

DESCRIPTION OF A SESSION TO ASSESS THE EFFECTIVENESS OF APPS IN MEDICAL EDUCATION

Laura Briz-Ponce, Juan Antonio Juanes-Méndez, Francisco José García-Peñalvo

University of Salamanca (SPAIN)

Abstract

Nowadays, there are few researches about the effectiveness of apps in education. With this paper, we want to describe the methodology and the initial results obtained with the first session performed with undergraduate students of Medical Schools in University of Salamanca.

The session lasted roughly two hours and was formed by ten students. We used a descriptive correlational method and the instruments utilized to obtain the qualitative data were a questionnaire and the images recorded and observed during the session. Besides it was necessary to divide the participants in two groups in order to contrast the results between them: control group and experimental group. One of them had to attend a normal class and the other one was going to use an app, which describes the anatomical parts of a brain. The results were measured with a knowledge test performed before and after the session. Both groups obtained better results after the experiment, however with the participants of the experimental group, the results were higher than the students of the control group.

It is very complex to measure the effectiveness of mobile apps in education, and although the study suggests that the mobile apps could be a tool to use in higher education, it is necessary to repeat the sessions with different students in order to have more samples, which allow showing and verifying the effectiveness of apps in education.

Keywords: Higher Education, mobile apps, mhealth, mobile devices, qualitative research.

1 INTRODUCTION

The use of new technologies in Higher education has becoming a tendency growing in different institutions as Medical Schools in United States [1]. However, not all Schools have adopted the cutting-edge technologies so easily. The reasons could be basically three:

- An unawareness of the potential they could achieve in learning. Besides, there are a lot of apps, so it is very complex to know the best one to fulfil the requirements you are looking for [2].
- The effectiveness of these tools in learning has still to be proved.
- Barrier or a clear tendency against change from faculty members, leaderships and sometimes by the students.

These reasons were then main reason of designing this research. The outreach of this study wants to measure the effectiveness of the mobile apps in learning and teaching context.

In order to establish this objective, it was necessary to design the methodology to conduct the research and the instruments and the hypothesis to cover it.

When we analyse the different aspects of the research, it is very relevant to take into account what it is the app selected for the study. We had different options, all of them developed within the "Image and Technology of the Biomedical Knowledge Centre" (CITEC-B), in Madrid, all together with the "Recognized Investigation Group" of the University of Salamanca, and Advanced Visualization Medical Systems (VisualMED System). Finally, we decided that the application had to fulfil the four following requirements:

- It should be available in App Store for iPhone and for iPad.
- It should be an app that had been used by students in the university school of Medicine in Salamanca.
- It should be suitable for self-study.
- It should be free or at least accessible for a base fee charge (two euros or less).

We sought the mobile applications that met these criteria, and at the end, we selected the Brain System 3D application. This app allows the students to learn about the structure and function of the human brain by interacting with high-resolution rotation 3D images in real time [2].

This paper is structured in four parts: first an overview of the methodology used with this experimental session and all the resources used. Then, we describe the process of the session conducted in University of Salamanca. Third, we will show the main results of the investigation and we finally draw the conclusion of the research.

2 METHODOLOGY

The methodology used for this section is an experimental correlational method with a pre-test and post-test design. According to the reference [3], this type of method establishes the relationship between different observed variables. Besides, the definition for experimental method is: "The experimental research is a process that consists of one object or group to be subjected to specific conditions (independent variable) in order to observe the different effects produced (dependent variable)". In this case, we control the independent factor in order to measure the dependent variables. This type of method requires two different groups: Experimental Group (G_e), which is treated with the experiment, and the control group (G_c), which is not treated with the experiment and it is used to compare the results.

In our study, the students that participated in the session were determined by the convenience or availability sample [4] but the groups were formed randomly, so we consider that it is an experimental method.

One of the advantages of using an experimental method is precisely that the researcher is able to control external factors [5], which he considers relevant in order to determine the causality.

Besides, we obtain the pre-test or initial measure and the post-test or final measure, for both types of groups.

According the variables, the independent variable is the variable controlled by the researcher, and the dependent or criterion variables are those measured in the experiment [3]. In our case, the independent variable is the profile of the participant (if it is an student, for example) and the dependent variables are the scores assigned by the participants for each block that forms the quality protocol and the score they obtained in the knowledge test.

In our case, the quality protocol is divided in five blocks: Content quality, navigation, Credibility, Design or User Interface and security and privacy.

The principal instruments used to collect data were the following:

- Video and photos. The output data allows observing the different interactions of participants when performing the tasks. They signed previously consent information of being recorded during the session.
- Knowledge questionnaire. It consisted on a questionnaire to measure the level of knowledge for a material. In this case, anatomy class. This questionnaire was filled by both groups (G_e and G_c)
- A task items. We selected different tasks to perform with the mobile app. The G_e must complete a form to indicate the tasks they were able to perform or not with the app.
- Validation protocol questionnaire. After performing the task paper, the G_e must complete as well a questionnaire based on a 5-likert point scale assessing the different blocks considered for the quality protocol: Content Quality, Navigation, Credibility, Design or User Interface and security and privacy. These dimensions were selected based on the collected data of a previous phase of the same research [2] [6].

3 PROCESS

As it is mentioned earlier, the session of the experimental group was divided in two separated groups. On one hand, we have the G_e and on the other hand, we have the G_c .

The session started first with a brief introduction of the session, we explained the participants that we will divide the group in two and each of them would follow different activities. Then, we provided the participants with all these information written and a knowledge test. In this test we collect also gender,

age and profile in order to be sure if these factors could affect the outcomes as it is suggested in [5]. All the data is collected anonymously so we request the participants to give their three last numbers of ID in order to identify the answers and match them afterwards with the rests of the tasks of the session.

After that, all of them filled the same knowledge test. Then, we bring the G_c to another class. In the meantime, a professor gave a brief overview to G_e about the use of the app that we selected for the session.

When he finished the speech, the G_e had to continue with the session performing some additional tasks and in the meantime, this same teacher gave a normal anatomy class to the G_c. We provide a list of tasks to the G_e in order to perform all of them. They have to fill the report pointing out if they have completed the task without problems or they had found some troubles to do it. Then, the G_e had to complete a questionnaire with a 5-point Likert scale from “strongly disagree” to “strongly agree”. The answers were transformed into a numeric form with a range of 1-5. So, “strongly disagree” category was mapped to the number 1 and “strongly agree” mapped to the value of 5. The survey consisted in questions to measure the quality of the app. The dimensions considered in the survey, as we mentioned earlier were: content quality, navigation, credibility, design, user interface, security and privacy. The reason to select these one were mainly because we conducted a survey with 124 participants of undergraduate students and medical professionals [2] [6] to ask the most important factors they consider when they use an app. The most picked factors were the one considered for this study.

Once we finish with these actions in G_e and the normal class was given to the G_c by same professor in order to minimise the external factors as it is one of the objectives that the experimental researchers must try to achieve [5], we distributed the same knowledge test to both groups in order to compare results.

The session lasted about two hours and was definitely very productive. The students were very participative. The participants of G_c did not have the chance to use the app. However, after the session was finished, they even want to test it and to know how it worked. This means that undergraduate students were very keen on cutting edge technologies and are very open-minded to use new gadgets, even in learning.

4 RESULTS

The outcome data was analysed using the SPSS v21 program and the main results of the study could be outlined in three different points:

- First, the study reflects the knowledge test results before and after the session for each group.
- Second, the study provides a comparative result between the G_e and the G_c.
- Third, the study describes the evaluation of the app assigning a score to the evaluable different parts.

4.1 Knowledge Test

The table 1 provides an overview of the descriptive statistics of the score of the knowledge test: mean, standard error, Median, typical deviation and variance.

Table 1. Descriptive statistics of the samples and the results of the Wilcoxon Test.

		Descriptive Statistics					Wilcoxon Test	
		\bar{X}	SE	Md	SD	σ^2	Z	p
G _e	Pre-test	1,8000	0,58310	2,000	1,700	1,700	-2,032	0,042
	Post-test	4,6000	0,24495	5,000	0,300	0,300		
G _c	Pre-test	2,5000	0,95743	3,000	3,667	3,667	-1342	0,180
	Post-test	3,2500	0,75000	4,000	2,250	2,250		

Before selecting the suitable method to analyse the data, a Shapiro Will Test [7] was used in order to examine if the variables are normal or not. We do not use Kolmogorov-Smirnov Test as the number of samples is <50 [8]. The results suggested that the variables are normal ($p > 0,05$). However, the number of samples is so small, that we consider that non-parametric tests were more appropriate. Besides, the variables are paired so it is necessary to use the Wilcoxon test to determine if there was any significant difference in the knowledge score variable between the pre and post test scores for each group. For this test, the hypotheses are the following:

The null hypothesis H_0 : There is no difference in scores before and after the treatment

Alternative Hypothesis H_a : There is a significance difference before and after the treatment.

After performing the results, at the $\alpha = 0,05$ level of significance, there exists enough evidence to conclude that we reject the null hypothesis that considers no difference between pre-test and post-test score variables for G_e ($p = 0,042$, $p < 0,05$). On the contrary, we have that at the $\alpha = 0,05$ level of significance, there is no evidence enough to reject the null hypothesis ($p = 0,180$, $p > 0,05$), so there is no statistically significance between the score variable before and after the test for G_c .

In summary, although in both groups, there was an increase of the results before and after performing the session as it is shown in Fig. 1, we can suggest that there was a statistically significant increase from the pre-test scores to the post-test scores in G_e , but not so significant in G_c .

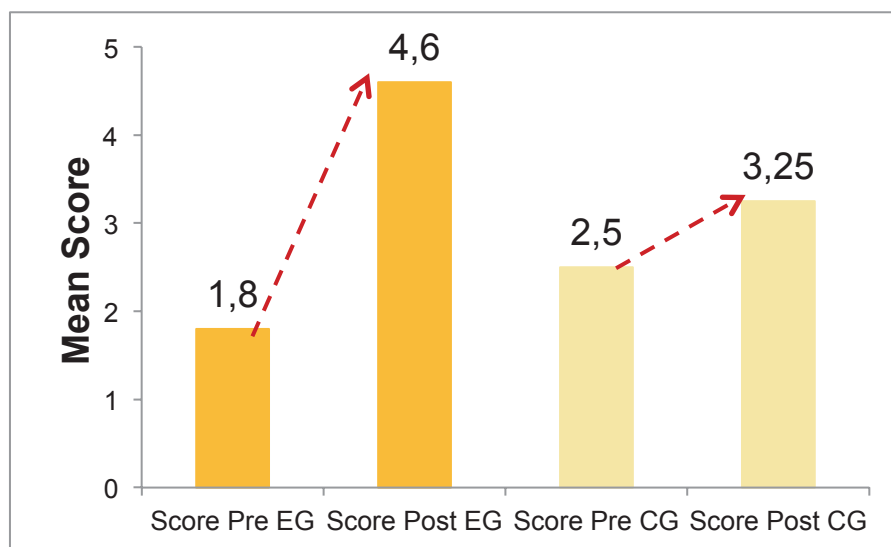


Fig. 1. Mean Score for Experimental Group and for Control Group.

4.2 Comparative Results

In order to examine the differences in the score for both groups, a Mann-Whitney U test was used [9]. This test is the most popular test used to determine the differences between two groups with no specific distribution.

As a result, the value $U = 7,000$ and $\rho = 0,455 > 0,05$ mean that no meaningful difference between the G_e and G_c was seen in the pre-test scores of the knowledge test.

There are some researches [10] [11] that consider that the comparison between the G_e and G_c is not relevant as there are many external factors that could influence in the differences obtained. For example, in [10], they claims “findings of studies that have used a non-equivalent control group design with pre-test and post-test should be interpreted cautiously since control and treatment group may differ due to selection bias”. In [11], the authors explain “researchers should not interpret study results by simply studying differences in post-test scores of the control and treatment groups since such differences could be attributed to differences in participants’ characteristics and/or differences in participants’ experiences during the experiment

It seems that external factors may influence in the differences for both groups, because of that, we has conducted the Mann-Whitney U test considering different characteristics of the sample in order to evaluate if other factors could influence in the final results. As most part of the variables are nominal, it

is necessary to use Chi Square Test and the Phi and Cramer's V value as well [12] [13]. When it is appropriated (2x2 table), we include the Fisher correction. The outcome results are presented in Table 2.

Table 2. Output data of Chi Square Test and Mann Whitney U Test.

	Chi Square Test (Nominal variables)				Mann-Whitney U Test	
	Chi Square	Fisher's exact test	Phi (V Cramer)	ρ	U	ρ
Knowledge Score Pre test					7,000	0,455
Age					9,000	0,777
Gender	0,9	0,405	0,316 (0,316)	0,343		
Device	0,09	0,643	-0,1 (+0,1)	0,704		
Operating System	2,07		0,478 (0,478)	0,358		
Use navigation	0,032	0,722	0,060 (0,060)	0,858		
Use for work	1,406	0,444	-0,395 (0,395)	0,236		
Use for studies	2,057	0,278	-0,478 (+0,478)	0,151		
Use for looking information	0,225	0,595	-0,158 (+0,158)	0,635		
Use for email	1,406	0,444	0,395 (0,395)	0,236		
Other uses	1,406	0,444	-0,395 (+0,395)	0,236		
Use of apps for learning	0,9	0,405	0,316 (0,316)	0,343		

In view of table 2, no meaningful differences ($p > 0,005$) are seen between the different variables and the experimental and control group before the session.

4.3 Evaluation of the app

The results drawn of this study could suggest that the mobile apps are a powerful tool for learning. The experiment also includes a 5-likert point survey with 45 items to evaluate the main dimensions considered relevant for undergraduate students.

Previous to this experiment, we conducted the first phase of the research to collect data from 124 participants among students and medical professionals. We obtained that the most important factors that they took into account to download apps were the following: Content Quality, Navigation, Credibility, Design or User Interface and security and privacy [2] [6].

In order to evaluate the reliability of this experiment we calculate the α Cronbach that it is shown in Table 3

Table 3. Number of items assigned to different dimensions considered for the survey and the value of α Cronbach.

Dimensión	Ítems	N	α Cronbach
Reliability	1,2,3,37,38,39,41	7	0,741
Navigation	4,5,6,7,8,9,10,11,12,13,14,15	12	0,877
Design/User Interface	16,17,18,19,20,21,22,23,24,25,26,27,28,29,45	14	0,862
Security/Privacy	30,31,32,33	4	0,844
Content Quality	34,35,36,40,42,43,44	7	0,686

The global value of α Cronbach that form the construct is 0,911. The article [14] establishes some rules to categorize the value of α . If $\alpha > 0,9$ the reliability is excellent, if $\alpha > 0,80$, the reliability could be considered as good. If $\alpha > 0,7$, the reliability is acceptable. If $\alpha > 0,60$ it is questionable. If $\alpha > 0,50$, it is poor and $\alpha < 0,5$ it is not acceptable.

As we can see in the table 3, the dimensions have a good reliability. Only the reliability and the content quality dimension have the least value, obtaining an acceptable and a questionable value respectively. However, the global value is 0,911, which it means an excellent valuation of the reliability

The table 4 shows the different scores obtained for the participant, the sum for all the participants, and the descriptive statistics as mean, standard deviation and the variance. Also, it includes, the maximum score that each participant could assign to each dimension in order to compare with the mean value. The dimension that obtained the least value is the security/privacy, which it means that this dimension should be improved by the developers.

Table 4. Descriptive statistics of the dimensions.

Dimensión	Sum	Max score/person	Mean	Standard Deviation	Variance
Reliability	26	7	5,2000	1,30384	1,700
Navigation	53	12	10,6000	2,07364	4,300
Design/User Interface	54	14	13,5000	1,29099	1,667
Security/Privacy	4	4	0,8000	1,09545	1,200
Content	32	7	6,4000	0,89443	0,800

5 CONCLUSION

The results drawn of this study could suggest that the mobile apps are a powerful tool for learning. The experiment could be a good starting to emphasize the importance of new technologies and how these could be used in different educational contexts making the learning a productive and useful goal. In fact, comparing results, it seems that using mobile devices and the app designed to analyse different parts of the brain is a tool that stir curiosity among students and make them interact with this tool, even to find all the possibilities, not only the tasks requested in this study. It is important to notice that the participants of G_c wanted to know more of the app although they were not in the G_e . It seems that it is an innate curiosity of the new generations as they want to be more and more prepared for the new digital environment where they have to live in and they will work on.

However, there are some limitations of the study as the sample size is very small to have more precise conclusion, so it is necessary to collect more data of these sessions and another limitation is that the study sample is limited to students of University of Salamanca.

ACKNOWLEDGEMENTS

This research work is made within University of Salamanca PhD Programme on Education in the Knowledge Society Scope.

REFERENCES

- [1] Briz, L., Juanes, J. A., & García, F. J. (2014). Analysis of mobile devices as a support tool for professional medical education in the university school. En *EDULEARN14 Proceedings* (pp. 4653-4658). Valencia: IATED Academy.
- [2] Briz Ponce, L., Juanes Méndez, J.A., & García-Peñalvo, F. J. (2014a). Analysis of certificated mobile application for medical Education purposes. In *Proceedings of the Second International Conference on Technological Ecosystem for Enhancing Multiculturality - TEEM'14* (pp. 13-17). ACM New York. doi: <http://dx.doi.org/10.1145/2669711.2669871>
- [3] Arias, F. G. (2006). El Proyecto de Investigación. Introducción a la Metodología Científica. 5ta. Edición. Caracas, Venezuela: Editorial Episteme C.A.
- [4] Lavrakas, P. (2008). *Encyclopedia of Survey Research Methods*. 2455 Teller Road, Thousand Oaks California 91320 United States: SAGE Publications, Inc. Retrieved from <http://srmo.sagepub.com/view/encyclopedia-of-survey-research-methods/n105.xml>
- [5] Muijs, D. (2010). *Doing Quantitative Research in Education with SPSS*. SAGE.
- [6] Briz-Ponce, L., Juanes-Méndez, J. A., & García-Peñalvo, F. J. (2014). First Approach of Mobile Applications Study for Medical Education purposes. In *Proceedings of the Second International Conference on Technological Ecosystem for Enhancing Multiculturality - TEEM'14* (pp. 647-651). doi: <http://dx.doi.org/10.1145/2669711.2669968>

- [7] Chakravarti, Laha, and Roy, (1967). Handbook of Methods of Applied Statistics, Volume I, John Wiley and Sons, pp. 392-394.
- [8] Vinacua, B. V. (2007). *Análisis estadístico con SPSS 14: estadística básica*. McGraw-Hill Interamericana de España S.L.
- [9] Mann, H. B., & D. R. Whitney. (1947). On a Test of whether one of two random variables is stochastically larger than the other. *The Annals of mathematical statistics*, 50-6
- [10] Best, J. & Kahn, J. (2006). *Research in education*. Boston: Pearson Education, Inc
- [11] Quasi-Experimental Design in Education. (s. f.).
- [12] Chakravarti, Laha, and Roy, (1967). Handbook of Methods of Applied Statistics, Volume I, John Wiley and Sons, pp. 392-394.
- [13] Alan Agresti. (2002). *Categorical Data Analysis* (Second Edition.). New york: John Wiley & Sons.
- [14] George, D., & Mallery, P. (2003). *SPSS for Windows step by step: a simple guide and reference*, 11.0 update. Allyn and Bacon.