclusions

In this study, the use of simulations effectively improved students' written production, regardless of the student's initial level, for students in the experimental and control groups. Progress in written production was greater for students who participated in the simulation-based instruction in the experimental groups. Progress in *organising and linking ideas* was statistically higher for students in the experimental groups. It may be inferred that the great exposure to written input in the target language, the critical dialogical exchanges about the different simulation issues, the elaboration of a written

Fig. 9 Average assessments by the three external examiners of written production post-treatment
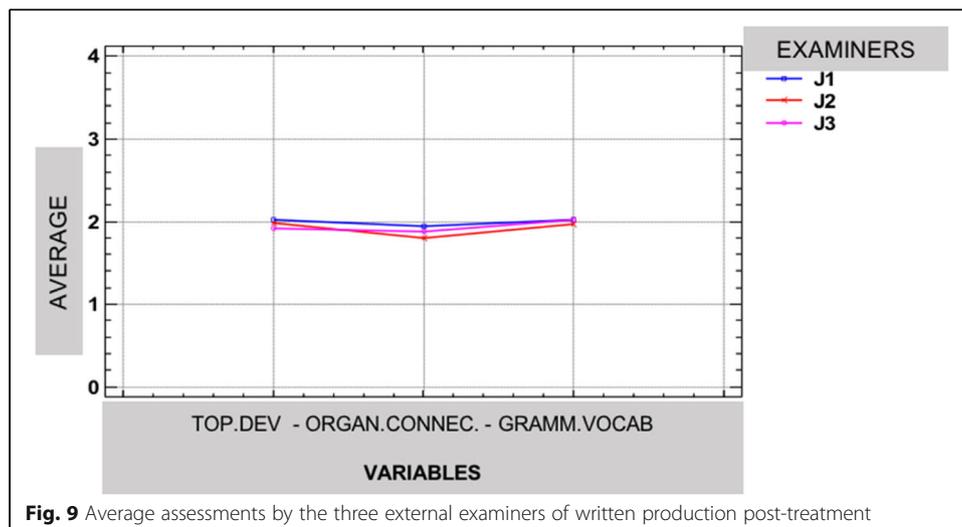
proposal to be later negotiated by other participants have led students to organize their ideas coherently and cohesively. The control groups, instead, did progress in the organization of ideas, grammar and variety of expressions though not as much as the experimental group. It may be inferred that the control group was more focused on dealing with the topics and written models presented by the course book. Thus, their written production was well-structured, there was good control of grammar though the ideas seemed similar to some of the written texts in the course book. However, notably, students' progress in *variety and accuracy in grammar and vocabulary;* and *language development* was non-significant for students in the experimental group. By establishing a knowledge base that would support the realization of the target language, students should have enriched their content knowledge of the topics using written and video material outside the classroom and simulations, debates and forums in class. Results indicate that these students were more inclined to use and overuse vocabulary and structures they were familiar with. A deeper interpretation, though, is linked to Wells (1999) and Lipman (2003) who supported the idea of developing thinking skills to be revealed through language use within a 'community of inquiry' in the classroom. Mastering the content knowledge of a specific topic did not guarantee language creativity in the present study.

In a future study, an ANOVA will report differences between the experimental and control groups by comparing the means of two or more variables at different times, between the two groups to clearly identify the differences pre and post treatment; and within the same group pre and post treatment. Furthermore, In the future, lessons will integrate simulations and an inquiry-based model that enhance reflections on the simulation experience and the students' learning in an attempt to reach a common reflection that favours inter-subjectivity and language development.

Additional files

Additional file 1: 1. Material used by the Experimental Group. (DOCX 43 kb)

Additional file 2: Essay Writing Assessment Criteria. (DOCX 32 kb)

Additional file 3: Letter of Consent. (DOCX 19 kb)

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

Andreu-Andrés, M. A., & García-Casas, M. (2011). Perceptions of gaming as experiential learning by engineering students. *International Journal of Engineering Education*, *27*(4), 795–804 Tempus Publications.

Author (2011). Student perceptions of collaborative work in telematic simulation. *Journal of Simulation/Gaming for Learning and Development*, *1*(1), 1–12.

Beckem, J. M. (2012). Bringing life to learning: Immersive experiential learning simulations for online and blended courses. *Journal of Asynchronous Learning Networks*, *16*(5), 61–70.

Burke, H., & Mancuso, L. (2012). Social cognitive theory, metacognition, and simulation learning in nursing education. *The Journal of Nursing Education*, *51*(10), 543–548.

Chang, Y. C., Peng, H. Y., & Chao, H. C. (2010). Examining the effects of learning motivation and of course design in an instructional simulation game. *Interactive Learning Environments*, *18*(4), 319–339.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*, (2nd ed.,). Hillsdale: Lawrence Earlbaum Associates.

Council of Europe. Council for Cultural Co-operation. Education Committee. Modern Languages Division (2001). Common European framework of reference for languages: Learning, teaching. In *assessment*. Cambridge: Cambridge University Press.

Crookall, D. (2010). Serious games, debriefing, and simulation/gaming as a discipline. *Simulation and Gaming*, *41*(6), 898–920.

Crookall, D., & Oxford, R. L. (1990). *Simulation, gaming, and language learning*. New York: Newbury House Publishers.

Ekker, K. (2000). *Changes in attitude towards simulation-based distributed learning. Project DoCTA*, (pp. 112–120). Oslo: Design and use of Collaborative Telelearning Artefacts.

Ekker, K. (2004). User satisfaction and attitudes towards an internet-based simulation. In D. Kinshuk, G. Sampson, & P. Isaías (Eds.), *Proceedings of the IADIS international conference cognition and exploratory learning in digital age*, (pp. 224–232). Lisbon: IADIS.

Gegenfurtner, A., Quesada-Pallarès, C., & Knogler, M. (2014). Digital simulation-based training: A meta-analysis. *British Journal of Educational Technology*, *45*(6), 1097–1114.

Halleck, G., & Coll-García, J. (2011). Developing problem-solving and intercultural communication: An online simulation for engineering students. *Journal of Simulation/Gaming for Learning and Development*, *1*(1), 1–12.

Jones, K. (2013). *Simulations: A handbook for teachers and trainers*. London: Routledge.

Klabbers, J. H. (2001). The emerging field of simulation and gaming: Meanings of a retrospect. *Simulation and Gaming*, *32*(4), 471–480.

Klabbers, J. H. (2009). *The magic circle: Principles of gaming and simulation*. Rotterdam: Sense Publishers.

Kriz, W. C. (2003). Creating effective learning environments and learning organizations through gaming simulation design. *Simulation and Gaming*, *34*(4), 495–511.

Levine, G. (2004). Global simulation: A student-centered, task-based format for intermediate foreign language courses. *Foreign Language Annals*, *37*(1), 26–36.

Lipman, M. (2003). *Thinking in education*. Cambridge: Cambridge University Press.

Michelson, K., & Dupuy, B. (2014). Multi-storied lives: Global simulation as an approach to developing multiliteracies in an intermediate French course. *L2 Journal*, *6*(1), 21–49.

O'Flaherty, J., & Phillips, C. (2015). The use of flipped classrooms in higher education: A scoping review. *The Internet and Higher Education*, *25*(1), 85–95.

Ranchhod, A., Gurău, C., Loukis, E., & Trivedi, R. (2014). Evaluating the educational effectiveness of simulation games: A value generation model. *Information Sciences*, *264*(1), 75–90.

Reeve, J. (2013). How students create motivationally supportive learning environments for themselves: The concept of agentic engagement. *Journal of Educational Psychology*, *105*(3), 579–595 https://doi.org/10.1037/a0032690.

Rising, B. (2009). Business simulations as a vehicle for language acquisition. In V. Guillén-Nieto, C. Marimón-Llorca, & C. Vargas-Sierra (Eds.), *Intercultural business communication and simulation and gaming methodology*, (pp. 317–354). Bern: Peter Lang.

Strayer, J. F. (2007). The effects of the classroom flip on the learning environment: A comparison of learning activity in a traditional classroom and a flip classroom that used an intelligent tutoring system. PhD dissertation, Ohio State University. https://etd.ohiolink.edu/!etd.send_file?accession=osu1189523914. Accessed 27 Apr 2018.

Strayer, J. F. (2012). How learning in an inverted classroom influences cooperation, innovation and task orientation. *Learning Environments Research*, *15*(2), 171–193.

Thatcher, D. C., & Robinson, M. J. (1985). *An introduction to games and simulations in education*. Hants: Solent Simulations.

Tourón, J., Santiago, R., & Diez, A. (2014). *The Flipped Classroom: Cómo convertir la escuela en un espacio de aprendizaje*. Spain: Grupo Océano.

Tucker, B. (2012). The flipped classroom: Online instruction at home frees class time for learning. *Education Next*, *12*(1), 82–84.

University of Cambridge. ESOL Examinations (2012). *Research Notes* [PDF] Accessed 26 January 2018. http://www.cambridgeenglish.org/images/23166-research-notes-49.pdf.

Wedig, T. (2010). Getting the Most from classroom simulations: Strategies for maximizing learning outcomes. *PS: Political Science and Politics*, *43*(3), 547–555.

Wells, G. (1999). *Dialogic inquiry: Towards a socio-cultural practice and theory of education*. Cambridge: Cambridge University Press.

Wiggins, B. E. (2012). Toward a model of intercultural communication in simulations. *Simulation & Gaming*, *43*(4), 550–572. https://doi.org/10.1177/1046878111414486.

Woodhouse, T. (2011). Thai University Students' Perceptions of Simulation for Language Education. https://absel-ojs-ttu.tdl.org/absel/index.php/absel/article/view/3026.