The Effectiveness Evaluation among Different Player-Matching Mechanisms in a Multi-Player Quiz Game

Fu-Hsing Tsai

Teacher Education Center, National Chiayi University, Taiwan // fhtsai@mail.ncyu.edu.tw

(Submitted July 30, 2015; Revised December 22, 2015; Accepted March 13, 2016)

ABSTRACT

This study aims to investigate whether different player-matching mechanisms in educational multi-player online games (MOGs) can affect students' learning performance, enjoyment perception and gaming behaviors. Based on the multi-player quiz game, TRIS-Q, developed by Tsai, Tsai and Lin (2015) using a free player-matching (FPM) mechanism, the same game, but incorporating an automatic player-matching (APM) mechanism and Elo rating system, was developed in order to compare its effectiveness with the original TRIS-Q. The research findings indicate that students using this new player-matching mechanism of TRIS-Q acquired more knowledge of energy, experienced more enjoyment and exhibited more favorable gaming behaviors than did the students using the original TRIS-Q with FPM mechanism. Besides, after comparing the other TRIS-Q with APM mechanism, but using an automatic random player-matching mechanism, this study found that using Elo scoring for automatic matching was the more favorable player-matching mechanism in educational MOGs.

Keywords

Player matching, Educational multi-player online game, Effectiveness evaluation, Game-based learning

Introduction

Competition is ubiquitous in daily life. Political campaigns, sports events, business rivalries and academic contests hosted at schools are all competition-related (Czaja & Cummings, 2009). Competition has been defined as a social process that occurs when people are rewarded on the basis of how their performance compares with that of others performing the same task or participating in the same event (Coakley, 1997). Therefore competitive activities are typically associated with the ambition to win and outperform opponents (Kilduff, Elfenbein, & Staw, 2010). To win, participants draw upon their potential to perform more exceptionally (Tauer & Harackiewicz, 2004). For example, bicycle racers perform better when competing with other racers in a competition than when practicing alone (Triplett, 1898). Competition also triggers the intrinsic motivation of participants and the intention to continue to participate in competitive activities (Deci & Ryan, 1985; Deci, Ryan, & Koestner, 1999). Additionally, competition enables people to learn from their failures, stimulates the learning motivation of participants and prompts the losers to improve themselves and continually strive toward victory (Lam, Yim, Law, & Cheung, 2004). In other words, competition exerts positive effects on performance improvement, participation motivation and learning motivation.

Although competition may enable losers to understand their flaws, and motivate them to learn, it may also reduce their self-confidence and learning motivation due to continual defeats (Füllöp, 2009). Thus, any forms of competition, whether direct, indirect, or cooperative, can also exert negative effects among the losers (Graham, 1976). However, according to Csikszentmihalyi’s flow theory, the flow state, which causes people to thoroughly immerse themselves in the activities, leading to optimal experience (Csikszentmihalyi & LeFevre, 1989), occurs when people concentrate on a competitive activity, and only when people’s skills and the difficulty of competition achieve a certain level of balance (Csikszentmihalyi, 1975; Rheinberg, 2008). In other words, if we attempt to get people immersed in a competitive activity, the skill level among competitors must be equal; when the disparity between skill levels is too large, it prevents the competitors from entering a state of flow, and the competitor with inferior skills can lose interest and confidence in the activities because of continual defeats and frustration. Hence, the ideal approach to reduce the negative effects of competitions is to create a fair competitive environment in which the skills of competitors are balanced, so everyone has a chance to win; that is, a fair player-matching mechanism is critical in competitions.

Because all games involve competition (Bright & Harvey, 1984; Crookall, Oxford, & Saunders, 1987), multi-player online games (MOGs) are also a typical competitive activity that includes multi-player competition against players. Also, MOGs include the general characteristic of digital games, such as fantasy, mystery and control, so they can elicit enjoyment from participants; thus, they have currently become an important recreation activity among young people. In other words, MOGs also reflect the positive effects which competition exerts, and attract many students’ involvement. Therefore, in recent decades, since Prensky (2001) began to advocate...
the use of digital game-based learning (DGBL), more and more researchers have studied educational MOGs; they integrate instructional contents and the features of MOGs, to promote student learning motivation and performance with numerous positive outcomes (Cheng, Kuo, Lou, & Shih, 2012; Tsai, Tsai, & Lin, 2015; Tsai, Yu, & Hsiao, 2012). All of these studies expected educational MOGs to achieve the positive effects of competition; that is, students can repeatedly improve their knowledge while engaging in games, in order to win, or outperform other players.

For example, Tsai et al. (2015) developed an educational MOG, TRIS-Q, in an online learning environment to promote students’ self-assessment and learning about energy by playing the MOG TRIS-Q is a multi-player quiz game which combines the tic-tac-toe game with multiple choice tests; that is, the first player to create a row of three pieces in a nine-square grid game board wins. Players need to answer random multiple choice questions related to energy knowledge at each step, when placing their piece on the game board (the game will be introduced in detail later). The game was developed so that students can continually and repeatedly improve their energy-related knowledge and participate in the game in order to defeat their competitors. The results of that study revealed that this game enhanced learning motivation and performance; in addition, providing immediate elaborated feedback resulted in optimal student learning. However, when the researchers also noted that the learning performance of the students playing the game might be affected by inappropriate player matching, they recommended further investigation of this problem. In other words, Tsai et al. (2015) also support the view that appropriate player matching is crucial in competitions.

Numerous scholars (Graepel & Herbrich, 2006; Véron, Marin, & Monnet, 2014) also maintain that providing a fair player-matching mechanism is critical in commercial MOGs. However, few studies have investigated this issue in educational MOGs. Therefore, the research suggestion by Tsai et al. (2015) is worth investigating. This study aims to follow their suggestion and research whether providing different player-matching mechanisms in TRIS-Q can positively affect learning effectiveness. To investigate this issue, the player-matching mechanism used in the TRIS-Q was first analyzed. This study found that TRIS-Q applied a lobby-based matching mechanism incorporated in past commercial MOGs: a game lobby is provided in which players can freely configure their own match (Berman & Bruckman, 2001). In this free player matching (FPM) mechanism, gamers can play with their friends or select opponents whose skills match their own. However, players may need to invest much time to search for opponents (Aikawa, Hei, Ogishi, Niida, & Hasegawa, 2013). Moreover, some players can repeatedly compete with the same opponents, resulting in cheating, whereby players gain points by conspiring to take turns beating each other.

Hence, many commercial MOGs currently adopt an automatic player-matching (APM) mechanism according to player skill, in order to provide a fair and effective competitive environment; these games always have a suitable skill rating system (Graepel & Herbrich, 2006). For example, a current popular commercial MOG, League of Legends, uses an APM mechanism and matches players who have nearly equal skill levels, primarily by using the Elo rating system (Shores et al., 2014) developed by Arpad Elo for assessing the relative skill levels of chess players, based on statistics (Elo, 1978). The traditional Elo rating system displays the score of each chess player from 0 to 3,000, based on the assumption that the performance of each player is a normally distributed random variable (Hacker & Von Ahn, 2009), so that the average player’s score is 1500. Because the initial skill level of each player is unknown, each player starts with a score of 1,500, which is then gradually modified according to the player’s performance in subsequent games (Glickman, 1995). The higher the score becomes, the higher the skill level a player possesses. Therefore, when two contestants compete, the competitor with the higher score is predicted to have the higher probability of winning the game. Assume \( R_A \) and \( R_B \) represent the rating scores of Contestants A and B, respectively. According to the Elo rating theory, the expected score of Contestant A winning the game can be expressed in the following formula (Elo, 1978):

\[
E_A = \frac{1}{1 + 10^{(R_B - R_A)/400}}
\]

Assume the original scores of Contestants A and B are 1,700 and 1,500, respectively. According to the aforementioned formula, the expected score of Contestant A (Contestant B) is .76 (.24); in other words, the higher the rating score, the greater the predicted probability that a player will win. Because the real skill level of each contestant is an unknown variable, the Elo rating system gradually adjusts the value representing the skill level of each contestant according to the following formula and the aforementioned expected scores:

\[
R'_A = R_A + K(S_A - E_A)
\]

where \( R'_A \) represents the new rating score of Contestant A after a game; \( R_A \) represents the original rating score of Contestant A. If Contestant A wins the game, then \( S_A = 1.0 \); if the game is a draw, then \( S_A = 0.5 \). If Contestant A
A relatively stable rating score representing the skill level of each contestant is analytically stable. To ensure fair play, each group was assigned a rating system, with one group that played the TRIS Q, another that played the TRIS P, and the third that played the TRIS R. This design allowed for the examination of various mechanisms' effects. However, since players cannot choose their opponents, a disadvantage is that players can rapidly compete with random opponents, while the advantage is that players cannot choose their opponents by personal preference. For example, Claypool, Decelle, Hall and O’Donnell (2015) found that League of Legends players think that most games lack balance and are not fun, although an analysis of game data showed that most games are balanced.

In summary, competitive activities have both advantages and disadvantages. To enable people to immerse themselves in competitive activities and minimize the negative effects of such activities, creating a fair competitive environment is necessary. Therefore, appropriate player-matching mechanisms are crucial in MOGs. Currently, two types of player-matching mechanisms are generally incorporated in commercial MOGs, namely the FPM and APM mechanisms, both of which have strengths and weaknesses. However, few studies have examined which of these two types of mechanisms is more appropriate for application in an educational MOG. Therefore, in this study, the effectiveness of the FPM and APM mechanisms was investigated in depth. On the basis of Tsai et al.’s (2015) TRIS-Q, which is a multi-player quiz game based on the FPM mechanism (hereafter referred to as TRIS-Q-FPM), this study developed the same game, but incorporated an APM mechanism (hereafter referred to as TRIS-Q-APM), and subsequently compared the TRIS-Q-APM and TRIS-Q-FPM. In addition, according to the approaches employed in commercial MOGs, the Elo rating system was incorporated into TRIS-Q-APM as a game scoring mechanism and the basis for automatic player matching. However, to verify the applicability of the Elo rating system in the APM mechanism, two types of TRIS-Q-APM, one game involving APM based on Elo rating scores (TRIS-Q-APM-Elo) and the other game incorporating APM based on random matching (TRIS-Q-APM-Ran), were designed for this study. The TRIS-Q-APM-Ran game was used as an experimental control group involving an unfair player-matching mechanism. Specifically, the objectives of this study were as follows:

- Compare the effectiveness of the TRIS-Q-FPM and TRIS-Q-APM on the knowledge acquisition of students playing the games.
- Compare the degree of enjoyment among students who played the TRIS-Q-FPM game with those who played the TRIS-Q-APM game.
- Compare the gaming behavior of students who played the TRIS-Q-FPM game with those who played the TRIS-Q-APM game.
- Examine the applicability of the Elo rating system in the TRIS-Q-APM game.

Methods

Research design, participants and procedures

A non-equivalent pre-test–post-test control group design was adopted in this study to investigate how the multi-player quiz games with different player-matching mechanisms influence the learning effectiveness. Hence, the independent variable was the player-matching mode divided into three groups: one group that played the TRIS-Q-FPM (FPM group) and the other two groups that played the TRIS-Q-APM-Elo (APM-Elo group) and TRIS-Q-APM-Ran (APM-Ran group) games, respectively. The dependent variables were students’ acquisition of energy knowledge, enjoyment perception and gaming behaviors. Knowledge acquisition was measured based on the score that each participant acquired from the energy knowledge test. The enjoyment perception was assessed based on the score that each participant obtained on the game enjoyment scale. The gaming behaviors were measured based on each participant’s gaming records, including the number of games participated in, the number of questions answered in the game and the correct answer ratio in the game. In addition, the applicability of the Elo rating system in the TRIS-Q-APM game was explored. Thus, the participants’ Elo game scores and tic-tac-toe skills between two APM groups were specifically analyzed. The tic-tac-toe skill was assessed based on the score that each participant obtained from the tic-tac-toe ability test.
The study participants comprised students from three randomly sampled 12th grade classes (Classes A, B and C with 40, 41 and 33 students, respectively) at a high school in Kaohsiung, Taiwan. After excluding some participants who could not complete the experiment, the FPM group comprised 28 students from Class A (11 boys and 17 girls), the APM-Elo group included 27 students from Class B (9 boys and 18 girls) and the APM-Ran group consisted of 25 students from Class C (17 boys and 8 girls). In correspondence with the energy technology unit of the living technology course in these three classes, the experiment lasted a total of 4 weeks (one 50-min class per week). Before the experiment, a pre-test was conducted on student knowledge of energy. At the beginning of the experiment, the teachers instructed the participants on how to log into the online learning systems and play the multi-player quiz game. Subsequently, because all of the participants had acquired basic knowledge related to energy in a junior high school course before the experiment, the participants were asked to read the online learning materials constructed by Tsai et al. (2015), including three energy knowledge topics (sources of energy, application of energy, and energy conservation and new energy) for at least 20 min. After that, all of the participants began playing the game assigned to them. After the experiment ended, all of the participants were administered an energy knowledge post-test and tic-tac-toe ability test, and participants rated their perceived enjoyment when playing the game by using a scale.

**Instruments and materials**

**TRIS-Q-FPM**

To examine the effectiveness of the game-based assessment in online learning, Tsai et al. (2015) developed a TRIS-Q game. This game, referred to as TRIS-Q-FPM in this study, was administered to the FPM group in this study, and included in Tsai et al.’s (2015) online learning system. After logging into the online learning system and entering the game via a general web browser, a game lobby is displayed (Figure 1). This interface enables participants to freely search for opponents. Participants can investigate the performance (number of wins and losses) of online participants by clicking the user name in the user list, to interact with other online players publicly or privately, and to challenge someone to compete by typing in text. Subsequently, the participants can create a “game room” with or without password and wait for other participants to join, or join a game room created by other participants in the competing list, and begin playing.

![Figure 1. The interface of the game lobby in TRIS-Q-FPM](image)

When a game starts, a tic-tac-toe board is displayed on the computer screen (step 1 of Figure 2) and used by two competing players. The basic game rule is similar to that of a conventional tic-tac-toe game, in which the first player to create a row of three pieces wins. However, each time a player places a piece on the board, a random multiple-choice question appears selected from the database, which had 149 test items related to the three primary energy knowledge topics from the online learning contents in the online learning system (step 2 of Figure 2). The player must answer the question correctly to retain the piece on the desired board location (step 3-2 of Figure 2); if the player answers incorrectly, the opponent’s piece is placed at the location instead (step 3-1 of Figure 2). Also, when completing a turn and waiting for their opponent’s turn, the left side of the game screen will show the previous question, immediate feedback and elaborated feedback to show whether their answer is correct, and to offer clues relevant to the question for the player (step 3-1, 3-2 of Figure 2). Additionally, as depicted in Figure 1, the function of the ranking board and answer history are provided in the TRIS-Q-FPM interface. The ranking board lists the top 10 players with the most favorable performances in the tic-tac-toe game and the most correct answer ratios; the answer history function enables participants to review all of the test questions and clues they have answered.
Hence, if the participants want to be the top 10 players, they must try to repeatedly defeat their competitors for acquiring the game scores. At the same time, in order to defeat their competitors, players need to try to answer the energy questions correctly and constantly improve their energy knowledge in multiple ways. For example, the participants can improve their energy knowledge by reading each question’s clue or by rote when competing or using the answer history function. Students can also acquire energy knowledge by exiting the game for reading the online learning contents in the online learning system. Accordingly, this game can make learning happen.

Figure 2. The process of taking a turn in TRIS-Q-FPM

TRIS-Q-APM

According to the TRIS-Q developed by Tsai et al. (2015), this study created an APM type of TRIS-Q, TRIS-Q-APM, for the APM groups. This game was designed to be embedded in Tsai et al.’s (2015) online learning system and can also be run in general web browsers. After entering the game, participants are shown the main game menu (Figure 3, left); they can click on the button at the lower right to begin the game.

Figure 3. The process of starting a new competition in TRIS-Q-APM

The game system then automatically matches participants with opponents, and the screen depicted on the right of Figure 3 appears. When two competitors both press the start button, the game begins. As mentioned, to verify the
applicability of the Elo rating system for player matching, the TRIS-Q-APM games were divided into two versions according to the automatic player-matching approaches: In the TRIS-Q-APM-Elo, the computer matches a player automatically with an opponent who possesses a similar Elo score to that of the player, and in the TRIS-Q-APM-Ran, a player is randomly matched with an opponent.

The game rules and the playing process of TRIS-Q-APM are the same as TRIS-Q-FPM: that is, during a game, a tic-tac-toe board is also displayed (Figure 4, left) and used by two competitors. The players are also required to answer questions (Figure 4, right) randomly selected from the same test items database with TRIS-Q-FPM, and feedback is generated according to the answers the players provide. In addition, as depicted in Figure 3, similar to the TRIS-Q-FPM, the TRIS-Q-APM features a ranking list and answer history function. In other words, the competition strategy and the multiple ways for making players obtain energy knowledge are also the same as for TRIS-Q-FPM. However, because the TRIS-Q-APM employs the Elo rating system to calculate game scores, the game performance ranking is based on the Elo scores of players rather than on the number of wins and losses.

Figure 4. The screenshot of playing TRIS-Q-APM

Energy knowledge test

A self-developed energy knowledge test was administered to understand the changes in the knowledge acquired by the students after they played the game. The test questions were created according to the content of the quiz game questions on energy knowledge. The test comprised 25 multiple-choice questions, and the full score was 100. The validity of the questions was verified by two high school teachers. According to a pre-test administered to 36 12th grade students, the Kuder–Richardson reliability of the test was .67, the average difficulty was .65 and the average discrimination was .31.

Tic-tac-toe ability test

A tic-tac-toe ability test revised by Tsai et al. (2015), according to the version created by Crowley and Siegler (1993), was administered to evaluate the tic-tac-toe skills of the participants. The test has 16 questions, each of which comprises an unfinished tic-tac-toe game; the participants select what they believe to be the optimal move required to finish the game. Each question has a correct answer regarding the optimal move. Thus, the tic-tac-toe skill of each participant is ascertained. The Kuder–Richardson reliability of the test was .78 through a pre-examination study with 36 12th grade students.

Enjoyment perception scale

A self-developed scale was used to evaluate the enjoyment the participants perceived when they played the multi-player quiz games. In reference to the scale on enjoyment developed by Downs and Sundar (2011), this scale was used to evaluate participant feelings on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). The participants responded to the following eight questions: (1) I liked this game; (2) I found this game boring; (3) I enjoyed the competition while playing this game; (4) This game constantly frustrated me; (5) This game provided me with an enjoyable experience; (6) I think this was an exciting game experience; (7) I found this game challenging; and (8) I like this type of competitive game. The Cronbach α of this scale was an acceptable .85.
Results

Comparisons among different player-matching groups on students’ knowledge acquisition

To compare whether different player-matching mechanisms in a multi-player quiz game can affect students’ learning performance, the player-matching mechanisms (FPM, APM-Elo and APM-Ran) were designated as the independent variables, and the pre-test and post-test scores from the energy knowledge test were set as the covariates and dependent variables, respectively, for a one-way analysis of covariance (ANCOVA). Before the ANCOVA was performed, a test of within-class regression coefficient homogeneity was conducted; the result obtained was $F(2, 74) = .293, p = .747$, indicating that it fulfilled the basic assumption of ANCOVA. The result of ANCOVA was $F(2, 76) = 8.350, p = .001, \eta^2 = .180$, revealing that after the effect of the covariate was removed, the effect of different player-matching mechanisms on the energy knowledge post-test scores differed significantly. As revealed in the post hoc test (Table 1) using the LSD (least significant difference) method, the adjusted post-test average scores of the APM-Elo group (74.35) and the APM-Ran group (78.02) were both significantly higher than those of the FPM group (68.84). However, the average scores of the APM groups differed nonsignificantly. This finding revealed that the students who accepted the TRIS-Q with the APM mechanism acquired more energy knowledge than those who played the FPM mechanism game after the period of the experiment. It implies that the APM mechanism is more appropriate for application in TRIS-Q to promote students’ acquisition of energy knowledge.

Table 1. Post hoc comparisons (LSD) of mean differences on energy knowledge post-test

<table>
<thead>
<tr>
<th></th>
<th>Adjusted mean</th>
<th>FPM</th>
<th>APM-Elo</th>
<th>APM-Ran</th>
</tr>
</thead>
<tbody>
<tr>
<td>FPM</td>
<td>68.84</td>
<td>-</td>
<td>.015*</td>
<td>&lt;.001*</td>
</tr>
<tr>
<td>APM-Elo</td>
<td>74.35</td>
<td>-</td>
<td>-</td>
<td>.114</td>
</tr>
<tr>
<td>APM-Ran</td>
<td>78.02</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Note. *p < .05.

Comparisons among different player-matching groups on students’ enjoyment perception

To compare whether different player-matching mechanisms in a multi-player quiz game can affect students’ enjoyment perception, the player-matching mechanisms (FPM, APM-Elo and APM-Ran) were designated as the independent variables, and the scores (average) from the game enjoyment scale were defined as the dependent variables for one-way analysis of variance (ANOVA). The result of Levene’s test for homogeneity of variances was $F(2, 77) = 2.872, p = .063$ and did not contradict the assumption of the homogeneity of variance. The ANOVA result indicated that the degrees of enjoyment experienced by participants in the three groups did not differ significantly, $F(2, 77) = 2.123, p = .127$. The LSD post hoc comparisons (Table 2) revealed that the average enjoyment score of the APM-Elo group (3.80) was significantly higher than that of the FPM group (3.47), but did not differ significantly from that of the APM-Ran group (3.58). This result revealed that most students had positive perception regarding the multi-player quiz game, no matter what player-matching mechanism was used. It means that game-based assessment can bring enjoyment for most students. However, the result also implies that using the Elo rating system for game scoring and automatic player matching is more appropriate for application in TRIS-Q to promote students experiencing enjoyment because the enjoyment perceived by the APM-Elo group was the highest among the three groups, and was significantly higher than that of the FPM group.

Table 2. Post hoc comparisons (LSD) of mean differences on enjoyment perception

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>FPM</th>
<th>APM-Elo</th>
<th>APM-Ran</th>
</tr>
</thead>
<tbody>
<tr>
<td>FPM</td>
<td>3.47</td>
<td>-</td>
<td>.045*</td>
<td>.487</td>
</tr>
<tr>
<td>APM-Elo</td>
<td>3.80</td>
<td>-</td>
<td>-</td>
<td>.202</td>
</tr>
<tr>
<td>APM-Ran</td>
<td>3.58</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Note. *p < .05.

Comparisons among different player-matching groups on students’ gaming behaviors

To compare whether different player-matching mechanisms in a multi-player quiz game can affect students’ gaming behaviors, the player-matching mechanisms (FPM, APM-Elo and the APM-Ran) were designated as the independent variables, and the gaming behaviors observed in this study (e.g., the number of games participated
in, the number of questions answered in the game and the correct answer ratio in the game) were defined as the dependent variables for one-way ANOVA.

First, regarding the total number of games participated in, the result of Levene’s test before the ANOVA was $F(2, 77) = 2.817$, $p = .066$ and corresponded with the assumption of the homogeneity of variance. The result of ANOVA was $F(2, 77) = 9.380$, $p < .001$, indicating that the number of games participated in differed significantly among the three groups. The LSD post hoc comparisons (Table 3) revealed that the average numbers of games participated in by participants in the APM-Elo group (17.22) and APM-Ran group (18.96) were significantly higher than that in the FPM group (11.96); however, the number of games engaged in by participants in the two APM groups differed nonsignificantly. This result showed that the students who accepted multi-player quiz game with APM mechanism participated in more competitions than those who played the FPM mechanism game during the period of the experiment. It implies that the APM mechanism is more appropriate for application in TRIS-Q to enhance student opportunities or motivations involving game-based assessment.

<table>
<thead>
<tr>
<th>Table 3. Post hoc comparisons (LSD) of mean differences on the number of games participated in</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>FPM</td>
</tr>
<tr>
<td>APM-Elo</td>
</tr>
<tr>
<td>APM-Ran</td>
</tr>
</tbody>
</table>

Note. $p < .05$.

Second, regarding the number of questions answered in the quiz game, the result of Levene’s test before the ANOVA was $F(2, 77) = 1.161$, $p = .319$, and corresponded with the assumption of the homogeneity of variance. The result of the ANOVA was $F(2, 77) = 8.922$, $p < .001$, indicating that the number of questions answered by participants in the three groups differed significantly. The LSD post hoc comparisons (Table 4) revealed that the average numbers of questions answered in the APM-Elo group (54.70) and APM-Ran group (62.80) were significantly higher than those answered in the FPM group (41.36); however, the number of questions answered in the two APM groups differed nonsignificantly. This finding revealed that the students who accepted the multi-player quiz game with the APM mechanism answered more questions than those who played the FPM mechanism game during the same duration. It also implies that the APM mechanism is more appropriate for application in TRIS-Q to enhance student opportunities or motivations in participating in online tests.

<table>
<thead>
<tr>
<th>Table 4. Post hoc comparisons (LSD) of mean differences on the number of questions answered in the game</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>FPM</td>
</tr>
<tr>
<td>APM-Elo</td>
</tr>
<tr>
<td>APM-Ran</td>
</tr>
</tbody>
</table>

Note. $p < .05$.

Finally, regarding the correct answer ratio in the quiz game, the results of Levene’s test before the ANOVA was $F(2, 77) = 9.397$, $p < .001$ and did not support the assumption of the homogeneity of variance; therefore, a Brown–Forsythe F test was applied for analysis (Field, 2009). The result revealed that the correct answer ratio in the multi-player quiz game differed significantly among the participants in the FPM and APM groups ($Brown–Forsythe F(2, 47.38) = 13.070$, $p < .001$). The Games–Howell post hoc comparisons (Table 5) showed that the average ratio of correct answers in the APM-Elo group (.85) and the APM-Ran group (.84) were both significantly higher than those of the FPM group (.73), but the two APM groups differed nonsignificantly regarding this gaming behavior. This finding revealed that the students who accepted the multi-player quiz game with the APM mechanism answered more questions correctly than those who played the FPM mechanism game during the same duration. It implies that the APM mechanism is more appropriate for application in TRIS-Q to enhance students’ correct answer ratios in game-based assessment.

<table>
<thead>
<tr>
<th>Table 5. Post hoc comparisons (Games–Howell) of mean differences on the correct answer ratio in game</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>FPM</td>
</tr>
<tr>
<td>APM-Elo</td>
</tr>
<tr>
<td>APM-Ran</td>
</tr>
</tbody>
</table>

Note. $^*p < .05$. 220
Applicability of the Elo rating system in the multi-player competitive quiz game

To examine the applicability of the Elo rating for player matching in multi-player quiz games, an analytical comparison was performed specifically on the two APM groups; it incorporated the Elo rating system to calculate the game scores. In both TRIS-Q-APM-Elo and TRIS-Q-APM-Ran, each player started with a game score of 1,500, and the k value was designated as 30. According to the theory of the Elo rating system, the overall average game score of the participants should be 1,500 and the scores should be normally distributed. Therefore, the score statistics and distributions of the two APM groups were first compared. The average score of the APM-Ran group was 1,517.88 (SD = 43.22), and that of the APM-Elo group was 1,500.67 (SD = 47.09); thus, the average score of the APM-Elo group was closer to 1,500. Figure 5 presents the frequency and normal distribution graphs of the game scores obtained by the two APM groups, in which the scores of the APM-Elo and APM-Ran groups have a standard normal distribution and negative skewness, respectively. According to the average scores and score distributions of these two groups, the scores acquired by the players of TRIS-Q-APM-Elo correspond more closely to the theory of the Elo rating system.

Moreover, according to the theory of the Elo rating system, the Elo score obtained by each player from multiple games should finally reflect the player’s gaming skill. Because TRIS-Q involves board-gaming skills and energy knowledge, theoretically, a player with more advanced tic-tac-toe skills and energy knowledge should perform more outstandingly in TRIS-Q; in other words, the tic-tac-toe skill and energy knowledge of a player in TRIS-Q affects the gaming skill of the player. Therefore, if the Elo game score can reflect each player’s gaming skill, then each player’s tic-tac-toe skill and energy knowledge should affect their Elo game score. To determine which group’s game scores reflected the gaming skills of the participants more accurately, a multiple regression analysis was conducted. A simultaneous regression was executed on the two APM groups by using the tic-tac-toe ability test scores and the energy knowledge post-test scores as the independent variables, and the Elo scores of the participants as the dependent variables. The results (Table 6) revealed that in the APM-Ran group, the two independent variables had little effect on the Elo scores; the explanatory power of the regressions (R-squared) was only .040, and the regression effect was nonsignificant (F(2, 22) = .453, p = .641). Conversely, in the APM-Elo group, the two independent variables exerted a greater effect on the Elo scores; the explanatory power of the regressions (R-squared) was .257, and the regression effect achieved significance (F(2, 24) = 4.156, p = .028). In summary, compared to that of the APM-Ran group, the Elo game score of the APM-Elo group reflected the gaming skill of the participants more accurately, and corresponded more closely to the theory of the Elo rating system. It implies that using the Elo rating system for game scoring and automatic player matching is more appropriate for application in TRIS-Q to provide a fair competitive environment.

Table 6. Comparisons of regression analysis between two APM groups for the relationships among Elo score, tic-tac-toe ability and energy knowledge

<table>
<thead>
<tr>
<th></th>
<th>APM-Ran</th>
<th></th>
<th>APM-Elo</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Beta</td>
<td>t</td>
<td>p</td>
<td>Beta</td>
</tr>
<tr>
<td>tic-tac-toe ability</td>
<td>-.198</td>
<td>-.948</td>
<td>.353</td>
<td>.327</td>
</tr>
<tr>
<td>energy knowledge</td>
<td>.016</td>
<td>.075</td>
<td>.941</td>
<td>.346</td>
</tr>
<tr>
<td>R²</td>
<td>.040</td>
<td>.257</td>
<td>.</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>F (2, 22) = .453, p = .641</td>
<td>F (2, 24) = 4.156, p = .028</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Discussion and conclusion

Based on a literature review, this study found that an appropriate player-matching mechanism is vital in educational MOGs, and also that the two types of player-matching mechanisms (FPM and APM) generally incorporated in commercial MOGs have their respective strengths and weaknesses. However, few studies have examined whether these two types of mechanisms in educational MOGs can affect students’ learning performance, enjoyment perception and gaming behaviors. Because the original TRIS-Q adopted an FPM mechanism for player matching, this study developed the same game, but incorporated an APM mechanism to compare its effectiveness with the original TRIS-Q. In addition, this study developed two types of APM mechanism for TRIS-Q: automatic player-matching by the Elo score and automatic random player-matching, to facilitate comparing the applicability of the Elo rating system in multi-player quiz games.

The experimental results indicated that students using an APM mechanism of TRIS-Q, whether based on Elo score or random matching, acquired more energy knowledge, experienced more enjoyment, invested more time playing the games and answering the questions, and answered more questions correctly than did the students using the original TRIS-Q with the FPM mechanism. The findings confirmed that the player-matching mechanism is important and can affect the learning effectiveness in educational MOGs. It was also consistent with the results of previous studies (Manweiler, Agarwal, Zhang, Roy Choudhury & Bahl, 2011), which had indicated that the FPM mechanism cost the participants excessive time in finding opponents, because compared with the TRIS-Q-APM, TRIS-Q-FPM severely reduced the number of games participated in and the number of questions answered. Besides, the most important finding is that it implied that the APM mechanism is more appropriate for application in TRIS-Q than is the FPM mechanism. Because this new player-matching mechanism accelerated the speed of finding the opponent in TRIS-Q, students using the APM mechanism of TRIS-Q perceived more enjoyment than did the students using the FPM mechanism of TRIS-Q. Accordingly, it further enhanced student motivation to spend more time playing the game. Because the more frequently that participants played the game, the more likely that the game helped students to improve their energy knowledge; therefore, this important finding seems to be reasonable.

However, according to the research findings, the players’ score distribution in TRIS-Q-APM-Elo was more consistent with the theory of the Elo rating system than was the players’ score distribution in TRIS-Q-APM-Ran, and the players’ Elo scores in TRIS-Q-APM-Elo represented their gaming skill more accurately. In other words, using the Elo scores to match players automatically was more meaningful than randomly matching opponents because this approach enabled a fairer competitive environment. This could be the reason why students using the TRIS-Q-APM-Elo experienced more enjoyment than those using TRIS-Q-APM-Ran. Hence, after comparing the two APM mechanisms of TRIS-Q, the finding of this study further implied that using Elo score matching for the APM mechanism is more appropriate for application in TRIS-Q than using the random matching for the APM mechanism. In summary, to create an effective player-matching mechanism for TRIS-Q, using APM mechanism and applying Elo scores as the standard for player matching constitute the best way. Because the TRIS-Q-APM-Elo not only provided a fair competitive environment by automatically matching players with similar Elo scores, but also created an effective player-matching mechanism for rapidly matching a new competition, the players’ enjoyment in TRIS-Q-APM-Elo was significantly greater than those who played the TRIS-Q-FPM. Consequently, students in TRIS-Q-APM-Elo were willing to spend more time playing the game repeatedly and thereby gaining more energy knowledge than those in the original TRIS-Q.

Although the research results in this study apparently can be applied widely in various educational MOGs, improvements upon this study in further studies are still needed. For example, the more accurate reasons why TRIS-Q-APM-Elo is the best player-matching mechanism to promote higher learning performance need to be researched further in future studies. The major method by which students acquire energy knowledge from TRIS-Q-APM-Elo also needs to be explored in depth in the future. Moreover, the Elo rating system using in TRIS-Q-APM-Elo can be studied and improved further. For example, the problem of freezing and inflation are well-known phenomena in the Elo rating system, especially when some players stop playing once they reach the top score or re-create multiple new accounts once they lose interest (Regan, Macieja & Haworth, 2012). Hence, in the future, larger samples are required to verify the results, and the duration of the experiment may be extended to expand the Elo score range of the participants, in addition to further examining the accuracy of the Elo scores and verifying if the problems of freezing or inflation still exist in TRIS-Q-APM-Elo. Alternatively, a dynamic k value in the Elo rating system is crucial to prevent the problem of inflation, and to accelerate the speed in finding players’ correct Elo value (Véron, Marin & Monnet, 2014). Therefore, the TRIS-Q-APM-Elo can be further enhanced by dynamically adjusting the k value for finding a suitable algorithm to rapidly converge players’ Elo scores and prevent the problems of the Elo rating system.
Acknowledgements

The work was supported by the Ministry of Science and Technology, Taiwan under contract numbers NSC 102-2511-S-415-013-MY2 and MOST 104-2511-S-415-010-MY2.

References


