Effect of whole-body movement on performance and efficiency: A comparison of three controlling methods for a math game

Antero Lindstedt[1[0000-0001-5906-7654]] and Kristian Kiili[2[0000-0003-2838-6892]]

[1] Tampere University, Information Technology and Communication, Pori , Finland
[2] Tampere University, Faculty of Education and Culture, Tampere, Finland
{antero.lindstedt, kristian.kiili}@tuni.fi

Abstract. During the last decade, the number of studies investigating learning effectiveness and motivational aspects of game-based learning has increased. Nevertheless, research that considers the meaning of the User Interface (UI) in game-based learning has been sparse. This paper reports a within-subject study in which we investigated how the implementation of the UI affects students' performance (accuracy), training efficiency (task completion duration), and user experience in a number line based math game. Ninety-three fifth graders played the same math game with three different UIs in a counter balanced order. The results revealed that the implementation of the UI influenced significantly on performance and training efficiency. Students' estimation performance with the chair-based exertion UI (whole-body movement) was significantly worse than with the tilting UI (controlled with hands) and virtual directional pad UI (controlled with fingers). Nevertheless, the players felt that the controlling of the game was equally easy with the gaming chair than with the other two controlling methods. Actually, the majority of the students named the chair as the most preferable controlling method. The results suggest that a whole-body movement can be an engaging and viable controlling method for learning games, but its effects on performance and efficiency should be considered, especially in game-based assessment context.

Keywords: User interface, Human-computer interaction, Whole-body movement, Game-based learning, User experience, Number line estimation.

1 Introduction

The introduction of digital games in the field of mathematics education has presented new ways to improve mathematical skills. The games can contain simple drills, similar to pen-and-paper tasks, but also more inventive learning content. For example, Devlin [1] has argued that video games can provide an alternative presentation to the traditional, symbol-based expressions that is easier and more natural to use. According to Moeller and colleagues [2], the latest technological developments have created new possibilities to train mathematical competencies with, for example, digital game-based learning solutions that provide interactive math training or allow the player to solve mathematical tasks in an embodied fashion. Digital games can also offer immediate feedback to the student without the need for the teacher to check the answers. For the teacher, the game could offer data about the students' progress. In fact, digital game-based learning have the potential to support learning assessment (e.g. [3] [4] [5]).

1.1 Movement-based user interfaces in games

During the past decade we have witnessed more attempts at using body movement in digital games [6] and exertion interfaces have been utilized also in game-based learning field (e.g. [7]). So-called exergames use player's own physical movement for game controlling. According to Staiano and Calfert [8], this can lead to physical and cognitive development, but also have other benefits by providing opportunities for social interaction. For an educational game, the physical component could help to alleviate cognitive load (embedded cognition) and promote learning by drawing on previous sensorimotor experiences (embodied cognition) [9]. There has also been evidence that exergame interventions in schools can reduce negative classroom behaviors and absenteeism while improving academic performance [10]. Exergames have been implemented using special devices, such as Kinect™ [11] or dance mats [12], but as mobile devices like phones and tablets have become better and more common, their accelerometers have been used to implement embodied number line trainings [3] [13].

It is important to implement controls based on physical movement properly, or otherwise it can distract the players from the actual gameplay [7]. A well-designed UI allows the player to master the game controls and lets the player focus on the game content. This might not be satisfying in itself, but the mastery of controls is necessary for meeting the player's psychological needs [14]. In fact, previous research has argued that the increase of physical movement in games increases positive emotional and social responses [15] [16].

Although the gamification of learning and exertion interfaces have the potential to increase engagement, we should also consider their consequences to other learning dimensions such as learning effectiveness and learning analytics. Game elements that increase enjoyment may in turn make performance less certain. This may be irrelevant in the training context if the learning is not impeded, but could become more problematic if the results are used for assessment (grading) purposes [17]. Kiili and Ketamo [3] considered game-based math test to be a valid and fair assessment approach regarding gender and students' previous gaming experience even though they also found that the performance was significantly worse with a game-based math test than with a paper-and-pencil based test. Game-based assessment in education is relatively new and only a little evidence exists on maximizing the validity and effect of it without losing engagement [18]. Thus, it is important to study how the user interfaces influence on students' performance in game-based learning solutions.

1.2 Present study and hypotheses

Despite the increasing interest in utilizing movement in digital games, very little is known how exertion interfaces influence on students' performance and attitudes in game-based learning. The aim of the current study is to shed light on this open question. In order to study the meaning of the user interface, we developed a number line based math game and three different controlling user interfaces: Chair UI (whole body), Tilt UI (hands), and virtual D-pad (directional pad) UI (fingers). In this within-subject study, we compared these user interfaces with respect to the number line estimation accuracy (performance), answering duration (efficiency), and students' opinions about the user interfaces (user experience).

Based on previous research we formed two hypotheses. As large whole body movements are not as precise as smaller hand and finger gestures, we hypothesize that answering accuracy is worse and task completion duration is longer with the Chair UI than with the Tilt and the D-pad UI (Hypothesis 1). Because previous research has shown that when the controller requires increased body movement, the player's engagement level improves [15], we hypothesize that the students will select the chair as the best user interface (Hypothesis 2).

2 Methods

2.1 Participants

A total of 93 fourth graders (age 11-12) from three Finnish schools participated in this study. 50 were male and 43 were female.

2.2 Game-based learning solutions

A mathematics learning game, created using a number line based game engine developed by the researchers, was used in this study (a screenshot in figure 1). Using the game engine, three versions of a rational number learning game were built. The games were played on an iPad device and they utilized the device's accelerometer and touch input. The contents of the game were exactly the same with each version, but three different controlling user interfaces were developed:

- *D-pad UI*. The game character was controlled with fingers using an onscreen virtual Directional pad and answering was done with an onscreen button.
- *Tilting UI*. To make the game character move, the tablet had to be tilted by hand movement towards left or right, which made the character walk in that direction. Tapping the screen made the character jump. An onscreen button was used for answering.
- *Chair UI*. This version required a special chair to play (figure 1). The player sat on the chair with the tablet device fixed to the chair in front. The player had to lean his/her body towards one of the four directions. Tilting the chair to either side made the character walk towards that direction. Tilting the chair backwards resulted in a jump and tilting forwards was interpreted as an answer. These controls required upper-body movement, but the exertion applied to legs, arms, and midriff too, even if these parts remained quite stationary.

The game content was divided into levels, each containing 10 tasks. The first task of each level was a tutorial task that provided a soft start for the student, and explained if there was some new game mechanic introduced in the level. The first level had trivial to solve whole number tasks with the correct answer visible. This was used as an onboarding level to measure how fast the player could get acquainted with the controlling method without mathematical competency affecting the results. The game characters' walking distance to the correct answer was the same in each of these tasks. The second level contained basic fraction number tasks. The third level introduced traps, which required additional estimating. The fourth level had enemies that had to be evaded or destroyed by jumping on them. The fifth level and all the levels afterwards mixed the same elements, but increased gradually in difficulty.

Fig. 1. On the left: A screenshot of the game. In this task, the player had to estimate a fraction $\frac{1}{5}$ while having to avoid a trap at $\frac{3}{5}$. On the right: Students playing with the gaming chair.

2.3 Measures and analyses

Data was collected using online questionnaires and in-game data storing. All the data was anonymized. The answer accuracy and duration for each task were stored in a database. The player had a maximum of two attempts per task. If the first attempt did not reach 90% or better accuracy, a second attempt was allowed. Some tasks showed additional hints on the second attempt. Therefore, only the accuracy of the first answer was stored. Additionally, the playing sessions were observed by the supervisor.

Statistical analysis was performed in IBM SPSS Statistics, version 25. A one way repeated measures ANOVA was conducted to determine whether there were differences in the accuracy and duration on per level basis. All the analyzed data were tested using Mauchly's sphericity, and in case the sphericity assumption was violated, the resulted F-test values were adjusted using Greenhouse-Geisser correction. In pairwise comparisons, a Post hoc analysis with Bonferroni adjustment was used.

2.4 Procedure

For the gaming sessions, a separate space was arranged where the gaming chairs were set up. Five students took part in each session, while the rest of the students continued in their own class normally. Each session lasted about 1.5 hours. This consisted of starting instructions, playing of each of the three game versions and filling out the required questionnaires. The students had 15 minutes to play each game version. The rest of the actions lasted until everybody was finished.

During a session, the students played the same game version at a time and then filled in a questionnaire about that game design. Then the same procedure was repeated for the other two designs. In the end, there was still one questionnaire to be filled. The order in which the game designs were played within each group was distributed so that each version was equally often the first, the second and the last one to be played. Since the behavior of students may differ at different times of a school day, the order of the game designs was similarly distributed evenly to morning, midday and afternoon.

3 Results

There was a significant effect of user interface (UI) on the number of completed tasks, $F(2, 184) = 40.195$, $p < .001$, $\eta_p^2 = .304$. Most tasks were completed with the D-pad UI (47.55 tasks, SD = 9.60), followed by the Tilt UI (44.31, SD = 10.68), and least tasks were completed with the Chair UI (38.18, SD = 8.98). Pairwise comparisons showed that students completed 9.37 (95% CI, 6.68 to 12.05) tasks fewer with the Chair UI than with the Tilt UI, $p < .001$, and 6.13 (95% CI, 3.46 to 8.80) tasks fewer than with the D-pad UI, $p < .001$. With the Tilt UI students completed 3.24 (95% CI, 0.84 to 5.63) tasks fewer than with the D-pad UI, $p < .001$.

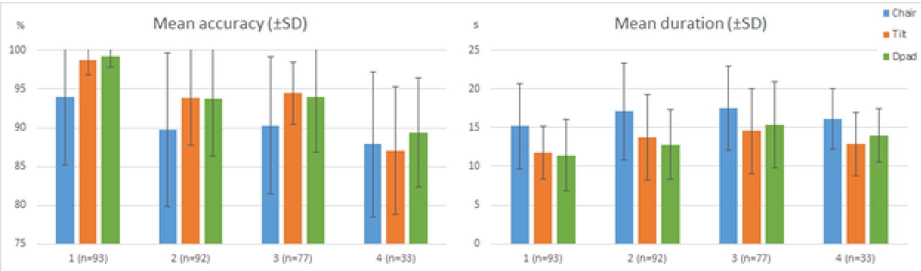

Fig. 2. The mean accuracy (left) and duration (right) for each of the game's four first levels. The horizontal axis shows the level number and participant count who completed it

As the students completed a differing amount of levels and the levels varied in design, in the following sections, students' performance is handled level by level (Figure 2). This means that only those students that completed the level with each of the three user interfaces were included in the level-based analyses.

3.1 Analyses of the onboarding level

Every student completed the first level (onboarding) successfully (N=93). There was a significant effect of the user interface both on accuracy, $F(1.11, 101.90) = 28.34$, $p <$

.001, $\eta_p^2 = .235$ and on answering duration, $F(1.70, 156.46) = 24.21$, $p < .001$, $\eta_p^2 = .208$. Pairwise comparisons indicated that the mean accuracy with the Chair UI was 4.78 (95% CI, 2.47 to 7.09) percentage points lower than with the Tilt UI, $p < .001$, and 5.34 (95% CI, 3.07 to 7.61) percentage points lower than with the D-pad UI, $p < .001$. In line with this, the mean answering duration with the Chair UI took 3.47 (95% CI, 1.96 to 4.98) seconds longer than with the Tilt UI, $p < .001$, and 3.80 (95% CI, 2.10 to 5.45) seconds longer than with the D-pad UI, $p < .001$. On the other hand, there was no statistically significant difference in mean accuracy ($p = .084$) or in mean duration ($p = 1.000$) between the Tilt and D-pad UIs. The rest of the levels showed no difference either. Thus, pairwise comparison results between Tilt and D-pad UIs are not reported in analyses of the rest of the levels.

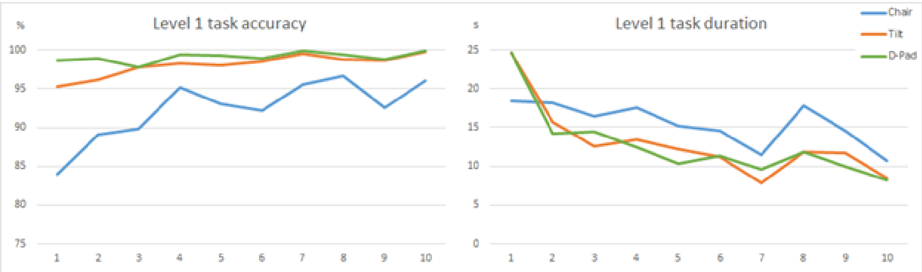

Fig. 3. The answer accuracy (left) and duration (right) for each task on the first level. Task number on horizontal axis

The more detailed analysis of the ten onboarding tasks revealed that in the very beginning, the students had more difficulties with the Chair UI than with the Tilt or the D-Pad UI. Figure 3 shows that the accuracy with the Tilt and D-Pad UIs are good (over 95%) from the very beginning, but the mean accuracy of the first task with the Chair UI was only 84%. This is rather poor considering that the correct answer was visible. The accuracy with the Chair UI improved during the following tasks, but still remained worse than with the other two UIs.

Task durations follow a similar trend. The figure 3 shows that answering took longer with the Chair UI than with other UIs. An exception to this is the first task, in which the duration was smaller with the Chair condition. This is likely due to accidental answering that some of the players faced in the beginning. The same was not experienced to the same extent with the Tilt or D-pad UI. These accidental answers lowered the mean accuracy and explain much of the poor accuracy on the first task with the Chair UI. After the first task, the unintended answers diminished with the Chair UI too.

3.2 Analyses of the basic levels

The second level contained basic fraction estimation tasks without any obstacles. One student had already dropped from the analyses at this point (n=92). There was a significant difference in accuracy between the user interfaces, $F(1.68, 152.64) = 10.99$, $p < .001$, $\eta_p^2 = .108$. Pairwise comparisons indicated that the mean accuracy with the Chair

UI was 4.14 (95% CI, 1.84 to 6.45) percentage points lower than with the Tilt UI, $p < .001$, and 3.99 (95% CI, 1.08 to 6.91) percentage points lower than with the D-pad UI, $p < .004$. There was a significant difference also in answering duration between the user interfaces, $F(2, 182) = 23.27$, $p < .001$, $\eta_p^2 = .204$. Answering with the Chair UI took 3.33 (95% CI, 1.65 to 5.02) seconds longer than with the Tilt UI, $p < .001$, and 4.29 (95% CI, 2.56 to 6.02) seconds longer than with the D-pad UI, $p < .001$.

In the third level, in addition to the basic task, the students had to estimate the locations of traps and jump over them. Level 3 was completed by 77 students. Again there was a significant difference in accuracy between the user interfaces $F(1.78, 135.44) = 12.34$, $p < .001$, $\eta_p^2 = .140$. The mean accuracy with the Chair UI was 4.15 (95% CI, 1.63 to 6.68) percentage points lower than with the tilt UI, $p < .001$, and 3.67 (95% CI, 1.37 to 6.00) percentage points lower than with the D-pad UI, $p < .001$. Like in level 2, there was a significant difference also in answering duration between the user interfaces, $F(2, 152) = 8.74$, $p < .001$, $\eta_p^2 = .103$. Answering with the Chair UI was 2.91 (95% CI, 1.12 to 4.71) seconds slower than with the Tilt UI, $p < .001$, and 2.13 (95% CI, .33 to 3.94) seconds slower than with the D-pad UI, $p = .015$.

In the fourth level, the students also faced enemies that had to be avoided or destroyed by jumping on them. Only a third of the students (n=31) were able to finish this level. This time, accuracy did not reveal any statistically significant differences between the user interfaces, $F(2, 60) = 1.92$, $p = .16$, $\eta_p^2 = .060$. Nevertheless, like in the previous levels, there was a significant difference in answering duration between the user interfaces $F(2, 60) = 8.73$, $p < .001$, $\eta_p^2 = .225$. With the Chair UI, answering was 3.24 (95% CI, 1.36 to 5.12) seconds slower than with the Tilt UI, $p < .001$, and 2.14 (95% CI, .08 to 4.19) seconds slower than with the D-pad UI, $p = .015$.

Only four students completed level 5 with all the user interfaces. It is not meaningful to analyze such a small sample.

3.3 User experiences

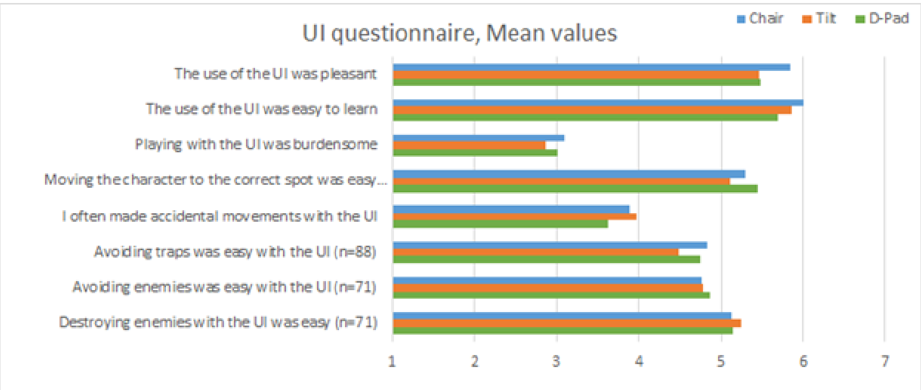

Fig. 4. Questions asked after each UI version. Answers on a 7-point Likert scale (1 = totally disagree, 7 = totally agree).

While the analysis showed that the Chair UI was less accurate and took more time to use, the question of preference was much more positive for the Chair UI. When the students had to name the best and the worst UI, the one receiving the most "best" votes was the Chair UI (77.6%). The D-pad UI (12.2%) and the Tilt UI (10.2%) followed at quite a distance. The "worst" votes followed a similar trend as the Tilt UI (44.9%) and the D-pad UI (43.9%) received votes roughly evenly and the Chair UI (11.2%) got much less.

Based on the questionnaires about each UI (Figure 4), there was very little difference between the three UIs, and none of the question showed any statistical significance. Sample size was a bit smaller on the three last questions as answers for those were omitted if the player had not actually faced traps and/or enemies.

4 Discussion and conclusions

In the current study, we aimed at evaluating how the user interface implementation influences on playing performance, efficiency, and user experience of a mathematical game-based learning solution. Particularly, we explored how whole-body game controller influenced on these metrics. In the following, we will first discuss the results concerning the player performance and efficiency before we will elaborate results about students' user experiences and adoption of the different user interface implementations.

4.1 Player performance and efficiency

One objective of the current study was to evaluate the usefulness of the whole-body user interface in a math game founded on the number line estimation task. The results revealed that the implementation of the user interface influenced significantly on answering accuracy and task completion duration as the effect sizes varied from medium to large. The students' answering accuracy with the Chair UI (whole-body movement) was significantly worse than with the other user interfaces that did not involve whole-body movement as a controlling method. In line with this, the Chair UI was also significantly slower to use than the other user interfaces. These results are consistent with our first hypothesis. One exception was the fourth level, where the accuracy no longer showed statistical significance. An explanation could be that the players who managed to complete the level 4 (n=31) were generally more competent with the UI and therefore its effect to the accuracy was diminished. The smaller sample size could also have an effect. The Tilt (hand movement) and the D-Pad (finger movement) UIs produced very similar results throughout the whole study. The only statistically significant difference was the completed tasks count, where the D-Pad condition fared slightly better.

The lowered performance and efficiency has clear implications to educational practices. Educators should be aware of how user interface solutions affect students' performance so that they can utilize the learning analytics or in-game metrics that the games may provide in a reasonable manner. We argue that game controlling methods that may influence negatively on performance (educational competences) suit best for training or for formative assessment context. That is, if the game is used in summative assessment, the user interface should be optimized with respect to controlling accuracy and fairness. The results also showed that the whole-body user interface was inefficient compared to other studied user interfaces as the students managed to complete fewer tasks with the Chair UI. With that respect, when selecting games, educators should consider whether the amount of training of the learning subject is more important than the physical training and other beneficial effects gained at the same time through the game controls.

4.2 User experiences and user interface adoption

Although exact answering was harder and it took more time with the whole-body user interface than with the other user interfaces, the students appreciated it more. As we expected in our second hypothesis, the students were enthusiastic to play with the gaming chair and it was the most popular controlling method by far despite its limitations, with 77.6% of the students naming it the best UI. The user experience questions did not show any significant difference on the perceived easiness of use between the UIs.

The onboarding tasks showed that the two first tasks produced many unintended answers for the Chair UI, some for the Tilt UI and none for the D-pad UI. The rest of the onboarding tasks had only a few unintended answers meaning the players got the hang of the controls quickly. This highlights the usefulness of a soft start to playing, like having a couple of tutorial tasks first where the players are free to experience the controls of the game. This means that the results from the tutorial tasks should be omitted if the game is used for assessment purposes. The soft start is especially important for controlling methods using whole-body movement, as the player has to familiarize her/himself with not only the execution of the commands but also the required intensity of the exertion.

4.3 Limitations and further research

One limitation of the study is that much of the Chair UI's appeal could be attributed to the novelty effect of getting to play with the gaming chair. On the other hand, it is a good sign if the players feel like they are playing with a toy when they are controlling an educational game. While this case required special equipment, it still shows that physical movement based educational games can be engaging. Secondly, the treatment duration of 15 minutes per UI might not be enough to reveal all the aspects of the UIs. To further investigate the subject, a between-subject study in which students play the game for a longer period could be conducted. This could negate the effect of having to play the same content for multiple times and the longer playing period could decrease the novelty effect. Moreover, longer playing sessions could reveal how physical demands of the chair (exhaustion) influence on students' user interface preferences and answering accuracy. By having the player focus on only one UI, the player could proceed further in the game content, which would open up possibilities to study learning gains.

References

1. Devlin, K.: The music of math games. In: American Scientist, vol. 101(2), pp 87-91 (2013).
2. Moeller, K., Fischer, U., Nuerk, H.C., Cress, U.: Computers in mathematics education – Training the mental number line. In: Comput. Human Behavior, vol. 48, pp. 597–607 (2015).
3. Kiili, K., Ketamo, H.: Evaluating cognitive and affective outcomes of a digital game-based math test. In: IEEE Transactions on Learning Technologies, vol. 11(2), pp. 255-263 (2018).
4. Shute, V. J., Wang L., Greiff, S., Zhao, W., Moore, G.: Measuring problem solving skills via stealth assessment in an engaging video game. In: Comput. Human Behavior, vol. 63, pp. 106–117 (2016).
5. Serrano-Laguna, A., Torrente, J., Moreno-Ger, P., Fernandez-Manjon, B.: Application of learning analytics in educational video-games. In: Entertainment Comput., vol. 5(4), pp. 313–322 (2014).
6. Márquez Segura, E., Waern, A., Moen, J., Johansson, C.: The design space of body games: technological, physical, and social design. In: Proceedings of the SIGCHI conference on Human Factors in computing systems, pp. 3365-3374. ACM. (2013).
7. Lindstedt, A., Kiili, K., Tuomi, P., Perttula, A.: A user experience case study: two embodied cognition user interface solutions for a math learning game. In: Seminar.net - International journal of media, technology and lifelong learning, vol. 2(12), (2016).
8. Staiano, A., Calvert, S.: Exergames for physical education courses: Physical, social, and cognitive benefits. In: Child Development Perspectives, vol. 5(2), pp. 93–98 (2011).
9. Pouw, W. T., Van Gog, T., Paas, F.: An embedded and embodied cognition review of instructional manipulatives. In: Educational Psychology Review, vol. 26(1), pp 51-72 (2014).
10. Lieberman, D., Chamberlin, B., Medina, E., Franklin, B., Sanner, B., Vafiadis, D.: The power of play: Innovations in getting active summit 2011: A science panel proceedings report from the American heart association. In: Circulation, 123, pp. 2507–2516 (2011).
11. Link, T., Moeller, K., Huber, S., Fischer, U., Nuerk, H.C.: Walk the number line – An embodied training of numerical concepts. In: Trends in Neuroscience and Education, vol. 2, pp. 74-84 (2013).
12. Moeller, K., Fischer, U., Nuerk, H.C., Cress, U.: Computers in mathematics education – Training the mental number line. In: Comput. Human Behavior, vol. 48, pp. 597–607 (2015).
13. Riconscente, M.M.: Results from a controlled study of the iPad fractions game Motion Math. In: Games and Culture, vol. 8(4), pp. 186-214 (2013).
14. Przybylski, A. K., Rigby, C. S., Ryan, R. M. A.: motivational model of video game engagement. In: Review of general psychology, vol. 14(2), pp. 154. (2010).
15. Bianchi-Berthouze, N., Kim, W.W., Patel, D.: Does Body Movement Engage You More in Digital Game Play? and Why?. In: Proc. ACII'07, Springer-Verlag, pp. 102-113. (2007).
16. Lindley, S.E., Le Couteur, J., Berthouze, N.L.: Stirring up experience through movement in game play: effects on engagement and social behaviour. In: Proc. CHI'08, ACM press, pp. 511-514. (2008).
17. Greipl S., Ninaus M., Bauer D., Kiili K., Moeller K. A Fun-Accuracy Trade-Off in Game-Based Learning. In: Gentile M., Allegra M., Söbke H. (eds) Games and Learning Alliance. GALA 2018. Lecture Notes in Computer Science, vol. 11385. Springer, Cham. (2019).
18. Kim, Y. J., Shute, V. J.: The interplay of game elements withpsychometric qualities, learning, and enjoyment in game-based assessment. In: Comput. Educ., vol. 87, pp. 340–356 (2015).