Design and Evaluation of a Game for Mobile Platforms about Periodic Properties of the Chemical Elements

Gabriela Trindade Perry
Marcelo Leandro Eichler
Danilo Marcondes Filho

ABSTRACT

This paper presents the results of two studies on Xenubi, a computer/mobile game about periodic properties of the chemical elements. These studies were aimed at evaluating Xenubi in 3 dimensions: interface usability, user preferences and learning outcomes. The usability and user preferences study focused on the understanding of the game mechanics by students from a public state school and from a private school in Brazil. The learning outcomes evaluation study was conducted in a public technical school, and sought for differences in post-test score between groups that played Xenubi as printed cards and on the computer. The results point to a consistent better understanding of the game mechanics by students from the private school and a higher performance improvement for the students who played Xenubi on the computer.

Keywords: Human-Computer interface. Learning objects. Educational games. Evaluation methodologies.

Projeto e Avaliação de um Jogo para Dispositivos Móveis sobre as Propriedades Periódicas dos Elementos Químicos

RESUMO

Este artigo apresenta os resultados de dois estudos sobre Xenubi, um jogo para computador e dispositivos móveis (celulares e tablets) sobre as propriedades periódicas dos elementos químicos. Os estudos realizados buscaram avaliar o jogo didático Xenubi em 3 dimensões: usabilidade de interface, preferências do usuário e resultados de aprendizagem. O estudo de usabilidade e preferências do usuário teve como foco a compreensão da mecânica do jogo por alunos de uma escola pública estadual e de

Gabriela Trindade Perry é Bacharel em Design, Mestre em Ergonomia e Doutora em Informática na Educação. Atualmente é Professora Adjunta da UFRGS, no curso de Design, ligado à Faculdade de Arquitetura e Professora permanente do Programa de Pós-Graduação em Informática na Educação da UFRGS.
E-mail: gabriela.perry@ufrgs.br

Marcelo Leandro Eichler é Licenciado em Química, Mestre em Psicologia e Doutorado em Psicologia do Desenvolvimento. Atualmente é Professor Adjunto do Departamento de Química Inorgânica da Universidade Federal do Rio Grande do Sul (UFRGS), docente permanente do Programa de Pós-graduação em Educação (UFRGS) e docente colaborador do Programa de Pós-graduação Educação em Ciências: Química da Vida e Saúde (UFRGS/UFSM/FURG). E-mail: exlerbr@gmail.com

Danilo Marcondes Filho é Graduado em Estatística, Mestre e Doutorado em Engenharia de Produção. Atualmente, é professor associado no Departamento de Estatística da Universidade Federal do Rio Grande do Sul.
E-mail: marcondes.danilo@gmail.com

uma escola particular. O estudo de avaliação de resultados de aprendizagem foi conduzido em uma escola técnica pública e por meio de pré e pós-testes se buscou analisar as diferenças na pontuação dos testes entre grupos que jogaram Xenubi com cartões impressos e com o computador. Os resultados apontam para uma melhor compreensão da mecânica do jogo entre alunos da escola privada e uma melhoria do desempenho dos alunos quando jogaram Xenubi no computador.

Palavras-chave: Interface homem-computador. Objetos de aprendizagem. Jogos educativos. Metodologias de avaliação.

INTRODUCTION

Chemistry is often seen as a complex and difficult subject to understand. However, through the use of educational games, learning chemistry concepts can be improved (Russell, 1999; Rastegarpour & Marashi, 2012). Games are excellent tools to introduce the topics and supplement lectures; such activity is much more stimulating to students than the traditional approach based on memorization (Franco-Mariscal et al., 2016). Generally, teachers report that game-based teaching motivated students to learn, then teaching with games engage students and influence their learning outcomes (Huizenga et al., 2017).

Cards constitute a type of game successfully used to teach chemistry. Franco-Mariscal, Oliva-Martínez and Bernal-Márquez (2012) quote 3 examples: a card game used to correlate the names and symbols of the most important chemical elements; a card game where elements have their own personality and challenge each other using their properties and oxidation states; a card game whose main purpose is to recognize the structure of the periodic table of elements and their groups and families.

In a study aimed at investigating the effect of card and computer games in learning chemistry concepts with high school students, Williams and Pence (2011) found that this type of teaching activity provided improved student performance in assessments. Its results indicated that there was a significant difference between teacher-made card games and computer games and the traditional teaching method. However, there was no significant difference between teacher-made card games and computer games. Playing had a significant role in learning abstract concepts, and understanding chemistry concepts was facilitated by creating excitement and joy, as well as interaction amongst students.

Furthermore, regarding mobile technologies, the changes introduced by them are legitimate and should be considered as valid ways of communicating and interacting (McCarthy & Wright, 2007). In this way, students use smartphones as learning aids due many reasons such as they provide convenience, portability, comprehensive learning experiences, multi sources and multitasks, and environmentally friendly (Anshari et al., 2017).

Further progress in mobile computing will continue to act as a stimulus and opportunity for innovative approaches to chemistry education. In this sense, we are working on producing digital content for teaching and learning chemistry through mobile devices such as smartphones and tablets (Perry & Eichler, 2015; Eichler et al., 2017). In this article, we present a game about the periodic properties of the chemical elements and its evaluation.

Designing learning interfaces

The design of learning interfaces [games included] is inherently interdisciplinary. In interdisciplinary design teams, people must integrate information, techniques, tools, concepts and theories from more than one body of knowledge to solve problems whose solutions are beyond the scope of a single discipline (Tress et al., 2005).

The ideal educational interface design team has several agents, for example: teachers and instructional designers; analysts and programmers from varied backgrounds; graphic, motion and interaction designers; writers and text-reviewers. Ideally, these team members would also share knowledge on relevant issues such as technology limitations and possibilities; design practices; classroom routines; teaching practice and learning theories. But that is not always the case. In research on usability evaluation for educational software, Zaharias and Poylymenakou (2009, p.76) mention an "ellipsis of research-validated usability evaluation methods that address the user as a learner in a holistic way, which includes the consideration of cognitive and affective learning factors". Hinostroza and Mellar (2001, p.27) asserted that "in addition to having knowledge of learning theories, educational software designers should take teaching practices into consideration". Despite the difficulty in managing interdisciplinary teams, such integration is highly desirable.

In addition to methodological proposals, there are also research studies on guidelines for designing learning interfaces. The aim of these studies is to propose and validate a set of guidelines that, if followed, would help design better interfaces. For example: Jones (2008) advises the use of storyboards as a means to bootstrap the design of learning tasks; Ariga and Watanabe (2008) designed a web design course to enhance the "visual expression" of students who are not in art or design courses. Malone (1980) proposed a set of heuristics for designing educational games with items grouped according to fantasia, challenge and curiosity. It is the result of 4 studies: one about students' gaming preferences; two about comparisons of slightly different versions of two games and one about an instructional game on fractions.

The point of this presentation is to show that there is an effort to create better ways to manage the process of designing learning interfaces. They represent the crystallized knowledge of a community and, as such, are of great value. Regarding our research, it provides insights about how the collaboration of actors from several different backgrounds might occur and how we should set up the roles in our design team. In this study, we divided the evaluation into 3 dimensions: usability, user preferences and learning outcomes. Each is presented in the following sections.

EVALUATING LEARNING INTERFACES

Usability inspection

The first dimension of evaluation in the present study corresponds to interface usability. According to Sharp et al. (2001, p.14), "usability is generally regarded as

ensuring that interactive products are easy to learn, effective to use and enjoyable from the user's perspective". The ISO 9241-11 (p.6) defines usability as "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction within a specified context of use"; therefore usability inspections are about objective qualities of an interface.

The most widely employed technique of usability inspection is through heuristics sets – Molich and Nielsen (1990) proposed a set with 9 heuristics that is very well known. Stanton and Young (2003, p.94) consider this technique has disadvantages – highly subjective and unstructured; lack reliability, comprehensiveness and auditability – and advantages – simple to execute, requiring little previous knowledge; efficient on resources (both time and materials). Addressing the issue of evaluators' qualification, Kirmani (2008) proposed a score test which defines levels of expertise permissible to conduct a usability evaluation using heuristics; the result is based on the number of unique, valid issues identified by the evaluators as well as the severity of each issue (p.50).

There are other techniques i.e. questionnaires, think-aloud protocols and cognitive walkthrough (Nielsen & Mack, 1994). Nevertheless, user testing is the most powerful method for usability inspection, because it can be suited to every task and use context, and to every phase in the design and development cycle – the downside is the cost in time and resources. The reader is headed to Hornbæk (2006) for an exhaustive review of procedures to assess effectiveness, efficiency and satisfaction – the 3 pillars of usability.

Evaluating usability of any software is an important task; however, there is a need for approaches that take user experience into account. The challenge, in this case, is how to measure something that is so fluid, dynamic and hard to define.

User experience evaluation

The second dimension of evaluation in this study is related to emotions and affect, which have received much attention from the human-computer interaction (HCI) research and practice community since Piccard's (1997) book "Affective computing". After all, the affection users feel for games and mobile devices might be as important as usability aspects, i.e. [effectiveness, efficiency and satisfaction]. Hassenzahl (2001), for example, confirmed the validity of a model which associates Hedonic Qualities; HQ [e.g. interesting, original, and innovative] with Ergonomic Qualities; EQ [e.g. simplicity, reliability and predictability] to perceived appeal [e.g. pleasant, motivating]. Hassenzahl's research brings evidence that aspects related to affection are as important as usability related aspects.

However, as Boehner et al. (2007) point out, measuring and defining emotion is challenging. In the introduction to their paper, Boehner et al. emphasize two theories about emotion: one rooted in the "cognition is computation" tradition – illustrated by Newell & Simon's (1972) book "Human problem solving" – the other rooted in the social and context embedding of cognition – illustrated by Winograd's (1996) book "Bringing design

to software". Boehner et al. claim that the way we choose to understand/explain emotion has implications in the way we design and measure user experience [UX].

The issue of where to test mobile application – "in the field" or in the lab? – has been addressed by Kaikkonen et al. (2005), who compared field and laboratory user testing of the same mobile application. They concluded that field tests are worthwhile only if user behavior needs to be investigated in a natural context. Regarding tools for mobile field testing, Stark and Riebeck (2008) present a log tool for unobtrusive data gathering, and Oulasvirta and Nyyssönen (2009) describe equipment for capturing the interactional relationship between the user and the environment in an unobtrusive manner.

Evaluating students' learning outcomes

Lastly, the third dimension of educational software [mobile games included] evaluation concerns students' learning outcomes. In this case, defining what is meant by "learning outcome" is one of the keys to designing the evaluation procedure. For example, one learning outcome could be students' ability to relate concepts with real world phenomena; to use information to solve a problem or to make a statement; to carry a procedure or to find the correct answer.

The other key would be the researcher's beliefs in how it can be assessed, which would translate into a qualitative study or into a controlled experiment. Both approaches have advantages and disadvantages, but in any case it is important to look for threats to the validity of the assessment because there will be many independent variables interacting, which are difficult to control – quoting Campbell and Stanley (1963, p.175), "internal validity is the basic minimum without which any experiment is uninterpretable. External validity asks the question of generalizability". Campbell and Stanley were talking specifically about controlled experiments, but even if the method of choice were qualitative, the concern for validity is still necessary. Recognizing the difficulty of achieving high reliability when analyzing textual matter, Krippendorff (2004, p.3) argues that "the mathematical complexity of analyzing variably unitized text, while an unquestionable hurdle for replicating research, is no justification for creating the methodological schism between quantitative and qualitative approaches". Next, we present the game for mobile devices

DESIGNING XENUBI

The design and development approach reported in this paper depicts our efforts to integrate technology, education, aesthetics, user experience and usability goals into a design and research framework. We believe that this way our projects would be useful to students and teachers. The most perfected stage of this design and development approach would be Design Based Research (Reeves & McKenney, 2015), which seeks to integrate user centered design and research data in continuous cycles of design, enactment, analysis,

and redesign. Next, we present the main design cycles of Xenubi, a card-based game for mobile platforms.

Learning/educational interface design

Xenubi is a card-based game for mobile platforms about Periodic Properties, a traditional theme in high school and undergraduate chemistry courses. It is a casual game: easy to learn, with simple rules and easy to remember how to play. Casual games are characterized by fast and short missions, which do not require the player to spend several hours playing to advance. We decided that this would be the most efficient approach, and chose to develop the game for mobile devices because we consider that this medium is optimized for the nature of casual games: the player could play anywhere, anytime. We decided not to make use of internet connection because the majority of teenagers have limited internet access.

The didactical strategy was to relate the periodic table structure to atomic models, so students could understand why it is called "periodic". "Xenubi" is not an existing word in Portuguese: it was created by the conjunction of the slang "newbie; noob" – used by gamers – and "chem." – short for "chemistry". In Xenubi, the player receives 6 cards, each representing a chemical element. Each card has the [numerical] value of 6 properties: atomic radius, first ionization potential, electron affinity, electronegativity, melting point, and density. These are the properties most teachers use to address the subject of periodic table of elements or chemical periodicity.

The player sees, in a periodic table, his and the opponent's [computer's] element, and has to chose among those 6 properties the one he thinks his element has a higher value. The player should infer this by using the knowledge about periodicity and the location of both elements in the periodic table.

Interface design and development

The design and development of the interface happened in iterative cycles, so the concept design evolved, as depicted in Figures 1 and 2. Figure 1 as a literal interpretation of the "notebook" concept, with stick figures and hand-drawn lettering. Figure 2 still holds this concept, but the graphics have a different treatment, with a more polished look.
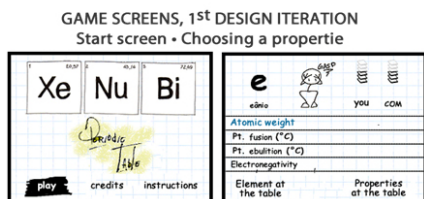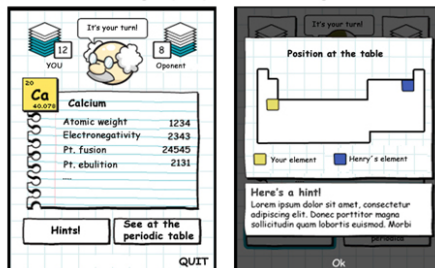
Figure 1. First graphical concepts.

Figure 2. The graphical concept at the second design iteration.

Figure 3 shows the final design. The main changes are: the periodic table is always visible and and the character that represents the opponent [the computer] was changed from an estereotyped scientist [see figure 2, for a charicature of Henry Moseley (born in 1887 – died in 1915), whose discoveries resulted in a more accurate positioning of elements in the Periodic Table by closer determination of atomic numbers (Scerri 2007).
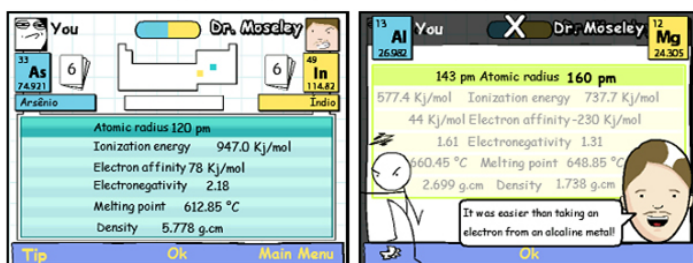

Figure 3. A design change motivated by Nokia's (2006) Interaction Guidelines.

Xenubi's prototypes were built using Adobe Flash© because it is easier to make changes in a Flash Lite© file than in other programming IDEs. Maiti (2010) also justifies the use of Flash Lite© in terms of being easy to edit. However, we would advice its adoption only for rapid prototyping and usability inspections. Once the interace and interaction design was approved, Xenubi was ported to HTML5, CSS3 and Javascript, using Adobe PhoneGap©, which is compatible with Android 2.1+ and iOS.

Discount usability evaluation

"Discount usability" is a term coined by Nielsen (1995) to refer to user testing, early prototyping and heuristic evaluation. In Xenubi's case study, it pointed to issues related to label readability and game control visibility.

One user study took place in a pre-medical course, which prepares students for Medical Science college tests. Six students [3 male and 3 female] aged between 18 and 20, regular smartphone users and with broad game habits individually tested the game in their classroom during a break between classes. The game mechanics was not explained – we just handed them a smartphone with the game loaded and told them it was about the Periodic Table of Chemical Elements. After playing the game, they answered an interview focused on game play and interface comprehension. None of the students had difficulty playing the game, and 4 of them repeatedly won the matches. It is important to note that these students are not Xenubi's target audience, since they are a bit older and much more motivated to study chemistry than regular students [because the entrance tests for public Medical colleges are very demanding].

During this evaluation, we realized that (1) there was no sign of how much it would take to win or lose the game and (2) the elements on the result screen – shown at the end of each round – looked like buttons. All students tried clicking on these elements, which were mere informative text labels. We also added "internet memes", which are very popular among teenagers nowadays. In the next section, we report the evaluation we conducted in private and public high schools, Xenubi's target audience.

XENUBI'S USABILITY AND EDUCATIONAL EVALUATION

Usability and player preferences evaluation

The hypothesis is that students from private schools – who are more familiar with smartphones – would understand the game mechanics better than students from a public school. 37 students [aged between 14 and 18; 25 females and 12 males] played Xenubi at the public school. The average match duration was 212 seconds (std = 63); the average round duration was 24 seconds (std = 8.4) and the average ratio between the amount of winning rounds and total amount of rounds was 0.54 (std = 1.9). 6 testers took part in the evaluation. Data from 14 students could not utilize data from 14 students, because one of the evaluators did not follow the test protocol (N = 37 – 14 = 23). In the private school, 4 testers conducted the evaluation with 20 students [10 male and 10 female, ages between 14 and 16]. The average match length was 275 seconds (std = 134); the average round length was 25 seconds (std = 14) and the average ratio between the amount of winning rounds and total amount of rounds was 0.49 (std = 1.9).

Despite the differences in ages in public to private schools, all students were in the same grade, and had just been taught about periodic properties and the periodic table. The classes had different teachers. The interaction between students and cell phones was recorded with HD cameras. The evaluation protocol was the same at both schools:

✓ The testers introduced themselves and explained the purpose of the evaluation, emphasizing that the goal was to evaluate Xenubi's interface, not students' knowledge about chemistry.

- ✓ The evaluation protocol was explained, emphasizing that the testers would not explain how to play the game and that the students could play it for as long as they liked. The goal was to make the use context closest to reality,

- ✓ While waiting to play Xenubi, students enrolled in a research activity, not related to periodic properties, in the computer laboratory [figure 4-a],

- ✓ Students were called in groups [of 6 in the public school and of 4 in the private school] to an adjacent room [figure 4-b]. They received a cell phone [Nokia C3, Nokia E63 or LG C570] with Xenubi loaded, in the game's 1st screen. After a minimum 3 minutes of playing, the tester would ask questions regarding understanding of Xenubi's interface elements. This interaction was recorded with a full HD camera.

- ✓ After the student decided to stop playing, he returned the cell phone to the tester and answered a questionnaire in the same room [figure 4-c]. Students were asked about their preferences regarding chemistry, Xenubi and games in general. The questions were in a Likkert scale format, with two anchors representing negative/positive extremes.
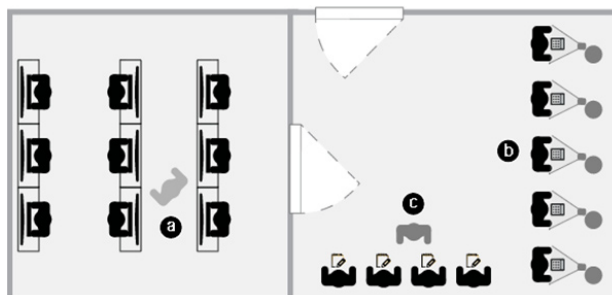
TEST ROOM LAYOUT

Figure 4. The computer lab of both schools were similar, as depicted.

Differences in learning outcomes

The second study is related to the efficacy of playing Xenubi in a digital version. Efficacy was measured as the difference in score for two multiple question tests.

Students from two classes from a public federal technical school took part in the evaluation. The classes comprised 25 students from an Informatics class [ages 14 to 16; 15 male, 10 female] and 22 students from a Management class [ages 14 to 16; 8 male, 14 female]. Both classes had the same teacher and the same curriculum on chemistry. A pretest-posttest experimental design was used.

The pretest was conducted right after the teacher presented the subject for each class. The pretest and the posttest were very similar, with 15 multiple choice questions,

as depicted by figure 5. Both tests were elaborated by the teacher and the second author of this paper. They were conducted one month apart from each other. Both were part of the quarterly evaluation of students, so they were committed to their performance on these tests.

EXCERPTS FROM THE PRETEST AND FROM THE POSTTEST

PRETEST	PROSTTEST
1) Which of the following element has the biggest atomic radius?	1) Which of the following element has the biggest atomic radius?
( )Mg ( )Sr (x)Ba	( )Na (x)Cs ( )Li

Figure 5. The 1st questions of the pretest and of the posttest.

We decided to use PCs because the class lasts 50 minutes – the time available to complete the test – and we did not have the required amount of cell phones. Besides, there would be problems if a given student was not familiar with the cell phone model [Perry et al, 2012], which would hardly be the case with a computer. All the tasks/commands of the cell phone version could be reproduced in the PC version. We made the printed card version as similar as possible to the PC version: the composition, colors, fonts and images were nearly the same. To simulate the position of the opponent's chemical element in the periodic table [as in the mobile and PC version of the game], we gave a periodic table along with the card game, and explained that both players should say the name of their elements out loud. The posttest protocol is described as follows:

✓ Introduce the research team and restate the research purposes [which were presented by the teacher on the pretest day, one month before];

✓ Explain how to play Xenubi. Since the goal was not to evaluate game mechanics, but how the game could contribute to students' learning, it was necessary for them to know how to play.

✓ Randomly assign students to play Xenubi with printed cards or on the PC.

✓ Take the students who were to play Xenubi on the PC to the computer laboratory, where they played for 50 minutes, returning to the classroom after this period.

✓ Ask the students that would play with cards to pair up with their colleagues, and give the pairs a deck with 12 printed cards and a periodic table.

✓ Conduct the posttest individually.

STATISTICAL ANALYSIS

Usability and player preferences evaluation

Thirteen variables were collected [table 1]. The impact of these variables, according to the type of school [public state or private] was analyzed via a logistic regression model (Hosmer, & Lemeshow, 2000).

Table 1
Description of the variables.

VARIABLE	MEANING	TYPE
Variable related to school type		
SCHOOL_TYPE	Private or public state school	Categorical binary [0-public; 1–private]
Variables related to interface comprehension		
SAW_TIPS	Student stated he saw the tips screen	Categorical binary [0-no; 1–yes]
SAW_WON	Student stated he knows who won the round	
SAW_PROP	Student stated he knows which properties were chosen.	
KNOWS_N_CARDS	Student stated he knows how many cards he has left.	
Variables related to game mechanics comprehension		
UNDST_MCHNICS	Student stated he understood how to play.	Categorical binary [0-no; 1–yes]
UNDST_STATUS	Student stated he knows who is winning and what it will take to loose/win the game.	
Variables related to player preferences. Measured on a two-point anchored Likkert scale.		
LIKES_CHEM	How much he likes Chemistry.	Continuous [scale from 0; "full disagreement" to 10; "full agreement"]
KNOW_PERIODIC	How much he thinks he knows the subject [periodic properties].	
GAME_FUN	How fun Xenubi is.	
GAME_PRETTY	How aesthetically pleasant Xenubi is.	
GAME_USEFUL	How useful Xenubi is.	
PLAY_OUTSIDE	Would he play Xenubi if it was not a school activity?	

The existence of difference among the variables related to game mechanics and interface comprehension was analyzed via a Chi-Square test for independence (Agresti, 2007) and via a *t*-test for comparing means (Mood et al., 1974). The significance level is α=0.05.

Table 2 presents the frequencies of the dichotomous variables by school type. From this set, the UNDST_MCHNICS variable is the most important, as it measures the overall feeling of the user about his understanding of the game.

Table 2
Binary profile responses.

Variables observed and (%)		School type		p-value
		Public state (N=23)	Private (N=20)	
SAW_TIPS	no	11 (47.8)	5 (25)	0.109
	yes	12 (52.2)	15 (75)	
SAW_WON	no	13 (56.5)	14 (70)	0.277
	yes	10 (43.5)	6 (30)	
SAW_PROP	no	8 (34.8)	6 (30)	0.498
	yes	15 (65.2)	14 (70)	
KNOWS_N_CARDS	no	4 (17.40	2 (10)	0.403
	yes	19 (82.6)	18 (90)	
UNDST_STATUS	no	6 (26.1)	4 (20)	0.459
	yes	17 (73.9)	16 (80)	
UNDST_MCHNICS	no	22 (95.7)	13 (65)	0.013*
	yes	1 (4.3)	7 (35)	

Significance level "Fisher exact test": ($p \leq 0.05$)*

Analyzing the percentages of the "public state" and "private" columns in table 2, it is possible to note that, except for the variable SAW_WON, most students from both schools understood the meaning of interface elements. For this variable, 56.5% of the students from the public sate school did not know who won the round, while 70% of the students from the private school did not know. Another negative finding was that few students understood how to play, with a significant difference according to school type: 95.7% of students from the public school did not understand how to play, while 65% in the private school did not understand how to play.

Next, table 3 presents the description of the variables related to player preferences, separated by school type. Analyzing each variable, only GAME_FUN, GAME_PRETTY and GAME_USEFUL [highlighted in Table 3] show significant differences in favor of the private school. This might point that students from the private school were more receptive to the game and therefore had more inclination to try to figure out how to play it.

Table 3
Mean and standard deviation of the variables related to player preferences.

Variables	School type		p-value
mean and (sd)	Public state (N=23)	Private (N=20)	
TIME_TOTAL	212.22 (65.35)	275.5 (133.86)	0.065
TIME_PER_ROUND	23.79 (8.41)	25.23 (14.06)	0.681
WIN_LOOSE_RATIO	0.54 (0.195)	0.49 (0.196)	0.384
LIKES_CHEM	6.02 (4.66)	5.15 (3.27)	0.478
KNOW_PERIODIC	4.3 (4.26)	5.87 (3.26)	0.187
GAME_FUN	*6.5 (4.65)*	*9.05 (3.61)*	*0.05**
GAME_PRETTY	*7.15 (5.12)*	*10.4 (2.74)*	*0.012**
GAME_USEFUL	*8.48 (5.21)*	*11.3 (2.27)*	*0.025**
PLAY_OUTSIDE	6.96 (5.57)	8.78 (4.11)	0.227

Significance level "T test": (p ≤ 0.05)*

Next, the impact of the variables related to interface understanding [table 1] and player preferences [table 2] were collectively analyzed in relation to the chances of success in discriminating school type. For this goal, models of logistic regression were defined. The initial model follows:

$$\text{Ln} \left[\frac{P(ESCOLA = 1)}{P(ESCOLA = 0)} \right] = \beta_0 + \beta_1 SAW\_TIPS + \beta_2 SAW\_WON + \beta_3 SAW\_PROP + \beta_4 KNOW\_CARDS + \quad [1]$$
$$\beta_5 UNDST\_STATUS + \beta_6 UNDST\_MCHNICS + \beta_7 LIKES\_CHEM + \beta_8 KNOWS\_PERIODIC + \beta_9 GAME\_FUN$$
$$+ \beta_{10} GAME\_PRETTY + \beta_{11} GAME\_USEFUL + \beta_{12}\ PLAY\_OUTSIDE$$

The right side expression of linear equation [1] represents the natural logarithm of the ratio of odds of belonging to the private school or to the public state school. If both sides of the equation [1] are powered to the *e* constant, the result is the equation shown in [2].

$$\left[\frac{P(ESCOLA = 1)}{P(ESCOLA = 0)} \right] = e\ (\beta 0 + \beta 1 SAW\_TIPS + \beta 2 SAW\_WON + \beta 3 SAW\_PROP + \beta 4 KNOW\_CARDS + \quad [2]$$
$$\beta 5 UNDST\_STATUS + \beta 6 UNDST\_MCHNICS + \beta 7 LIKES\_CHEM + \beta 8 KNOWS\_PERIODIC + \beta 9 GAME\_$$
$$FUN + \beta 10 GAME\_PRETTY + \beta 11 GAME\_USEFUL + \beta 12\ PLAY\_OUTSIDE)$$

The left side term of equation [2] represents the odds ratio of belonging to the private school taking as reference the public school. Through equation [2], the impact of the regression variables can be analyzed. Table 4 points 3 variables with significant

discriminative power on school type: LIKES_CHEM, GAME_PRETTY and UNDST_ MCHNICS [highlighted].

Table 4
Binary Logistic Regression Coefficients.

	Dependent Variable: SCHOOL						
Model	**Initial model [2], with all variables**		**Second model, with all 3 significant variables from the 1st model**		**Final model, with all significant variables from the 2nd model**		
Regressor variable	β	p-value	β	p-value	β	p-value	Exp(β)
Intercept	-7.646 (4.076)	0.061	-3.824 (1.936)	0.048*	-3.755 (1.664)	0.024*	
SAW_TIPS	3.532 (1.833)	0.054					
SAW_WON	-1.782 (1.374)	0.195					
SAW_PROP	0.433 (1.423)	0.76					
KNOW_CARDS	1.256 (1.819)	0.49					
UNDST_STATUS	1.1 (1.416)	0.437					
UNDST_MCHNICS	5.383 (2.279)	0.018*	3.414 (1.429)	0.017*	3.345 (1.336)	0.012*	28.36
LIKES_CHEM	-0.653 (0.322)	0.042*	-0.208 (0.113)	0.066			
KNOW_PERIODIC	0.112 (0.19)	0.556					
GAME_FUN	0.312 (0.29)	0.282					
GAME_PRETTY	0.811 (0.375)	0.031*	0.453 (0.189)	0.017*	0.3219 (0.146)	0.027*	1.38

Dependent Variable: SCHOOL			
Model	**Initial model [2], with all variables**	**Second model, with all 3 significant variables from the 1st model**	**Final model, with all significant variables from the 2nd model**
GAME_USEFUL	-0.464 0.166 (0.335)		
PLAY_OUTSIDE	0.057 0.803 (0.229)		

Maximum likelihood method (standard errors in brackets)
Significance level "Wald test": ($p \leq 0.05$)*
Goodness of fit Tests (Pearson, Deviance, Hosmer-Lemeshow) – all models have significant fit

The results for the first model confirm the previous analysis of the impact of each variable regardless of school type. With the exception of LIKES_CHEM, UNDST_ MCHNICS and GAME_PRETTY showed significant difference in the Chi-Squared [table 2] and T-Test [table 3]. The variables GAME_FUN and GAME_USEFUL, though significantly different in the individual analysis [table 3] had no impact on school discrimination.

The second model presents the estimate of the impact considering only the significant variables of the initial model [the highlighted variables on table 4]. We note that the impact of LIKES_CHEM is not significant in this improved model. The final model presents the coefficients of the adjusted equation after removing the LIKES_CHEM variable. Both remaining variables are still significant [UNDST_MCHNICS and GAME_PRETTY]. The $Exp(\beta_{10}) = 1.38$ on the final model value relative to GAME_PRETTY indicates that at each increment on a scale of "game prettiness", the chances of the student belonging to the private school increases by 38% [the values of the other variables should remain unchanged]. Again, on the final model, the $Exp(\beta_6) = 28.36$ value indicates that, if the student understands game mechanics, the chance of belonging to the private school is approximately of 28 against 1. Table 4 shows this was the most important variable of the final model: it has the smallest *p-value*, suggesting it is the most important variable to discriminate the school type.

Some last words about this first experiment, it was divided into two parts: the first sought to evaluate usability as comprehension of interface elements, instead of using more traditional measurements such as task completion times and error rates, while the second focused on user preferences. There were two reasons for this choice. First, since Xenubi is a learning game, we thought that task related measurements would be of lesser value, since the tasks were, per se, very simple. Besides, completing a task might not mean that the player made the best choice [based on his knowledge of the subject matter] or even

that he had a clear goal in mind [motivated by understanding the mechanics of the game]. Data on task completion and success rates [TIME_TOTAL, TIME_PER_ROUND and WIN_LOOSE_RATIO] were not significant in any of the individual analysis or models [see tables 2 and 3 for p values]. For these reasons, we decided to focus on the interface element understanding. Besides, knowing what a given interface element represents does not mean the player knows its role in the game. This understanding is summarized by the variable UNDST_MCHNICS, the most important variable of this study, as all the models confirm. The lack of specific guidelines for evaluating learning games should also be mentioned: the game heuristics sets found in the literature review were considered too broad.

The second reason the experiment on usability was divided into two parts was Hassenzahl's (2001) model of appeal, where ergonomic and hedonic aspects would have the same importance. However, we chose not to use any existing models of emotion evaluation because when it comes to learning, motivation is regarded as an important factor (Reeve, 1997). We designed a questionnaire asking whether the player: likes chemistry; thinks the interface is pretty; finds the game fun; thinks the game is useful and thinks he knows the subject matter. We considered that maybe players who thought the game was good [pretty, useful, fun] and who liked Chemistry might be the ones who would better understand how to play. We conducted the same test with students from a private school and from a public state school. The reasoning behind this choice is that we expected that private school students would understand better how to play the game because they are more familiar with smartphones and mobile technology.

This hypothesis was confirmed by the logistic regression models [table 4], which pointed to the variables UNDST_MCHNICS, GAME_PRETTY and LIKES_CHEM as the variables with highest discriminative power – they are the best to tell the odds of a random student belonging to the public or private school. Regarding the individual analysis [Chi-Square test, table 2 and T-test, table 3], these were among the variables which had significant difference. This partially confirms the regression models – we say "partially" because not all variables that had significant difference in the individual analysis were significant in the logistic regression models. These results could mean that (1) there were important effects that were not measured and (2) that we might have not measured with enough subjects.

Learning outcomes evaluation

The goal of the second study was to test the hypothesis of the difference in student performance at the posttest according to the game version [card or PC]. Students from technical courses [Management and Informatics] from the same public federal school, with the same teacher, played Xenubi with printed cards and on a PC. A multiple regression model (Kutner, 2004) was used, with a significance level of $\alpha=0.05$.

The dispersion graph shown in figure 6 correlates pretest and posttest data – PC data is in blue; printed card data in red. The multiple regression models are the following:

POS(cards) = β_0 + β_1 PRE(cards) + ERROR   [3]
POS(PC) = β_0 + β_1 PRE(PC) + ERROR   [4]
POS=β_0 + β_1 *PRE*+ β_2 *GAME*+ β_3 *PRE*GAME* + ERROR   [5]

On these 3 models ERROR is a random normally distributed variable, with average of zero [0] and constant variance along the domain of the regressor variable PRE [homoscedastic errors]. Model [3] describes a linear relation between the response variable [POST] and the regressor variable [PRE], considering that the groups played Xenubi with printed cards. Model [4] describes the same relation for the group who played Xenubi on the PC. Model [5] is the combination of models 3 and 4 using the dummy variable GAME, where GAME = 0 indicates the group that played with printed cards and GAME = 1 indicates the group that played on the computer. The estimate for the $\beta3$ coefficient represents the learning rate on both methods. The graph on figure 6 suggests that the students who played on the PC had a positive variation on performance – the blue line has a steeper incline than the red line.
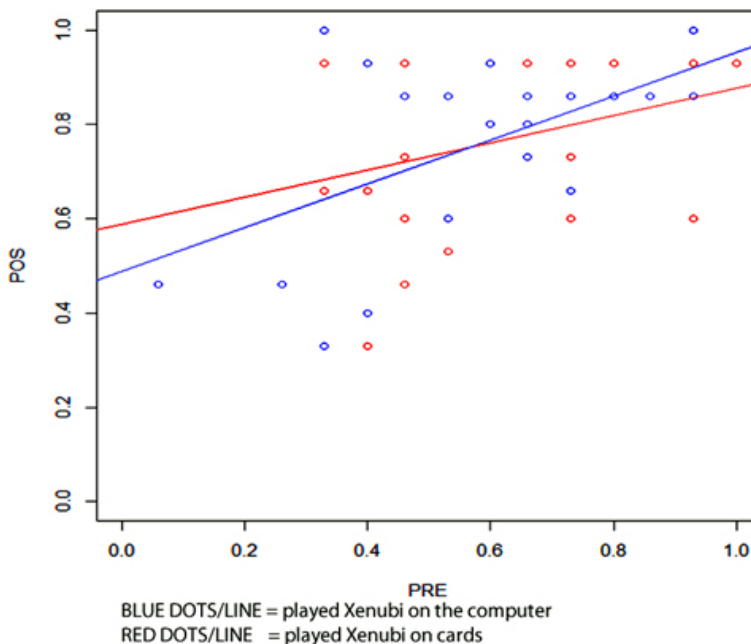
CORRELATION BETWEEN PRETEST AND POSTTEST SCORES

BLUE DOTS/LINE = played Xenubi on the computer
RED DOTS/LINE = played Xenubi on cards

Figure 6. Correlation between performance at pretest and posttest of the students who played Xenubi with printed cards [red] and on the PC [blue].

In figure 6, it is possible to see that the variation of dots around their respective lines is inverse [the lower the value on the x axis, the higher the value on the y axis], violating the premise of error homocedastivity. This is due to the fact that dependent variable POS represents the amount of correct answers in the [0,1] interval; for better students, the improvement rate tends to be lower. To work around this issue and make variations uniform along the variable PRE, equations [3], [4] and [5] were multiplied by the squared root of the PRE value, therefore assuming the following structure.

$$POS(cards) * \sqrt{PRE} = \beta_0 * \sqrt{PRE} + \beta_1 \, PRE(cards) * \sqrt{PRE} + ERROR * \sqrt{PRE} \qquad [6]$$

$$POS(PC) * \sqrt{PRE} = \beta_0 * \sqrt{PRE} + \beta_1 \, PRE(PC) * \sqrt{PRE} + ERROR * \sqrt{PRE} \qquad [7]$$

$$POS * \sqrt{PRE} = \beta_0 * FV + \beta_1 \, PRE * \sqrt{PRE} + \beta_2 \, GAME * \sqrt{PRE} + \beta_3 \, PRE*GAME * \sqrt{PRE} + ERROR' \qquad [8]$$

In these 3 models, $ERROR' = ERROR * \sqrt{PRE}$ represents the random variable ERROR described in models [3], [4] and [5] as the adjustment for homoscedastic variance.

Table 5 presents the estimated regression values for models [6], [7] and [8]. It is important to note that the corrected equations [models 6, 7 and 8] present a substantial data adjustment [see the compared R^2 for each model], justifying the refinement.

Table 5
Normal regression coefficients.

	Dependent variable: POST					
Model	**Printed cards [3and6]**		**PC [4and7]**		**Mixed [5 and 8]**	
Regressor	Non-corrected	Corrected for homoscedastic variances	Non-corrected	Corrected for homoscedastic variances	Non-corrected	Corrected for homoscedastic variances
Intercept	0.5874* (0.1085)	0.5753* (0.1126)	0.4868* (0.1021)	0.5344* (0.1153)	0.5874* (0.1102)	0.5753* (0.1092)
PRE	0.2887 (0.1653)	0.3064 (0.1582)	0.4662* (0.1653)	0.3941 (0.1682)	0.2887 (0.1680)	0.30637 (0.1534)
GAME					-0.1006 (0.1490)	-0.0409 (0.1622)
*PRE*GAME*					0.1774 (0.2325)	0.0877 (0.2327)
R^2	0.1127	0.9675	0.2846	0.9705	0.2039	0.9688

Ordinary last square method (standard errors in brackets). Significance level "Wald test": ($p \leq 0.05$)*

On the individual non-adjusted equations of each group it is possible to note that the estimated learning rate of the group that played with printed cards is not statistically

significant. On the other hand, the learning rates of the group that played on the PC is higher [0.4662, highlighted, versus 0.2887] and statistically significant.

In the adjusted equations, the learning rate of the group that played on the PC is also higher, although not statistically significant [0.3941 versus 0.3064]. However, despite there being a relatively steep difference in the improvement rates [0.1774 in the non-adjusted model and 0. 0.0877 in the adjusted model] for both groups, for the equation of model [8] – which considers pretest values for both game types – no statistical significance was found. This might be due to a need for a bigger sample size: since model [8] has more parameters, to maintain the same accuracy of models with less parameters more data is needed. In other words, the amount of information that is possible to extract from more complex models demands more data.

CONCLUSIONS

In this paper we presented the process of designing Xenubi, a game [about periodic properties of the chemical elements] designed for smartphones. The design approach was to make it a casual game that could be played anywhere, requiring little effort to learn and to advance along the game levels. On Xenubi, the player has a set of 6 cards, each representing a chemical element. The player has to choose from 6 chemical properties the one he thinks his card's element has a higher value. To do that, the player should rely on his knowledge of the periodic properties, since he can see the location of the computer's chemical element on the screen.

Xenubi was designed by an interdisciplinary team, comprising two designers who also programmed the game (one with a PhD and one undergraduate) and a chemistry professor. While designing Xenubi, we conducted what is called "discount usability inspections", in our case we realized a user test with pre-medical students. This pilot study was important to assure the test protocol was appropriate and the questions were easy to understand.

Two experiments were designed: one to measure usability and user preferences aspects and the other to measure learning outcomes. The first experiment sought the usability of the educational game Xenubi. The evaluation involved both the understanding of the interface elements and the user's own preferences. The second experiment tested the hypothesis of difference in performances of students who played Xenubi with printed cards and on the computer.

With this goal, a pretest-posttest experiment was conducted with two different groups. We are aware that this design is not as powerful as the Solomon four group design [Solomon four group design considers the effect of a group taking only the pretest], but that was our choice because it requires less test subjects. We analyzed the posttest as the respondent variable of linear models: models 3, 4 and 5 represent the model with the actual data set, while models 6, 7 and 8 are the models with data corrected for homocedastivity. Table 5 summarizes the results for these 6 models. Our results point that, when data is

not corrected for homocedastivity [models 3, 4 and 5], the group that played Xenubi on the computer had a significantly higher performance gain than the group that played with printed cards. However, when the models are corrected – a necessary step, given the nature of the data – the difference is no longer significant.

ACKNOWLEDGEMENT

REFERENCES

Agresti, A. (2007). *An introduction to categorical data analysis*. Wiley-Interscience.

Anshari, M., Almunawar, M.N., Shahrill, M., Wicaksono, D.K., & Huda, M. (2017). Smartphones usage in the classrooms: Learning aid or inteferecence? *Education and Information Technologies, 22*(6), 3063-3079.

Ariga, T. & Watanabe, T. (2008). Teaching materials to enhance the visual expression of Web pages for students not in art or design majors. *Computers & Education, 51*(2), 815-828.

Boehner, K. et al. (2007). How emotion is made and measured. *International Journal of Human-Computer Studies, 65*(4), 275-291.

Campbell, D. T. & Stanley, J. C, (1963). Experimental and Quasi-Experimental Designs for Research on Teaching. In N. L. Gage (ed.), *Handbook of Research on Teaching*. Chicago: Rand McNally.

Eichler, M. L., Perry, G. T., Lucchesi, I. L., & Melendez, T. T. (2017). Mobile Game-Based Learning in STEM Subjects. In M. Khosrow-Pour (ed.), *Encyclopedia of Information Science and Technology, Fourth Edition* (p.6376-6387). Hershey (PA): IGI Global.

Franco-Mariscal, A.J., Oliva-Martínez, J.M., & Bernal-Márquez, S.B. (2012). An educational card game for learning families of chemical elements. *Journal of Chemical Education, 89*, 1044-1046.

Franco-Mariscal, A.J., Oliva-Martínez, J.M., Blanco-López, A., & España-Ramos, E. (2016). A game-based approach to learning the idea of chemical elements and their periodic classification. *Journal of Chemical Education, 93*, 1173-11190.

Godoi, T. A., Oliveira, H. P., & Codognoto, L. (2010). Tabela periódica – um super trunfo para alunos do ensino fundamental e médio. *Química Nova na Escola, 32*(1), 22-25.

Hassenzahl, M. (2001) The Effect of Perceived Hedonic Quality on Product Appealingness. *International Journal of Human-Computer Interaction, 13*(4), 481-499.

Hinostroza, J. E. & Mellar, H. (2001). Pedagogy Embedded in Educational Software Design: Report of a Case Study. *Computers & Education, 37*(1) 27-40.

Hornbæk, K. (2006). Current practice in measuring usability: Challenges to usability studies and research. *International Journal of Human-Computer Studies, 64*(2), 79-102.

Hosmer, D. W. & Lemeshow, S. (2000). *Applied logistic regression*. John Wiley & Sons.

Huizenga, J.C., ten Dam, G.T.M., Voogt, J.M., & Admiraal, W.F. (2017). Teacher perceptions of the value of game-based learning in secondary education. *Computers & Education, 110*, 105-115.

ISO 9241-11: *Ergonomic requirements for office work with visual display terminals (VDTs)* – Part 11: Guidance on usability.

Jones, I. (2008). Storyboarding: A method for bootstrapping the design of computer-based educational tasks. *Computers & Education, 51*(3), 1353-1364.

Kaikkonen, A., Kallio, T., & Kankainen, A. (2005). Usability testing of mobile applications: Comparison between laboratory and field testing. *Journal Usability Studies, 1*(1), p.4-16.

Kirmani, S. (2008). Heuristic Evaluation Quality Score (HEQS): Defining Heuristic Expertise. *Journal Usability Studies, 4*(1), 49-59.

Krippendorff, K. (2012). *Content Analysis: An Introduction to Its Methodology*. Thousand Oaks: Sage Books.

Kutner, M., Neter, J., & Nachtsheim, C. (2004). *Applied linear statistical models*. McGraw-Hill/Irwin.

Maiti, A. (2010). Adobe Flash Lite–Based Online Laboratory in Mobile Phone. *International Journal of Interactive Mobile Technologies (iJIM), 4*(4), 45-48.

Malone, T.W. (1980a). What makes things fun to learn? heuristics for designing instructional computer games. *Challenge, 162*, 162-169.

Martí-Centelles, V. & Rubio-Magnieto, J. (2014) ChemMend: A card game to introduce and explore the periodic table while engaging students' interest. *Journal of Chemical Education, 91*, 868-871.

McCarthy, J. & Wright, P. (2007) *Technology as Experience*. The MIT Press.

Molich, R. & Nielsen, J. (1990). Improving a human-computer dialogue. *Communications of the ACM, 33*(3), 338-348.

Mood, A.M., Graybill, F. A., & Boes, D. C. (1974). *Introduction to the theory of statistics*. Mcgraw-Hill.

Newell, A. & Simon, H. A. (1972). *Human Problem Solving*. Englewood: Prentice Hall.

Nielsen, J. & Mack, R. (1994). *Usability inspection methods*. New York: John Wiley & Sons.

Nielsen, J. (1995) Applying discount usability engineering. *IEEE Software, 12*(1), 98-100.

Oulasvirta, A. & Nyyssönen, T. (2009). Flexible hardware configurations for studying mobile usability. *Journal Usability Studies, 4*(2), 93-105

Perry, G. T. & Eichler, M. L. (2015). Science Learning Games for Mobile Platforms. In: Z. Yan. (Org.). *Encyclopedia of Mobile Phone Behavior* (p.562-574). Hershey (PA): IGI Global.

Perry, G. T., Kulpa, C., & Eichler, M. (2012). Lessons from an Educational Game Usability Evaluation. *International Journal of Interactive Mobile Technologies (iJIM), 6*, 23-28.

Piccard, R. (1997). *Affective computing*. MIT Press.

Rastegarpour, H. & Marashi, P. (2012). The effect of card games and computer games on learning of chemistry concepts. *Procedia – Social and Behavioral Sciences, 31*, 597-601.

Reeve, J. (1997). *Understanding Motivation and Emotion*. Harcourt Brace.

Reeves, T. & McKenney, S. (2015). Design-based research. In J. Spector (Ed.), The SAGE encyclopedia of educational technology (pp.189-191). Thousand Oaks, CA: SAGE Publications

Russell, J. (1999). Using games to teach chemistry: an annotated bibliography. *Journal of Chemical Education, 76*, 481-484.

Scerri, E. R. (2007). *The Periodic Table: Its Story and Its Significance*. New York: OUP.

Sharp, H., Rogers, Y., & Preece, J. (2007). *Interaction Design: Beyond Human-Computer Interaction*. Wiley.

Stanton, N.; Young, M. S. (2003). *A Guide to Methodology In Ergonomics: Designing For Human Use*. Taylor & Francis.

Stark, A. & Riebeck, M., (2008). A New Approach to Analyze Human- Mobile Computer Interaction. *Journal Usability Studies, 3*(2), 90-98

Tress, B., Tress, G., & Fry, G. (2005). Defining integrative research concepts and process of knowledge production. In: Tress, B., Tress, G., Fry, G., Opdam, P. (eds.). *From landscape research to landscape planning: Aspects of integration, education and application* (pp.13-26). Heidelberg: Springer.

Williams, A.J. & Pence, H.E. (2011). Smart Phones, a powerful tool in the chemistry classroom. *Journal of Chemical Education, 88*, 683-686.

Winograd, T. (1996). *Bringing Design to Software*. Addison-Wesley Professional.

Zaharias, P. & Poylymenakou, A. (2009). Developing a Usability Evaluation Method for e-Learning Applications: Beyond Functional Usability. *International Journal of Human-Computer Interaction, 25*, 1, 75-98.