Evaluating Playing Experience and Adoption of a Math Learning Game

Antero Lindstedt, Kristian Kiili
Tampere University of Technology
Finland
antero.lindstedt@tut.fi, kristian.kiili@tut.fi

Abstract: The purpose of this study was to explore the playing experience and adoption of a math learning game. Fifty-four Finnish fourth grade students learned rational numbers by playing a number line based math game for five 30 minutes sessions. The game was played with either a laptop or a tablet device. The experienced flow and the perceived playability of the game was evaluated after the first playing session and later when they had achieved a good level of control over the game. Flow and playability were both at good level, but the results showed that they declined over time. There was a strong correlation between flow and playability. The analysis revealed that the outcome of the last game did not cause any significant biases in experienced flow level, but instead the players tended to evaluate their playing experience based on a longer playing period supporting the use of self-reporting flow questionnaires.

Keywords: Flow experience, Playing experience, Game-based learning, Mathematics, User interface

1. Introduction

The enjoyment level that an educational game offers is a key factor in determining whether the player will be engaged in the gameplay, perform according to their skills, and achieve the objectives of the game. Thus, the ability to quantify the playing experience and ability to identify whether the game elements engage or disengage players is an important goal for both industry and academia. The concept of flow is one of the most popular constructs used to describe the playing experience (Procci, Singer, Levy, & Bowers 2012). In fact, Kiili (2005) has argued that games are most successful and engaging if they can produce flow experiences. This study considers playing experience in the game-based learning context. To be more precise, we explored how fourth graders experienced a math learning game with respect to flow experience and perceived playability of the game. In this paper, the playability concept refers to the usability and the mastery of game controls. In particular, we consider how flow experience and perceived playability change over a period of time.

1.1 Rational Number Game Engine

The present work is a part of an ongoing project in which we are developing a rational number game engine called Semideus. The games created with the Semideus engine are based on recent findings on numerical development (McMullen, Laakkonen, Hannula-Sormunen, & Lehtinen, 2015; Ni, & Zhou, 2005; Siegler, Thompson, & Schneider, 2011; Torbeyns, Schneider, Xin, & Siegler. 2015; Vamvakoussi, 2015) and theories that provide an account for the use of manipulatives in digital learning materials (Pouw, Van Gog, & Paas, 2014). Several rational number tasks founded on number lines can be included into the Semideus games, for example, number line estimation, magnitude comparison, and magnitude ordering.

In general, the game engine can be used to create games that aim to support the development of rational number conceptual knowledge and conceptual change processes. In this study, we used the engine to configure a game prototype, Semideus School, which was targeted to 10-12-year old students. The created game, as well as all games created with the Semideus engine, are designed to work as non-invasive assessment instruments (Kiili & Ketamo, 2017; Manuel et al., 2017).

1.2 Flow Experience

Flow describes a state of complete absorption or engagement in a specific activity in which a person excludes all irrelevant emotions and thoughts (Csikszentmihalyi, 1975; Csikszentmihalyi, 1991; Csikszentmihalyi, 2002). During the optimal experience, a person is in a positive psychological state where s/he is so involved with the goal-driven activity that nothing else seems to matter. An activity that produces such experiences is so pleasant that the person may be willing to do something for its own sake, without being concerned with what s/he will get out of the action. This kind of intrinsic motivation is very important especially in learning games that usually require different cognitive investments compared to entertainment games. A systematic literature review about flow experience in in game-based learning indicated that quantitative measurement of flow has been the most common method to study flow in serious games and that flow can be used as a game quality measure in serious games context (Perttula, Kiili, Lindstedt, & Tuomi, 2017). Furthermore, the review identified some evidence that flow has a positive relation with game-based learning outcomes.

1.3 Present Study

The paper reports results from an ongoing project that aims to develop an engaging and effective digital game for training conceptual rational number knowledge. We report the results of an iteration in which we studied how students experienced the first prototype of Semideus School game at two different time points: after a 30 minutes onboarding phase and after playing the game for approximately 2.5 hours.

The following Research questions were investigated:

1. Does the experienced flow level change over time?
2. What is the relation between perceived playability of the game and experienced flow level?
3. Does the result of the game played just before the measurement of flow have a clear influence on the experienced flow level?

Our hypothesis for Research question 1 (Hypothesis 1) is that the experienced flow level decreases over time. This hypothesis is based on our assumption that at the beginning the player may be excited about an alternative for basic math exercises, but the novelty effect probably wears off eventually. Furthermore, we assume that players may get bored with the core game mechanics in the long run as later levels offer no other changes to the game than more difficult tasks. Regarding Research question 2, we hypothesized that perceived playability correlates with flow experience (Hypothesis 2). This hypothesis is based on a motivational model of video game engagement (Przybylski, Rigby, & Ryan, 2010) emphasizing that good playability and mastery of game controls is not implicitly satisfying, but it unlocks game's potential to meet player's psychological needs by letting the player to focus on the actual gameplay instead of the game controls. For Research question 3, we hypothesized that the playing occasion of a game that is played just before flow measurement distorts the reported flow level (Hypothesis 3). A successfully completed game level may raise the flow level. In contrast, a failed attempt at a level may influence negatively on experienced flow level. Furthermore, a level that has been manually quitted could reflect the answers in a slightly negative way as the player may feel frustrated that s/he had to forfeit the level because of a lack of time or because the level has started so poorly that continuing it seems futile.

2. Method

2.1 Participants

Four Finnish fourth grade classes participated in the study. Altogether 98 students were involved in the beginning, but only 54 followed the designed research protocol (played the game enough and completed the required questionnaires) and were included in the study (N = 54). Thirty-two of the participants were females. Based on schools' infrastructure, nineteen of the participants played the game with Chromebooks (Keyboard user interface) and 35 played with iPads (Tilting user interface).

2.2 Description of the Semideus School Game Prototype

The Semideus School game prototype featured six worlds with a total of 62 levels. Each level, a trail to the mountaintop, contained either 10 or 12 tasks. At first levels, the player faced tasks with different kinds of hints to help the player to familiarize her/himself with the game mechanics and rational numbers. The hints were for example coins that marked the correct answer, animated lianas that divided the number line into sections or cave paintings that showed the player what to do (like jump over a trap). After a couple of these introduction levels, the difficulty of the levels raised steadily. The hints were visible occasionally also on later levels when some new concept was introduced. The game also offered the player some tutor features like a possibility to request a goat to show the right answer, a bird to expand the fraction numbers to common denominators, or a bird to place a visual marker on the certain point of the number line. The player was able to activate these tutor features at will by using in-game currency - diamonds. Diamonds were collectible items the player could gather from the game.

The game had two types of levels (Figure 1): number line estimation (34 levels) and magnitude comparison/ordering (28 levels). In the estimation tasks the player was given a value and had to estimate its position on a number line. The player got rewarded with points based on estimation accuracy. In addition, some of the tasks included traps that were hidden and the player had to avoid stepping on them by estimating their position in the same way as with the correct answer. In comparison levels, the player had to compare two stones with values on them, and arrange those into an ascending order by carrying them into positions. The exact spot did not matter as long as the order was correct. Alternatively, the tasks might have included several stones that required ordering. Occasionally the tasks contained equivalent values that required the stones to be piled up. Both types of tasks contained different kinds of numbers: fractions, decimals, whole numbers, mixed numbers and combinations of the aforementioned. Some of the numbers were presented graphically, for example as pie charts.

Figure 1. On the left: an estimation task with some torches as visual hints and a trap as an obstacle. On the right: a comparison task in which an enemy character provides time pressure.

The game was a web based application played through a browser. It was available on two platforms: tablet (iPad) and computer. The versions were equal in every aspect other than the controlling method. With the tablet version, player character’s movement was handled by tilting the device. The movement speed was dependent on the angle of tilting with a steeper angle resulting in faster walking speed. Also the jumping length was dependent on the angle. The keyboard version had keyboard keys for every action, including alternatives for the onscreen buttons. The player character was moved using the arrow keys. The walking speed and jump length was always static. The possibility to adjust the walking speed on tablets was required as the tilting interface would have been less precise without it.

2.3 User Interface Checkpoint Tasks

In order to evaluate players’ game controlling skills (user interface mastery) we included User Interface (UI) checkpoint tasks in the Semideus School game. UI-checkpoint tasks are very simple, almost obvious tasks, that the players should be able to complete accurately regardless of their rational number knowledge level (conceptual understanding about whole numbers was required). In the number line estimation checkpoint task the player had to estimate value ½ on a 0-1 number line within 12 seconds. The correct location was marked with a coin, so it should be very easy to estimate the task accurately and fast. The aim of the comparing checkpoint tasks was to determine whether a player can arrange the stones from smallest to largest (requires picking up, carrying, and dropping stones). Again the tasks were made mathematically very simple by using whole numbers (for example, 5 and 7) in the task. Finally, the piling checkpoint tasks required the player to pile equivalent stones with simple whole numbers (e.g. 5 and 5).

2.4 Measures and Analyses

Flow was measured with a modified version of Flow Short Scale (Engeser, & Rheinberg, 2008). Flow Short Scale measures the flow experience with ten items (7-point scale). We modified the statements of the scale from present tense to imperfect and added references to game playing activity (see Appendix A). Perceived playability was measured with two items developed by the authors: "Controlling of the game character was clear and intuitive." and "Game was easy to use." (7-point scale).

In order to explore how the playing experience changed over time the flow and perceived playability scores were measured at three time points with digital questionnaire. The first and the last measure were used in the analysis. In order to explore whether the previous game had any effect to the answers, the playing data from the last game played within 20 minutes of the submittal of the questionnaire was used in the analysis. Because some of the participants did not play a game within the 20 minutes' time period, 18 participants were omitted from this analysis ($n = 36$) at the first time point and 28 were omitted from the analysis at the second time point ($n = 26$). The game data was logged and stored into a secured server in a format defined by the semantic model of the game's learning analytics engine.

2.5 Procedure

In the beginning each participant was given a randomized id which was used to login to the game and was also required in the questionnaires. This way the game data and questionnaire data could be connected while maintaining participant anonymity. Students played the game for 5 x 30 minutes during a three-week period. The first session for each class was supervised by the authors, who also demonstrated how the game works. After that the teachers of the class conducted the later playing sessions by themselves. The teachers were given instructions on how to run the sessions. A flow questionnaire was filled after the first, the third, and the fifth session.

3. Results

In this chapter T1 refers to the time when the playing experience questionnaire was filled for the first time and T2 to the time when participants filled the questionnaire for the last time.

3.1 Flow and Playability

The internal consistencies of flow short scale at T1 ($\alpha = .85$) and T2 ($\alpha = .90$) were good. The internal consistency of the playability scale at T1 was also good ($\alpha = .86$), but at T2 it was only satisfactory ($\alpha = .70$).

We employed a Wilcoxon signed-rank test to investigate whether players' flow level decreased over time (Research question 1). The difference scores were symmetrically distributed, as assessed by a histogram. Consistent with our Hypothesis 1, a signed-rank test determined that there was a statistically significant median decrease in flow score from T1 (Mdn = 6.1) to T2 (Mdn = 4.6), $z = -5.313$, $p < .001$. The flow score at T1 ($M = 5.82$, $SD = 0.86$) lied clearly above the overall Flow Short Scale mean (4.7) attained with various activities and across various previous studies (Engeser, & Rheinberg, 2008). However, the flow score at T2 ($M = 4.55$, $SD = 1.36$) lied a little bit below the overall Flow Short Scale mean and there was more variation among the players. Furthermore, the discussions with the players revealed that number line estimation tasks were more engaging than comparison and ordering tasks.

Players' first impression of playability (T1) was really good ($M = 5.72$, $SD = 1.35$, Mdn = 6), but when players progressed in the game and faced more complex tasks (T2), like the ones requiring carrying of stones and knocking out enemies, the playability decreased ($M = 4.92$, $SD = 1.66$, Mdn = 5). A wilcoxon signed-rank test showed that playability score decreased statistically significantly, $z = -3.547$, $p < .001$. We employed correlation analyses to investigate the relation between playability and experienced flow level (Research question 2). In line with our Hypothesis 2, flow correlated strongly with playability at both times (T1 and T2), $r = .64$, $p < .001$. This indicates that the playability of the game can have a strong influence on the overall playing experience and engagement level.

In order to address Research question 3, we explored how the result of the last played game influenced on experienced flow level. A Kruskal-Wallis test showed that the end condition of the last game did not influence on the experienced flow level (T1: z =2.067, p = .559, n = 36 [win (18); lose (15); quit (3)]; T2: z = 2.174, p = .537, n = 28 [win (8); lose (14); quit (6)]). These results indicate that the success of the last game did not distort players' answers for Flow Short Scale items (conducted within 20 minutes of the last game), but players tended to evaluate the flow items more generally based on their overall playing experience with the game. This was in contradiction with our Hypothesis 3 where we expected to see more significant effects of the last played game. This result indicates that questionnaires can be used in repeated flow measures.

3.2 User Interface Mastery

We used user interface checkpoint tasks to explore the adoption of the user interface. Figure 2 shows the development of players' accuracy on estimating, ordering and piling UI checkpoint tasks at two time points: newbie (based on the first two checkpoint tasks) and regular (based on the last two UI checkpoint tasks). The term newbie refers to a player who is new to the game and is still getting used to the controls. Regular refers to a player who has already had enough time to familiarize her/himself to the game. As we can see the players' accuracy in estimating checkpoint tasks was very good already at the beginning (M = 98.7%, SD = 2.7%; T2: M = 98.68%, SD = 3.4%). However, the players made a lot of mistakes in the first two comparison checkpoint tasks (M = 63.2%, SD = 38.17%). The comparison accuracy got significantly better over time (M = 86.7%, SD = 26.22%), but was still surprisingly low. The accuracy was better in the piling checkpoint tasks (M = 92.6%, SD = 18.1%: same at both time points) than in the comparison tasks. A likely reason for this was that when the players faced the first piling checkpoint task they were already familiar with picking and dropping stones activities (needed in the comparison tasks) that was also required in the piling tasks.

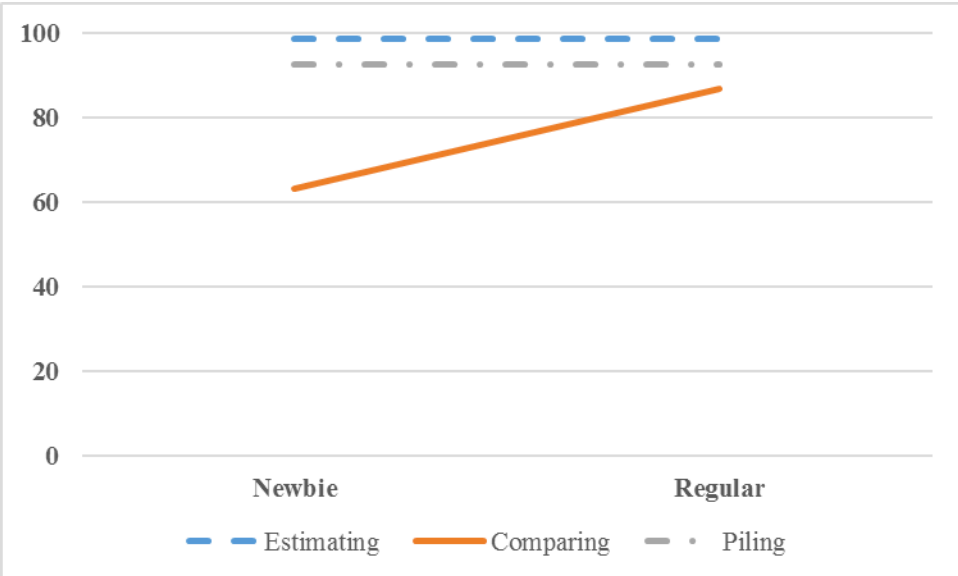

Figure 2. Mean accuracy of UI checkpoint tasks at two time points.

In order to explore reasons for low accuracy in the comparison checkpoint tasks, players' comparison answers were analyzed. The analysis revealed that 13% of the mistakes in comparison checkpoint tasks at the regular phase were wrong because the players ordered the stones from largest to smallest. This is a common careless error that happens also in the paper based tasks every now and then. Nevertheless, the players who ordered stones into wrong order were able to pick up and drop stones indicating that they knew how to use the game controls. The rest of the mistakes were detected to be caused by the stones not being moved at all (stones remained piled). The mean answering duration in these tasks (M = 4.6s, SD = 1.9s) was smaller than the mean answering duration in successfully completed tasks (M = 6.8s, SD = 4.8s). We assume that players have navigated the game character to the stones and then accidentally pressed the answer button instead of the carry button resulting in a wrong answer. These kind of mistakes were not observed with the keyboard version of the game at all, indicating that the user interface of the iPad version was more error prone at least during the newbie phase. Based on this finding we decided to make some modifications for the user interface of the game.

4. Discussion

In general, the participants enjoyed playing the Semideus School game and experienced high levels of flow. This was particularly apparent at the beginning of the study when the game was new for the players. Later the experienced flow level as well as the perceived playability clearly dropped, but remained still moderately high. The decrease of flow may have been a result of the increasing complexity of the tasks and game controls. It is also possible that the core game mechanics of the game did not engage players in the long run. However, also the repetitive filling of the same questionnaire may have had an influence on the answers. Furthermore, the results indicated that the success of the last game did not distort players' answers for Flow Short Scale, but players tended to evaluate their playing experience more generally.

4.1 User Interface Mastery and Game Mechanics

The user interface mastery checkpoints revealed that some of the players had difficulties to adopt the game controls that was needed to order the stones. In the tablet version of the game, some unintended actions were noticed with the carry and answering buttons. The same was not observed in the keyboard version where the keyboard keys did not get mixed up. This might require a redesign of the onscreen answer button or the carry button so that they do not get mixed up. For example, the answer button could be changed to require a double tap. Alternatively, the carry button could be removed from the screen and the activity could be triggered with touch gestures. For example, the carriable stones could be picked up with a swipe gesture upwards and likewise dropping the stone could be handled with a similar motion downwards. However, if players do not feel comfortable to do such gestures with their thumbs and tend to use forefinger instead, these gestures might disturb controlling of game character by tilting the tablet. Thus, the possible controlling schemes have to be studied with controlled experiments in the future.

The discussions with the players revealed that number line estimation mechanic was more engaging than comparison and ordering mechanic. Some of the players explained that the number line estimation is more engaging because you can get better in estimation task all the time (hard to achieve 100% accuracy), but once you master comparison and ordering tasks you always got the full points. In terms of flow, the players who master rational number magnitudes get easily bored with comparison and ordering tasks, but the estimation tasks still provides appropriate challenge for the players. This is to say that the players understood that they can dream about developing sufficient skills to reduce the estimation error close to zero that makes the experience enjoyable. In general, this means that the number line estimation is very promising mechanic for math games.

The study revealed that the end game content was not as engaging to the players as the beginning of the game. To improve this the tutor skill system will be redesigned. More skills will be made available and they are introduced gradually over the game. This way the game offers new features even in the latter stages of the game. The skills will also be represented by different kinds of animal companions that will hopefully add some personality to the skills. In order to make the skills more integral part of the game, we will change their cost so that the use of a skill does not depend on the players collected diamond count. Now the skills are available to every player regardless of the time they have spent with the game. Also using the collected diamonds as currency might have made some players avoid using the skills as they do not want to spend their hard earned diamonds. Now each level offers a set of diamonds that will afford a couple of skill uses depending on the cost of the skill.

4.2 Limitations

There are several limitations in this study. Since most of the playing sessions were administered by the teachers without researchers' supervision, there were some inconsistencies in the playing sessions. For example, some of the questionnaires were not filled at the planned moments. Some players did not fill the questionnaires at all or the questionnaires were submitted so late that they had to be excluded from the analysis decreasing the sample size. In fact, the valid sample size of the study was quite small and thus for example the statistical comparison of differences between tilting and keyboard conditions was not reasonable. Also the qualities of a particular teacher might have influenced on the playing experience. Furthermore, a web based game application was used in this study that may have caused some problems. While the functionality was the same as in a native application, there were some issues that might have occurred if the web application was not handled properly. The game was instructed to be played in full screen mode but it was technically possible to access the game with browser interface still visible. The game worked otherwise fine but an essential part of the UI, the number line values, were cut of the screen as the navigation bar occupied a portion of the screen. Also the tablet version was supposed to be played in a landscape position with a rotation lock on. However, as the tablets were in common use at the school, someone might have turned the lock off. This might have resulted in a lowered game experience as the screen could rotate unintentionally for example while tilting the device. In addition the rotating screen might have left the game in an incorrectly scaled down state where all the game elements appeared smaller, surrounded by some white borders. These issues might have had an effect on the playing experience and playing performance for some participants that we were unable to track. In order to fix above mentioned user interface issues, we have already made a native iOS version of the Semideus School game and published it in Apple store for iPads.

5. Conclusions

In this paper, we studied the adoption of a math game in a primary school. Overall, the study demonstrated that flow experience, playability, and user interface mastery checkpoints can be used to evaluate playing experience and adoption of games in game-based learning context. The results showed that players were highly engaged with the Semideus School game prototype at the beginning, but the level of students' flow experience decreased over time. This indicates that continuously increasing difficulty is not enough to maintain engagement but the game needs to offer varying content and new game mechanics throughout its duration as well. Also the perceived playability decreased over time as the players faced more complex challenges. As the tasks get more demanding, it is even more important that the user interface is good enough to let the players focus on the solving of the task instead of having to concentrate on the UI.

The user interface checkpoint tasks worked in revealing some UI shortcomings such as the occasional presses of the answer button instead of the carry button. The checkpoint tasks could be valuable tools to improve the playing experience and game controlling accuracy. Especially, if the game is used to assess the player's competence in a subject, such as mathematics, the checkpoint tasks could be used to validate the results by eliminating the user interface's effect.

Acknowledgements

This work is a part of a projects funded by Academy of Finland (grant numbers: 274240 and 289140).

References

Csikszentmihalyi, M. (1975). Beyond boredom and anxiety: Experiencing flow in work and play. *Jossey-Bass, San Francisco*.

Csikszentmihalyi, M. (1991). Flow: The Psychology of Optimal Experience. *Harper Perennial, New York.*

Csikszentmihalyi, M. (2002). Flow: The psychology of optimal experience (2nd edition). *Harper & Row, New York.*

Engeser, S., and Rheinberg, F. (2008). Flow, performance and moderators of challenge-skill balance. *Motivation and Emotion*, *32*(3), 158-172.

Kiili, K. (2005). Digital game-based learning: Towards an experiential gaming model. *The Internet and higher education*, *8*(1), 13-24.

Kiili, K., Devlin, K., Perttula, T., Tuomi, P., and Lindstedt, A. (2015). Using video games to combine learning and assessment in mathematics education. *International Journal of Serious Games*, *2*(4).

Kiili, K., & Ketamo, H. (2017). Evaluating Cognitive and Affective Outcomes of a Digital Game-Based Math Test. IEEE Transactions on Learning Technologies, http://doi.org/10.1109/TLT.2017.2687458.

McMullen, J., Laakkonen, E., Hannula-Sormunen, M., and Lehtinen, E. (2015). Modeling the developmental trajectories of rational number concept(s). *Learning and Instruction*, *37*, 14-20.

Ni, Y., and Zhou, Y. D. (2005). Teaching and learning fraction and rational numbers: The origins and implications of whole number bias. *Educational Psychologist*, *40*(1), 27-52.

Ninaus, M., Kiili, K., McMullen, J., & Moeller, K. (2017). Assessing fraction knowledge by a digital game. *Computers in Human Behavior*, *70*, 197-206.

Perttula, A., Kiili, K., Lindstedt, A., Tuomi, P. (2017). Flow experience in game based learning – a systematic literature review. *International Journal of Serious Games, 4*(1), 57-72.

Pouw, W. T., Van Gog, T., and Paas, F. (2014). An embedded and embodied cognition review of instructional manipulatives. *Educational Psychology Review*, *26*(1), 51-72.

Procci, K., Singer, A. R., Levy, K. R., and Bowers, C. (2012). Measuring the flow experience of gamers: An evaluation of the DFS-2. *Computers in Human Behavior*.

Przybylski, A. K., Rigby, C. S., & Ryan, R. M. (2010). A motivational model of video game engagement. *Review of general psychology*, *14*(2), 154.

Siegler, R. S., Thompson, C. A., and Schneider, M. (2011). An integrated theory of whole number and fractions development. *Cognitive psychology*, *62*(4), 273-296.

Torbeyns, J., Schneider, M., Xin, Z., and Siegler, R. S. (2015). Bridging the gap: Fraction understanding is central to mathematics achievement in students from three different continents. *Learning and Instruction*, *37*, 5-13.

Vamvakoussi, X. 2015. The development of rational number knowledge: Old topic, new insights. *Learning and Instruction*, *37*, 50-55.

Appendix A. Modified Flow Short Scale (7-point Scale)

- The game provided just the right amount of challenge.
- My thoughts/activities ran fluidly and smoothly.
- I didn’t notice time passing.
- I could concentrate on playing.
- My mind was completely clear.
- I was totally absorbed in playing.
- The right thoughts/movements occurred of their own accord.
- I knew what I had to do in the game.
- I felt that I had everything under control.
- I was completely lost in thought.