Development and Test of Survey Scale for University Humanities and Social Sciences Teachers' Satisfaction with Scientific Research Performance Evaluation Mechanism

Ruishu Wang[1],*

[1]*Foreign Studies College, Northeastern University, Shenyang, Liaoning, 110819, China*
*Corresponding author: wxmzml@126.com

Abstract: *Scale of satisfaction with the scientific research performance evaluation mechanism of humanities and social sciences in universities was developed to investigate the current teachers' satisfaction with the scientific research performance evaluation mechanism of humanities and social sciences in universities, and to identify the problems in the scientific research performance evaluation mechanism of humanities and social sciences in universities in order to promote targeted improvements in universities, and further improve the scientific research performance evaluation mechanism of humanities and social sciences in universities. In developing the scale, the initial scale was compiled on the relevant literature and semi-structured interviews. The scale was developed through the steps of item development, draft scale development, expert validity establishment, data collection of pre-test scale, item analysis, exploratory factor analysis and confirmatory factor analysis. The quality of the scale was assessed through reliability test and validity test. The scale includes two key dimensions: evaluation mechanism elements and evaluation mechanism operation. The scale provides a feasible tool for the measurement of satisfaction with the evaluation mechanism of university humanities and social sciences research performance, proposes a new perspective for the study of the evaluation mechanism of university humanities and social sciences research performance, and expands new ideas for comprehensive understandings of the evaluation mechanism of university humanities and social sciences research performance.*

Keywords: *humanities and social sciences; scientific research performance; evaluation mechanism; scale development*

1. Introduction

Universities play an important role as the main force in the development of humanities and social sciences. Teachers, as the main body of university research, are the most central force in promoting the development of university research. The scientific research performance evaluation mechanism is the most concrete manifestation of university teachers' scientific research performance evaluation. It is the main way to stimulate teachers' research creativity. Also, it is important to mobilize humanities and social sciences teachers to ensure the growth of social sciences talents which plays a vital role in humanities and social sciences management of universities. It analyzes and judges the performance of humanities and social sciences teachers to strengthen their strengths and improve their weaknesses which could provide coordination and guarantee for the improvement and development of the university's overall research level and the construction of its disciplines.

However, with the development of university research in the humanities and social sciences, the performance evaluation mechanism of university research in the humanities and social sciences has also revealed some problems. In order to figure out the current situation and problems of the university humanities and social sciences scientific research performance evaluation mechanism, it is necessary to understand the current teachers' satisfaction with the university humanities and social sciences scientific research performance evaluation mechanism and the influencing factors. Therefore, it is important to develop a scientific and reasonable survey scale. The purpose of this study is to develop a scientific and reasonable survey scale to investigate the current teachers' satisfaction with the evaluation mechanism of university scientific research performance in humanities and social sciences, and then to identify the problems in the evaluation mechanism of university scientific research

performance in humanities and social sciences which could promotes universities to make targeted improvements and further improve the scientific research performance evaluation mechanism of university in humanities and social sciences.

2. Related work

Deng (2004) claimed that the evaluation mechanism of humanities and social sciences research scientific research needed to be improved. The improvement tools included the adoption of the citation analysis method, the off-site review method, the online public review method, the establishment of a pool of experts and the improvement of the supervision mechanism, the establishment of a system for checking the scientific research and the importance of objections to the evaluation scientific research [1]. Li (2010) believed that institutional innovation in the evaluation mechanism of humanities and social sciences research scientific research included giving full play to the advantages between different evaluation mechanisms and combining qualitative evaluation with quantitative evaluation; establishing a comprehensive, objective and fair scientific evaluation system which in turn promotes the development of humanities and social sciences; establishing scientific evaluation concepts and adhering to the correct evaluation orientation; and strengthening the academic ethics of teachers [2]. However, there are some problems of the evaluation mechanism of scientific research in humanities and social sciences. Zhang et al. (2022) stated that the evaluation mechanism of scientific research in humanities and social sciences was inadequate in terms of evaluation subjects, evaluation objects, evaluation purposes, evaluation standards, evaluation systems and evaluation methods [3]. Ye et al. (2015) believed that the shortcomings of the evaluation mechanism for humanities and social sciences research mainly included unclear evaluation subjects, difficulty in selecting evaluation criteria, difficulty in judging the value of utility evaluation, and poor atmosphere for academic evaluation. In order to solve these problems, it was necessary to further explore the theory of academic evaluation, improve the evaluation mechanism, pay attention to the basic work of academic evaluation, improve various databases related to academic evaluation, consolidate the realistic foundation of scientific evaluation, and strengthen the system construction related to academic evaluation in academic management [4]. Jiang et al. (2016) stated that there were shortcomings in the evaluation standards, evaluation subjects, evaluation methods and evaluation procedures of the scientific research evaluation mechanism of humanities and social sciences. In order to solve these problems, it was necessary to implement categorical evaluation, advocate quality priority, and dilute management evaluation under the establishment of a stable support mechanism [5]. According to relevant studies, it can be found that there are some problems with the current evaluation mechanism of humanities and social sciences research which is not satisfactory. Therefore, it is necessary to investigate the current teachers' satisfaction with the university humanities and social sciences research evaluation mechanism, and analyze the problems in order to improve the mechanism. However, the current survey scale on teachers' satisfaction with university evaluation mechanisms in the humanities and social sciences is limited. Thus, the study aims to develop a survey scale on teachers' satisfaction with the university humanities and social sciences research evaluation mechanism for promoting universities to make targeted improvements and further improving the evaluation mechanism in humanities and social sciences.

3. Research design and scale development

3.1 Research object

The research object of the study is teachers of humanities and social sciences in universities, including teachers of philosophy, literature, law, history, education, economics and management.

3.2 Research methodology

This study was conducted according to the research process of scale development by Zhang et al. (2021) and Meng (2021), following the steps of item development, draft scale development, expert validity establishment, data collection of pre-test scale, item analysis, exploratory factor analysis, confirmatory factor analysis, reliability analysis and validity analysis [6,7].

3.3 Item development

3.3.1 Clarifying the conceptual scope

Scientific research performance. Performance is defined as achievement. Based on a management perspective, performance is the result of organizational expectations and effective outcomes at different levels to achieve organizational goals. Scientific research performance is an important element in the management of research staff which could contribute effectively to the enhancement of researcher competence and the production of research results. Scientific research performance is a comprehensive judgment of researchers' research ability and level of performance within the framework of research activities. Scientific research performance is a reflection of the completion of research tasks. Usually, research projects and research awards are used to indicate scientific research performance.

Scientific research performance evaluation. Evaluation means that the subject measures, compares, judges, assesses, evaluates, deliberates, appraises or reviews the object. Performance evaluation refers to adopting scientific and standardized evaluation methods to make scientific, objective, fair and comprehensive measurement and judgments on the economy, efficiency, effectiveness and fairness of the process of public expenditure behaviour and its results, based on the expected objectives. It is constantly cycled and dynamically adjusted for improvement, and is continuously enhanced and developed in the cycle. The meaning of scientific research performance evaluation is that in order to accomplish the overall objectives, scientific research organizations adopt scientific methods to judge the scientific research results and scientific research capabilities of scientific researchers with regard to the scientific research activities they have engaged in and the achievements they have achieved. In the process of scientific research performance evaluation, the evaluation subject makes comments on the research achievements of the evaluation object and points out the strengths and weaknesses.

Scientific research performance evaluation mechanism. The mechanism of scientific research performance evaluation refers to a series of systems, rules, policies and procedures that need to be formulated to ensure that scientific research evaluation is carried out scientifically and effectively.

3.3.2 Identifying key dimensions

This study summarizes the key dimensions that should be included in the evaluation mechanism for the scientific research performance of humanities and social science teachers based on the related literature, and obtains two dimensions including the evaluation mechanism elements and the evaluation mechanism operation.

The first dimension is the evaluation mechanism elements (EME). Evaluation mechanism elements are the factors that constitute the evaluation mechanism, including evaluation criteria, evaluation indicator, evaluation subject, evaluation object, evaluation purpose, evaluation process, evaluation method, evaluation periodicity, etc. Since the research object of the study is the evaluation object, in the survey of the elements of the evaluation mechanism, researcher only selects evaluation criteria, evaluation indicator, evaluation subject, evaluation purpose, evaluation process, evaluation method and evaluation periodicity as the measured items. The second dimension is the evaluation mechanism operation (EMO). Evaluation mechanism operation refers to the systems, policies and rules that universities conduct to evaluate teachers' scientific research performance. In this study, the researcher focuses on the system for follow-up research. Based on relevant research, four elements are selected, including research project ranking system, thesis ranking system, performance appraisal system and scientific research achievements transformation system.

3.4 Initial scale development

The study conducted two key dimensions of evaluation mechanism elements and evaluation mechanism operation as the framework to collect measured questions that reflect each dimension as sample items in the initial scale. The items were based on related scale items and interviews to gather relevant content to generate measured items.

Two key dimensions were identified for this study based on the relevant research. By referring to the measured items in related literature scales, 11 sample items were generated for this study. To verify the applicability of the measured items collected in the study from the related literature, the researcher conducted semi-structured interviews with university humanities and social sciences teachers to collect their opinions or suggestions on the relevant measured items. The researcher conducted online interviews with 10 humanities and social sciences faculty members of different disciplines and genders

at different times from January 5 to 31, 2022. The results of the interviews showed that the interviewees confirmed the applicability of the sample items derived from the literature. In addition, the interviewees proposed a new system of dimension of the evaluation mechanism operation, which the researcher added to the collection of measured items. Through the above process, the researcher generated 12 measured items under 2 key dimensions (shown in Table 1) to form the initial scale.

Table 1: Initial scale measured items

Dimension	Item number	Measured Items
Evaluation Mechanism Elements (EME)	EME-1	The clarity of the evaluation criteria
	EME-2	The comprehensiveness of the evaluation indicator
	EME-3	The professionalism of the evaluation subject
	EME-4	The scientific nature of the evaluation purpose
	EME-5	The fairness of the evaluation process
	EME-6	The richness of the evaluation method
	EME-7	The flexibility of the evaluation periodicity
Evaluation Mechanism Operation (EMO)	EMO-1	The research project ranking system
	EMO-2	The thesis ranking system
	EMO-3	The performance appraisal system
	EMO-4	The scientific research achievements transformation system
	EMO-5	The scientific research awards ranking system

4. Test and analysis of the scale

4.1 Expert validity test

To ensure the validity of the items, researcher invited five experts in the field of performance evaluation to evaluate the scientific validity of the scale items. Experts reviewed whether the initial scale was difficult to understand and whether there were any ambiguities in the scale. In addition, experts also examined the extent to which the dimensional concepts expressed by the question measures. Experts agreed with the overall structure of the scale and only revised the statements in the introductory part of the scale, while no deletions or adjustments were made to the measured items.

4.2 Data collection of pre-test scale

Before the survey, researcher invited 10 university teachers to fill in the scale to ensure that the respondents could accurately understand the content of the scale. Researcher distributed 300 electronic scales to the humanities and social sciences teachers at N University and recovered 243 scales, with the recovery rate of 81.0%. After eliminating invalid scales, 231 valid scales were obtained. Therefore, the recovery rate of valid scales was 77.0%.

4.3 Item analysis

Table 2: The CITC of the scale

Measured Variable	Measured Items	CITC	Cronbach's Alpha if Items Deleted	Cronbach's Alpha
1	Q1	0.769	0.934	
	Q2	0.766	0.935	
	Q3	0.845	0.927	
	Q4	0.740	0.937	0.940
	Q5	0.804	0.931	
	Q6	0.867	0.925	
	Q7	0.839	0.928	
2	Q8	0.814	0.935	
	Q9	0.835	0.932	
	Q10	0.818	0.934	0.943
	Q11	0.884	0.923	
	Q12	0.884	0.922	

Researcher examined the discrimination of the items by the critical ration. The top 27% of the total scores after ranking were calculated for the high group. The bottom 27% of the total scores after the

ranking were calculated for the low group. Independent sample t-tests were conducted to calculate the significant difference between the means of the high group and low group on each item. The significance of the independent sample t-test for each item was less than 0.001 which indicated that the items were distinguishable from each other. In addition, researcher calculated CITC by SPSSAU to test the correlation of the items. Meanwhile, the reliability of the items was analyzed according to Cronbach's Alpha coefficients. If the Cronbach's Alpha coefficient was higher than 0.7 or the CITC was less than 0.5, then the item should be deleted. It can be seen from Table 2 that the Cronbach's Alpha coefficients for the two measured variables were 0.940 and 0.943, both higher than 0.7. The initial CITC for all items was higher than 0.5. Thus, all items of the initial scale all passed the test.

4.4 Exploratory factor analysis

After item analysis, researcher explored the internal dimensional composition of the scale through exploratory factor analysis. In order to test the suitability of the scale for factor analysis, the Kaiser-Meyer-Olkin (KMO) and Barlett's sphericity test were conducted on the data to explore whether each item of the scale was relevant. It is generally believed that factor analysis can be conducted when the KMO coefficient is higher than 0.6 and the p coefficient of the Barlett's sphericity test is lower than 0.001. KMO and Barlett's sphericity tests were conducted on the scales by SPSS 26. The results of the tests are shown in Table 3. According to Table 3, the Bartlett's test coefficient was 660.291, p coefficient was lower than 0.001, and KMO coefficient was 0.877. It indicates that there is a significant correlation between the variables and factor analysis could be conducted.

Table 3: The KMO and Bartlett's test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		.877
Bartlett's Test of Sphericity	Approx. Chi-Square	660.291
	df	66
	Sig.	.000

SPSS 26 was used to extract factors from the 12 items by principal component analysis and maximum rotated variance to conduct exploratory factor analysis on the initial scale. The results are shown in Table 4. It is generally believed that if the factor loadings are higher than 0.4, it indicates that the quality of measurement of the items is acceptable. According to Table 4, it can be seen that two factors were obtained with a cumulative variance contribution of 77.645%. The measurements showed that the factor loadings all met the criterion of 0.6. Therefore, in this study, two factors were extracted from the 12 items.

Table 4: The rotated component matrix

Measured Items	Measured Content	Component		Cumulative Variance Contribution Rate
		1	2	
Q1	EME-1	0.707		77.645%
Q2	EME-2	0.673		
Q3	EME-3	0.816		
Q4	EME-4	0.812		
Q5	EME-5	0.810		
Q6	EME-6	0.882		
Q7	EME-7	0.845		
Q8	EMO-1		0.832	
Q9	EMO-2		0.838	
Q10	EMO-3		0.838	
Q11	EMO-4		0.850	
Q12	EMO-5		0.868	

Factor 1 contained seven measured items. The factor loadings for each measured item ranged between 0.673 and 0.882. The items included are: evaluation criteria, evaluation indicators, evaluation subject, evaluation purpose, evaluation process, evaluation method and evaluation periodicity. Factor 1 reflects the elements of the evaluation mechanism. Therefore, these seven measurement items are grouped into a factor named "evaluation mechanism elements". Factor 2 contained 5 measured items. The factor loadings for each measured item ranged between 0.832 and 0.868. The items included are: research project ranking system, thesis ranking system, scientific performance appraisal system, scientific research achievements transformation system and scientific research awards ranking system. Factor 2 reflects the operation of the evaluation mechanism. Therefore, these five measures are grouped into a factor named "evaluation mechanism operation".

4.5 Confirmatory factor analysis

After exploratory factor analysis, the stability of the internal structure of the scale was examined through confirmatory factor analysis by SPSSAU. A structural equation model was constructed with 12 items generated by the exploratory factor analysis as observed variables and 2 factors as latent variables. It is generally believed that if $x_2^{/df}$ is lower than 5, RMR is lower than 0.1, RMSEA is lower than 0.1, NFI is higher than 0.9, CFI is higher than 0.9 and GFI is higher than 0.8, then the model fitting is acceptable. The analysis of each fitting index is shown in Table 5. As can be seen from Table 5, the indicator values for $x^{2/df}$, RMR, RMSEA, CFI and GFI all met the requirements. The indicator value for NFI was 0.846. However, according to Meng (2021), when the rest of the indicator values are good, the value of NFI (>0.8) is acceptable as well [7]. Therefore, the scale can be used as a formal survey scale.

Table 5: The results of confirmatory factor analysis

Fitting Index	Results
$x^{2/df}$	2.364
RMR	0.076
RMSEA	0.086
NFI	0.846
CFI	0.902
GFI	0.873

5. Scale evaluation

5.1 Reliability test.

Table 6: The results of internal consistency analysis

	Cronbach's Alpha	Number of Items
Overall Scale	0.950	12
EME	0.940	7
EMO	0.943	5

The researcher analyzed the data and tested the internal consistency of the scales by measuring the overall consistency reliability and the composite reliability with SPSS 26. It is generally believed that if the Cronbach's Alpha coefficient of the overall scale is higher than 0.8, the Cronbach's Alpha coefficient of the sub- scale is higher than 0.7, then the reliability of the scale is good. As can be seen from Table 6, the Cronbach's Alpha coefficient for the overall scale was 0.950, and the Cronbach's alpha coefficients for "evaluation mechanism elements" and "operation of the evaluation mechanism" were 0.940 and 0.943, respectively, indicating that the internal consistency of the scale items was high and the reliability was good.

Table 7: The results of composite reliability

Dimension	Measured Items	Std. Estimate	CR
EME	Q1	0.782	0.941
	Q2	0.780	
	Q3	0.868	
	Q4	0.763	
	Q5	0.844	
	Q6	0.903	
	Q7	0.887	
EMO	Q8	0.843	0.944
	Q9	0.846	
	Q10	0.841	
	Q11	0.929	
	Q12	0.925	

In terms of the composite reliability, a high composite reliability indicates a high consistency of the underlying constructs. It is generally accepted that if the composite reliability (CR) coefficient is higher than 0.7, it indicates that reliability of the scale is good. The measurements were conducted by SPSSAU. And the results of the measurements are shown in Table 7. From Table 7, it can be seen that all composite reliability coefficients were higher than 0.7 which indicated that the reliability was good.

5.2 Validity test

5.2.1 Content validity

Researcher constructed the scale from two dimensions, and the sample items were synthesized from the related literature and interviews. Researcher combined with expert opinions to validate and revise the content of the scale to ensure the content validity of the scale.

5.2.2 Convergent validity

Convergent validity was tested by factor loadings, average variance extracted (AVE), and CR with SPSSAU. It is generally accepted that if factor loading coefficient is higher than 0.7, AVE coefficient is higher than 0.5, and CR coefficient is higher than 0.7, then it indicates that the convergent validity is good. The factor loading coefficient, AVE coefficient and CR coefficient of each dimension are shown in Table 8. From Table 8, it can be seen that all coefficients met the standards which indicated that the convergent validity was good.

Table 8: The results of convergent validity

Dimension	Measured Items	Std. Estimate	CR	AVE
EME	Q1	0.782	0.941	0.696
	Q2	0.780		
	Q3	0.868		
	Q4	0.763		
	Q5	0.844		
	Q6	0.903		
	Q7	0.887		
EMO	Q8	0.843	0.944	0.771
	Q9	0.846		
	Q10	0.841		
	Q11	0.929		
	Q12	0.925		

5.2.3 Discriminant validity

The AVE square root value can indicate the aggregation of factors and correlation coefficient indicates correlations. If the AVE square root value of a factor is higher than the absolute value of the correlation coefficient between the factor and the other factors, and all factors comply the above request, then it indicates that the discriminant validity is good. The specific results are presented in Table 9. The diagonal lines in Table 9 are the AVE square root values and the other values are the correlation coefficients. For the discriminant validity analysis, the AVE square root value for EME was 0.834 which was higher than the maximum absolute value of the inter-factor correlation coefficient (0.677) which indicated that the discriminant validity was good. For EMO, the AVE square root value was 0.878 which was higher than the maximum absolute value of the inter-factor correlation coefficient (0.677), which indicated that the discriminant validity was good. The results of the reliability test and validity test indicate that the scale could be applied in practice with good quality.

Table 9: The results of discriminant validity

	EME	EMO
EME	0.834	
EMO	0.677	0.878

6. Discussion and conclusion

The scientific research performance evaluation mechanism plays a vital role in humanities and social sciences management of universities. It analyzes and judges the performance of humanities and social sciences teachers to strengthen their strengths and improve their weaknesses which could provide coordination and guarantee for the improvement and development of the university's overall research level and the construction of its disciplines. Thus, it is necessary to understand the current teachers' satisfaction with the university humanities and social sciences scientific research performance evaluation mechanism and the influencing factors in order to make targeted improvements and further improve the scientific research performance evaluation mechanism of university in humanities and social sciences.

Based on relevant literature and semi-structured interviews, the study constructed key dimensions

of scientific research performance evaluation mechanism for humanities and social sciences teachers, including evaluation mechanism elements and evaluation mechanism operation. Then, the study developed a scale of satisfaction with the scientific research performance evaluation mechanism of humanities and social science teachers through the steps of item development, draft scale development, expert validity establishment, data collection of pre-test scale, item analysis, exploratory factor analysis, confirmatory factor analysis, reliability analysis and validity analysis. The data results showed that the internal consistency coefficient of the scale was 0.950, and the internal consistency coefficients of the dimensions ranged from 0.940 to 0.943, and the correlation coefficients of all validity tests were at good levels, which indicated that the scale could effectively measure the current situation of satisfaction with the scientific research performance evaluation mechanism of humanities and social sciences teachers.

The scale is of practical value. Firstly, the scale could be used for the survey of current situation of satisfaction with the scientific research performance evaluation mechanism of humanities and social sciences teachers, and for identification of the problems of the scientific research performance evaluation mechanism of humanities and social sciences in universities. Secondly, the scale could be used to explore solutions to improve the scientific research evaluation mechanism of university humanities and social sciences research performance. For example, the scale could be used to analyze the influencing factors of the scientific research evaluation mechanism of university humanities and social sciences research performance, and to propose new methods of improvement from the influencing factors.

However, there are some limitations of the study. The source of data collection for the study was limited to humanities and social sciences teachers at N university, which might affect the generalisability of the scale. The researcher will expand the scope of data collection and expand the sample balance in order to iterate and revise the existing scale in future studies. Secondly, the study only developed the scale based on part of perspectives of scientific research performance evaluation mechanism. The measurement content is not that comprehensive. Researcher will collect sample items in a more comprehensive perspective in future studies.

Acknowledgement

References

[1] Deng Y. (2004). Establish a scientific assessment and improve achievement awarding. Journal of South China Normal University (Social Science Edition), 5, 7-11.
[2] Li H. (2010). The research on humanities and social science achievement evaluation mechanism in universities. Journal of Beijing Information Science & Technology University, 12, 13-15. DOI: 10. 16508/ j.cnki.11-5866/n. 2010.s1.010.
[3] Zhang E. M., Shi W. B. (2022). Study of scientific research performance evaluation of humanities teachers in local universities. Scientific Management Research, 40(03), 76-82. DOI: 10. 19445/ j. cnki. 15-1103/ g3. 2022.03.009.
[4] Ye J. Y., Yuan X. L. (2015). The past and future of Chinese academic evaluation. China Social Science Review, 1, 65-77.
[5] Jiang F., Huang F. (2016). Suggestions for outcomes about humanities and social science. R&D Management, 28(04), 126-131. DOI:10.13581/j.cnki.rdm. 2016. 04.010.
[6] Zhang W. L., Li S. S. (2021). Development and test of online course learning experience scale. Modern Educational Technology, 31(02), 65-72.
[7] Meng G. H., Liu C. (2021). Development and application of the personal academic information organization level scale for college students. Library and Information Service, 65(12), 73-82. DOI: 10. 13266/ j.issn.0252-3116. 2021.12. 007.