MEEGA+KIDS: A Model for the Evaluation of Games for Computing Education in Secondary School

Christiane Gresse von Wangenheim, UFSC, c.wangenheim@ufsc.br
Giani Petri, UFSM, giani.petri@ufsm.br
Adriano Ferreti Borgatto, UFSC, adriano.borgatto@ufsc.br

Abstract. *Teaching computing in schools has become a worldwide trend. This is operationalized through diverse instructional methods, including educational games. As it is essential to systematically evaluate the quality of such games for computing education in school, the objective of this article is to present the design and evaluation of the MEEGA+KIDS model, a customization of the prominent method MEEGA+, a reliable and valid method to evaluate games. It has been tailored to the specific target audience (secondary school) through a participatory design approach by decomposing evaluation goals into measures and defining a standardized measurement instrument in the form of a self-assessment questionnaire. Results of a reliability and validity analysis of the model, based on a sample of 90 students, show evidence of its reliability (Cronbach's alpha α=.882) and a first indication of its validity. The MEEGA+KIDS model can, thus, support game creators, instructors and researchers during the game design process and contribute to their improvement and effective adoption in practice.*

Keywords: *Game; Model; Evaluation; Computing Education; Secondary School*

MEEGA+KIDS: Um Modelo para a Avaliação de Jogos para o Ensino de Computação na Educação Básica

Resumo. *O ensino de computação nas escolas se tornou uma tendência mundial. Isso é operacionalizado por diversos métodos instrucionais, incluindo jogos educacionais. Assim, é essencial avaliar sistematicamente a qualidade desses jogos usados para o ensino de computação nas escolas. O objetivo deste artigo é apresentar o design e a avaliação do modelo MEEGA+KIDS. O modelo é uma customização do MEEGA+, um método confiável e válido para avaliar jogos educacionais. Ele foi adaptado ao público-alvo específico (educação básica) por meio de uma abordagem de design participativo, decompondo as metas de avaliação em medidas e definindo um instrumento de medição padronizado na forma de um questionário de auto-avaliação. Os resultados de uma análise de confiabilidade e validade com base em uma amostra de 90 alunos, mostram evidências de sua confiabilidade (alfa de Cronbach α = 0,882) e uma primeira indicação de sua validade. O modelo MEEGA+KIDS pode, assim, apoiar criadores de jogos, instrutores e pesquisadores no processo de design de jogos e contribuir para sua melhoria e adoção na prática.*

Palavras–Chave: *Jogo; Modelo; Avaliação; Ensino de Computação; Educação Básica.*

1. Introduction

Teaching computing in K-12 has become a worldwide trend to popularize computing competencies as well as increase the awareness and interest of the students (Garneli et al., 2015). In this context, diverse instructional strategies are adopted, including also game-based learning by learning through gameplay (Kazimoglu et al., 2012). Such games typically aim at teaching basic concepts of algorithms and programming, in which players execute actions (representing moves, etc.) to command ("program") a game character to achieve a goal. There exist diverse digital games for teaching computing in K-12, such as Lightbot (https://lightbot.com/flash.html), Programming your Robot (Kazimoglu et al., 2012), CodeCombat (https://br.codecombat.com), etc. Also, non-digital games are adopted as unplugged activities (Battistella & Gresse von Wangenheim, 2016). This

includes board games such as SplashCode (Gresse von Wangenheim et al., 2019), Robo Rally (https://en.wikipedia.org/wiki/RoboRally), Code Island Monkey (http://codemonkeyplanet.com) as well as card games such as FlexiCard (http://www.computacional.com.br/index.php#FlexiCard) or Logirunner (Casarotto et al., 2018), among others.

Taking into consideration the growing number of games for teaching computing, it is important to obtain evidence on the expected benefits as a basis for their systematic selection, adoption, and improvement (Decker et al., 2016). Although games are believed to be motivational and educationally effective, the empirical evidence to support this assumption is still limited and contradictory, particularly regarding the effectiveness of games for concrete educational purposes, given that prior studies have focused more on motivational aspects than on curricular content aspects and core academic benefits (Papastergiou, 2009). Specifically, concerning computing education, empirical research is scarce (Kazimoglu et al., 2012). Very few studies demonstrate how gameplay can be associated with computing competencies and how the education of introductory programming can be supported by playing games. Yet, these studies typically target higher education (Ibrahim et al., 2011; Petri & Gresse von Wangenheim, 2017) and not K-12. A possible reason may be the lack of evaluation models focusing on this educational stage. Existing models for the evaluation of educational games, such as MEEGA+ (Petri & Gresse von Wangenheim, 2019; Petri et al., 2019) focus on higher education only (Petri & Gresse von Wangenheim, 2016). Thus, these models may not provide adequate support for evaluating educational games on the school level, taking into consideration the specific characteristics of the target audience and context.

Observing this gap, we have tailored the MEEGA+ method for the evaluation of games for teaching computing on the secondary school level. The MEEGA+ method has been chosen as it is currently one of the most prominent evaluation methods that has been systematically developed and widely validated through a series of case studies (Petri & Gresse von Wangenheim, 2019). MEEGA+ provides systematic support to evaluate games in terms of player experience and usability. It has been systematically developed by decomposing evaluation goals into measures and defining a standardized measurement instrument in the form of a self-assessment questionnaire. Thus, the MEEGA+ measurement instrument has been adapted by revising the wording in accordance with the specific target audience (secondary school students) through a participatory design approach. As a result, this article presents the design and evaluation of the MEEGA+KIDS model (in English and Brazilian Portuguese). We expect that the MEEGA+KIDS model provides game creators, instructors, and researchers with a measurement instrument to evaluate the quality of games used in secondary school and, thus, contribute to their improvement and effective adoption in practice.

2. Related Work

While game-based learning has become popular also for computing education, investigations on its potential educational benefits are still scarce (Kim & Ifenthaler, 2019) and focus mostly on higher education (Petri & Gresse von Wangenheim, 2017). Very few studies demonstrate how gameplay can be associated with the learning of computing competencies (Kazimoglu et al., 2012; Ibrahim et al., 2011). Chaffin et al. (2009) studied the students' ability to write algorithms to generate data structures as part of the gameplay. They measured data through a multiple-choice pre- and post-test as well as a post-survey concerning game qualities. Although lacking empirical evidence, their initial feedback suggests that students who played the game were better able to visualize how data structures work than the students who did not play the game.

Vlahu-Gjorgievska et al. (2018) evaluated instructional alternatives applying questionnaires concerning cooperative game-based learning with primary school students measuring the students' attitudes toward learning to code with different tools. Initial results show that the students are very keen to adopt the technology offered by programming tools and games enabling them to be active participants.

Other studies evaluated the learning behaviors of students in addition to their motivation in learning programming. Liu, Cheng & Huang (2011) created a simulation game and analyzed the feedback and problem-solving behaviors of students. During a case study with higher education students they collected data on the perceived learning using the learning experience survey proposed by Pearce, Ainley, & Howard (2005) as well as motivation adopting the Motivated Strategies for Learning Questionnaire (MSLQ), in addition to activity logs. They found that students motivated by the game frequently used analytical strategies such as critical thinking to discover available solutions, but, on the other hand, when felt bored with the game, solved problems only at a superficial level. Papastergiou (2009) evaluated the learning effectiveness and motivational appeal of a computer game for learning computer memory concepts. Results of an experimental study with pre- and post-test with multiple choice questions and questions concerning the various attributes of the application and their perception with high school students indicated that the gaming approach was both, more effective in promoting students' knowledge of computer memory concepts and more motivational than the non-gaming approach.

Casarotto et al. (2018) evaluated the unplugged game Logirunner using MEEGA+ as part of a case study with higher education students. They identified as strengths of the game-based learning approach the social interaction between students, the learning of the concept of algorithms as well as the fun provided by the game.

In summary, these results indicate an overall positive game experience of various games. However, currently, most evaluations use a simple research design in which, typically, the game is used, and afterward subjective feedback is collected via questionnaires from the learners sometimes in combination with multiple-choice tests. Most of the studies adopt non-standardized questionnaires, which have not rigorously analyzed concerning their reliability or validity (e.g., Vlahu-Gjorgievska et al. (2018)). Another shortcoming is that several evaluations were carried out with participants different from the main target audience (e.g., Casarotto et al. (2018)), involving only students from higher education and/or other participants not representing K-12 students, leaving the generalization of the results to younger students for whom the games are designed questionable. This shows that there is a need for more rigorous evaluations as well as methodological support to improve the games as well as to support decisions on when or how to include them in instructional units.

3. Research Method

The MEEGA+KIDS model has been developed based on the MEEGA+ method (Petri & Gresse von Wangenheim, 2019) conducting a multi-method research (Figure 1).
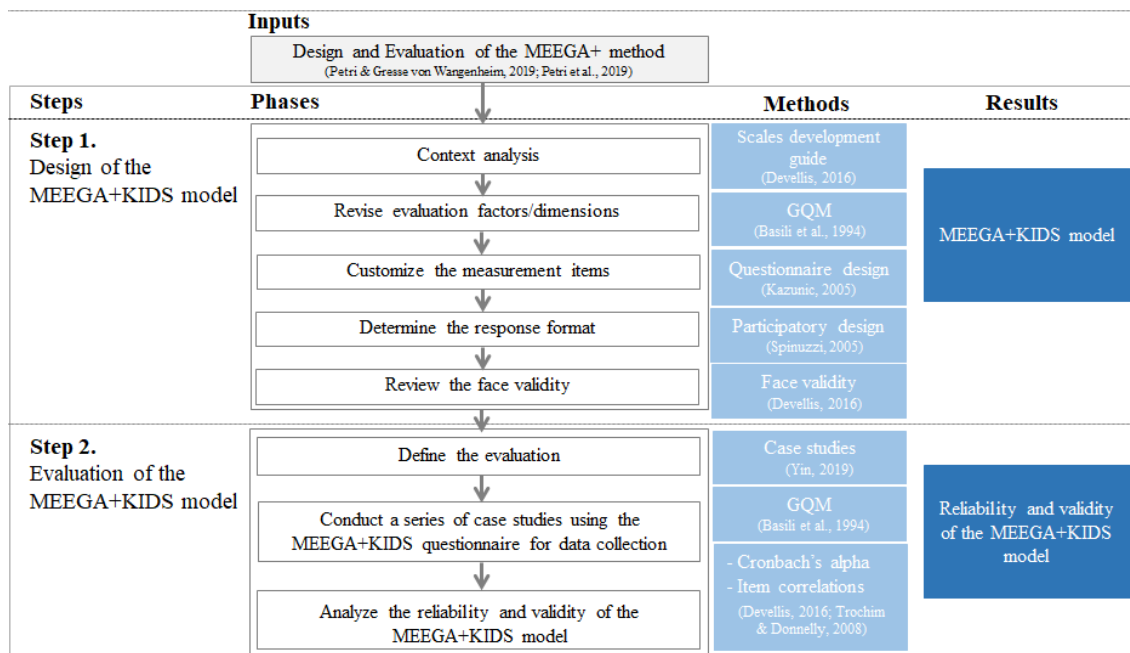
Figure 1. Research method

Step 1. Design of the MEEGA+KIDS model. Based on a context analysis of the expected application of the model in computing education in secondary school, we revised the analysis questions and quality dimensions. We customized the measurement instrument of the MEEGA+ model (Petri & Gresse von Wangenheim, 2019) by adopting a participatory design approach (Spinuzzi, 2005), involving representatives of the target audience in the adaptation of the measurement items of the self-assessment questionnaire to assure a formulation of the items understandable by the target audience.

Step 2. Evaluation of the MEEGA+KIDS model. To evaluate the MEEGA+KIDS model, we conducted an evaluation study through a series of case studies (Wohlin et al., 2012) applying the model in computing education in secondary schools. The evaluation study aims to analyze the reliability and validity of the MEEGA+KIDS. The study objective is defined using the GQM goal template (Basili et al., 1994) and decomposed into analysis questions. A series of case studies applying a game for computing education in secondary school is conducted in a one-shot post-test only design. We use a non-probability sampling technique in each case study applying the convenience sampling method (Wohlin et al., 2012), in which our sample is composed of students of secondary school. Data collected in each case study are pooled in a single sample to answer our analysis questions in terms of reliability and construct validity following the definition of Trochim and Donnelly (2008).

4. Design of the MEEGA+KIDS Model

The objective of the MEEGA+KIDS model is to evaluate the quality of educational games in terms of usability and player experience from the students' perspective in the context of computing education in secondary school.

Context analysis. The target audience for this game is secondary school students, aged typically between 8 and 14 years. Typically, most students at this age already have knowledge and skills in using IT for social networks and entertainment. The teaching of computing in K-12 is being introduced in schools as part of regular classes or extracurricular activities either by focusing explicitly on teaching computing or in a multidisciplinary way integrated into other disciplines. Yet, as computing is still not part

of the basic curriculum in several countries, these instructional units are often limited to a short duration. Classes typically take place in computer labs, but may not have sufficient computers for each student, requiring them to work in teams or groups. Specifically, in the context of public schools, there may be few resources available for the acquisition of consumables such as paper, etc.

Revision and customization of the MEEGA+ model. Based on the MEEGA+ model (Petri & Gresse von Wangenheim, 2019), the MEEGA+KIDS model is decomposed into two quality factors and their dimensions (Table 1). In this context, usability is defined as the degree to which a product (educational game) can be used by specified users (students) to achieve specified goals with effectiveness and efficiency in a specified context of use (computing education), being composed of the following dimensions: aesthetics, learnability, operability, and accessibility. Player experience is a quality factor that covers a deep involvement of the student in the gaming task, including its learning improvement, feelings, pleasures, and interactions with the game, environment, and other players (Petri & Gresse von Wangenheim, 2019).

Definition of the research design. We maintained the research design defined in the MEEGA+ method, a case study design, which the study is conducted as a one-shot post-test only design, in which the case study begins with the application of the treatment (educational game) and then the data are collected through a standardized measurement instrument (questionnaire). The questionnaire is answered by the students (self-assessment) to collect data on their perceptions about the game and through a set of multiple-choice questions measuring the knowledge covered by the game.

Definition of the MEEGA+KIDS measurement instrument. Customizing the MEEGA+ questionnaire by adopting a participatory design methodology (Spinuzzi, 2005), the MEEGA+KIDS questionnaire has been proposed as presented in Table 1. Due to the shorter attention span of students at this education level less important items that are partly covered by other items have been excluded as part of the customization to shorten the questionnaire. We also changed the wording in some cases using a language more closely to the one used by the target audience to facilitate understanding. As a result, we created the questionnaire in English and Brazilian Portuguese.

Table 1. MEEGA+KIDS decomposition and questionnaire items

Decomposition of the MEEGA+KIDS		MEEGA+ questionnaire	MEEGA+KIDS questionnaire		
Quality factor	Dimension	Description	No.	English version	Brazilian Portuguese version
Usability	Aesthetics	The game design is attractive (interface, graphics, cards, boards, etc.).	1	The game design is attractive (game board, cards, etc.).	O design do jogo é atraente (tabuleiro, cartas, etc.).
		The text font and colors are well blended and consistent.	2	The font and colors of the game match.	As cores e fontes do material do jogo combinam.
	Learnability	I needed to learn a few things before I could play the game.		--	--
		Learning to play this game was easy for me.	3	Learning to play this game was easy for me.	Aprender a jogar este jogo foi fácil para mim.
		I think that most people would learn to play this game very quickly.		--	--
	Operability	I think that the game is easy to play.	4	I think that the game is easy to play.	Eu considero que o jogo é fácil de jogar.
		The game rules are clear and easy to understand.	5	The game rules are clear and easy to understand.	As regras do jogo são claras e compreensíveis.
	Accessibility	The fonts (size and style) used in the game are easy to read.	6	The size and style of fonts used in the game are easy to read.	O tamanho e estilo de letras utilizadas no jogo são legíveis.
		The colors used in the game are meaningful.	7	The colors used in the game are meaningful.	As cores utilizadas no jogo são compreensíveis.
Player Experience	Confidence	The contents and structure helped me to become confident that I would learn with this game.	8	The organization of the content helped me to become confident that I would learn with this game.	A organização do conteúdo me ajudou a estar confiante de que eu iria aprender com este jogo.
	Challenge	This game is appropriately challenging for me.	9	This game is appropriately challenging for me.	Este jogo é desafiador suficiente para mim.
		The game provides new challenges (offers new obstacles, situations, or variations) at an appropriate pace.	10	The game provides new challenges (offers new obstacles, situations, or variations) at an appropriate pace.	O jogo oferece novos desafios (novos obstáculos, situações ou variações) com um ritmo adequado.

	Satisfaction	The game does not become monotonous as it progresses (repetitive or boring tasks).	11	The game does not become monotonous as it progresses (repetitive or boring tasks).	O jogo não se torna monótono nas suas tarefas (repetitivo ou com tarefas chatas).
		Completing the game tasks gave me a satisfying feeling of accomplishment.	12	Completing the game tasks gave me a satisfying feeling of accomplishment.	Completar as tarefas do jogo me deu um sentimento de satisfação.
		It is due to my personal effort that I managed to advance in the game.	13	It is due to my personal effort that I managed to advance in the game.	É devido ao meu esforço pessoal que eu consigo avançar no jogo.
		I feel satisfied with the things that I learned from the game.	14	I feel satisfied with the things that I learned from the game.	Me sinto satisfeito com as coisas que aprendi no jogo.
		I would recommend this game to my colleagues.	15	I would recommend this game to my colleagues.	Eu recomendaria este jogo para meus colegas.
	Social Interaction	I was able to interact with other players during the game.	16	I was able to interact with other people during the game.	Eu pude interagir com outras pessoas durante o jogo.
		The game promotes cooperation and/or competition among the players.	17	The game promotes cooperation and/or competition among the players.	O jogo promove momentos de cooperação e/ou competição entre os jogadores.
		I felt good interacting with other players during the game.	18	I felt good interacting with other players during the game.	Eu me senti bem interagindo com outras pessoas durante o jogo.
	Fun	I had fun with the game.	19	I had fun with the game.	Eu me diverti com o jogo.
		Something happened during the game (game elements, competition, etc.) which made me smile.	20	Something happened during the game that made me smile.	Aconteceu alguma situação durante o jogo que me fez sorrir.
	Focused Attention	There was something interesting at the beginning of the game that captured my attention.	21	There was something interesting at the beginning of the game that captured my attention.	Houve algo interessante no início do jogo que capturou minha atenção.
		I was so involved in my gaming task that I lost track of time.	22	I was so involved in the game that I lost track of time.	Eu estava tão envolvido no jogo que eu perdi a noção do tempo.
		I forgot about my immediate surroundings while playing this game.		--	--
	Relevance	The game contents are relevant to my interests.	23	The game's content is of my interest.	O conteúdo do jogo me interessa.
		It is clear to me how the contents of the game are related to the course.	24	It is clear to me how the contents of the game are related to the course.	É claro para mim como o conteúdo do jogo está relacionado com a disciplina.
		This game is an adequate teaching method for this course.	25	I learned content of the course with this game.	Eu aprendi conteúdo da disciplina com este jogo.
		I prefer learning with this game to learning through other ways (e.g. other teaching methods).	26	I prefer learning with this game than through other ways (e.g. expositive lectures given by the teacher).	Eu prefiro aprender com este jogo do que de outra forma (p.ex. aula no quadro pelo professor).
	Learning	The game contributed to my learning in this course.		Descriptive question "What did you learn playing the game?""	Pergunta descritiva" O que você aprendeu jogando esse jogo?"
		The game allowed for efficient learning compared with other activities in the course.		--	--

Response format. We maintained the response format as defined in the MEEGA+ method, adopting a 5-point Likert scale with response alternatives ranging from strongly disagree to strongly agree. As part of a self-assessment of the perception of the learning effect of the game we included also the descriptive question "What did you learn by playing the game?". Yet, differently to the MEEGA+ measurement instrument in which the achievement of the learning goals is measured through a self-assessment of the participants, this measurement is done as part of the MEEGA+KIDS questionnaire through a set of multiple-choice questions assessing the achievement of the learning goals. These questions have to be carefully defined in accordance with the specific learning goals and the competence level to be achieved based on the revised version of Bloom's taxonomy.

Templates of the MEEGA+KIDS questionnaire are available in English and Brazilian Portuguese on the website: http://www.gqs.ufsc.br/quality-evaluation/meega-plus/

5. Evaluation of the MEEGA+KIDS Model

5.1 Definition and Execution

The objective of this evaluation is to analyze the MEEGA+KIDS measurement instrument to evaluate its reliability and construct validity from the viewpoint of the senior researchers in the context of computing education in secondary schools. In this study, we understand reliability as the degree of consistency of instrument items. Reliability is typically measured through its internal consistency, which measures the consistency of results across items within a questionnaire (Trochim & Donnelly, 2008) through the Cronbach's alpha coefficient (Cronbach, 1951). Construct validity is the ability of an instrument to measure what it purports to measure, including convergent and

discriminant validity, which is measured through the degree of item correlation (Trochim & Donnelly, 2008).

Following the GQM approach, the study objective is decomposed into quality aspects and analysis questions to be analyzed based on the evaluation of measurement instruments (Trochim & Donnelly, 2008):

AQ1-Reliability: Is there evidence for the internal consistency of the MEEGA+KIDS measurement instrument?

AQ2- Construct Validity: Is there evidence of the convergent and discriminant validity of the MEEGA+KIDS measurement instrument?

From 2018 to 2019, we collected data from six case studies, evaluating an educational game (SplashCode) using the MEEGA+KIDS model. In each of these case studies, the MEEGA+KIDS measurement instrument (Brazilian Portuguese version) was applied after the game session (treatment) to collect the students' perceptions about the game quality. In total, responses from 90 secondary school students in Brazil were collected as summarized in Table 2. Participating students and their responsible signed an informed consent and authorization for image use in academic publications. The evaluations have been approved by the Ethics Committee of the Federal University of Santa Catarina (Certificate No. 2.677.698).

Table 2. Summary of the conducted case studies

Game	Game session time	Course/Institution	City	Sample size
Splashcode A low-cost board game to reinforce basic algorithms and programming concepts (Gresse von Wangenheim et al., 2019).	15 minutes	Oficina 1 Escola E.E.B. Almirante Carvalhal	Florianópolis/SC	18
		Treinamento Jovens Tutores Escola E.E.B. Almirante Carvalhal		10
		Oficina Computação na Escola/UFSC		2
		Oficina 2 Escola E.E.B. Almirante Carvalhal		11
		Oficina 3 Escola E.E.B. Almirante Carvalhal		24
		Oficina 4 Escola E.E.B. Almirante Carvalhal		25
			Total	**90**

5.2 Analysis

To evaluate the MEEGA+KIDS model, we performed a statistical evaluation using the IBM SPSS Statistics trial version 23.

Reliability: Is there evidence for the internal consistency of the MEEGA+KIDS measurement instrument?

To measure the internal consistency of the MEEGA+KIDS questionnaire we adopted the Cronbach's alpha coefficient (Trochim & Donnelly, 2008). Cronbach's alpha coefficient (Cronbach, 1951) describes the degree to which a set of items measure a factor, such as whether the MEEGA+ measurement instrument measures the quality of an educational game for computing education in the context of secondary schools. Values of Cronbach's alpha between $0.8 > \alpha \geq 0.7$ are acceptable, between $0.9 > \alpha \geq 0.8$ are good, and $\alpha \geq 0.9$ are excellent (Trochim & Donnelly, 2008), indicating an internal consistency of the instrument. Analyzing the 26 items of the MEEGA+KIDS questionnaire, the value of Cronbach's alpha is considered good ($\alpha = .882$). We also analyzed the Cronbach's alpha if an item was deleted, as result we obtained that no item (if deleted) causes a substantial increase in the Cronbach's alpha. We, thus, can conclude that the responses to the items are consistent and precise, indicating the reliability of the items of the MEEGA+KIDS questionnaire.

Construct Validity: Is there evidence of the convergent and discriminant validity of the MEEGA+KIDS measurement instrument?

Convergent and discriminant validity are subcategories of construct validity. To obtain evidence of convergent and discriminant validity of the questionnaire items of the MEEGA+KIDS model, item correlation is calculated (Trochim & Donnelly, 2008).

Convergent validity shows that items that should be related are, in fact, related. On the other hand, discriminant validity shows that items that should not be related are, in fact, not related (Trochim & Donnelly, 2008). To analyze the correlations between the items, we used the nonparametric Spearman correlation matrices for each quality factor (Tables 3 and 4). The matrices show the Spearman correlation coefficient, indicating the degree of correlation between two items (item pairs). We used this correlation coefficient, as it is the most appropriate for Likert scales (Trochim & Donnelly, 2008). The correlation coefficients between the items within the same dimension are colored. Following Cohen (1988), a correlation between items is considered satisfactory, if the correlation coefficient is greater than 0.29, indicating that there is a moderate correlation, or a strong correlation if the coefficient is greater than 0.50. A coefficient of about 0.10 indicates a low correlation between de items (Cohen, 1988). Satisfactory correlations are marked in bold.

Table 3. Spearman correlation coefficient of the quality factor Usability

Item/ Dimension	1	2	3	4	5	6	7
	Aesthetics		**Learnability**	**Operability**		**Accessibility**	
1	1.000						
2	**.592**	1.000					
3	.228	**.355**	1.000				
4	.144	**.327**	**.384**	1.000			
5	.040	.193	.261	.158	1.000		
6	.023	**.325**	**.345**	.213	**.430**	1.000	
7	.128	**.301**	**.335**	.187	**.495**	**.401**	1.000

Analyzing the correlations between the items of the quality factor usability (Table 3), we can observe that most of the item pairs (11 item pairs) (marked in bold) present a moderate or strong correlation. Other item pairs present a correlation coefficient slightly close to the moderate degree of correlation (e.g., item pair 3-5). No item pairs present a low correlation coefficient under 0.10. Thus, although a higher correlation coefficient between the items was expected, which may have been caused by the size of the sample analyzed (n=90), we can observe a first indication that the questionnaire items of the quality factor usability tend to be correlated since there are no negative or low correlation coefficients and most of the item pairs present a moderate or strong correlation.

Concerning the quality factor player experience (Table 4), most of the item pairs (87 item pairs) are correlated. In this case, we can observe that items belonging to the same dimension (e.g. relevance, fun, social interaction, focused attention, and relevance) present a moderate or high correlation coefficient. Again, indicating evidence of convergent validity.

Table 3. Spearman correlation coefficient of the quality factor Player Experience

Item/ Dimension	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
	Confidence	Challenge			Satisfaction				Social Interaction			Fun		Focused Attention		Relevance			
8	1.000																		
9	**.336**	1.000																	
10	**.543**	**.553**	1.000																
11	.214	.243	**.483**	1.000															
12	**.419**	**.298**	**.451**	.098	1.000														
13	.275	.026	.264	.096	.285	1.000													
14	**.540**	**.381**	**.606**	.120	**.534**	.299	1.000												
15	.229	**.292**	**.405**	.117	.281	.190	**.491**	1.000											
16	.112	.038	.138	.050	.270	.059	**.349**	.232	1.000										
17	**.316**	.070	.189	.049	.192	.087	.274	**.301**	**.353**	1.000									
18	.268	.170	**.420**	.222	**.333**	.274	**.505**	**.294**	**.553**	**.373**	1.000								
19	.273	.069	**.340**	.107	**.311**	.197	**.447**	**.528**	.091	**.430**	**.404**	1.000							
20	.046	.145	.250	.062	**.369**	.042	**.323**	.246	.266	**.376**	**.421**	**.380**	1.000						
21	**.323**	.240	**.430**	.095	**.336**	.124	**.527**	**.483**	.206	**.317**	**.351**	**.290**	**.354**	1.000					
22	.234	.228	**.373**	**.373**	**.341**	.221	**.403**	**.367**	.169	**.311**	.196	**.301**	**.303**	**.410**	1.000				
23	**.334**	.262	**.461**	.171	**.341**	.150	**.453**	**.381**	.137	**.363**	**.310**	**.446**	.250	.287	**.561**	1.000			
24	**.377**	.267	**.385**	.155	**.303**	-.081	**.478**	**.417**	.269	**.341**	.299	**.347**	**.377**	**.494**	.282	**.385**	1.000		
25	**.299**	.133	.275	.112	.245	.045	**.319**	**.517**	.253	.269	.167	**.353**	.174	**.399**	.304	**.448**	**.665**	1.000	
26	.199	.147	.173	.041	.072	.218	.217	**.306**	.030	.245	.171	.185	.233	**.293**	.327	.248	**.323**	**.361**	1.000

Analyzing the correlation coefficients between items of different quality factors, we can observe that most of the items do not correlate, confirming that they are measuring

different quality factors (usability and player experience). Thus, indicating evidence of discriminant validity.

        Summarizing, we can observe that in general, the results show that most of the item pairs present a moderate or strong correlation coefficient, indicating evidence of convergent validity. Evidence of discriminant validity was also obtained, showing that items of different quality factors present a small correlation coefficient. However, considering the sample size used in this study, we need to confirm these results by adopting a larger data set in further analysis.

5.3 Threats to Validity

Like any kind of empirical research, due to limitations, our study is subject to threats to validity. We, therefore, identified potential threats and applied mitigation strategies to minimize their impact on our research. Some threats are related to the design of the study. To mitigate this threat, we defined and documented a systematic research method, based on the methodology adopted by the MEEGA+ method (Petri et al., 2019). Besides, the MEEGA+KIDS measuring instrument has been customized from the MEEGA+ model adopting a participatory design methodology (Spinuzzi, 2005), including representatives of the target audience in the definition of the measurement instruments items. One limitation of our study refers to evaluating learning using multiple-choice questions. Adopting a non-experimental research design (case studies), only a post-test using self-assessment, through multiple-choice questions, has been applied to evaluate the students' achievement of the learning goals. A pre-test has not been applied and, therefore, it was not possible to accurately the learning difference promoted by the games. However, regarding the self-assessment, although there is no consensus, there is evidence that self-assessment provides reliable, valid, and useful information for this type of study (Sharma et al., 2016), mainly when using a systematic, reliable and valid measurement instrument. In terms of external validity, a threat to the possibility to generalize the results is related to the sample size and diversity of the data used for the evaluation. In respect to sample size, our evaluation used data collected from six case studies evaluating one educational game, involving a population of 90 students. Although it is a considerable sample size, allowing the generation of first statistical evidence, further analysis with a larger data set including data from other educational games is required to confirm our results.

6. Conclusion

In this article, we present the design and evaluation of the MEEGA+KIDS as a model to evaluate games used for computing education in secondary school. First evaluation results indicate that the questionnaire has satisfactory reliability and validity. Concerning reliability, a Cronbach's alpha $\alpha=.882$ indicates a good internal consistency, which means that the responses between the questionnaire items are consistent and precise. Regarding the construct validity, analyzing the Spearman correlation coefficient between the items, we obtain evidence of discriminant validity, showing that items of different quality factors do not correlate. In the same way, results show that most of the item pairs present a moderate or large correlation coefficient. Thus, providing a first indication of convergent validity. However, further analysis with a larger data set is required to obtain more significant results in terms of construct validity. As the next steps, we plan to continue to conduct game evaluations to increase the sample size.

Acknowledgments

References

Basili, V. R., Caldiera, G., & Rombach, H. D. Goal, Question Metric Paradigm. In J. J. Marciniak, **Encyclopedia of Software Engineering**. New York: Wiley-Interscience, 1994. p 528-532.

Battistella, P. & Gresse von Wangenheim, C. Games for teaching computing in higher education – a systematic review. **IEEE Technology and Engineering Education Journal**, v.9, n.1, p 8-30, 2016.

Casarotto, R. I., Bernardi, G., Cordenonsi, A. Z., & Medina, R. D. Logirunner: um Jogo de Tabuleiro como Ferramenta para o Auxílio do Ensino e Aprendizagem de Algoritmos e Lógica de Programação. **RENOTE – Revista Novas Tecnologias na Educação**, v.16, n.1, p 1-10, 2018.

Chaffin, A., Doran, K., Hicks, D., & Barnes, T. Experimental evaluation of teaching recursion in a video game. In: **Proc. of the ACM SIGGRAPH Symposium on Video Games**, New Orleans, EUA, 2009.

Cohen, J. **Statistical Power Analysis for the Behavioral Sciences**. New York: Routledge Academic, 1988.

Cronbach, L. J. Coefficient alpha and the internal structure of tests. **Psychometrika**, v.16, n.3, p 297-334, 1951.

Decker, A., McGill, M. M. & Settle, A. Towards a common framework for evaluating computing outreach activities. In: **Anais do ACM Technical Symposium on Computer Science Education**. Memphis, TN, EUA, 2016, p 627-632.

Garneli, V., Giannakos, M. N., & Chorianopoulos, K. Computing education in K-12 schools: A review of the literature. In: **Proc. of the IEEE Global Engineering Education Conference**, Tallin, Estonia, 2015.

Gresse von Wangenheim, C. et al. Desenvolvimento e Avaliação de um Jogo de Tabuleiro para Ensinar o Conceito de Algoritmos na Educação Básica. **Revista Brasileira de Informática na Educação**, v.27, n.3, p 310-335, 2019.

Ibrahim, R., Yusoff, R. C. M., Omar, H. M., & Jaafar, A. Students Perceptions of Using Educational Games to Learn Introductory Programming. **Computer and Information Science**, v.4, n.1, p 205-216, 2011.

Kazimoglu, C., Kiernan, M., Bacon, L., & MacKinnon, L. Learning programming at the computational thinking level via digital game-play. **Procedia Computer Science**, v. 9, p 522-531, 2012.

Kim Y.J. & Ifenthaler D. Game-Based Assessment: The Past Ten Years and Moving Forward. In: Ifenthaler D., Kim Y. (eds) **Game-Based Assessment Revisited**. Advances in Game-Based Learning. Springer, 2019.

Liu, C. C., Cheng, Y. B., & Huang, C. W. The effect of simulation games on the learning of computational problem solving. **Computers & Education**, v. 57, n.3, p 1907-1918, 2011.

Papastergiou, M. Digital Game-Based Learning in high school Computer Science education: Impact on educational effectiveness and student motivation. **Computers & Education**, v. 52, n.1, p 1-12, 2009.

Pearce, J. M., Ainley, M., & Howard, S. The ebb and flow of online learning. **Computers in Human Behavior**, v.21, n.5, p 745–771, 2005.

Petri, G. & Gresse von Wangenheim, C. How to evaluate educational games: a systematic literature review. **Journal of Universal Computers Science**, v. 22, n.7, p 992-1021, 2016.

Petri, G. & Gresse von Wangenheim, C. How games for computing education are evaluated? A systematic literature review. **Computers & Education**, v.107, p 68-90, 2017.

Petri, G. & Gresse von Wangenheim, C. A Method for the Evaluation of the Quality of Games for Computing Education. In: **Anais do Congresso Brasileiro de Informática na Educação**. Brasília, Brasil, 2019, p 951-960.

Petri, G., Gresse von Wangenheim, C., & Borgatto, A. F. MEEGA+: Um Modelo para a Avaliação de Jogos Educacionais para o ensino de Computação. **Revista Brasileira de Informática na Educação**, v. 27, n.3, p 52-81, 2019.

Sharma, R., Jain, A., Gupta, N., Garg, S., Batta, M., & Dhir, S. K. Impact of self-assessment by students on their learning. **Journal of Applied and Basic Medical Research**, v.6, n.3, p 226–229, 2016.

Spinuzzi, C. The Methodology of Participatory Design. **Technical Communication**, v.52, n.2, p 163-174, 2005.

Trochim, W. M., & Donnelly, J. P. **Research methods knowledge base** (3rd ed.). Mason: Atomic Dog Publishing, 2008.

Vlahu-Gjorgievska, E., Videnovik, M., & Trajkovik, V. Computational Thinking and Coding Subject in Primary Schools. In. **Proc. of the IEEE Int. Conf. on Teaching, Assessment, and Learning for Engineering**, Wollongong, Australia, 2018.

Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., & Wesslén, A. **Experimentation in Software Engineering**. New York: Springer-Verlag Berlin Heidelberg, 2012.