Evaluating the effects of mobile applications on course assessment: A quasi-experiment on a macroeconomics course

Universities are facing the need to rethink their educational strategies, especially due to the emergence of new technologies, such as mobile applications, which have had great expectations. Previous studies have been focused on changes in student engagement from using mobile applications in the classroom, whereas there has been little research on the impact of mobile applications on student assessment. This research uses a quasi-experimental study to examine the relationship between student assessment and the use of a mobile application. Two groups of students (a control and an experimental group) were tested in the same academic semester with the same lecturer. Two analyses were carried out (t-test and difference-in-differences) to evaluate this relationship. Contrary to the general expectations, the results showed that there is no significant difference on assessment when comparing the two groups' scores. However, students showed a positive attitude in engaging with the mobile application. Although there has been an increase on the use of mobile applications in classrooms, they do not directly affect student scores. This research shows that mobile applications should be used as a complement to traditional education, and not as a substitute to it.

Keywords: higher education; course delivery; assessment; student response systems; teaching and learning; Millennials

1. Introduction

Universities are facing the need to rethink their educational strategies, especially due to the emergence of new technologies (Lucena, 2016; Riaza & Rodríguez, 2016). In this context, mobile technology is being increasingly used in educational institutions, and in higher education in particular, with great expectations (see Balta, Perera-Rodríguez, & Hervás-Gómez, 2017). Applications, such as Socrative.com, remind.com, or Kahoot.it, are examples of mobile technology, also called m-learning development (Kokina & Juras, 2017; Onodipe, 2017; Zou & Lambert, 2017). These applications are used in universities as: Student Response Systems (SRS) (Lim, 2017; Onodipe, 2017; Zou & Lambert, 2017); "clickers" (Guarascio, Nemecek, & Zimmerman, 2017), and as online homework platforms (Balta et al., 2017).

Numerous studies (e.g. Fotaris, Mastoras, Leinfellner, & Rosunally, 2016; Guarascio et al., 2017; Lim, 2017; Onodipe, 2017; Wuthisatian & Thanetsunthorn, 2019) have focused on assessing the effect of SRS' use on student engagement, with most of them showing an increase of classroom motivation. However, little research has been done on the relation between the use of mobile applications and assessment. An example is presented by Arain et al (2018), who compared students' final exam scores between an experimental and a control group.

The introduction of such new technologies in the classroom implies extra efforts, on top of than those of traditional teaching (Savec, Hrast, Šuligoj, & Avsec, 2018; Sefo, Granados Romero & Fernández-Larragueta, 2017). Because of this, it is important to analyse if this new technologies leads to better student's results, i.e. assessment. The present research evaluates the effect of a SRS, through a quasi-experiment, on assessment.

The structure of the paper is as follows: Sections 2 and 3 provides an overview of the main literature on this field; Section 4 presents the quasi-experiment context and design; Section 5 shows the main results, and Section 6 presents the discussion and Section 7 the conclusions.

2. Literature Review

A key component of the learning process is the type of interaction selected (Vrasidas, 2000). Moore (1989) suggested three types of interaction of the learning process: (1) learner-content; (2) learner-instructor; and (3) learner-learner interaction. Hillman, Willis, and Gunawardena (1994) complemented this with a new kind of interaction: the learner-interface interaction. In a context with numerous innovations in teaching methodology and the use of ICTs for learning, this type of interaction has shown its vital importance in distance education (Danesh, Baily, & Whisenand, 2015), as well as in traditional face-to-face courses (Sinha, Khreisat, & Sharma, 2009), through mixing conventional pedagogical approaches with online learning (Mattheos & Lic, 2004).

The need to design and implement new teaching methodologies at a university is directly related to students' characteristics (Lim, 2017). Most of the current students are part of the Millennial generation, which is composed of digital natives who have developed a great capacity to perform multi-task activities (Picault, 2019; Varela-Candamio, Novo-Corti, & Barreiro-Gen, 2014). Millennials use technology constantly, desire immediate feedback (Montenery et al., 2013), and are continuously connected to the digital world (Lim, 2017; Manning, Keiper, & Jenny, 2017). Moreover, student engagement is shown to decline as student pass to higher levels of education (Anderson et al., 2019). Due to theses challenges, innovative teaching methods at the University level have been proposed to better engage them in the classroom (Sinha et al., 2009). Engagement can be understood as an individual's involvement with the educationally relevant activities and conditions that are instrumental to their learning (Jennings & Angelo, 2006).

The use of m-learning applications, as SRS in the classroom, makes the students more engaged (Guarascio et al., 2017; Kokina & Juras, 2017; Wuthisatian &

Thanetsunthorn, 2019), since the students can express themselves in different ways than with traditional SRS methods, such as oral question-and-answer reports, or pen-and-paper questionnaires (Zou & Lambert, 2017).

There have been a number of SRS applications, such as Socrative.com, remind.com, and Kahoot.it (Kokina & Juras, 2017; Onodipe, 2017; Zou & Lambert, 2017). Among these, Socrative has had great acceptance and use amongst teachers (Balta et al., 2017; Haintz & Ebner, 2014; Kokina & Juras, 2017). Socrative is a SRS that is used to engage, assess and personalize classes. Educators can carry out formative assessments through quizzes, quick question polls, exit tickets and space races using their Socrative Teacher app. Socrative let teachers instantly grade, aggregate and provide visuals of results to help them identify opportunities for further instruction. The teacher has to register for a Socrative Teacher account and the students can connect to his/her unique room by opening their apps or joining his/her room on Socrative from any device (e.g. smartphones, tablets, and laptops)[1].

Guarascio et al. (2017) conducted a survey to measure student preferences between standard SRS methods compared to Socrative, where the students preferred the use of the Socrative application to the traditional SRS methods because they felt Socrative helped them to be more active in class. The Kokina and Juras (2017) study shows similar results for using m-learning applications. Zou and Lambert (2017) emphasised that the acceptance in the classroom of Socrative was very positive, as well as of TodaysMeet and Google Drive. However, these three papers did not take into consideration a control group to compare their results. Fotaris et al. (2016) carried out a comparison of students and instructors in two consecutive years, where they applied m-

[1] For more information refer to www.socrative.com

learning technologies in one year but not in the other. They found that the gamified approach mixed with m-technology applications motivated and enriched students and instructors.

The previous examples highlight the positive effects of m-learning applications on engagement. However, there have been few carried out on the links between the use of m-learning applications and assessment. A quasi-experiment evaluates the relation between an intervention and an outcome, comparing an experimental/treatment group and a control group (Handley, Lyles, McCulloch, & Cattamanchi, 2018). The most common ways to reduce bias in quasi-experiments are, for example, assigning members partially randomly to the treatment and control groups, and controlling that both groups have the same characteristics before the experiment, that is to say, if there are no differences in the composition of the two groups. An example of a quasi-experiment that evaluated the link between the use of m-learning applications and assessment is the study of Balta et al. (2017), who used Socrative as an online homework platform and found that students' exam scores were positively influenced by the use of Socrative. In this research, the experimental group comprised just students who volunteered. Another example is the research by Chui, Martin, and Pike (2013) that offers an empirical analysis of links between innovative SRS and assessment in Accounting using a single-difference method. Their results did not show significant differences in examination performance between the control and the experimental group. It should be noted, that this study did not measure the knowledge of the students at the beginning of the course and thus it could not assess the differences between the beginning and end of the course. Golenhofen et al (2019) attempted to measure the effect of a SRS (eMed-App) on assessment. They concluded that the app itself did not result in better outcomes, but they did not have control group. Arain et al (2018) carried out a quasi-experiment using a single-difference method. Their results showed positive effects on assessment from using a M-Learning App (Darsgah) in a Communication Skills course for engineers.

3. Linking course delivery and assessment

In a course, there are three key elements: learning outcomes; delivery of the course; and assessment; and there must be consistency between these (Norton, 2009). A learning outcome is "*a written statement of intended and/or desired outcome to be manifested by student performance*" (Prøitz, 2010). Course delivery includes lectures, active and cooperative learning, instructional technology, and other techniques (Felder, & Brent, 2003; Lozano, Ceulemans, & Scarff Seatter, 2015). Assessment is linked to predetermined learning objectives, and through an appropriate delivery of the course (Combs, Gibson, Hays, Saly, & Wendt, 2008; Streveler, Smith, & Pilotte, 2012). Changes in delivery of the course may have effects on assessment (Balta et al., 2017).

According to Norton (2009), assessment is a key element since is the most effective way to change student learning. Assessment has four main purposes: (1) pedagogy; (2) measurement or evaluation of student knowledge, understanding, abilities or skills; (3) standardisation (establishing the student's performance), and (4) certification of level of achievement (Norton, 2009). Assessment is "*a machine for reasoning about what students know, can do, or have accomplished, based on a handful of things they say, do, or make in particular settings"* (Mislevy, Steinberg, & Almond, 2003). Two actors, the administrator and the participant, are involved in the assessment cycle, which has four different processes: (1) Activity Selection; (2) Presentation; (3) Response Processing, and (4) Summary Scoring (Almond, Steinberg, & Mislevy, 2002).

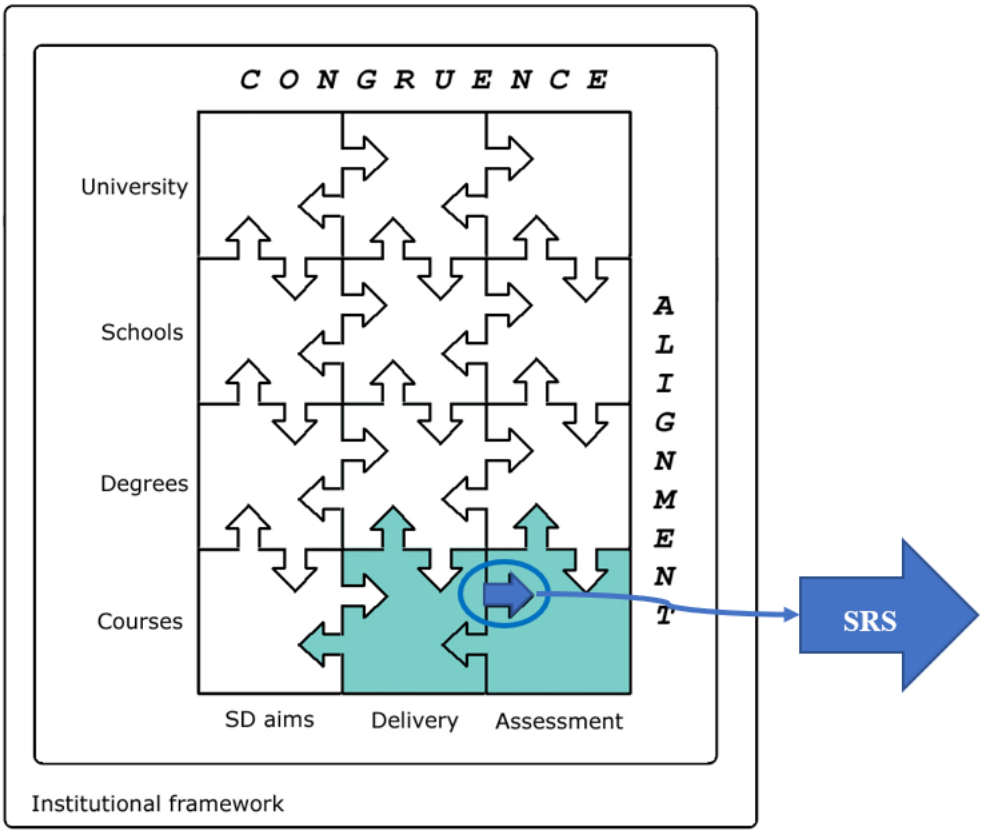

Figure 1. Changing the course delivery using SRS and its link with the assessment (adapted from Lozano 2010)

This paper is focused on the link between the delivery and the assessment (Figure 1), and, in particular, in changing the course delivery and its link with the assessment. Since, the aims of the course had already been developed during the previous year, specific changes in the way of teaching (using SRS) were made in order to detect positive or negative impact on assessment, that is to say, if these changes could have influenced their performance at the examination.

4. Methods

Students for the Law Degree usually have to learn basic principles of Economics in their first course and this is often complicated, due to the need to use a vocabulary and tools (mathematics, graphics) that are not commonly used in law. Teaching

Macroeconomics is based on the use of analytical models that require good mathematical ability (as discussed by Barnett, 2009), and when using traditional methods it is not easy to achieve student engagement, especially in subjects with a high level of difficulty (see Fotaris et al., 2016).

For these reasons, a new way to deliver the course was designed and implemented in an experimental group in order to detect effects on assessment.

The study was carried out with students of the subject "Principles of Economics and Public Finance" during the second semester of the academic year 2017-2018. This subject was taught at the first year in the Law Degree at University of A Coruna, Spain (University of A Coruna, 2018). The students were organised in two groups (morning and afternoon). The subject was divided into two parts: Principles of Economics, belonging to the area of Economic Analysis, and Public Finance, which was part of the area of Applied Economics. Therefore, there were two people from different areas who had to teach each part and the examinations were done separately. The author of this paper taught the Principles of Economics part during the period under study upon which this study is based.

4.1 Learning outcomes, course delivery and assessment

Four course objectives or learning outcomes were set up, as presented in Table 1.

Table 1. Learning outcomes

Learning Outcomes:	**Description:**
Learning Outcome 1:	To be familiar and able to discuss the general principles of Economics.
Learning Outcome 2:	To understand the concept of Macroeconomics and its general features and principles.
Learning Outcome 3:	To be aware of the links between inflation and unemployment.
Learning Outcome 4:	To be able to work with documents, graphics and texts related to topics taught during the course.

The five-week course contained topics related to the basic concepts of Macroeconomics, such as unemployment, inflation, economic growth and economic development.

The course was designed with different types of interaction (see Table 2). There were theoretical and practical classes in this course. Theoretical hours (two per week) were normal lecturing with the help of a set text (learner-instructor and learner-content) in both groups. The practice hours focused on student participation, with three types of interaction in the morning group: (1) learner-content; (2) learner-learner (collaboration with peers), and (3) learner-instructor; while in the afternoon group were used (1) learner-content; (2) learner-learner; (3) learner-instructor (just giving instructions), and (4) learner-interface interaction with feedback. That is to say, practical questions (similar to questions in the final exam) were completed in a different way in the two groups: in the morning in a traditional way; and in the afternoon using Socrative.

Table 2. Interactions types

Class type	**MORNING GROUP (control group)**	**AFTERNOON GROUP (experimental group)**
Theoretical classes	learner-instructor learner-content	learner-instructor learner-content
Practical classes	learner-content learner-learner learner-instructor	learner-content learner-learner learner-instructor (just giving instructions) **learner-interface interaction: *Socrative***

Students were doing multiple choice questions in practical classes to prepare for the exam. Questions for both groups were the same, but using different techniques.

Students had to pass a multiple choice test at the end of the semester. Each question had four potential answers, but just one was correct.

4.2 Intervention and sample

This study employed a quasi-experimental design to measure the m-learning application's effects on students' assessment. The experiment was designed to isolate the effects of using this new technology, and then used the "double-difference" method to test the value added of the instruction.

The first day of class all the students (N=134) did a pre-test. This pre-test contained eight questions related to the topics of the subject, such as unemployment, inflation, or GDP. The main purpose of this initial test was to assess their average scores before the beginning of the course.

The control group (untreated, they did not participate in the innovation) was the group taught in the morning, and the experimental group during the afternoon sessions. The groups were divided according to their last name, where the A-L are in the first group and M-Z in the second. There was not possible any switching between groups during the course. The "treatment" or intervention consisted of using Socrative in the practical classes instead of traditional methods to solve the questions raised. Therefore, students in the treatment group used an interactive response device to answer their questions during practical class time. Attendance at practice classes was controlled during the five weeks.

The students from both groups did a simulation of the final exam on the last day of class, composed of eight questions about the topics taught, and in the same format as the final exam. Students did not know beforehand that they would have to sit this mid-term exam.

The students also completed a survey composed of two parts. The first part, done by the whole sample, focused on statements related to Economics, ICTs and time distribution (see Table 3). Their degree of agreement or disagreement was scored using a Likert scale of 5 points in which 1 indicates the position most in disagreement with the proposed statement, and 5 the one most in agreement. The main objectives of this survey were to detect differences between the control and experimental groups, and to identify their attitudes towards these topics.

The second part, done just by the experimental group, was focused on Socrative' questions. All the questions were taken from previous literature, as Table 4 shows.

Table 3. Survey. Part I. Macroeconomics, ICTs and time distribution. Questions

Q1.1	I liked learning issues related with Macroeconomics.	
Q1.2	I have regularly attended the theoretical classes in the Principles of Economics.	
Q1.3	I like using new technologies (computer, mobile phone, etc.).	
Q1.4	I believe that new technologies facilitate my learning.	
Q1.5	I like using new technologies in the classroom.	
Q1.6	We use new technologies (computer, PPoint, mobile phone, etc.) in other subjects in the classroom.	
Q1.7	Time distribution: Indicate the hours you dedicate per week to the following activities:	Social networks Doing sport Sleeping Going out with friends Doing private study (outside of class)

The final sample was composed by 85 students, aged from 18 to 24 years, that met the following requirements (with the main objective of having the same sample in each group for all the tests): (1) participating for the first time in this course; (2) agreed to do the pre-test; (3) to attend the practical lessons; (4) to do the mid-term exam; (5) to complete the survey, and, (6) to do the final test. Thus, the sample included a control group (CG) of 45 students (26 females) who attended the Economics course in the morning classes, and an experimental group (EG) of 40 students (21 females), who attended the afternoon classes. There were no significant differences in demographic characteristics of interest between these two groups.

Table 4. Survey. Part II. Socrative. Questions

	Question	Based on
Q2.1	I liked Socrative.	(Balta et al., 2017)
Q2.2	Socrative tool helps in learning material.	(Kokina & Juras, 2017)
Q2.3	The usage of Socrative tool motivates me in my learning.	(Lim, 2017)
Q2.4	The usage of Socrative tool encourages me to stay focus in the classroom.	(Lim, 2017)
Q2.5	Socrative tool encourages student participation in the classroom.	(Kokina & Juras, 2017)
Q2.6	Socrative is more fun than traditional methods (Pen-and-paper questionnaires).	(Zou & Lambert, 2017)
Q2.7	Socrative is more stressful than traditional methods (Pen-and-paper questionnaires).	(Zou & Lambert, 2017)
Q2.8	Overall, I am satisfied with the usage of Socrative tool in my classroom.	(Lim, 2017)
Q2.9	Socrative should be used when teaching other subjects also.	(Balta et al., 2017)

4.3 Data analysis

A quantitative analysis was done using different techniques:

First, t-test for Equality of Means was carried out using IBM SPSS Statistics 24, in order to find statistically significant differences in the pre-test, in the mid-term review, and in the final test, among the two groups.

Second, a difference-in-differences (DD) analysis was performed. The difference-in-difference analysis, also known as the "double difference" method, compares the changes in the results over time between experimental and control group in order to estimate the impact of the treatment (Pomeranz, 2011; White & Sabarwal, 2014), in this case, the use of Socrative as SRS. The DD method provides a stronger estimated impact than the single-difference method, which only compares the difference in outcomes between the experimental and control groups after the intervention (at t+1). The application of the DD analysis eliminates the difference in the results between the two groups at the beginning of the course (White & Sabarwal, 2014).

Third, a complementary analysis was conducted to analyse their perceptions of Economy, ICTs and Socrative, using survey results.

4.4 Limitations of the methods: reliability, validity, and generalisability

This research may be affected by the following limitations. Some barriers exist when a quasi-experiment with students is designed. On the one hand, the research has to be designed to avoid, as much as possible, potential bias, using a control and an experimental group; and, on the other hand, equity in the treatment of students is mandatory. Before the beginning of the course, the average scores of the both groups have been measured in order to avoid comparing heterogeneous groups. The quasi-experiment was designed to isolate the effects of using this new technology, and to test just the value added of the instruction.

All the students were evaluated the same day in each of the three tests that they did in order to reduce the participant error. The test could not be anonymous, which could imply the existence of participant bias when they had to give with their opinions (e.g. they could provide more positive opinions than the real one because they could want to please the lecturer), but not when they had to answer questions about the content of the course, in the assessment questions, in which all of them tried to do their best. Each of the tests were composed by multiple-choice questions in order to obtain objective results.

A more complex way to evaluate the students could get more information about their real skills and knowledge.

In practice, the lecturer's attitude toward the m-application can affect the relation that the students have with this new way to learn.

The course was taught at a European Higher Education Establishment, with subjects lasting for just one semester or less. This research is based on a single course at one university which lowers the generalisation of gathered results. The research was performed in a specific context and discipline. This implies that it is not possible to generalise the results.

There were few problems in the use of Socrative, related to WiFi connection and lack of space in mobile phones to download the application. However, these obstacles were overcome, for instance, lending them laptops.

5. Results

5.1 T-test

T-test for Equality of Means were performed with the aim of comparing the averages of the experimental and control groups in the pre-test, in the mid-term exam and in the final exam.

The classification variable was, for both groups, group membership, so the scores of the control and experimental groups in each test were compared, in order to detect possible differences in the marks of both. As a prelude to the t-test, a Levene's test was conducted and it indicated that it was possible to assume homogeneous variances.

Table 5. T-test for equality of means

Question	Group	N	Mean	Std. Deviation	Statistic	Sig.	Degrees of Freedom	Differences on averages

Pre-test score	CG*	45	4.200	1.199	-1.444	0.152	83	-0.400
	EG*	40	4.600	1.355				
Mid-term exam score	CG	45	4.600	1.558	0.222	0.825	83	0.075
	EG	40	4.525	1.552				
Final exam score	CG	45	4.211	1.384	2.720	0.008	83	0.819
	EG	40	3.450	1.170				

*CG = Control group; EG = Experimental Group.

The mean score of the control group in the pre-test was 4.2 (minimum = 0; maximum = 8), as Table 5 shows. Meanwhile, the mean for the experimental group in that test was 4.6. According to this analysis, the differences were not significant at a 95% level. Therefore, at the beginning the two groups had a similar level of knowledge about the topics that they will study.

In the second case, the average score of the control and experimental groups did not differ significantly ($\overline{Y_{CG}}(1) = 4.6$; $\overline{Y_{EG}}(1) = 4.525$).

In the final exam, the mean score of the control group was 4.211 and the mean for the experimental group in that test was 3.450. In that case the differences were significant at a 99% level (Sig. = 0.008).

In order to analyse this issue more deeply, a difference-in-differences analysis was performed.

5.2 Difference-in-differences (DD) analysis

After applying the DD method using data from the pre-test and the mid-term test, it was perceived that the morning group (CG) improved its average by 0.4, while the afternoon group (EG) remained stagnant (see Table 6).

Table 6. DD analysis between the pre-test and mid-term exams. Results.

Means	**Pre-test (t-1)**	**Mid-term test (t)**	**Change**
EG	4.6	4.525	-0.075
CG	4.2	4.6	+0.4

Differences	Selection bias or difference at the starting point: $\overline{Y_{EG}}(t-1) - \overline{Y_{CG}}\ (t-1) = +0.4$	Single-difference method (SD): $\overline{Y_{EG}}(t) - \overline{Y_{EG}}\ (t-1) =$ -0.075	Difference-in-differences (DD) method: $\boldsymbol{DD} = \overline{Y_{EG}} - \overline{Y_{CG}} =$ -0.475

If $\overline{Y_{EG}}(t)$ is the average' scores of the experimental group in the mid-term test; $\overline{Y_{EG}}\ (t-1)$ is the average' scores of the experimental group in the pre-test; and $\overline{Y_{CG}}\ (t-1)$ is the average' scores of the control group in the pre-test; DD analysis is equal to (1):

$$[\overline{Y_{EG}}(t) - \overline{Y_{EG}}\ (t-1)] - [\overline{Y_{EG}}(t-1) - \overline{Y_{CG}}\ (t-1)] \quad (1)$$

Taking into account the results at the starting point and at the mid-term, DD was equal to - 0.475:

[4.525 – 4.6] – [4.6 – 4.2] = -0.075 – 0.4 = - 0.475

Table 7. DD analysis between the pre-test and final test. Results.

Means	**Pre-test (t-1)**	**Final test (t +1)**	**Change**
EG	4.6	3.45	- 1.15
CG	4.2	4.211	+0.011
Differences	Selection bias or difference at the starting point: $\overline{Y_{EG}}(t-1) - \overline{Y_{CG}}\ (t-1) = +0.4$	Single-difference method (SD): $\overline{Y_{EG}}(t+1) - \overline{Y_{CG}}\ (t+1) =$ - 0.761	Difference-in-differences (DD) method: $\boldsymbol{DD} = \overline{Y_{EG}} - \overline{Y_{CG}} =$ -1.161

After applying the DD method using data from the pre-test and the final test, the control group remained stagnant and for the experimental group its average was worse by 1.15 (see Table 7).

If $\overline{Y_{EG}}(t+1)$ is the average' scores of the experimental group in the final test; $\overline{Y_{EG}}\ (t-1)$ is the average' scores of the experimental group in the pre-test; and $\overline{Y_{CG}}\ (t-1)$ is the average' scores of the control group in the pre-test; DD analysis in this case is equal to (2):

$$[\overline{Y_{EG}}(t+1) - \overline{Y_{EG}}\ (t-1)] - [\overline{Y_{EG}}(t-1) - \overline{Y_{CG}}\ (t-1)] \quad (2)$$

Taking into account the results at the starting point and at the final test, DD was equal to - 1.161:

[3.45 – 4.211] – [4.6 – 4.2] = -0.761 – 0.4 = - 1.161

5.3 Survey results

5.3.1 Part I. General opinions

The purpose of these general questions was to compare attitudes towards ICT and towards Macroeconomics in both groups and to be sure that the only difference between the control and experimental group was the use of Socrative for the latter. Students (in both groups) were not used to incorporating ICTs in class (see means in Q1.6). There were no significant differences between their mean opinions about economics, ICT, and ICT in class (Table 8), and homogeneous variances could be assumed according to Levene's test for Equality of variances. Differences between time distribution were not found, even in studying outside the class. In general, they had positive opinions about all these issues in both groups, and they spend many hours on social networks.

Table 8. T-test for equality of means

Q	Group	N	Mean	Std. Deviation	Statistic	Sig.	Degrees of Freedom	Differences in averages
Q1.1	CG	45	4.00	0.640	0.494	0.4842	83	-0.10
	EG	40	4.10	0.672				

Q1.2	CG	45	4.22	0.823	1.173	0.2820	83	-0.22
	EG	40	4.44	0.672				
Q1.3	CG	45	4.47	0.815	0.010	0.9224	83	-0.02
	EG	40	4.45	0.749				
Q1.4	CG	45	4.22	0.876	0.068	0.7948	83	-0.04
	EG	40	4.18	0.781				
Q1.5	CG	45	3.98	1.011	1.288	0.2596	83	-0.22
	EG	40	4.20	0.758				
Q1.6	CG	45	3.71	0.920	0.113	0.7382	83	0.06
	EG	40	3.65	0.740				

5.3.2 Part II. Socrative

Table 9 shows the main results of the students' answers about Socrative. The Cronbach-alpha value ($\alpha = 0.811$) reflects high internal consistency of the instrument. Analysing these results, the experimental group produced the highest means (over 4) in Q2.5, Q2.6, Q2.8, and Q2.9 (Table 9). Students considered that Socrative enhances participation in the classroom (4.18), that this application is more fun than traditional methods (4.25), they were satisfied with their use (4.20) and they thought Socrative should be used in other subjects as well (4.05). Questions related with helping in learning (Q2.2), engagement (Q2.3) and staying focused (Q.2.4) presented moderately high means (3.97; 3.85, and 3.60, respectively). They agreed that Socrative is less stressful than traditional methods, such as pen-and-paper questionnaires.

Table 9. Part II. Socrative Questions. Descriptive statistics.

	Question	**Mean**	**Std. Deviation**
Q2.1	I liked Socrative.	4.00	0.679
Q2.2	Socrative tool helps in learning material.	3.97	0.660
Q2.3	The usage of Socrative tool motivates me in my learning.	3.85	0.864
Q2.4	The usage of Socrative tool encourages me to stay focused in the classroom.	3.60	0.955
Q2.5	Socrative tool encourages student participation in the classroom.	4.18	0.844

Q2.6	Socrative is more fun than traditional methods (Pen-and-paper questionnaires).	4.25	0.809
Q2.7	Socrative is more stressful than traditional methods (Pen-and-paper questionnaires).	2.13	1.159
Q2.8	Overall, I am satisfied with the usage of Socrative tool in my classroom.	4.20	0.516
Q2.9	Socrative should be used in other subjects as well.	4.05	0.932

6. Discussion

The majority of students in Higher Education Institutions are now millennials, who are digital natives and multi-taskers. Millennials are accustomed to a faster-paced life and quicker flow of information than previous generations. At the same time, engagement is shown to decline as student pass to higher levels of education (Anderson et al., 2019). These two phenomena present challenges that suggest instructors the need to adapt their teaching methods (Picault, 2019). A way to address these challenges has been the incorporation of new technologies, such as apps and media, to teaching (Montenery et al., 2013; Picault, 2019). HEIs have invested in mobile learning applications, such as SRS, to deliver course content in an innovative way and address the needs of this new generation of students. Such applications have shown that student engagement increases; however, the effect of these technologies on student assessment has not been investigated in depth.

This paper examines the relation between assessment and the use of mobile technologies as SRS in the classroom. Contrary to the general expectations, the results of this study show that mobile learning applications have no positive effects on assessment in the short-term, which concurs with the findings of Chui et al. (2013) and Harmon and Tomolonis (2019). However, positive attitudes towards engaging with new technologies, technologies in the classroom and, specifically, favourable towards Socrative, were found (concurring with Guarascio et al., 2017; Lim, 2017; Zou & Lambert, 2017). The students highlighted that Socrative was to fun learn with, it helped to improve engagement, and stay focused on the given tasks. Students also highlighted that Socrative was less stressful than traditional methods, such as pen-and-paper questionnaires, which disagrees with the findings of Zou & Lambert (2017).

7. Conclusions

The results of this study show that the expectations of the results of mobile learning technologies may be exaggerated, since no positive effects on student assessment were found while using Socrative. However, it should be noted that mobile learning applications improve engagement, and this could be help deliver content in a better way to the new generation of students.

Lecturers should continue adapting traditional methods to the Millennials' profile by incorporating, for instance, m-applications, but, such technologies should be used as a complement to traditional education, and not as a substitute to it.

Further research should be carried out in order to check if the results are the same in different contexts, such as other disciplines, and the effects on assessment in the long-term from the use of Socrative as SRS. It also could be interesting to use a more complex assessment design, in order to evaluate the students' skills.

References

Almond, R. G., Steinberg, L. S., & Mislevy, R. J. (2002). Enhancing the design and delivery of assessment systems: A four process architecture. *Journal of Technology, Learning, and Assessment*, *1*(5), 1–64.

Anderson, R. C., Graham, M., Kennedy, P., Nelson, N., Stoolmiller, M., Baker, S. K., &

Fien, H. (2019). Student agency at the crux: Mitigating disengagement in middle and high school. *Contemporary Educational Psychology*, *56*(January), 205–217. https://doi.org/10.1016/j.cedpsych.2018.12.005

Arain, A. A., Hussain, Z., Rizvi, W. H., & Vighio, M. S. (2018). An analysis of the influence of a mobile learning application on the learning outcomes of higher education students. *Universal Access in the Information Society*, *17*(2), 325–334. https://doi.org/10.1007/s10209-017-0551-y

Balta, N., Perera-Rodríguez, V. H., & Hervás-Gómez, C. (2017). Using socrative as an online homework platform to increase students' exam scores. *Education and Information Technologies*, 1–14. https://doi.org/10.1007/s10639-017-9638-6

Barnett, L. (2009). Key aspects of teaching and learning in economics. In S. K. and S. M. H. Fry (Ed.), *A Handbook for Teaching and Learning in Higher Education* (Routledge, pp. 405–423). New York.

Chui, L., Martin, K., & Pike, B. (2013). A quasi-experimental assessment of interactive student response systems on student confidence, effort, and course performance. *Journal of Accounting Education*, *31*(1), 17–30. https://doi.org/10.1016/j.jaccedu.2013.01.002

Combs, K. L., Gibson, S. K., Hays, J. M., Saly, J., & Wendt, J. T. (2008). Enhancing curriculum and delivery: Linking assessment to learning objectives. *Assessment and Evaluation in Higher Education*, *33*(1), 87–102. https://doi.org/10.1080/02602930601122985

Danesh, A., Baily, A., & Whisenand, T. (2015). Technology and Instructor-Interface Interaction in Distance Education. *International Journal of Business and Social Science*, *6*(2), 29–47.

Felder, M.R.; Brent, R. (2003). Designing and Teaching Courses for ABET.pdf. *Journal of Engineering Education*, *92*(1), 7–25.

Fotaris, P., Mastoras, T., Leinfellner, R., & Rosunally, Y. (2016). Climbing Up the Leaderboard : An Empirical Study of Applying Gamification Techniques to a Computer Programming Class. *Electronic Journal of E-Learning*, *14*(2), 94–110. https://doi.org/10.1007/s11528-016-0070-3

Golenhofen, N., Heindl, F., Grab-Kroll, C., Messerer, D. A. C., Böckers, T. M., & Böckers, A. (2019). The Use of a Mobile Learning Tool by Medical Students in Undergraduate Anatomy and its Effects on Assessment Outcomes. *Anatomical Sciences Education*. 1-11. https://doi.org/10.1002/ase.1878

Guarascio, A. J., Nemecek, B. D., & Zimmerman, D. E. (2017). Evaluation of students' perceptions of the Socrative application versus a traditional student response system and its impact on classroom engagement. *Currents in Pharmacy Teaching and Learning*, *9*(5), 808–812. https://doi.org/10.1016/j.cptl.2017.05.011

Haintz, C., & Ebner, M. (2014). Developing a Web-Based Question-Driven Audience. *Journal of Universal Computer Science*, *20*(1), 39–56.

Handley, M. A., Lyles, C. R., McCulloch, C., & Cattamanchi, A. (2018). Selecting and Improving Quasi-Experimental Designs in Effectiveness and Implementation Research. *Annual Review of Public Health*, *39*(1), 5–25. https://doi.org/10.1146/annurev-publhealth-040617-014128

Harmon, O. R., & Tomolonis, P. A. (2019). The effects of using Facebook as a discussion forum in an online principles of economics course: Results of a randomized controlled trial. *International Review of Economics Education*. 30, 100157. https://doi.org/10.1016/j.iree.2019.100157

Hillman, D. C. A., Willis, D. J., & Gunawardena, C. N. (1994). Learner-interface in distance education: An extension of contemporary modelas and startegies for practitioners. *The American Journal of Distance Education*, *8*(2), 30–42.

Jennings, J. M. ., & Angelo, T. (2006). Introduction. In *Student Engagement: Measuring and enhancing engagement with learning* (pp. 38–43). Retrieved from http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0104-40602016000100277&lng=pt&tlng=pt

Kokina, J., & Juras, P. E. (2017). Using Socrative to Enhance Instruction in an Accounting Classroom. *Journal of Emerging Technologies in Accounting*, *14*(1), 85–97. https://doi.org/10.2308/jeta-51700

Lim, W. N. (2017). Improving Student Engagement in Higher Education through Mobile-Based Interactive Teaching Model Using Socrative. In *Global Engineering Education Conference (EDUCON), 2017 IEEE* (pp. 404–412).

Lozano, R., Ceulemans, K., & Scarff Seatter, C. (2015). Teaching organisational change management for sustainability: Designing and delivering a course at the University of Leeds to better prepare future sustainability change agents. *Journal of Cleaner Production*, *106*, 205–215. https://doi.org/10.1016/j.jclepro.2014.03.031

Lucena, S. (2016). Culturas digitais e tecnologias móveis na educação. *Educar Em Revista*, (59), 277–290. https://doi.org/10.1590/0104-4060.43689

Manning, R. D., Keiper, M. C., & Jenny, S. E. (2017). Pedagogical Innovations for the Millennial Sport Management Student: Socrative and Twitter. *Sport Management Education Journal*, *11*(1), 45–54. https://doi.org/10.1123/smej.2016-0014

Mattheos, N., & Lic, O. (2004). *Information Technology and Interaction in Learning Studies of Applications in Academic Oral Health Education*. *PhD. diss.* University of Mälmo.

Montenery, S. M., Walker, M., Sorensen, E., Thompson, R., Kirklin, D., White, R., & Ross, C. (2013). Millennial Generation Student Nurses' Perceptions of the Impact of Multiple Technologies on Learning. *Nursing Education Perspectives*, *34*(6), 405–409. https://doi.org/10.5480/10-451

Moore, M. G. (1989). Editorial: Three Types of Interaction. *American Journal of Distance Education*, *3*(2), 1–7. https://doi.org/10.1080/08923648909526659

Norton, L. (2009). Assessing Student Learning. In *A Handbook for Teaching and Learning in Higher Education* (Vol. 1, pp. 132–149). https://doi.org/10.1187/cbe.02-03-0007

Onodipe, G. O. (2017). Engaging and Empowering Dual Enrollment Students: A Principles of Economics Course Example. *Empowering Learners With Mobile Open-Access Learning Initiatives*, 1–6. https://doi.org/10.4018/978-1-5225-2584-4.ch058

Picault, J. (2019). The economics instructor's toolbox. *International Review of Economics Education*, *30*(September 2018), 100154. https://doi.org/10.1016/j.iree.2019.01.001

Pomeranz, D. (2011). Métodos de Evaluación. *Harvard Business School*, 1–12.

Prøitz, T. S. (2010). Learning outcomes: What are they? Who defines them? When and where are they defined? *Educational Assessment, Evaluation and Accountability*, *22*(2), 119–137. https://doi.org/10.1007/s11092-010-9097-8

Riaza, B. G., & Rodríguez, A. I. (2016). Students' Perception of the Integration of Mobile Devices as Learning Tools in Pre-Primary and Primary Teacher Training Degrees. *International Journal of Human Capital and Information Technology Professionals*, *7*(2), 19–35. https://doi.org/10.4018/IJHCITP.2016040102

Robert J. Mislevy, Linda S. Steinberg, & Russell G. Almond. (2003). On the Structure of Educational Assessments. *Measurement: Interdisciplinary Research and Perspectives*, *1*(1), 3–62. https://doi.org/10.1207/S15366359MEA0101

Savec, V. F., Hrast, Š., Šuligoj, V., & Avsec, S. (2018). The innovative use of ICT in STEM teacher training programmes at the University of Ljubljana. *World Transactions on Engineering and Technology Education*, *16*(4), 421–427.

Sefo, K., Granados Romero, J. M., Lázaro, M.-N., & Fernández-Larragueta, S. (2017). La formación del profesorado para un uso innovador de las tic: Un estudio de caso en la educación obligatoria en la provincia de Almería. *Profesorado, Revista de Currículum y Formación Del Profesorado*, *21*(4), 241–258.

Sinha, N., Khreisat, L., & Sharma, K. (2009). Learner-Interface Interaction for Technology-Enhanced Active Learning. *Innovate: Journal of Online Education*, *5*(3), Article 3.

Socrative. (2019). Socrative. Retrieved from www.socrative.com

Streveler, R. a., Smith, K. a., & Pilotte, M. (2012). Aligning Course Content, Assessment, and Delivery: Creating a Context for Outcome-Based Education. *Outcome-Based Science, Technology, Engineering, and Mathematics Education: Innovative Practices*, 1–26. https://doi.org/10.4018/978-1-4666-1809-ch001

University of A Coruna. (2018). Degree in law. Retrieved from http://estudos.udc.es/en/study/start/612G01V01/2018

Varela-Candamio, L., Novo-Corti, I., & Barreiro-Gen, M. (2014). Do studies level and age matter in learning and social relationship in the assessment of web 3.0? A case study for “digital natives” in Spain. *Computers in Human Behavior*, *30*, 595–605. https://doi.org/10.1016/j.chb.2013.07.048

Vrasidas, C. (2000). Constructivism versus objectivism: implications for interaction, course design, and evaluation in distance education. *International Journal of Educational Telecommunications*, *6*(4), 339–362. https://doi.org/10.1080/0031383890330103

White, H., & Sabarwal, S. (2014). Diseño y métodos cuasiexperimentales. *Sinopsis Metodológicas. Sinopsis de La Evaluación de Impacto N°. 8*, 1–16. Retrieved from https://www.unicef-irc.org/publications/pdf/MB8ES.pdf

Wuthisatian, R., & Thanetsunthorn, N. (2019). Teaching macroeconomics with data: Materials for enhancing students’ quantitative skills. *International Review of Economics Education*. 30, 100151. https://doi.org/10.1016/j.iree.2018.11.001

Zou, D., & Lambert, J. (2017). Feedback methods for student voice in the digital age. *British Journal of Educational Technology*, *48*(5), 1081–1091.

https://doi.org/10.1111/bjet.12522