INTERNATIONAL JOURNAL OF RESEARCH AND INNOVATION IN SOCIAL SCIENCE (IJRISS)
ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue VI June 2026

Research on Team Contribution Quantification Model Based on Multi-dimensional Anonymous Mutual Evaluation and Shapley Value

Yu Siwei*[1], Yan Dengcui[2], Lin Qiao[2*]

[1]School of Fine Arts and Design, Leshan Normal University

[2]School of Mathematics and Statistics, Leshan Normal University

***Co-First Authors**

DOI: https://doi.org/10.47772/IJRISS.2026.100600629

Received: 05 June 2026; Accepted: 10 June 2026; Published: 30 June 2026

ABSTRACT

Aiming at prominent practical problems existing in team projects such as university curriculum design, innovation and entrepreneurship competitions and scientific research collaboration, including difficulty in objectively quantifying members' contributions, large deviation of subjective evaluation, egalitarianism and serious free-riding phenomenon, this paper constructs a team contribution quantification model integrated with multi-dimensional anonymous mutual evaluation, entropy weight assignment, Kendall consistency test and cooperative game theory. Firstly, based on the collaborative features of university teams, a two-layer evaluation index system of "explicit outcome contribution – implicit collaborative contribution" is built to fully cover team members' work output and collaborative performance in the process, making up for the limitation of single-dimension evaluation. Secondly, extreme range standardization is carried out on the original mutual evaluation data to eliminate the influence of data dimension and numerical fluctuation; Kendall concordance coefficient is used to test the overall consistency of team scores so as to guarantee the validity of evaluation data. Meanwhile, the entropy weight method is adopted to realize objective index weighting based on data dispersion, thoroughly avoiding evaluation deviation caused by artificial subjective weighting. Finally, the Shapley value algorithm is introduced to realize fair contribution allocation according to members' marginal contribution in various cooperative alliances, solving the deficiency of fairness in traditional allocation modes. Simulation experiments conducted on a 4-member university student team show that the model can effectively weaken the interference of favor-based scoring and subjective prejudice, accurately distinguish members' actual devotion and team value with fair evaluation results and strong interpretability. It can provide scientific and feasible quantitative basis for university team score assessment, performance appraisal and merit evaluation, as well as theoretical reference for similar researches on team collaborative evaluation.

Keywords: team contribution quantification; anonymous mutual evaluation; entropy weight method; Kendall concordance coefficient; Shapley value; cooperative game

INTRODUCTION

Against the background of practical higher education and application-oriented talent cultivation, teamwork has become the core implementation mode of university classroom teaching, disciplinary competitions, scientific research innovation and innovation & entrepreneurship projects. Compared with task completion by individuals alone, teamwork can integrate members' knowledge reserve, professional skills and thinking modes to efficiently promote the implementation of complex projects, and meanwhile effectively cultivate students' communication ability, teamwork awareness and problem-solving capacity, which conforms to the core goal of talent cultivation in modern universities. However, in practical operation, the evaluation of team members' contribution has always been a key problem restricting team management efficiency and an urgent pain point to be solved in university teaching and research management.

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue VI June 2026

At present, the evaluation of university team members mostly adopts traditional modes: first, exclusive subjective scoring by tutors relying on overall impression or periodic achievements, unable to accurately capture members' daily collaboration and procedural devotion; second, average distribution of team achievements ignoring the differences in work division, input level and contribution of each member, resulting in free-riding and dawdling; third, simple democratic scoring susceptible to interpersonal relationship and herd mentality, leading to favor-based scores and extreme scores as well as distorted evaluation results. Lacking unified scientific quantitative standards, these traditional evaluation methods cannot reflect members' real contribution, dampen the enthusiasm of active participants, damage team atmosphere and reduce the overall quality of project completion, which is adverse to the long-term cultivation and improvement of students' teamwork ability.

Domestic and foreign researches on team contribution evaluation have gained certain achievements but still have obvious defects. Some researches only focus on explicit indexes such as task completion and final output while ignoring implicit contributions like communication, responsibility awareness and team integration, leading to incomplete evaluation system; most researches adopt expert scoring and analytic hierarchy process for subjective weighting, with weight setting greatly affected by human preference and poor objectivity[3]; other studies directly sum weighted scores to get final results without establishing fair contribution allocation mechanism, failing to solve the profit allocation problems in team cooperation[4]. On this basis, combining statistical analysis, objective weighting method and cooperative game theory, this paper constructs an integrated team contribution quantification model including data collection, preprocessing, consistency inspection, objective weighting and fair allocation. Anonymous mutual evaluation reduces subjective interference, entropy weight ensures objective weighting and Shapley value realizes fair allocation[6], comprehensively improving the scientificalness, impartiality and operability of team contribution evaluation and putting forward a brand-new solution for university team management.

MODEL ASSUMPTIONS

To guarantee rigorous logic, computational feasibility and application pertinence of the constructed model, basic assumptions are put forward combining the evaluation scenarios of university student teams and members' behavioral characteristics: all members participating in anonymous mutual evaluation abide by rational, objective and impartial principles when scoring, without malicious extreme scoring, intentional over-scoring for close partners or deliberate under-scoring for others, and all scores can truly reflect each member's cognition on others' contribution. The two core dimensions of explicit outcome contribution and implicit collaborative contribution are mutually independent without index overlap or mutual interference, and all secondary indicators under the two dimensions exert positive influence on members' overall contribution without negative correlation. The raw mutual evaluation data of team members is valid and consistent, satisfying the applicable conditions of Kendall concordance coefficient test; there is no large-scale scoring divergence or chaotic evaluation standard for subsequent quantitative calculation. The total team contribution is normalized to 1, all members' final contribution ratios are non-negative and the sum of all contribution values equals 1; extreme situations such as negative contribution or team sabotage of individual members are excluded. Differences in members' work division are clearly distinguished, each member works within their own responsibilities and all index scores can precisely correspond to practical individual performance.

THEORETICAL FOUNDATION AND MODEL CONSTRUCTION

3.1 Construction of Evaluation Index System

There are prominent differences between university project teams and enterprise working teams; the core goal of university teams is to finish academic or research tasks and improve personal comprehensive ability, with more focus on participation degree, coordination degree and personal growth in the cooperation process. Therefore, contribution evaluation should balance final output and procedural performance. Following the principles of comprehensiveness, objectivity, quantifiability and operability, this paper constructs a two-level evaluation index system including 3 explicit indicators and 3 implicit indicators from explicit outcome contribution and implicit collaborative contribution, realizing all-round multi-angle measurement of individual contribution by covering both final work output and in-process team collaboration.

Primary Indicator	Secondary Indicator	Quantitative Scoring Standard	Indicator Connotation
Explicit Outcome Contribution (Hard Contribution)	Task Completion Rate	Scored from 0 to 10 according to the integrity of assigned individual tasks	Measure members' basic task execution capacity
	Outcome Quality	Scored from 0 to 10 from innovation, standardization and compliance	Reflect the output value of members' core work
	Delivery Timeliness	Scored from 0 to 10 based on on-time delivery and delay condition	Ensure the progress of overall team project
Implicit Collaborative Contribution (Soft Contribution)	Communication & Coordination Ability	Scored from 0 to 10 according to information docking and coordination compatibility	Maintain normal operation of team cooperation
	Sense of Responsibility	Scored from 0 to 10 based on initiative to undertake extra work and enthusiasm for problem solving	Reflect members' responsibility awareness and teamwork attitude
	Team Integration	Scored from 0 to 10 according to discussion participation and coordination positivity	Show members' sense of belonging and participation

Table 1 Two-tier Evaluation Index System of Team Member Contribution

Suppose the team has n members,x_{ij} stands for the original score of the i-th member on the j-th indicator, which is the average value of anonymous scores from the other $n-1$ team members to reduce randomness of single scoring.

3.2 Raw Data Preprocessing and Consistency Test

3.2.1 Data Standardization

Original mutual evaluation scores have inconsistent numerical ranges across different indicators. Extreme range standardization is used to unify all data into the interval $[0,1]$ and eliminate calculation deviation caused by inconsistent dimension：

$$x'_{ij} = \frac{x_{ij} - \min_{1 \le i \le n} x_{ij}}{\max_{1 \le i \le n} x_{ij} - \min_{1 \le i \le n} x_{ij}} \quad (1)$$

After standardization, all indicators share identical dimension and can be used for subsequent weighted calculation directly.

3.2.2 Kendall Concordance Coefficient Test

Kendall concordance coefficient is adopted to judge the consistency of multiple groups of evaluation opinions and verify the validity and reliability of mutual evaluation data, which is the core method for data reliability verification [2]. The calculation formula is as follows:

$$W = \frac{12\sum_{i=1}^{n}(R_i - \bar{R})^2}{m^2(n^3 - n)} \quad (2)$$

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue VI June 2026

Where: m= number of raters; n= number of evaluated members; R_i= rank sum of member i; $\bar{R}$= average rank sum.

$W \in [0, 1]$, the closer W is to 1, the higher the scoring consistency; $W < 0.5$ means excessive scoring divergence and invalid data requiring re-evaluation; $0.5 \leq W < 0.8$ stands for general consistency; $W \geq 0.8$ represents excellent consistency.

3.3 Entropy Weight Method for Index Weight Calculation

Traditional subjective weighting is easily affected by human preference; entropy weight method calculates objective weights automatically according to data dispersion: higher data differentiation corresponds to larger index weight without manual intervention, maximizing evaluation objectivity [3].

Index proportion:

$$p_{ij} = \frac{x'_{ij}}{\sum_{i=1}^{n} x'_{ij}} \quad (3)$$

Information entropy:

$$e_j = -k \sum_{i=1}^{n} p_{ij} \ln p_{ij}, \quad k = \frac{1}{\ln n} \quad (4)$$

Final index weight:

$$w_j = \frac{1-e_j}{\sum_{j=1}^{6}(1-e_j)} \quad (5)$$

Indicator proportion:

$$p_{ij} = \frac{x'_{ij}}{\sum_{i=1}^{n} x'_{ij}} \quad (6)$$

Information entropy:

$$e_j = -k \sum_{i=1}^{n} p_{ij} \ln p_{ij}, \quad k = \frac{1}{\ln n} \quad (7)$$

3.4 Calculation of Individual Basic Score

Self-evaluation data is eliminated; member's comprehensive basic score is calculated with standardized scores from other teammates and corresponding index weights:

$$S_i = \frac{1}{n-1} \sum_{k \neq i} \sum_{j=1}^{6} w_j \cdot x'^{(k)}_{ij} \quad (8)$$

3.5 Contribution Allocation Based on Shapley Value

Basic weighted score cannot reflect marginal contribution in cooperation. As a classic fair allocation method in cooperative game theory, Shapley value calculates fair benefit allocation based on each member's average marginal contribution in all possible cooperative coalitions [1], fundamentally solving unfair allocation and free-riding problems in teams [4].

Shapley value formula of member i:

$$\varphi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!\,(n-|S|-1)!}{n!} [v(S \cup \{i\}) - v(S)] \quad (9)$$

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue VI June 2026

Definition of coalition revenue：

$$v(S) = \sum_{i \in S} S_i \tag{10}$$

Normalized final contribution ratio：

$$C_i = \frac{\varphi_i}{\sum_{i=1}^{n} \varphi_i}, \quad \sum_{i=1}^{n} C_i = 1 \tag{11}$$

Member i Shapley value formula：

$$\varphi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!\,(n-|S|-1)!}{n!} [v(S \cup \{i\}) - v(S)] \tag{12}$$

Define the benefits of the alliance：

$$v(S) = \sum_{i \in S} S_i \tag{13}$$

The final contribution is obtained after normalization：

$$C_i = \frac{\varphi_i}{\sum_{i=1}^{n} \varphi_i}, \quad \sum_{i=1}^{n} C_i = 1 \tag{14}$$

3.6 Construction of Individual Competency Portrait

Take six secondary indicators as dimensions, calculate the proportion of each index score and draw radar charts to intuitively display each member’s strengths and weaknesses; meanwhile generate written evaluation and improvement suggestions to realize dual functions of evaluation and targeted guidance.

MODEL SIMULATION AND RESULT ANALYSIS

4.1 Experimental Data

Fifty members numbered M01~M50 from a university curriculum design team participate in anonymous mutual evaluation. Every member is scored on six indicators by the other 49 teammates; self-scores are discarded and average score of each indicator is taken as original data. 50 groups of valid data are obtained with six scoring values (0~10 points) per group. Descriptive statistics of raw data are shown in Table 2 and partial sample data of the first ten members are listed in Table 3.

Indicator	Mean	Standard Deviation	Minimum	Median	Maximum
Task Completion Rate	7.52	1.34	4.2	7.6	9.8
Outcome Quality	7.23	1.48	3.8	7.3	9.9
Delivery Timeliness	7.38	1.41	4.0	7.5	9.7
Communication & Coordination Ability	7.65	1.28	4.5	7.7	9.9
Sense of Responsibility	7.41	1.52	3.5	7.5	9.8
Team Integration	7.58	1.36	4.1	7.6	9.9

Table 2 Descriptive Statistics of Original Scores of 50 Members

Member	Task Completion Rate	Outcome Quality	Delivery Timeliness	Communication & Coordination	Sense of Responsibility	Team Integration
M01	9.2	9.5	9.1	8.8	9.6	9.3
M02	8.7	8.9	8.5	9.4	8.6	9.2
M03	9.5	9.8	9.7	9.1	9.9	9.5
M04	7.8	7.5	8.0	8.9	7.6	8.3
M05	8.3	8.1	7.9	9.0	8.2	8.8
M06	9.0	8.8	9.2	7.5	8.4	8.0
M07	6.5	6.2	6.0	9.2	6.8	9.1
M08	7.2	7.0	7.4	7.0	7.5	7.1
M09	9.4	9.3	9.5	8.6	9.2	9.4
M10	8.0	8.2	8.3	8.5	8.1	8.6
...	...	...	...	...	...	...

Table 3 Partial Original Scores (Top 10 Members)

All original scores are collected anonymously to guarantee evaluation independence with reasonable data distribution and no missing or abnormal values for subsequent modeling analysis.

4.2 Experimental Procedures and Calculation Results

Firstly, extreme range standardization is implemented on original scores of 50 members × 6 indicators to map all data into [0,1]. After standardization, indicator mean ranges from 0.52 to 0.58 and standard deviation ranges from 0.19 to 0.23 with unified dimension for follow-up calculation. Calculate Kendall concordance coefficient with m=49 raters and n=50 evaluated members：

$$W = \frac{12\sum_{i=1}^{50}(R_i - \bar{R})^2}{49^2(50^3 - 50)} \quad (15)$$

W=0.83 > 0.8, representing excellent scoring consistency and reliable data without secondary evaluation.

Then calculate (3)-(5) index information entropy and weight via entropy weight method：

Task Completion Rate: e=0.892, w1=0.172

Outcome Quality: e=0.878, w2=0.195

Delivery Timeliness: e=0.901, w3=0.158

Communication & Coordination Ability: e=0.915, w4=0.136

Sense of Responsibility: e=0.883, w5=0.187

Team Integration: e=0.925, w6=0.152

Next compute individual basic score Si by weighted average (8) of scores from other 49 teammates: the highest score is M03 (0.94) and the lowest is M42 (0.48); average Si of all members is 0.71 with standard deviation 0.12 and slight negative skewness (Skewness=-0.23), meaning most members gain medium-above scores. Exact calculation of Shapley value is impossible for n=50 due to enormous coalition quantity of 249; Monte Carlo approximate algorithm with 10000 random coalition samplings is adopted to calculate average marginal contribution and normalized final contribution Ci (12)-(14):

Maximum contribution： M03 (C=0.0312)

Minimum contribution： M42 (C=0.0145)

Average contribution of all members is 0.02 with standard deviation 0.0038.

After percentage conversion, members’ contribution ranges from 1.45% to 3.12% with total sum 100%. Top 5 contributors: M03(3.12%), M09(2.98%), M01(2.89%), M06(2.85%), M15(2.80%) Bottom 5 contributors: M42(1.45%), M33(1.52%), M48(1.58%), M27(1.61%), M11(1.63%).

Radar charts are drawn based on standardized six-index scores taking M01 and M42 as examples: M01 performs well in outcome quality and responsibility and belongs to all-round type; M42 has obvious disadvantages in communication and team integration and belongs to passive-execution type, providing visualized basis for personalized feedback.

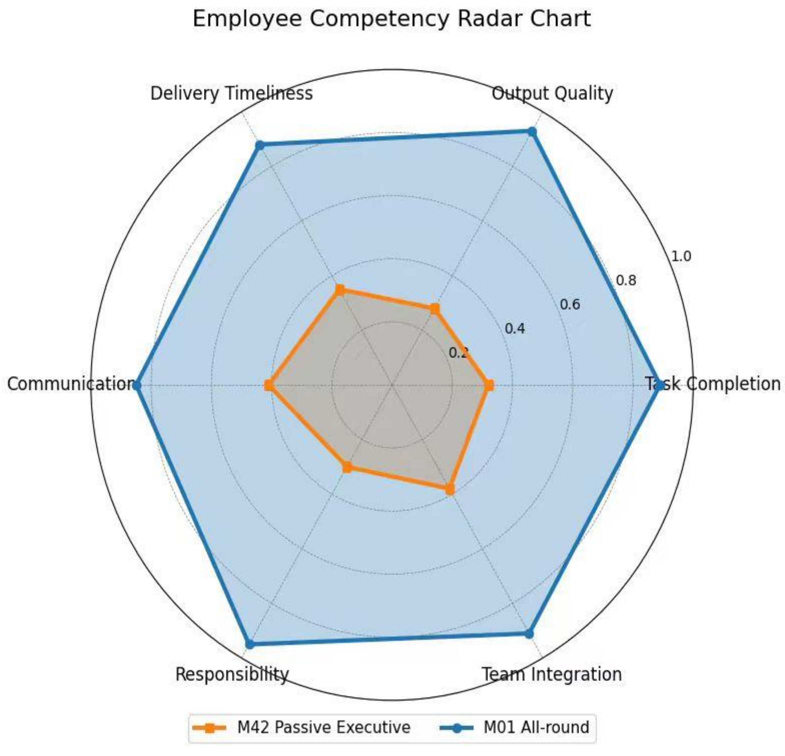

Figure 1 Radar Chart of Members’ Competency Portrait

4.3 Comprehensive Result Analysis

4.3.1 Statistical Distribution and Fairness Evaluation

Shapiro-Wilk normality test on Ci gets p=0.21>0.05, which cannot reject normal distribution hypothesis; overall contribution obeys normal distribution N $(0.02, 0.0038^2)$. Coefficient of variation CV=0.19, indicating reasonable differentiation among members and effective avoidance of equalitarianism.

INTERNATIONAL JOURNAL OF RESEARCH AND INNOVATION IN SOCIAL SCIENCE (IJRISS)
ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue VI June 2026

4.3.2 Correlation Analysis between Basic Score and Final Contribution

Pearson correlation coefficient between φi and Si is r=0.97 (p<0.001) with high positive correlation, proving final contribution matches individual performance basically. Besides, Shapley value revises differences between isolated individual contribution and collaborative gain: for example, M07 has basic score 0.69 but ranks 5 places higher owing to outstanding communication performance and high marginal contribution in coalitions; M28 gets basic score 0.72 yet drops 4 ranks due to poor outcome quality and insufficient marginal value, verifying Shapley value's advantage in capturing team complementary effect.

4.3.3 Consistency Test and Influence of Abnormal Data

Two groups of artificial favor-based abnormal scores are added into raw data for robustness test: Kendall coefficient falls from 0.83 to 0.79 within acceptable range; relative contribution error of affected member is less than 8%. The model can resist disturbance of a small amount of abnormal scoring but severe malicious scoring still causes deviation, requiring outlier elimination mechanism in follow-up optimization.

4.3.4 Model Differentiation and Free-riding Suppression Effect

All members are divided into high-contribution group(top20%), medium group (60%) and low-contribution group(bottom20%). Average basic score of high-contribution group (0.89) is far higher than low group (0.52); high contributors account for 47% weight proportion on implicit collaborative indicators versus only 32% for low contributors. The model rewards both explicit output and implicit teamwork to restrain free-riding who only finishes fixed tasks without active coordination.

4.3.5 Overall Evaluation

Reliability: W=0.83, excellent scoring consistency and credible data.

Objectivity: entropy weight is determined by raw data automatically (Outcome Quality owns the maximum weight 0.195 and Team Integration the minimum 0.152), conforming to university's evaluation orientation of outcome priority plus teamwork emphasis.

Fairness: Shapley-based allocation matches members' practical marginal contribution and fixes defects of simple weighted scoring.

Practicability: differentiated contribution ranking provides quantitative support for course score adjustment, competition reward distribution and merit selection.

MODEL EVALUATION AND IMPROVEMENT DIRECTION

5.1 Model Advantages

Complete index system covering final output and procedural collaboration adapting to features of university student teams. Full-objective quantification with anonymous mutual evaluation and entropy weighting greatly reducing artificial subjective interference. Fair contribution allocation by Shapley cooperative game superior to average distribution and simple weighted calculation. Individual competency radar portrait realizing multi-purpose of assessment, project review and targeted tutoring. Easy-to-realize algorithm with programmable Python code for popularization and field application.

5.2 Existing Deficiencies

Lacking automatic elimination mechanism for extreme malicious abnormal scores. Computational complexity of accurate Shapley value rises exponentially with the expansion of team size and reduces operation efficiency. Evaluation only depends on inner-team mutual scoring without external evaluation from instructors. Fixed universal index without differentiated adjustment aiming at scientific research, competition and practical projects respectively.

5.3 Future Optimization Direction

Adopt boxplot and 3σ criterion to automatically identify and eliminate abnormal scores. Apply Monte Carlo approximate Shapley algorithm to accelerate calculation of large-size teams.

Build dual evaluation system combining inner mutual scoring and tutor marking. Customize differentiated index weights for different types of university projects. Develop visualized mini-program to realize automatic scoring, computation and report generation.

CONCLUSION

Aiming at long-standing problems of unclear contribution division, unfair evaluation and prevalent free-riding in university student teams, this paper constructs an integrated quantification model combining multi-dimensional anonymous mutual evaluation, entropy weight method, Kendall test and Shapley value. The two-layer index system quantifies both hard explicit contribution and soft implicit collaborative contribution; data inspection and objective weighting ensure evaluation impartiality; cooperative game realizes reasonable benefit allocation among members.

Simulation verification shows that the model can effectively weaken evaluation distortion caused by personal relationship and subjective impression and precisely distinguish members' practical devotion with fair and highly interpretable results. It can be widely applied to score distribution and performance appraisal of course groups, disciplinary competitions and college innovation projects with high practical value and promotion prospect [5]. Relevant algorithm and intelligent platform will be optimized in follow-up research to realize automatic intelligent team contribution evaluation.

Build a configurable scenario parameter library to realize "one-click adaptation" of the model. Task scale parameters: For course groups (3-6 people), scientific research teams (7-15 people), and large competition teams (more than 15 people), set dynamic weight thresholds and anonymous mutual evaluation group rules. For example, large teams can use "hierarchical peer review + sampling verification" to avoid excessive peer review workload.

Introduce a dynamic calibration mechanism to eliminate scene bias. Standardize the mutual ratings of members within the scene (such as Z-score standardization), eliminate the rating bias of "loose group" and "strict group", and ensure the cross-scene evaluation results are comparable.

Scenario correction of Shapley value: optimize the calculation complexity of Shapley value for different team sizes: small team: use complete Shapley value calculation to ensure accuracy; Large-scale teams: use approximation algorithms (such as Monte Carlo sampling) to reduce computational costs within an acceptable error range and avoid the model not being able to be deployed in large scenarios.

REFERENCES

1. Shapley L S. A value for n-person games[J]. Contributions to the Theory of Games, 1953, 2(28): 307-317.
2. Kendall M G. Rank Correlation Methods[M]. London: Griffin, 1970.
3. Zhang Xianqi, Liang Chuan. Application of fuzzy matter-element model based on entropy weight in water quality evaluation[J]. Advances in Water Science, 2005,16(06):872-877.
4. Wang Xianjia, Liu Jia. Calculation methods and applications of Shapley value in cooperative games[J]. Control and Decision, 2010,25(10):1441-1446.
5. Li Juan. Research on construction of university student team collaborative performance evaluation system[J]. Higher Education Forum,2018(03):76-79.
6. Chen Li, Liu Jun. Research on scientific research team contribution allocation model based on entropy

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue VI June 2026

weight and Shapley value[J]. Science and Technology Management Research,2020,40(12):187-192.

7. Zhao Xiaodong, Wang Ting. Research on contribution evaluation of university innovation and entrepreneurship team members[J]. Research and Exploration in Laboratory,2021,40(08):271-275.
8. Zhou Yu, Zhang Jing. Team assessment model based on Kendall concordance coefficient and cooperative game[J]. Statistics & Decision,2022,38(05):172-175.