



Research on User Experience Optimization of Tutorial Design for Battle Royale Games Based on Grey AHP Theory

Jinghan Lin^(✉), Wei Zhou, and Si-Si Yuan

User Experience Research Center, Thunder Fire Studio, NetEase,
Hangzhou, People's Republic of China
linjinghan@corp.netease.com

Abstract. In recent years, the battle royale game has become one of the main leisure activities in people's daily life. Due to diverse and endless gameplays in all kinds of battle royale games, the novice tutorial plays a significant role in helping the player master the game and quickly get involved. A good novice tutorial can help players understand and start the game very quickly which also reduces the cognition bias. The user experience of the novice tutorials has a decisive influence on players' retention and payment ratio. In the high-frequency iterative game design, evaluating user experience can help designers know about the novice tutorial's quality effectively and promptly improve the potential problems that players may encounter.

For most battle royale games, their novice tutorials not only possess the attributes of guiding players, but they also own the game attributes. Thus, it is hard to measure all the factors that affect the user experience of novice tutorials. It is a system that exterior might seem clear while the inside is vague. On account of these characteristics, this study applies grey AHP theory to construct a system evaluating the user experience of novice tutorials in battle royale games. It helps game designers evaluate the quality of the user experience objectively and obtain valuable instructions for game design. It also provides support for design iteration.

The study is as follows. Several representative emotional words related to the battle royale games were selected and novice tutorials of three battle royale games were adopted as experimental samples for perceptual evaluation experiment. Then, a system evaluating the user experience of novice tutorials in the battle royale games was established based on the experimental data. The system consisted of 3 levels and 11 level 2 indicators. The interconnect degree of comparison sequence and reference sequence were calculated using the Grey relationship analysis method. Quantitative analysis about the importance of each experience dimension was then conducted according to the interconnect degree from small to big. At last, comprehensive evaluation was performed on the obtained interconnect degree of the design plans.

The conclusions are as follows. (1) We established a model evaluating the user experience of novice tutorials in the battle royale games so that designers can evaluate the user experience quality of the design plans during the development stage. (2) We verified the possibility that the Grey AHP Theory can be applied in evaluating the user experience of novice tutorials in the battle royale games as well as other games and provide strategic guidance for game design.

The research conclusions of this paper are mainly as follows: (1) Created a user experience evaluation model for the novice tutorial of the battle royal games. Designers can use this model to evaluate the quality of user experience of the design plans during game development and optimize the design in a scientific way. (2) Verified the possibility of analytic hierarchy process and grey correlation analysis for experience evaluation of novice tutorials in the battle royal games, and can be extended to experience evaluation of other games and provide strategic guidance for gaming experience design.

Keywords: Gaming experience · Perceptual evaluation · The novice tutorial · The grey AHP theory

1 Introduction

User experience (UE or UX) was firstly proposed by designer Donald Norman (1997) and widely recognized in the mid-1990s. According to ISO 9241-210 (1998), user experience refers to all emotions, beliefs, preferences, feelings, physical and psychological responses, behaviors, and sense of accomplishment in the processing of how users use products and services¹. That is, the physical and psychological feelings that users may have when interacting with a product, system or service. In recent years, user experience design has been widely used in the field of game design. Perspectives to study user gaming experience are diverse which also leads to the diversity of evaluation methods that are used. In general, there are two kinds of methods, one is the direct evaluation method, and another is to construct a user experience evaluation model at first. The direct evaluation method is usually qualitative that analyzing a series of physiological data of the player measured by an instrument as to evaluate the user experience level. Commonly used methods include eye movement and Electroencephalogram (EEG) experiments. While the method to construct an evaluation model is usually quantitative or a combination of quantitative and qualitative analyses. It mainly uses some mathematical and statistical methods, such as multiple regression analysis, to establish a model that assesses the relationship between user experience elements and ultimate goal. Since the concept of users' gaming experience is abstract and ambiguous and contains many elements, it is hard to fully reflect the real situation of user experience by a single method. Exploring the dimensions of how to evaluate users' gaming experience and making users' gaming experience specific becomes necessary. The basic idea is to divide a complex system into multiple indicators to reflect the overall situation. Bai (2019) applied theories about Kansei engineering and used the semantic differential method and Oneway ANOVA to study the differences of the player's perception influenced by each system in card games². Using League of Legends as an example, Wang and Zhao (2014) studied the user's gaming experience from the perspective of senses, operation, and interaction³. Combined with the previous studies, it is practicable to use quantitative methods to establish the model evaluating user's gaming experience and provide scientific and reliable data to support game design.

In recent years, the battle royale games such as PUBG and APEX were very popular among players. Fast-pace and intense competitiveness are basic characteristics of such games. Therefore, whether new players can quickly understand and master

basic operations of the gameplay determines the quality of the gaming experience. Hence, for the battle royale games, the novice tutorial plays a significant role in the user retention rate. For the game UX designers, it is necessary to screen out the best design plan using certain methods to evaluate and compare all the proposals during product designing. People are crucial to evaluate the game UX design. Perceptions of users about UX design obtained through some qualitative and quantitative research methods can help UX designers establish an evaluation system that gives help to guide and makes designers introspect their work. However, it is not all factors in user gaming experience can be measured. It is a complex system containing plenty of unknown factors that the exterior might seem clear while the inside is vague. In this regard, this study uses grey system related theories and the analytic hierarchy process to explore the construction of a model evaluating user experience in the battle royale games. Proposed by Prof. Deng Julong (2017) of China, grey system theory is a research method based on mathematical theory. Its main study object is usually a system with fewer samples and less information. By quantifying uncertain gray information, it solves problems that contain unknown factors⁴. The object of this study is the user experience of novice tutorials in the Battle Royale Games. It contains quantifiable and non-quantifiable information, which is a typical gray system. In such a complex system, not all factors that affect the player's experience are of equal weight.

The analytic hierarchy process (AHP) was a decision method proposed by Thomas Saaty (1980). This method treats decision making as a system and decomposes it into three layers, that is target, criterion and plan. It has good effect on solving complex decision-making problems and multi-plan optimizations. Grey AHP theory is a method combining the grey system theory and AHP absorbed the strengths and characteristics of both methods. It fits well with the object in this study.

2 Collection of Emotional Words

By emotional words, we refer to any word that conveys emotional connotations and typically is adjectives. For example, adjectives such as exquisite and smooth are emotional words we might use to describe porcelain. In the study of user experience evaluation, emotional words to describe a product are typically used to assess users' emotional dimensions about the product. Theoretically, emotional vocabulary has single adjectives as well as pairs of adjectives with opposite meanings. However, considering the similarities between adjectives, researchers usually choose pairs of adjectives with opposite meanings when constructing user experience evaluation models. For evaluation model established by grey AHP method, it is significant to select emotional vocabulary. It has a great impact on the accuracy of the evaluation results. The appropriateness of emotional vocabulary and the number of it are important factors affecting the evaluation results. Too many words may increase the burden of the participants, while too few emotional words may result in insufficient description of the evaluation object. Concerning the appropriateness, there are two types: (a) appropriateness of the evaluation object indicating whether the emotional words can accurately describe the objective evaluated, (b) appropriateness of the user referring to the frequency that users use these words as well as the familiarity with them. Therefore,

researchers need to choose an appropriate number of emotional words based on their knowledge and experience. In the collection process, we need to select adjectives that can describe the cognitive characteristics of the object in every aspect. This study uses two ways for collection: (a) through websites and forums of the battle royale games, (b) from interviews with the players of the battle royale games.

In the above two approaches, the study team collected 33 pairs of emotional words in the beginning. After filtering out the less frequently used words in daily life, the remaining 30 pairs are adopted and stated in the Table 1.

Table 1. Selected emotional words

Soothing-Intense	Safe-Dangerous	Easy-Hard
Active-Serious	Independent-Dependent	Single-Multiple
Clear-Vague	Specific-General	Beautiful-Ugly
Ordinary-Unique	Enthusiastic-Indifferent	Natural-Awkward
Straightforward-Implicit	Understandable-Obscure	Interesting-Monotonous
Accomplished-Frustrated	New-Old	Flexible-Stiff
Plain-Connotative	Mutable-Unchanging	Adaptive-Rigid
Fault-tolerant-Demanding	Impassive-Responsive	Convenient-inconvenient
Smooth-Obstructive	Efficient-Inefficient	Satisfied-Disappointed
Encouraging-Withdrawing	Orderly-Disorderly	Progressive-Rapid

To screen the most representative words, we further classify and reduce the number of the above words. The study team recruited 11 participants who had a certain understanding of the Battle Royale Game. There were 9 males and 2 females, and 8 participants were design practitioners. The study team asked the participants to classify the 30 groups of emotional adjectives initially collected. The number of classifications is not limited, and the number of words in each category is controlled within 4. By counting the number of pairwise words that occurred in the same group, we got a 30 * 30 matrix. Multi-dimensional scaling was adopted to analyze the above matrix, and the dimension was set up as 2–6. Multidimensional scaling (MDS) is a method of multivariate analysis. It is commonly applied in the area of sociology, psychology, marketing, etc. The reason that the research team using multidimensional scaling analysis is to fit the research data through a lower dimension and obtain the coordinate values of each emotional word in the spatial map. This experiment measures the goodness of fitting of the spatial map according to the empirical standard proposed by Kruskal (1964) shown in the Table 2. As shown in the Table 3, the stress value indicates that the two-dimensional model fits the observed data well in this study, which is 0.071. In this way, we got the coordinate distribution of each emotional word in the two-dimensional space.

Table 2. Relationship between Kruskal stress coefficient and goodness of fit

Stress index > 0.5	Poor fits
Stress index < 0.1	Good fits
Stress index < 0.05	Very good fits
Stress index < 0.025	Great fits
Stress index = 0	Perfect fits

Table 3. Emotional vocabulary coefficient

Dimension	Stress index	RSQ explained variation
2	0.07125	0.87954
3	0.09231	0.89776
4	0.18672	0.91263
5	0.24581	0.93658
6	0.34544	0.95531

Next, a cluster analysis was performed on the two-dimensional spatial coordinates of the obtained emotional words. First, a systematic clustering analysis was conducted and a line chart was drawn with “category number” as the abscissa and “aggregation coefficient” as the ordinate, as shown in the Fig. 1 below. The broken line became slow when the category number reached 11. In this way, we can assert that there were 11 categories.

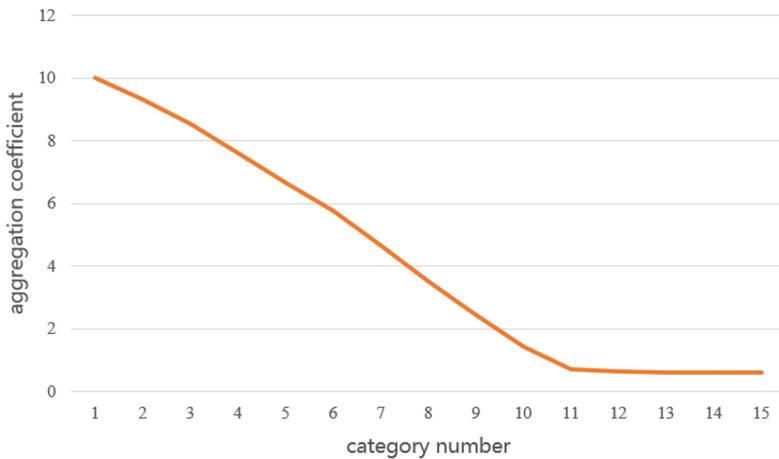


Fig. 1. Aggregation coefficient

Table 4. Categories of emotional words

Group	Emotional words in each group (pairs)
1	Soothing-Intense; Safe-Dangerous; Progressive-Rapid
2	Easy-Hard; Understandable-Obscure; Efficient-Inefficient
3	Active-Serious; Enthusiastic-Indifferent
4	Independent-Dependent; Responsive - Impassive
5	Interesting-Monotonous; Flexible-Stiff; Mutable-Unchanging; Smooth-Obstructive
6	Clear-Vague; Straightforward-Implicit; Specific-General; Plain-Connotative
7	Beautiful-Ugly; Ordinary-Unique; New-Old
8	Fault-tolerant-Demanding; Encouraging-Withdrawing
9	Natural-Awkward; Adaptive-Rigid; Convenient-inconvenient
10	Accomplished-Frustrated; Satisfied-Disappointed
11	Orderly-Disorderly; Single-Multiple

A K-means clustering analysis was then conducted on the two-dimensional coordinates of the emotional words according to the obtained aggregation coefficient, and the K value was set to 11. After calculation, the pair groups of the emotional words were sorted out and the distance of each group from the cluster center were obtained. Pairing words that were closest to the cluster center in each group turned out to be the most representative ones. The final grouping of the emotional words was shown in the Table 4. The study team renumbered the obtained 11 representative adjective pairs (E1–E9), and the results were also shown in Table 5 below.

Table 5. Representative emotional words

Group	Emotional words with closest distance to the clustering center (pairs)
1	Safe-Dangerous
2	Understandable-Obscure
3	Active-Serious
4	Responsive-Impassive
5	Smooth-Obstructive
6	Clear-Vague
7	Beautiful-Ugly
8	Fault-tolerant-Demanding
9	Natural-Awkward
10	Accomplished-Frustrated
11	Orderly-Disorderly

The interpretation of each representative emotional word is as follows.

E1 Safe: The experience of the novice tutorial makes players feel safe.

E2 Understandable: The introduction of the gameplay is easy for players to understand.

E3 Active: The teaching method of the novice tutorial is not rigid but lively.

E4 Responsive: Players are able to get feedback after finishing corresponding operation or achieving a goal.

E5 Smooth: Players will not get stuck when experiencing the novice tutorial.

E6 Clear: The interface guide of the novice tutorial is clear and straightforward.

E7 Beautiful: The interface and scene design of the novice tutorial can meet the aesthetic needs of players.

E8 Fault-tolerant: Players are allowed to misuse in the novice tutorial and the cost of mistakes is small.

E9 Natural: The interaction of the novice tutorial is natural and fit with players' interactive habits.

E10 Accomplished: Players can gain a certain sense of accomplishment after achieving the goals set in the novice tutorial.

E11 Orderly: The arrangement of the teaching modules of the novice tutorial is orderly.

3 Establish of the Evaluation Hierarchy

The analytic hierarchy process method usually consists of three layers from top to down. The uppermost layer is the target layer, which is the decision goal; the second layer is the criterion layer which mainly includes the criteria considered in the decision-making process, and there can be multiple sub-layers. The third is a plan layer, which includes several alternative plans for the decision. Combining the relative researches of user gaming experience, the target layer established in the evaluation hierarchy in this study was the user experience of novice tutorials in the battle royale games. Moreover, the study team divided the user experience into three criterion layers, that is sensory experience, behavioral experience and value experience. And the emotional words obtained above were classified into the sub-layers of the criterion layers. Since there were no design plans in this study, plan layer was not included for analysis. The schematic diagram of the final evaluation hierarchical structure is shown in the Fig. 2.

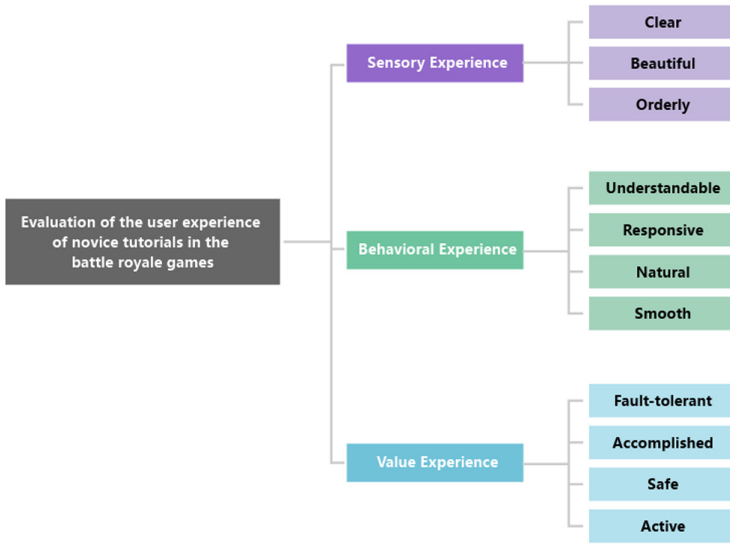


Fig. 2. Evaluation hierarchy of the user experience in novice tutorials in the Battle Royal Game

4 Perceptual Evaluation Experiment and Calculation of Subjective Weights

Before acquiring users’ perceptual needs through perceptual experiment, certain samples as well as questionnaires about perceptual evaluation of the above emotional words were to collect at first. We need to understand that the purpose of this experiment was to obtain users’ perceptual needs for the Battle Royale Games’ novice tutorials. Thus, novice tutorials of three battle royale games were selected as samples. In order to control variables, the three games were all FPS games, called PUBG, APEX, and CODwarzone, named as D1, D2, D3. First, participants were asked to experience the novice tutorials of these three games and then filled in the semantic difference questionnaires. Participants all had a certain understanding of the battle royale games or were the target players of those games. There were 22 participants in total, of which 17 were males and 5 were females. The age range of the participants were 20–25. Questionnaires of the user’s perception needs are shown in the Table 6 below.

Table 6. Perceptual questionnaires sample

Perception needs	-1	-2	-1	0	1	2	3	Perception needs
Vague								Clear
Ugly								Beautiful
Disorderly								Orderly
Obscure								Understandable
Serious								Active
Slient								Responsive
Unchanging								Mutable
Demanding								Fault-tolerant
Rigid								Adaptive
Frustrated								Accomplished
Dangerous								Safe

After calculating the average value of each sample in every emotional vocabulary group, a line chart of expected target results of users was obtained, as shown in the Fig. 3.

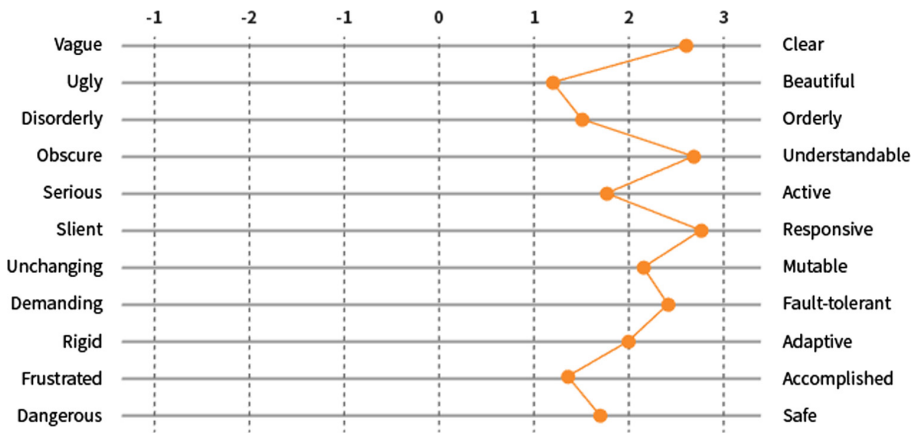


Fig. 3. Line chart of users' perceptual needs

After obtaining the users' perceptual needs for novice tutorials in the battle royale games, the study team aimed to calculate the weight of each evaluation index based on the evaluation hierarchy acquired. The study team compared the elements at the same level one by one with reference to the users' perceptual needs and produced a comparison matrix. The scaling method proposed by Professor Saaty (1980) was applied to score the comparison results⁵, as shown in the Table 7.

Table 7. Scoring scale

Importance scale	Implications
1	Equally important when comparing two items
3	The former is slightly more important when comparing two items
5	The former is obviously more important when comparing two items
7	The former is strongly important when comparing two items
9	The former is extremely important when comparing two items

When comparing the indexes, we assumed that there were n criteria (i.e., $a_1, a_2 \dots a_n$) in the next layer of the target C , it was necessary to use C as a standard to compare the importance of these criteria (i.e., $a_1, a_2 \dots a_n$) respectively and produce the comparison matrix A . The matrix A could be expressed as follows.

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{n4} \end{bmatrix}$$

$$A = (a_{ij})_{n \times n}, \quad i, j = 1, 2, \dots, n \tag{1}$$

And, a_{ij} is the score after comparing the importance of factor i and j .

After obtaining the matrix, a consistency test will be performed to obtain the consistency ratio (CR). The calculation method of CR is as follows.

$$CR = \frac{CI}{RI}$$

$$CI = \frac{\lambda - n}{n - 1} \tag{2}$$

and $\lambda = \frac{1}{n} \sum_{i=1}^n \frac{(AW)_i}{w_i}$ is the characteristic root of the matrix.

The values of RI are as follows according to research⁶.

If $CR = 0$, the matrix is perfectly consistent. If $CR = 1$, the matrix is perfectly inconsistent. Normally, if $CR < 0.1$, the consistency of the comparison matrix is considered satisfactory. When the matrix deviates from consistency, the accuracy of the index weights cannot be guaranteed. Based on the above methods, the consistency ratio of this experiment was calculated and $CR = 0.043 < 0.1$, which suggested the matrix was acceptable.

Table 8. Value of RI

n	1	2	3	4	5	6	7	8	9	10
RI	0	0	0.58	0.90	1.12	1.24	1.32	1.41	1.45	1.49

Table 9. The weight of each evaluation index

Target layer	Criterion layer	Weight	Sub-criterion layer	Weight
Experience of Novice tutorials in the Battle Royal Games	Sensory experience	0.256	Clear-Vague	0.102
			Beautiful- Ugly	0.071
			Orderly- Disorderly	0.083
	Behavioral experience	0.352	Understandable- Obscure	0.074
			Active-Serious	0.076
			Responsive - Impassive	0.113
			Safe-Dangerous	0.089
	Value experience	0.392	Mutable- Unchanging	0.105
			Fault-tolerant- Demanding	0.089
			Adaptive- Rigid	0.110
Accomplished- Frustrated			0.088	

At last, Yaahp was used to process the obtained data to calculate the weight index of each element relative to the upper layer. The player's perceptual needs can be recorded as follows.

5 Grey Relational Analysis

After obtaining the weight of each evaluation index, this paper used Grey relational analysis method to improve the hierarchical analysis model. The main purpose was to transform the data from grey to white, that is from uncertain information to definite information. As mentioned, evaluation of gaming experience is a system containing lots of fuzzy information and the cognition bias from different evaluators might also bring deviation to the experimental results. Thus, the average weight of each index obtained above is not yet an ideal result. It is necessary to combine the Grey relational analysis method to further optimize the obtained evaluation model.

- (1) First, we determined the reference sequence and the comparison sequence. The reference sequence reflected the characteristics of the system referring to the best scores of each indicator about the user experience of the novice tutorial in the battle royale games here. The comparison sequence referred to data that has impact on system and needed to be sorted. Here it referred to experimental data of the novice tutorials of three the battle royale games. According to the serial number of the

plans, the comparison sequence was remarked as x_i , and i indicated the serial number, $x_i(k)$ referred to the evaluation score of the i -th plans within the index k .

$$x_i = (x_{1i}, x_{2i}, \dots, x_{ji}), \quad i = 1, 2, \dots, m \tag{3}$$

- (2) Non-dimensionalization of the comparison sequences. Since dimensions impacting the evaluation indexes may be different, all indexes were made dimensionless before the comparison sequence was analyzed. All sequences were divided by the k -th adjustment sequence to ensure that the evaluation scores of all series fall in (0, 1). Yet the obtained matrix in the experiment was dimensionless already, non-dimensionalization was unnecessary.

$$f(x_i(k)) = \frac{x_i(k)}{A(k)}, \quad k = 1, 2, \dots, n, \quad i = 1, 2, \dots, m \tag{4}$$

- (3) Calculated the Grey relational coefficient (GRC). The GRC $\xi(x_0)$ refers to the degree of difference between the geometric shapes of the curves, that is, the difference between the curves can measure the degree of correlation. The calculation is as follows. The i refers to the serial number of the plans while k represents the index sequence.

$$\xi_i(k) = \frac{\min_{i \in m} \min_{k \in n} |x_0(k) - x_i(k)| + \rho \max_{i \in m} \max_{k \in n} |x_0(k) - x_i(k)|}{|x_0(k) - x_i(k)| + \rho \max_{i \in m} \max_{k \in n} |x_0(k) - x_i(k)|} \tag{5}$$

Among them, ρ is the discrimination coefficient which can improve the statistical significance between GRC. The smaller the ρ , the greater the discrimination. The general value range of ρ is (0, 1), and its general value is 0.5. According to relative research, the value of ρ is adopted as 0.5 in this study as well.

- (4) Calculated the Grey interconnect degree. Since the relational coefficient is the value of the interconnect degree of the comparison and reference sequences at each time, that is, each point in the curve, it has more than one value. To prevent the information from being too scattered, all relational coefficients at each time need to become centralized as one value, that is Grey interconnect degree r_i , making it feasible for comparison as a whole. The calculation formula is as follows.

$$r_i = \sum_{k=1}^n w_k^G \xi_{oi}(k), \quad k = (1, 2, \dots, n) \tag{6}$$

Design plans will be compared with each other for better selection according to the value of the Grey interconnect degree. If $r_i \geq r_j$, it suggests that the *plan* x_i is better than *plan* x_j . In this way, we compared three design plans in this study and selected the one with the biggest Grey interconnect degree as the best plan scores of Grey interconnect degree of three design plans were obtained, as shown in the Table 10.

Table 10. Score of the Grey interconnect degree

Design plan	Score of the Grey interconnect degree
D1	0.714
D2	0.825
D3	0.520

Combined with the weight of each evaluation index obtained, scores of weighted Grey interconnect degree of three design plans were calculated, as shown in the Table 11.

Table 11. Score of Weighted Grey interconnect degree

Design plan	Score of Weighted Grey interconnect degree
D1	0.2036
D2	0.2253
D3	0.1249

6 Conclusions and Limitations

- (1) This paper discusses a method of constructing the system evaluating the user experience of novice tutorials of the battle royal games in the process of game design. This method can help designers to fully understand the satisfaction degree of users in a scientific way in the process of game development and select the most superior design plan. It also provides a certain reference for the direction of further iterations. When we were making general evaluation about the design plans, the Analytic Hierarchy Process method and Grey Relational Analysis method were applied to calculate the weighted interconnected degree of the evaluation targets and ideal plans so that the plan with higher interconnected degree was screened out. This verified the auxiliary role of the grey AHP Theory in evaluating game experience. The ideas and methods to establish the model evaluating gaming experience in this paper can be extended to experience evaluation of other game systems other than the novice tutorial and other kinds of games. It is of great significance for game design.

- (2) Due to limited time and resources, only three novice tutorials of the battle royale games that were extremely popular recently were selected for evaluation in this study. In the follow-up research, this paper may adopt more samples or design plans of our own game projects for evaluation to improve the accuracy of the model. As for research methods, this study adopted online questionnaires to conduct the perceptual experiment due to physical constraints. The study team will consider applying eye-tracking equipment or other perceptual measurement methods in further research.

References

- ISO 9241-11: Ergonomic requirements for office work with visual display terminals (VDTs), Part 11. Guidance on Usability. ISO (1998)
- Nuan, B.: Research on Card Game Design Based on Kansei Engineer (2019)
- Yan, W., Xujiang, Z.: The study on the influence of user experience in game design—an instance of “LeAgue of Legends.” *Design* **02**, 118–119 (2014)
- WeiWei, S.: Research on the Evaluation Model of APP for the Elder Based on Grey AHP (2017)
- Saaty, T.L.: *The Analytic Hierarchy Process*. McGraw-Hill, New York (1980)
- Saaty, T.L.: *Fundamentals of Design Making and Theory with the Analytic Hierarchy Process*. Analytic Hierarchy Process (2000)