The impact of serious games in mathematics fluency: A study in Primary Education

Impacto de los juegos serios en la fluidez matemática: Un estudio en Educación Primaria

Dr. Fernando Fraga-Varela
Lecturer, Department of Pedagogy and Didactics, University of Santiago de Compostela (Spain)

Dr. Esther Vila-Couñago
Assistant Professor, Department of Pedagogy and Didactics, University of Santiago de Compostela (Spain)

Dr. Esther Martínez-Piñeiro
Professor, Department of Pedagogy and Didactics, University of Santiago de Compostela (Spain)

Abstract

In recent years there has been a renewal of educational software encouraged by the incorporation of specific designs based on serious games. Previous studies on their use do not provide conclusive data on the advancement in learning, both at a general level and in specific contents. The main objective of this work is to study the impact of the use of serious games in primary education classrooms, specifically on mathematics fluency, taking into account gamification variables and teaching experience. A quasi-experimental study was carried out with a pretest-posttest design, without a control group and with several experimental groups, involving 284 students from the first to the fourth grade. The results show a significant improvement in mathematics fluency with the use of serious games in the different grades and classroom groups studied. The gamification strategy promotes even greater progress over the classes where it has not been implemented. There is a similar time of use of serious games by both novice and experienced teachers, with better results in mathematics fluency in the case of the second group. It also shows the relationship between the results obtained and the school grades of the students. The findings point to the potential of using serious games designed specifically for school environments and challenge previous work on generational barriers in teachers.

Resumen

En estos últimos años ha habido una renovación del software educativo propiciada por la incorporación de diseños específicos basados en juegos serios. Los estudios previos sobre su uso no ofrecen datos concluyentes sobre el avance en el aprendizaje, tanto a nivel general como de contenidos específicos. El objetivo principal del presente trabajo es conocer el impacto del uso de juegos serios en las aulas de educación primaria, concretamente en la fluidez matemática del alumnado, atendiendo a variables de gamificación y experiencia docente. Se lleva a cabo un estudio cuasi-experimental con pretest-postest, sin grupo control y con varios grupos experimentales, en el que participan 284 estudiantes de primero a cuarto curso. Los resultados muestran una mejora significativa de la fluidez matemática con el uso de juegos serios en los distintos cursos y grupos-aula estudiados. La estrategia de gamificación promueve un progreso aún mayor respecto a las aulas en las que no se ha implementado. Se observa un tiempo de uso similar de los juegos serios por parte de profesores noveles y experimentados, con mejores resultados en fluidez matemática en el caso de los segundos. También se muestra la relación existente entre los resultados obtenidos y las calificaciones del alumnado. Las conclusiones señalan el potencial del uso de juegos serios diseñados específicamente para entornos escolares y cuestionan trabajos previos sobre las barreras generacionales en el profesorado.

Keywords / Palabras clave

1. Introduction and state of the art

In the past years, there has been a renewal of educational software thanks to the incorporation of specific designs based on serious games. "Serious game" refers to a game in which "education (in its various forms) is the primary goal, rather than entertainment" (Michael & Chen, 2006: 17), with the focus placed on specific contents, regardless of the form and the structure that are being used (Zagalo, 2010). These initiatives build on one of the possibilities of serious games, but we should not lose sight of the richness of all the modalities of serious games, based on their ludic structure, purpose or scope (Alvarez & Djaouti, 2012; Romero-Rodríguez & Torres-Toukoumidis, 2018). The design of experiences from this perspective, which has multiple parallels with the field of videogames, provides teachers in educational centers with a new way of working. In addition to these developments, we should mention the emergence of initiatives that include several gamification-based elements, placing the focus on the domain of motivation (Pérez-Manzano & Almela-Baeza, 2018). Although their implementation usually has positive effects, it is greatly dependent on the contexts of application (Hamari et al., 2014).

Some studies related to the use of serious games in the classroom reveal improvements in learning (Clark et al., 2016; Wouters et al., 2013), or identify progress in students' cognitive capacity (Lamb et al., 2018). This situation informs the school curriculum (Carvalho et al., 2018). However, there are also certain studies that, while reporting benefits in the level of student participation and involvement, do not provide conclusive data on learning progress in general (Chauhan, 2017; Fisher et al., 2020), or in relation to specific contents (Boendermaker et al., 2017; Mellado et al., 2018). Additionally, although teachers do not rule out the possibility of using serious games, there are only very sporadic experiences (Del-Moral & Fernández, 2015). Kaufman's work (2013) refers to the existence of a generational barrier (i.e., teacher not older than 35 years), as teachers above this age lack previous positive experiences with this type of software (Marín-Díaz et al., 2019). This situation reveals a strong relationship between teachers' attitude toward this type of software and its potential use in the classroom (Stieler-Hunt & Jones, 2015).

What all available studies suggest is how beneficial serious games are to students' attitude, with a remarkable impact on the domain of motivation (Filella et al., 2017; Gómez-García et al., 2016). This factor should be borne in mind when considering the use of these resources in the classroom, especially in areas of the curriculum deemed critical, such as languages and mathematics. Specifically, calculus and its automation are still some of the main challenges to be tackled at these levels of education (Baroody et al., 2009). It is not just one more element of content: continued improvement in these learnings throughout compulsory education serves as a guarantee that spreads to schoolwork in general (Duncan et al., 2007) and may become a good predictor of learning at higher levels (Geary, 2011). Therefore, this space is potentially useful for the implementation of serious games, which might facilitate learning that requires some sort of automation. There are currently several pieces of software that have been designed as serious games specifically for these contents and which, in some cases, also enable the use of gamification as a potential complement, as this is, precisely, a key strategy in the domain of motivation (McGonigal, 2011). Gamification may take place in parallel, reinforcing motivation in a non-ludic context (Teixes, 2015). It provides help in the progress toward a final goal, thereby increasing students' interest considerably (Zagal & Altizer, 2014).

At present, it is possible to access online serious games designed to be used in educational centers around the world, breaking the traditional barriers of publishing markets and opening the door to experimentation with initiatives available to other countries. The use of these products in the Spanish context may be challenging, since in some cases there are differences with the curricula of the country of origin, as happens with the learning of calculus. It is common to present addition and subtraction at the same time, in grades 1 and 2, and multiplication together with division in grade 3, as contained in the proposal of the U.S. Council of Chief State School Officers (2020) under what has been called "Common Core State Standards", compared with the much more linear guidelines of the curriculum applied in Spain.

Precisely, the contents relating to calculus have been a recurring focus of work in the field of serious games in the past years, with a direct impact on the field of mathematics fluency (Baroody et al., 2013). The latter is understood as a sign of skillfulness in algorithm solving. It is based on principles such as efficiency, precision, the use of strategies and flexibility (Kling & Bay-Williams, 2014). The development of fluency maximizes efficiency and precision, and making progress in fluency is a protective factor against failure in the areas of mathematics or reading (Meiri et al., 2019). Evidence of these processes can be identified as an indicator of performance according to the specific learning proposals.

Recent studies have explored the benefits that this type of software bring to mathematics fluency, and such benefits have become the reference to assess these products (Van-der-Ven et al., 2017). An example of this

is Reflex Math, with a substantial level of use in American classrooms at different educational levels (Cozad, 2019; Cress, 2019; Sarrell, 2014). The software generates a personalized learning route based on response times and error rates. It also provides help to students through a virtual trainer. The figure of the virtual trainer provides some guidance to develop strategies for overcoming any difficulties detected in the individual learning routes, adapted to the specific situation of the student. Errors are not seen as penalties; rather, they become opportunities for improvement. In this sense, errors are analyzed by the system to activate the support processes by the virtual instructor. In addition, the software has a built-in gamification system that rewards students by improving their avatars and awarding them diplomas, but never giving them academic benefits, through redeemable points based on the number of activities carried out, the frequency of use of the software or the correct completion of the tasks. This type of rewards that are external to the use of the software and the promotion of these rewards are a supplemental task that should be managed by teachers. Students carry out their work independently. Teachers can monitor the process through specific tools that provide information based on the data collected by the system. However, we should not lose sight of the fact that the use of software with these characteristics reveals, again, certain contradictions. A number of studies focused on mathematics have demonstrated benefit in learning (Fernández-Robles et al., 2019; Pires et al., 2019), while other studies suggest the contrary (Hieftje et al., 2017).

In this scenario, we designed a research study with the general purpose of understanding the impact of the use of serious games on mathematics fluency among primary education students. Our research, conducted in real classroom contexts, was based on several variables of interest, such as the grade level of the students, the application of gamification, the classroom group and teaching experience. We also intended to understand the relationship between the results derived from this learning initiative and the grades obtained by the students. The findings from this study may be useful to governmental agencies and educational centers in decision-making processes related to the provision of resources and the development of innovative strategies in the classroom-based methodologies used.

2. Materials and methods
2.1. Design

In response to the objective described, a pretest-posttest quasi-experimental design without a control group and with several experimental groups was used. Totally standardized conditions were sought, i.e., a common school scenario with classroom groups already formed and without any sort of randomness. Each group worked independently between the two tests that were applied, using the educational software Reflex Math. Those in charge of the implementation were, in each case, teachers in the area of mathematics, who were all given the same indications to use all the possibilities provided by the system. The study was carried out after obtaining informed consent from the school where it was conducted.

2.2. Participants

The research was carried out in a school that offered three lines for each level, with the participation of 12 primary classes between levels 1 and 4. This was a private school financed by public funds through an agreement with the State and located in an urban area of Galicia. The study sample was constituted by 284 students. A total of 54.2% were boys and 45.8% were girls. A total of 24.3% of them were in the first grade, 25.4% were in the second grade, 25.7% were in the third grade and 24.6% were in the fourth grade. Our proposal was previously submitted to the management team and the 25 members of the staff teaching at the primary education stage. They were given information explaining the characteristics of the program and the research proposal, as well as the implications of the use of the software for their teaching. Lastly, a positive evaluation was obtained for their participation in the research, particularly from the teachers responsible for the area of mathematics, who were expected to be more deeply involved in the process.

2.3. Tool

The Basic Math Operations Task (BMOT), developed by Foegen and Deno (2001), was used to assess the learning of calculus. Access to this tool was gained through Sarrell's work (2014). Its use required a full translation into Spanish of both the pretest and the posttest. In both cases, combined addition, subtraction, multiplication and division calculation operations were presented. This proposal perfectly suits the educational

level envisaged at this school in the third and the fourth grades. For the first and the second grades, our proposal was adapted to the curricular level and included only addition and subtraction operations. The correction of the test generates an indicator of individual performance based on the summation of all correct answers in a maximum of one minute. This indicator is one of the potential benefits that can be directly derived from the use of the software, hence the advisability of using it.

2.4. Procedure

The research was conducted in the first term of the year 2019-2020. The pretest was applied in September 2019, and the posttest was applied in December that year. The use of the software was integrated into the duration of the mathematics class; following the developer's guidance, it was used in three sessions per week. Also, student data were collected in relation to academic performance, specifically the students' grades in all the areas evaluated at this stage at the end of the term, at the time the posttest was applied. As to the teachers, data on the number of years of teaching experience were collected. By accessing the system's database using a tool intended specifically for teachers, it was possible to compile individual data on the number of days of use of the program, the volume of activities completed and the use of gamification strategies.

2.5. Data analysis

The statistical software SPSS v. 25 was used for data analysis. Univariate descriptive analyses based on measures of central tendency and dispersion were carried out. Because parametric assumptions were not met for related situations (comparison between the pretest and the posttest), the Wilcoxon signed-rank test was applied, for which a level of significance of 0.05 was set. For independent situations, the Mann–Whitney U test was used to find out whether there were significant differences ($p<0.05$) between the classes in which gamification had been applied and those that had not applied it. It was also used to find differences between the classes with novice teachers and those with experienced teachers. Also, the statistic $r=|z| / \sqrt{N}$ was calculated as measure of effect size (Field, 2018; Fritz et al., 2012). For the interpretation of r, we followed the criterion proposed by Cohen (1988), with the extension suggested by Rosenthal (1996). Spearman's ρ coefficient (r_s) was used to establish the degree of association between the different variables analyzed. Statistical significance ($p<0.05$) and relationship strength were jointly considered for its interpretation, for which the indications given in the specialized literature (including, among others, Sánchez-Huete, 2013) were followed. Lastly, the coefficient of determination (R^2) was calculated to have an approximation to the amount of variance of academic performance as explained by the results of the learning method based on serious games.

3. Analyses and results

First, we will present an overview of the effect of serious games on mathematics fluency to later analyze it according to the grade leven were the games were used. Additionally, the results of the proposal will be analyzed according to the application or non-application of gamification, and also with regard to each specific class, considering the distinction between novice teachers and experienced teachers in those classes with gamification conditions. We will finish with a study of the relationships between the results achieved through this learning proposal and the grades obtained by the students.

3.1. Serious games and improvement in mathematics fluency

The software was used for a mean of 27 days, over which the students carried out 5,747 activities on average. After using the software, a statistically significant improvement of great magnitude was seen in mathematics fluency ($n=284$, $Z=-14.291$, $p=0.000$, $r=0.60$), which went from an average score of 8.99 (pretest) to 17.79 (posttest). The time of use of the software was clearly related to the number of activities solved ($r_s=0.82$, $p=0.000$).
When we look at each educational level, it can be seen that the posttest score comes hand in hand with two aspects that turn out to be revealing, with a generally significant relationship that tends to be moderate in the case of time of use (grade 1, $r_s=0.20$, $p=0.108$; grade 2, $r_s=0.36$, $p=0.002$; grade 3, $r_s=0.39$, $p=0.001$; grade 4,

r_s=0.40, p=0.001) and moderate with regard to the number of activities solved (grade 1, r_s=0.37, p=0.002; grade 2, r_s=0.65, p=0.000; grade 3, r_s=0.48, p=0.000; grade 4, r_s=0.48, p=0.000).

Table 1 contains the main descriptive statistics for each level. Significant differences are found between the scores obtained in the pretest and those attained in the posttest, with a difference of great magnitude in all cases: in grade 1 (Z=-7.225, p=0.000, r=0.62), in grade 2 (Z=-7.378, p = 0.000, r=0.61), in grade 3 (Z=6.354, p=0.000, r=0.53) and in grade 4 (Z=-7.251, p=0.000, r=0.61).

Table 1. Descriptive statistics by grade						
	First grade (*n*=69)			Second grade (*n*=72)		
	M	Md	SD	M	Md	SD
Pretest score	2.97	2	3.16	8.64	8	3.651
Posttest score	12.80	13	5.43	20.99	20.5	8.139
Days of use	28.3	27	4.9	29.6	28.5	6.1
Activities solved	5,662.4	4,775	2,829.6	6,775.5	5,733.5	3,918.5
	Third grade (*n*=73)			Fourth grade (*n*=70)		
	M	Md	SD	M	Md	SD
Pretest scores	10.64	10	4.66	13.57	13	6.04
Posttest scores	15.45	15	5.71	21.84	21	6.52
Days of use	29.4	29	8.7	21.3	20.5	5.2
Activities solved	6,620.1	6,344	3,315.3	3,865.4	3,445	1,890.6

As Figure 1 shows, the scores were lower and tended to be more concentrated in the pretest, while the posttest scores reached higher values and had greater dispersion. Substantial progress was made between the baseline situation and the final situation at the end of the term at all educational levels.

Figure 1. Comparison between the pretest scores and the posttest scores by grade
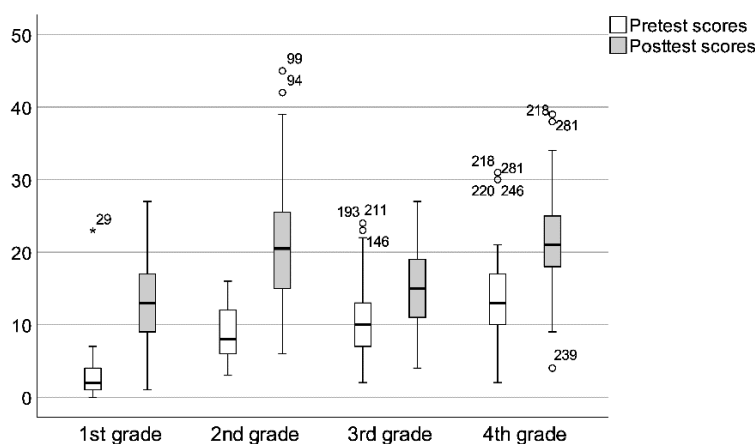
3.2. Serious games and gamification in the classroom

The software proposed included a gamification strategy to be used in the classroom as a complementary element, which was employed by some of the teachers. The pretest score was lower in the classes where gamification was implemented, and four points higher in the classes in which gamification was not applied (Table 2). This is a statistically significant difference of medium magnitude (*RYes*=123.24, *RNo*=181.43, *Z*=-5.628, *p*=0.000, *r*=0.33). It is important to note the variable "grade" in this result, since the classes that implemented no gamification were of the third and the fourth grades, while the classes that did implement gamification covered also the first and the second grades. However, the posttest reflected greater progress in the classes with gamification (a 10-point difference between the two means) compared with the classes in which gamification had not been applied (6.38 points). This difference is significant in the case of the progress made by the classes that had gamification versus the progress of the classes in which gamification was not

used, with an effect size that comes close to medium (RYes=158.68, RNo=109.80, Z=-4.729, p=0.000, r=0.28). In fact, the posttest score does no longer reveal any significant or remarkable differences between the classes that applied gamification and those in which it was not implemented (RYes=139.59, RNo=148.38, Z=-0.850, p=0.395, r=0.05). In addition, the classes with gamification attained more days of use of the program and showed a higher number of activities solved by the students. Statistically significant differences with a medium effect size were obtained for the days of use (RYes=164.82, RNo=97.39, Z=-6.517, p=0.000, r=0.39), and an effect size close to medium for the number of activities (RYes=157.58, RNo=112.02, Z=-4.399, p=0.000, r=0.26).

Table 2. Descriptive statistics according to use of gamification in the classroom					
	Gamification in the classroom (*n*=190)			No gamification in the classroom (*n*=94)		
	M	Md	SD	M	Md	SD
Pretest score	7.64	7	5.50	11.73	11	5.78
Posttest score	17.63	17	7.75	18.11	18	7.06
Days of use	29.1	28	6.6	23.4	23	7
Activities solved	6,257	5,220.5	3,399.6	4,718.7	3,892.5	2,781.8

At an intragroup level, there was a statistically significant improvement of great magnitude in mathematics fluency, both in the group that benefited from gamification in the classroom (n=190, Z=-11.959, p=0.000, r=0.61) and in the group in which gamification had not been implemented (n=94, Z=-7.638, p=0.000, r=0.56), although the effect size was greater for the former.

3.3. At classroom level: Novice teachers versus experienced teachers

All classroom groups made progress in mathematical calculus fluency after the serious gaming experience (Table 3). However, there were groups whose level of progress was clearly lower, as occurred in A8 and A9. Equally, this situation extended to apply to effect size, which was large in all classes (r=0.62) but lower in A8 (r=0.43) and A9 (r=0.49). We started from a similar scenario for all cases; however, diverse ways of use, appropriation and management can be seen among the teachers.

Table 3. Descriptive statistics by class									
Grade 1 (*n*=69)	A1 (*n*=24)			A2 (*n*=22)			A3 (*n*=23)		
	M	Md	SD	M	Md	SD	M	Md	SD
Pretest s.	2.42	2	2.10	3.86	2	4.76	2.70	2	1.82
Posttest s.	12.08	12.5	5.83	13.32	14	6.11	13.04	12	4.35
Days of use	28.5	27	6.1	28.4	27.5	4.4	28	28	4
Activities	5,940.8	4,566	3,515.6	6,059	5,557	3,147.5	4,992.5	4,604	1,288.4
Grade 2 (*n*=72)	A4 (*n*=23)			A5 (*n*=25)			A6 (*n*=24)		
	M	Md	SD	M	Md	SD	M	Md	SD
Pretest s.	7.70	8	3.02	9.96	9	3.91	8.17	8	3.67
Posttest s.	20.04	20	6.35	23.44	21	10.23	19.33	20.5	6.82
Days of use	29	28	4.7	32.1	32	7.3	27.7	28	5.3
Activities	6,314.6	5,708	3,072.1	8,206.4	6,140	5,357.6	5,726.6	5,145.50	2,198.8
Grade 3 (*n*=73)	A7 (*n*=25)			A8 (*n*=24)			A9 (*n* = 24)		
	M	Md	SD	M	Md	SD	M	Md	SD
Pretest s.	11.24	10	4.19	8.21	8	2.98	12.46	11	5.54
Posttest s.	17.92	18	5.62	11.29	10	4.71	17.04	18.5	4.44
Days of use	35.6	35	7.1	26.4	24	7.72	26	25.5	7.9
Activities	8,334.1	7,747	3,129.6	5,899.9	5,222.5	3,382.6	5,555	5,076	2,797.7
Grade 4 (*n* = 70)	A10 (*n*=23)			A11 (*n*=24)			A12 (*n*=23)		
	M	Md	SD	M	Md	SD	M	Md	SD
Pretest s.	13.43	12	6.99	14.29	13.5	5.55	12.96	13	5.67
Posttest s.	23.39	24	6.49	21.13	20	7.13	21.04	20	5.87
Days of use	21.4	22	5	22.8	22	5.8	19.6	19	4.4

Activities	4,171.4	3,668	1,778	4,247.5	3,768	1,694.7	3,160.6	2,434	2,066.6

At this point in the analysis, a potential relationship between these data and the teacher's professional experience emerges. In order to make comparisons between classes under this criterion, we took educational levels as reference, since the requirements of the curricula were similar. At the same time, there had to be a coincidence between the conditions of implementation or non-implementation of gamification in the classroom. In this sense, we saw that both the teachers with greater seniority (more than 30 years of teaching experience) and the novice teachers (between 1 and 5 years of teaching experience) used the software in a similar way during class time. Even in some cases, higher values were obtained for the teachers with greater seniority and, in addition, better posttest results were seen among the students who had more experienced teachers. In this regard, in the second grade, in which all the teachers used gamification, those who were more experienced (in A4 and A5) attained better mean results in the posttest than the novice teachers (A6) (M=21.81 versus M=19.33), made more intensive use of the program (M=30.6 days versus M=27.7) and obtained a higher rate of activities carried out (M=7,299.9 versus M=5,726.6). It was seen that there were no significant differences between the values obtained for the novice teachers and those for the experienced teachers, with small effect sizes (pretest: $Z=-0.665$, $p=0.506$, $r=0.08$; posttest: $Z=-1.005$, $p=0.315$, $r=0.12$; days of use: $Z=-1.580$, $p=0.114$, $r=0.19$; activities solved: $Z=-1.374$, $p=0.170$, $r=0.16$). As to grade 3, for the two classes that were compared (A8: novice teacher; A9: experienced teacher), in which gamification had not been used, similar values were obtained in the use of the software and the number of activities solved (Table 3). In fact, there were no significant differences in this regard (Table 4); in contrast, the pretest and posttest scores did show significant differences, with a notable effect size.

Table 4. Mean ranks and statistics of the Mann–Whitney U test for A8 and A9						
	R_{nov}	R_{exper}	U	Z	p	r
Pretest score	18.44	30.56	142.500	–3.011	0.003	0.43
Posttest score	16.69	32.31	100.500	–3.876	0.000	0.56
Days of use	24.75	24.25	282.000	–0.124	0.901	0.02
Activities solved	24.79	24.21	281.000	–0.144	0.885	0.02

3.4. Relationship between the posttest scores and the academic performance of the students

In general, a significant and moderate relationship can be seen between the posttest scores and the grades obtained in mathematics in the first term, except in the third grade (grade 1: $r_s=0.62$, $p=0.000$, $R^2=0.38$; grade 2: $r_s=0.53$, $p=0.000$, $R^2=0.28$; grade 3: $r_s=0.15$, $p=0.203$, $R^2=0.02$; grade 4: $r_s=0.43$, $p=0.000$, $R^2=0.18$). There was also a significant and moderate relationship between posttest scores and the academic grades in general, except in the third grade, where a lower correlation was obtained, although it was also significant (grade 1: $r_s=0.57$, $p=0.000$, $R^2=0.32$; grade 2: $r_s=0.63$, $p=0.000$, $R^2=0.40$; grade 3: $r_s=0.33$, $p=0.004$, $R^2=0.11$; grade 4: $r_s=0.46$, $p=0.000$, $R^2=0.21$). We should highlight the coefficient of determination in the second grade, which allows us to infer that the posttest scores account for up to 40% of the variance of the overall academic grades. Depending on the classes (Table 5), we can also notice, in general, a significant and moderate-high relationship between the posttest scores and the school grades, except for a number of specific groups in the third grade (A7, A8 and A9) and in the fourth grade (A12). Several coefficients of determination stand out which indicate that more than 50% of the variability of the grades in mathematics and of the overall academic grades can be accounted for by their relationship with the posttest scores.

Table 5. Correlation between the posttest scores and the academic performance of the students by class						
	Grade in mathematics			Overall grade (all subjects)		
	r_s	Sig.	R^2	r_s	Sig.	R^2
A1 posttest score	0.54	0.007	0.29	0.61	0.002	0.37
A2 posttest score	0.67	0.001	0.45	0.55	0.008	0.30
A3 posttest score	0.74	0.000	0.55	0.71	0.000	0.50
A4 posttest score	0.56	0.006	0.31	0.72	0.000	0.52

A5 posttest score	0.75	0.000	0.56	0.72	0.000	0.52
A6 posttest score	0.56	0.005	0.31	0.55	0.006	0.30
A7 posttest score	0.33	0.111	0.11	0.28	0.173	0.08
A8 posttest score	0.20	0.360	0.04	0.22	0.299	0.05
A9 posttest score	0.31	0.138	0.10	0.40	0.050	0.16
A10 posttest score	0.45	0.030	0.20	0.46	0.027	0.21
A11 posttest score	0.42	0.043	0.18	0.53	0.008	0.28
A12 posttest score	0.39	0.065	0.15	0.35	0.099	0.12

4. Discussion and conclusions

The purpose of this study was to investigate the impact of the use of serious games on the work carried out at primary education levels. Our study provides scientific evidence of the improvement in mathematical contents through the use of serious games. These results are consistent with the work of Clark et al. (2016), Carvalho et al. (2018) and Wouters et al. (2013). Our study reinforces this line of research with evidence of the benefit that serious games can provide in real-life contexts. Specifically, substantial progress in mathematics fluency can be seen in all four grades analyzed, in all 12 classes involved, and for all the participating students. These data are consistent with previous studies that used much smaller samples (Cozad, 2019) and students with learning difficulties (Sarrell, 2014). There is a statistically significant difference of great magnitude between the baseline situation and the situation that was seen after the serious gaming experience. We should bear in mind that there is an apparently limiting factor in these data: the fact that a piece of software designed for a US curriculum was used. This did not determine its potential. Additionally, greater progress in mathematics fluency as well as more intensive use of the program and a higher number of activities solved were observed in the classes that used the gamification strategy. This leads us to confirm the impact of the use of these strategies on the domain of motivation, clearly in line with previous studies (Filella et al., 2017; McGonigal, 2011; Zagal & Altizer, 2014), and to associate its use with an improvement in students' performance (Fernández-Robles et al., 2019; Pires et al., 2019).

A significant and moderate relationship is also found between the scores obtained in the performance tests applied after the serious gaming experience and the students' grades, with similar percentages of variation, which exceed 50% for some classes. In this regard, we should point out that this relationship between the different variables does not necessarily imply causality but refers to the degree of relationship. These data reveal the explanatory power of this type of curricular contents in relation to the work carried out in the classroom, which serves to update previous studies on the matter and highlights the potential influence of this sort of proposals on students' general performance (Duncan et al., 2007). The evidence of improvement shown by these data allows us to question the studies that suggest the existence of generational barriers against the potential use of this software by teachers (Kaufman, 2013) owing to a lack of previous positive experiences (Marín-Díaz et al., 2019). Adherence to work with serious games and the process to implement this in the classroom requires modulation by teachers. Rather than being a problem, as suggested by previous literature, age is seen here as an opportunity that complements the work carried out with the software, banishing the image of novice teachers as being more sensitive and inclined to the use of these programs. Experienced teachers utilize them with a level of intensity that is similar to that of novice teachers, and under equal conditions significant differences can even be found in favor of experienced teachers compared to younger ones. These results seem to indicate that these teachers value the use of serious games and gamification according to the perceived benefits, which drives them to reinforce this proposal by facilitating the conditions to carry out the tasks suggested, with knowledge and experience that lead to even better results. This situation opens up the possibility of investigating the elements that these teachers bring into play, and highlights the need to conduct research studies focused on core areas of schoolwork that have a high value for educational centers.

Everything suggests that contextual elements are determining factors in the study of serious games and gamification, and may lead to disparate results. The selection of contents that are relevant to teachers appears as a key factor when it comes to having an impact on schoolwork, bringing different skills—particularly mathematical and digital skills—into play. Additionally, students, who are clearly motivated and involved in this proposal, are offered digital gaming experiences with a high degree of isomorphism compared to the usual experiences of their leisure time, which helps to bridge the gap between the reality of the school and social reality and to add value to the role played by educational centers. This situation allows us to conclude that the

use of serious games and gamification specifically designed for school environments has potential in relation to students' performance.

Nevertheless, this work has some limitations. We should highlight the fact that it only included students and teachers from one single educational center, so the results should be used with caution in any potential generalizations beyond the context in which the study was carried out. Additionally, the reasons for the different forms of appropriation by teachers are unknown, especially in the case of experienced teachers, who, quite unexpectedly, joined the proposal with great intensity.

Author Contribution

Idea, F.F.V; Literature review (state of the art), F.F.V; Methodology, E.V.C., E.M.P; Data analysis, E.V.C., E.M.P; Results, E.V.C, F.F.V., E.M.P; Discussion and conclusions, F.F.V., E.V.C.; Writing (original draft), F.F.V., E.V.C; Final revisions, E.V.C., E.M.P; Design project and sporsonship, F.F.V.

References

Alvarez, J., & Djaouti, D. (2012). *Serious games: An introduction*. Éditions Questions Théoriques. http://bit.ly/3gVVAZY

Baroody, A.J., Bajwa, N.P., & Eiland, M. (2009). Why can't Johnny remember the basic facts? *Developmental Disabilities Research Reviews*, *15*(1), 69-79. https://doi.org/10.1002/ddrr.45

Baroody, A.J., Eiland, M.D., Purpura, D.J., & Reid, E.E. (2013). Can computer-assisted discovery learning foster first graders' fluency with the most basic addition combinations? *American Educational Research Journal*, *50*(3), 533-573. https://doi.org/10.3102/0002831212473349

Boendermaker, W.J., Veltkamp, R.C., & Peeters, M. (2017). Training behavioral control in adolescents using a serious game. *Games for Health Journal*, *6*(6), 351-357. https://doi.org/10.1089/g4h.2017.0071

Carvalho, C.V., Rodríguez, M.C., Nistal, M.L., Hromin, M., Bianchi, A., Heidmann, O., Tsalapatas, H., & Metin, A. (2018). Using video games to promote engineering careers. *The International Journal of Engineering Education*, *34*(2), 388-399. http://bit.ly/3qjVwET

Chauhan, S. (2017). A meta-analysis of the impact of technology on learning effectiveness of elementary students. *Computers & Education*, *105*, 14-30. https://doi.org/10.1016/j.compedu.2016.11.005

Clark, D.B., Tanner-Smith, E.E., & Killingsworth, S.S. (2016). Digital games, design, and learning: A systematic review and meta-analysis. *Review of Educational Research, 86*(1), 79-122. https://doi.org/10.3102/0034654315582065

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. L. Erlbaum Associates. https://bit.ly/3uE1kfh

Council of Chief State School Officers (Eds.) (2020). *Preparing america's students for college & career.* Common core state standards initiative. http://bit.ly/2XFRaMc

Cozad, L. (2019). *Effects of a digital mathematics fluency program on the fluency and generalization of learners*. [Doctoral Dissertation, The Pennsylvania State University]. http://bit.ly/3qkbnDE

Cress, T. (2019). *Influence of the reflex math fact fluency program on math scores*. [Doctoral Dissertation, Walden University]. http://bit.ly/3sryn5x

Del-Moral, M., & Fernández, L.C. (2015). Videojuegos en las aulas: Implicaciones de una innovación disruptiva para desarrollar las Inteligencias Múltiples. *Revista Complutense de Educación*, *26*, 97-118. https://doi.org/10.5209/rev_RCED.2015.v26.44763

Duncan, G.J., Dowsett, C.J., Claessens, A., Magnuson, K., Huston, A.C., Klebanov, P., Pagani, L.S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K., & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, *43*(6), 1428-1446. https://doi.org/10.1037/0012-1649.43.6.1428

Fernández-Robles, J.L., Gaytán-Lugo, L.S., Hernández-Gallardo, S.C., & García-Ruíz, M.A. (2019). La alfabetización cuantitativa en estudiantes de tercer grado de primaria a través de un juego serio. *Revista Latinoamericana de Tecnología Educativa*, *18*(1), 131-147. https://doi.org/10.17398/1695-288X.18.1.131

Field, A. (2018). *Discovering statistics using IBM SPSS statistics*. Sage.

Filella, G., Pérez-Escoda, N., & Ros-Morente, A. (2017). Evaluación del programa de educación emocional 'Happy 8-12' para la resolución asertiva de los conflictos entre iguales. *Electronic Journal of Research in Education Psychology*, *14*(40), 582-601. https://doi.org/10.25115/ejrep.40.15164

Fisher, D., Frey, N., & Hattie, J. (2020). *The distance learning playbook, grades K-12: Teaching for engagement and impact in any setting*. Corwin Press. https://bit.ly/33wqrEG

Foegen, A., & Deno, S.L. (2001). Identifying growth indicators for low-achieving students in middle school mathematics. *The Journal of Special Education*, *35*(1), 4-16. https://doi.org/10.1177/002246690103500102

Fritz, C.O., Morris, P.E., & Richler, J.J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology. General*, *141*(1), 2-18. https://doi.org/10.1037/a0024338

Geary, D.C. (2011). Cognitive predictors of achievement growth in mathematics: A 5-year longitudinal study. *Developmental Psychology*, *47*(6), 1539-1552. https://doi.org/10.1037/a0025510

Gómez-García, S., Planells-de-la-Maza, A.J., & Chicharro-Merayo, M. (2016). ¿Los alumnos quieren aprender con videojuegos? Lo que opinan sus usuarios del potencial educativo de este medio. *Educar*, *53*(1), 49-66. https://doi.org/10.5565/rev/educar.848

Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does gamification work? - A literature review of empirical studies on gamification. *2014 47th Hawaii International Conference on System Sciences*, 3025-3034. https://doi.org/10.1109/HICSS.2014.377

Hieftje, K., Pendergrass, T., Kyriakides, T.C., Gilliam, W., & Fiellin, L. (2017). An evaluation of an educational video game on mathematics achievement in first grade students. *Technologies, 5*(2), 30. https://doi.org/10.3390/technologies5020030

Kaufman, D. (2013). Videogames in education - comparing students', student teachers' and master teachers opinions and experiences. In O, Foley, M.T, Restivo, J. Uhomoibhi, & M. Helfert (Eds.), *Proceedings of the 5th International Conference on Computer Supported Education - Volume 1: CSEDU* (pp. 101-105). CSEDU. https://doi.org/10.5220/0004383701010105

Kling, G., & Bay-Williams, J.M. (2014). Assessing basic fact fluency. *Teaching Children Mathematics, 20*(8), 488-497. https://doi.org/10.5951/teacchilmath.20.8.0488

Lamb, R.L., Annetta, L., Firestone, J., & Etopio, E. (2018). A meta-analysis with examination of moderators of student cognition, affect, and learning outcomes while using serious educational games, serious games, and simulations. *Computers in Human Behavior, 80*, 158-167. https://doi.org/10.1016/j.chb.2017.10.040

Marín-Díaz, V., Morales-Díaz, M., & Reche-Urbano, E. (2019). Educational possibilities of video games in the primary education stage according to teachers in training. A case study. *Journal of New Approaches in Educational Research, 8*(1), 42-49. https://doi.org/10.7821/naer.2019.1.330

McGonigal, J. (2011). *Reality is broken: Why games make us better and how they can change the world*. Penguin. https://stanford.io/3xWkgaW

Meiri, R., Levinson, O., & Horowitz-Kraus, T. (2019). Altered association between executive functions and reading and math fluency tasks in children with reading difficulties compared with typical readers. *Dyslexia, 25*(3), 267-283. https://doi.org/10.1002/dys.1624

Mellado, R., Melgarejo, B., Velasquez, C., Cubillos, C., Roncagliolo, S., & Gonzalez, N. (2018). ROLE video game tool for teaching myths and legends to school basic students. *2018 37th International Conference of the Chilean Computer Science Society (SCCC)*, 1-8. https://doi.org/10.1109/SCCC.2018.8705258

Michael, D.R., & Chen, S.L. (2006). *Serious games: Games that educate, train and inform*. Thomson Course Technology. https://bit.ly/3uFzQWz

Pérez-Manzano, A., & Almela-Baeza, J. (2018). Gamificación transmedia para la divulgación científica y el fomento de vocaciones procientíficas en adolescentes. [Gamification and transmedia for scientific promotion and for encouraging scientific careers in adolescents]. *Comunicar, 55*, 93-103. https://doi.org/10.3916/C55-2018-09

Pires, A.C., González-Perilli, F., Bakała, E., Fleisher, B., Sansone, G., & Marichal, S. (2019). Building blocks of mathematical learning: Virtual and tangible manipulatives lead to different strategies in number composition. *Frontiers in Education, 4,* 1-11. https://doi.org/10.3389/feduc.2019.00081

Romero-Rodríguez, L., & Torres-Toukomidis, A. (2018). Con la información sí se juega: Los newsgames como narrativas inmersivas transmedias. En A. Torres-Toukomidis & L. Romero-Rodríguez (Eds.), *Gamificación en Iberoamérica. Experiencias desde la comunicación y la educación.* Abya-Yala. https://bit.ly/2TLCK8O

Rosenthal, J.A. (1996). Qualitative descriptors of strength of association and effect size. *Journal of Social Service Research, 21*(4), 37-59. https://doi.org/10.1300/J079v21n04_02

Sánchez-Huete, J.C. (2013). *Métodos de investigación educativa*. Punto Rojo. https://bit.ly/3tFYZ1X

Sarrell, D. (2014). *The effects of reflex math as a response to intervention strategy to improve math automaticity among male and female at-risk middle school students*. [Doctoral Dissertation, Liberty University]. http://bit.ly/3qiFdYV

Stieler-Hunt, C., & Jones, C.M. (2015). Educators who believe: Understanding the enthusiasm of teachers who use digital games in the classroom. *Research in Learning Technology, 23*, 1-14. https://doi.org/10.3402/rlt.v23.26155

Teixes, F. (2015). *Gamificación: Fundamentos y aplicaciones*. UOC. https://bit.ly/2SzJoEc

Van-der-Ven, F., Segers, E., Takashima, A., & Verhoeven, L. (2017). Effects of a tablet game intervention on simple addition and subtraction fluency in first graders. *Computers in Human Behavior, 72*, 200-207. https://doi.org/10.1016/j.chb.2017.02.031

Wouters, P., van-Nimwegen, C., van-Oostendorp, H., & van-der-Spek, E.D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology, 105*(2), 249-265. https://doi.org/10.1037/a0031311

Zagal, J.P., & Altizer, R. (2014). Examining «RPG elements»: Systems of character progression. In M. Mateas, T. Barnes, & I. Bogost (Eds.), *Proceedings of the 9th International Conference on the Foundations of Digital Games, FDG 2014, Liberty of the Seas, Caribbean, April 3-7, 2014.* https://bit.ly/3oEhli9

Zagalo, N. (2010). Alfabetización creativa en los videojuegos: Comunicación interactiva y alfabetización cinematográfica. [Creative game literacy. A study of interactive media based on film literacy experience]. *Comunicar, 35*, 61-68. https://doi.org/10.3916/C35-2010-02-06