A Systematic Analysis of the Quality Evaluation System for Graduate Education

Pengjun Li[1]

[1] *Graduate School, Jinan University, Guangzhou, Guangdong 510632, China*

ABSTRACT

The graduate education quality evaluation system is the application of a multi indicator comprehensive evaluation method in the field of educational science, and is a complex systematic work. It includes several links and steps such as the guiding ideology of evaluation, evaluation index system, statistical sampling, statistical testing, statistical investigation (subjective and objective indicators), weight determination, data processing, statistical analysis, etc. It is a systematic analysis method that pre-processes and statistically summarizes the different indicators of each evaluation object to obtain comprehensive evaluation values. As a complete system, it is comprehensive and systematic, and is a systematic and scientific testing process for the graduate education status of each evaluation object.

Keywords: *Graduate education, Quality evaluation, System analysis.*

1. INTRODUCTION

At present, China has entered a period of rapid development in graduate education, and establishing a scientific and standardized quality evaluation system is the fundamental guarantee for the sustainable development of degree and graduate education. The quality evaluation of graduate education is conducted in accordance with the Higher Education Law of the People's Republic of China, the Regulations on Academic Degrees of the People's Republic of China, and their provisional implementation measures.

Only by clarifying the guiding ideology for each evaluation, the reasonable evaluation index system, evaluation methods, and analysis methods can be designed to achieve fair and reasonable evaluation requirements in evaluation work, and truly play a role in promoting construction through evaluation.[1]

2. THE GUIDING IDEOLOGY FOR EVALUATING QUALITY OF GRADUATE EDUCATION

In the actual evaluation of graduate education quality, the guiding ideology of evaluation should be determined based on the nature of different types of evaluation activities and the characteristics of graduate training. The guiding ideology for determining the quality evaluation of graduate education should clarify the following main contents:

2.1 The Positioning of Graduate Education

The clear positioning of graduate education is equivalent to grasping the main characteristics of graduate education. According to the requirements of different types of evaluation work, the positioning of graduate education also has different characteristics. For example, If you want to evaluate the quality of doctoral programs in a certain discipline category, it is first necessary to clarify the educational positioning of doctoral students: that is, the training objectives of doctoral students are university teachers and scientific research talents, and the focus of training is to further strengthen the theoretical foundation and expand the scope of knowledge, strengthen scientific research activities, and cultivate the innovative ability of doctoral students. Once the educational positioning of doctoral students is clarified, when setting up an evaluation index system, it is necessary to highlight the

characteristics of this educational positioning, to highlight the assessment of scientific research ability, innovation ability, and other related factors. In short, a reasonable evaluation plan for graduate education can be formulated with a clear positioning of graduate education.[2]

2.2 The Knowledge Structure That Various Graduate Students Should Possess

The clear knowledge structure of various graduate students is crucial for accurately formulating graduate education evaluation plans. Similarly, different types of graduate students require different knowledge structures based on their training objectives. For example, for theoretical master's students, their knowledge structure should reflect the training policy of "wide caliber, thick foundation", that is, while deepening theoretical knowledge, expanding the scope of knowledge. Therefore, when evaluating the quality of theoretical master's education, it is important to focus on assessing the curriculum construction of such graduate students, including the curriculum system, teaching quality, textbook quality, and other influencing factors. For professional degree graduate students, their knowledge structure should reflect two characteristics: firstly, knowledge should be compound, strengthening the learning and mastery of comprehensive knowledge; secondly, knowledge should be operable, reflecting practicality. Therefore, when evaluating the quality of professional degree graduate education, it is important to focus on assessing the teaching characteristics, teaching quality, teaching courseware and case teaching quality, as well as other relevant influencing factors of this type of graduate student. An educational ideology can only be reflected after clarifying the knowledge structures with different characteristics that different types of graduate students should possess.[3]

2.3 The Influencing Factors of Graduate Education Quality

The key assessment content for each type of evaluation can only be determined based on the characteristics of different types of graduate education positioning and knowledge structure. In addition, a comprehensive evaluation index system for a system should carefully analyze other factors that affect the quality of graduate education, including student source quality, training quality, mentor guidance quality, academic environment of subject points, social influence of subject points, school books and equipment, experimental equipment, graduate education management, and other aspects.[4]

In short, the guiding ideology of graduate education quality evaluation is the core of the entire evaluation work. Only when the guiding ideology is clear can it provide fundamental guarantees for the smooth formulation of evaluation plans and implementation of evaluation work.

3. THE INDEX SYSTEM FOR EVALUATING QUALITY OF GRADUATE EDUCATION

The evaluation index system for graduate education is an important component of the evaluation plan, which reflects the guiding ideology of evaluation and is the key to achieving fair and reasonable evaluation. A reasonable and effective evaluation index system is not only a true reflection of the educational quality of graduate training units, but also the foundation of evaluation work. To establish an objective and reasonable evaluation index system, the following principles should be followed.

3.1 The Principle of Consistency

The principle of consistency refers to the common characteristics of different evaluation objects in the process of graduate education. To establish an evaluation index system, it is first necessary to comprehensively understand and grasp the common cultivation characteristics of each evaluation object in graduate education. Only by establishing an indicator system on this basis can the connotation of the indicator system be consistent, Otherwise, there will be bias, which will affect the objective fairness of the indicator system and make it difficult for the evaluation work to be fair, reasonable, and effective.

3.2 The Principle of Comprehensiveness

The designs of the indicator system should be as detailed and comprehensive as possible to avoid one-sided or even biased phenomena. The 'comprehensive' here includes the following aspects:

The first is the combining qualitative and quantitative indicators. Reflecting the status of graduate education includes both quantitative indicators that reflect the true level of cultivation, such as academic teams and teaching equipment, as

well as qualitative indicators that reflect the cultivation environment and quality, such as research direction and academic environment construction. Therefore, the establishment of an indicator system should include both objective evaluation indicators that reflect the current development situation and subjective evaluation indicators obtained from expert surveys, ensuring a combination of qualitative and quantitative indicators, as well as a combination of objective and subjective evaluation indicators.[5]

The second is the combining static indicators with dynamic indicators. The design of each indicator in the indicator system should not only reflect the existing scale and level, but also fully reflect the process of dynamic development and change. This indicator can only reflect the potential for development and the future in this situation. For example, if a school wants to assess the publication status of papers in a certain year, the cumulative number of published papers can reflect the scientific research strength of a discipline construction unit. If an annual average growth rate index of published papers is added, it can reflect the development status and potential of a discipline construction unit. According to evaluation theory, a winner in evaluation can only have true competitiveness if they have strong strength and strong development potential. Therefore, while designing level analysis indicators, designing speed analysis indicators will make the entire indicator system more convincing.[6]

3.3 The Principle of Independence

The design of the indicator system should not only be comprehensive and have overall comparability, but also follow scientific and standardized statistical rules. No matter how many levels of the indicator system are, each level must undergo a weighted average in order to be ranked and evaluated. If each evaluation aspect or specific indicator of each evaluation aspect does not have relative independence at a certain level, then no matter how scientific the weight is applied and how scientific the summary calculation is conducted, the results are actually meaningless. For example, what is the weighted average of the total number of mentors and the proportion of mentors under 50 years old to the total number of teachers? Unknown. Therefore, the principle of independence is a fundamental principle that must be followed when designing evaluation indicators.[7]

The principle of independence refers to the independence between various aspects or indicators within the same aspect at each level, with a theoretical correlation coefficient of O. If Xi and Xj are used to represent two different aspects or indicators in the same hierarchy, then:

$\gamma(X_i, X_j) \le 0$ i, j=1,2……n

In practical operations, as long as the correlation coefficient is less than a certain threshold, it can be considered to basically comply with the principle of independence. Namely:

$\gamma(X_i, X_j) \le a$, a is a threshold.

If it occurs $\gamma\left(\mathrm{X}_i, \mathrm{X}_j\right) > a$, it is necessary to consider the reduction indicators, and the specific reduction methods can refer to the fourth principle. The purpose of reducing provincial indicators here is to ensure the independence of the indicators.

3.4 The Principle of Simplicity

No matter how comprehensive evaluation is conducted, the complexity of the indicator system will not only affect the adherence to the principle of indicator independence, but also increase the evaluation workload, let alone the position and role of certain indicators in the entire evaluation is not significant. Therefore, specific indicators in the original evaluation index system need to be screened, and as many indicators with small effects as possible should be eliminated while retaining indicators with strong representativeness. There are two commonly used methods for screening indicators in comprehensive evaluation:

Firstly, the qualitative analysis is used to select evaluation indicators. The selection of evaluation indicators through qualitative analysis mainly relies on the analysis results of experts. Based on the evaluation purpose, guiding ideology, and principles, general indicators are removed based on expert evaluation experience, and representative indicators are selected as evaluation indicators.

Secondly, the quantitative analysis is used to select evaluation indicators. Quantitative analysis and selection of evaluation indicators refers to the use of statistical methods and the selection of indicators through numerical calculations. The principle of minimizing generalized variance is commonly used to select indicators, and multiple statistical analysis methods can also be used to select indicators. Here is only one commonly used method: combining multivariate statistical analysis

with correlation analysis to screen indicators. The specific approach is to perform cluster analysis on each selected indicator in multivariate statistical analysis, aggregate different indicators into different categories, and then perform correlation analysis on each aggregated indicator to calculate the mean of the determinable coefficient γ^2 (i.e. the square of the correlation coefficient) between the indicators γ^2.

$$\bar{\gamma}_i^2 = \sum_{k=1}^{n} \gamma^2 i,k/(n-1)$$

In the formula, n is the number of indicators in the class. Select $\bar{\gamma}^2$ the indicator Xi with the highest value as a typical indicator of this category, namely:

$$\bar{\gamma}_i^2 = \max_{1 \le i \le n} \bar{\gamma}_i^2$$

By conducting this process for each category, representative indicators can be selected in sequence. Based on the selected indicators, comprehensive evaluation not only ensures the representativeness of the evaluation results, but also makes the evaluation work more operable. After the initial determination of evaluation indicators, it is often necessary to consult with peer experts, send a consultation opinion form, return the consultation opinion form, and conduct comprehensive analysis before determining the final evaluation indicator system.

4. THE COLLECTION OF DATA FOR EVALUATING QUALITY OF GRADUATE EDUCATION

4.1 It Is Necessary to Determine the Sample

According to the basic theory of statistical sampling, a sample refers to a portion of the population, which is composed of individuals or sampling units selected according to a certain procedure from the population. In fact, conducting statistical sampling is to investigate or observe a portion (sample) of the entire (population) of the research object according to a certain procedure, obtain data, and make inferences (such as estimates) about a certain target quantity (parameter) of the population based on this. Therefore, the quality of the sample, in other words, the representativeness of the sample and the amount of overall information contained in the sample, directly affect the quality of inferences or estimates made about the population. Therefore, selecting appropriate samples as evaluation objects is an important step in ensuring the quality of evaluation data.

In the evaluation of graduate education quality, there are only two aspects involved in the sample: on the one hand, from the perspective of the data provider, the samples that need to be extracted are from each surveyed object. For example, when assessing the quality of teaching, it is necessary to randomly check the teaching classroom, teaching teachers, lesson plans, lecture notes, textbooks, student assignments, student exam papers, etc. When assessing the quality of a thesis, a certain number of graduates should be randomly selected for evaluation, and so on. On the other hand, from the perspective of evaluators, the samples that need to be extracted are various experts, such as the selection of peer experts for academic thesis evaluation, and the selection of experts for on-site evaluation. In fact, regardless of the type of evaluation conducted, extracting a certain number of samples as survey subjects or making evaluations is an important part of the evaluation work.[8]

According to statistical sampling theory, there are multiple sampling methods for a certain number of samples, such as simple random sampling, stratified random sampling, cluster sampling, multi-stage sampling, system sampling, and non-probability sampling. In practice, the relatively simple sampling method commonly used is simple random sampling, which means that the probability of each unit in the population being sampled is equal. When using a simple random sampling method to extract samples, there are two points to note: firstly, there should be a clear sampling box, that is, a list or inventory containing all sampling units. The second is to determine the value of the estimator based on sample values, that is, to estimate the overall eigenvalues, including mean and variance estimators. The basic formula is as follows:

Unbiased estimator of overall mean:

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^{n} y_i$$

Unbiased estimator of population mean variance:

$$V(\bar{y}) = \frac{N-n}{nN} S^2 = \frac{1-f}{n} S^2$$

Here i, j=1,2... n represents the n samples collected, represents the observed values of each sample, represents the sample mean, S^2 represents

the sample variance, f represents the sampling ratio, $V(\bar{y})$ represents the population mean variance.

In the actual sampling process, a sampling box can be used to randomly select a certain number of samples for investigation; By using estimators, the overall indicator value can be estimated.

4.2 It Is Necessary to Design a Survey Questionnaire

After determining the sample, the next step is to conduct a survey on the sample to obtain data. Designing a reasonable and effective survey questionnaire is the fundamental guarantee for obtaining accurate survey data. This includes the design of two types of survey questionnaires: one is the objective evaluation indicator data questionnaire in the indicator system; The second is the subjective evaluation indicator data questionnaire in the indicator system.

The objective indicator data questionnaire is a survey conducted on specific data of each objective indicator. Therefore, the survey form should indicate the survey time, survey scope, and survey object for each item of data. At the same time, the design of the survey form should be concise, concise, and clear, so as to minimize objections from the respondents when filling out the form. The subjective evaluation index data questionnaire is a questionnaire designed separately based on the survey information required for each subjective evaluation index. When designing such surveys, special attention should be paid to survey techniques. Because subjective evaluation is the main focus, each survey item in the questionnaire should not be idealized, and at the same time, the respondents should be able to make truthful judgments as objectively as possible. On the one hand, it is necessary to avoid the phenomenon of no response, and on the other hand, statistical sampling methods (such as Werner's randomized response model, Simmons' randomized response model, etc.) can be used to cleverly ask questions about third sex questions. If there is a phenomenon of no response (which often exists in actual investigation and evaluation), it should be resolved in a timely manner. The first solution is to replace the samples in a timely manner and conduct a re-investigation; The second is to use methods such as weight segmentation to minimize the impact of non-response when compiling estimates, in order to minimize non response errors.

4.3 It Is Necessary to Collect Survey Data

Based on the actual situation of graduate education, there are several ways to collect survey data:

The first is that the respondents provide objective evaluation indicators and survey data. The survey data of objective evaluation indicators provided by the respondents is based on the objective evaluation indicator questionnaire designed by the respondents and combined with the basic data provided by the actual situation. This part of data is objectively existing, and as long as the respondents fill it out carefully, the accuracy of their data can be guaranteed.

The second is the subjective evaluation index survey data provided by experts (or others). There are three commonly used data collection methods here: the first is through communication and evaluation, which involves sending the evaluated materials and evaluation forms to experts or other hired individuals, and experts (or others) make judgments based on the provided materials and evaluation form content. The second method is for experts (or others) to conduct on-site inspections and make judgments on the evaluated objects based on the evaluation table based on their impressions. The third type is to hold a symposium specifically between experts (or others) and all (or part) of the respondents. On the one hand, the symposium listens to the evaluation report of the surveyed unit, and on the other hand, they converse with each other to make judgments on the respondents based on the evaluation content.

The third is to access public information through the graduate education management information system. Evaluation data can be obtained through the local area networks of various graduate education institutions. This can not only reduce the workload of respondents filling in data, but also ensure the authenticity and effectiveness of the data. With the development of network information systems, this will become an important data source channel for graduate education evaluation.

In short, the collection of evaluation data is a very important aspect of evaluation work, and it is a complex statistical inference technique based on statistical sampling theory, which also includes some skills. Therefore, data collection is an important step in ensuring the fairness and rationality of evaluation results.

5. THE PROCESSING OF GRADUATE EDUCATION EVALUATION DATA

The processing of graduate education evaluation data includes three steps: determining the weight of statistical indicators, preprocessing statistical data, and summarizing statistical data. These steps are described as follows:

5.1 The Measurement and Statistical Testing of Statistical Indicator Weights

The implementation of comprehensive evaluation methods relies on a complete indicator system and the weight of the indicators. Therefore, the determination of weights is an essential and important step in the comprehensive evaluation process.

The so-called weight refers to the position or role played by an indicator (or aspect) in the indicator system, and also represents the relationship between this indicator (or aspect) and other indicators (or aspects). The different weights of different indicators (or aspects) directly determine the evaluation results and final evaluation results at each level. There are two commonly used weight measurement methods in the evaluation of graduate education quality both domestically and internationally: deterministic weight and fuzzy weight. Deterministic weights are commonly used in foreign countries using the Delphi method, which is characterized by consulting expert opinions to determine weights and not being influenced by others. Experts provide specific weight values for each evaluation indicator or aspect based on their evaluation experience, and then calculate a simple arithmetic mean of the weight values given by all experts for the same indicator or aspect. Namely:

$$\bar{\omega}_i = \sum_{j=1}^{n} \omega_{i,j} / n$$

This ω_i represents the weight of the i-th indicator, ω_{ij} represents the weight given by the jth expert to the i-th indicator. Fuzzy weights, in simple terms, divide the indicators or aspects to be weighted into five levels or several levels in order, A, B, C, D, E, each level representing the importance of the indicator or aspect in the evaluation. Experts, based on evaluation experience, draw one of the levels set after each indicator (or aspect), and use all expert votes in different importance levels of each indicator as weights for weighting calculation, Finally, normalize the results. Namely:

$$\omega_i^{'} = \omega_i \Big/ \sum_{i=1}^{n} \omega_i$$

The final adjusted weight u is the average weight value of each indicator.

The determination of weights plays a crucial role in comprehensive evaluation, and the representativeness of the measured weights is an important indicator of whether the weights are reasonable or not. Therefore, regardless of which method is used to determine weights, statistical testing of the measured weights is an essential and important step in weight determination. There are two simple statistical methods commonly used for weight testing. One is to use the idea of variance analysis to calculate the variance of the weights given by all experts for each indicator. Namely:

$$\delta_i^2 = \sum_{j=1}^{n} (\omega_{ij} - \bar{\omega}_i)^2 \Big/ n$$

Here, δ_i^2 represents the variance of the weights given by all experts in the i-th indicator, ω_{ij} represents the weights given by the jth expert to the i-th indicator, and $\bar{\omega}_i$ represents the average weight given by all experts to the i-th indicator.

Based on the calculation results, if the variance is large, it indicates that the distribution of weights given by different experts is scattered, and the average of the calculated weights is not representative. Further expansion of the sample size is necessary to conduct expert consultation again. If the variance is small, it indicates that the average of the calculated weights in the distribution set of weights given by different experts is representative and can be used as an indicator or aspect of weight. Another weight testing method is to calculate the discrete coefficients.

$$C.V = \frac{\delta}{\bar{X}}$$

Based on the size of the dispersion coefficient and the degree of concentration of the weight distribution given by each expert, a high degree of concentration indicates that the calculated average weight is representative and can be directly used as a comprehensive evaluation indicator or aspect weight. On the contrary, it is necessary to conduct a new investigation and consultation until the

distribution is concentrated and the average weight is representative.

Overall, the determination of weights is a very important aspect of comprehensive evaluation. The selection of weight determination methods is based on the characteristics of each method and objective circumstances. Testing the measured weights is a necessary condition to ensure the objectivity and impartiality of the weights.

5.2 The Preprocessing of Evaluation Data

Whether it is evaluating the quality of graduate education in various universities or conducting subject evaluations, when calculating the evaluation values layer by layer based on the collected data, the original data must be preprocessed to eliminate the impact of different indicator dimensions and convert dimensional indicators into dimensionless indicators. Only in this way can the synthesis of indicators be achieved in a scientific sense. For example, if the proportion of the total number of mentors and the number of mentors under 50 years old to the total number of teachers is weighted average, the result is neither a number nor a percentage, and weighted summary is meaningless. Therefore, data preprocessing is necessary before data synthesis. Common methods for data preprocessing include:

- The standardization transformation

$$x_{ij'} = \frac{x_{ij} - \bar{X}_i}{S_i}$$

Among them: $\bar{X}_i = \sum_{j=1}^{n} x_{ij} / n$

$$S_i = \sum_{j=1}^{n} (x_{ij} - \bar{X}_i)^2 / n - 1$$

i，j=1,2……n

Xi represents the average of the i-th indicator of each evaluated unit, and Si represents the sample standard deviation of the i-th indicator of each evaluated unit.

- The normalization transformation

$$x'_{ij} = \frac{x_{ij} - \min\limits_{1 \le j \le n} \{x'_{ij}\}}{\max\limits_{1 \le j \le n} \{x'_{ij}\} - \min\limits_{1 \le j \le n} \{x'_{ij}\}}$$

i，j=1,2……n

- The logarithmic transformation

$$x'_{ij} = InX_{ij}$$

i，j=1,2……n

The most commonly used method in practical work is data standardization transformation. After the standardization transformation of the original data, it becomes a dimensionless pure number that fluctuates around it, and is subjected to weighted arithmetic operations, in accordance with the arithmetic operation rules.

5.3 The Summary of Statistical Data

According to evaluation theory, statistical data aggregation refers to the weighted aggregation of preprocessed data layer by layer to obtain a comprehensive evaluation value. The basic formula is:

$$Y_j = \sum_{i=1}^{n} \omega_i g X'_{ij}$$

Here, Y_j represents the evaluation value of the jth evaluated unit, ω_i epresents the weight of the i-th indicator, and X_{ij} represents the standardized value of the i-th indicator of the jth evaluated unit.

Based on the evaluation values Y_j of each evaluated unit, evaluation and analysis can be made on the evaluated unit, thus completing the entire process of data processing.

6. CONCLUSION

After statistical data processing, the comprehensive evaluation values of each evaluated object were obtained, and the evaluated objects were ranked based on the comprehensive evaluation values. So far, it seems that the evaluation work has come to an end. In fact, evaluation itself is not the purpose, the true purpose is to "promote construction through evaluation". Only by conducting comprehensive, systematic, and scientific statistical analysis of the evaluation results can the educational status of all evaluated objects be truly revealed, and scientific decision-making basis be truly provided for "promoting construction through evaluation". T Therefore, statistical methods and evaluation results are important references for adapting to the needs of different subjects and making scientific decisions, and are a very important content in the quality evaluation of graduate education.

ACKNOWLEDGMENTS

Fund program: The 2018 National Education Science "13th Five-Year Plan" Key Project of the Ministry of Education "Research on Graduate Education Quality Assessment Based on Connotative Development" (No. DIA180376).

REFERENCES

[1] Report on the Development of Degree and Graduate Education in China from 1978 to 2003 [M] Beijing: Higher Education Press, 2006.

[2] KOFFMAN R G. A comparison of the perceptional organization of outstanding and randomly selected teachers in open and traditional classrooms [D]. University of Masschuselts, 1975:22-23.

[3] Compilation Group for "Report on the Development Strategy of Degree and Graduate Education in China". Report on the Development Strategy of Degree and Graduate Education in China (2002-2010) (Draft for Soliciting Opinions) [J] Degree and Graduate Education, 2002 (6): 5-11.

[4] Donald A. Schein. Reflection on Practitioners: How Professionals Think in Action [M] Translated by Xia Linqing Beijing: Education Science Press, 2012:85.

[5] John Dewey. The Search for Certainty [M]. Translated by Fu Tongxian Shanghai: Shanghai People's Publishing House, 2005: 5.

[6] Ralph Taylor. Fundamentals of Curriculum and Teaching [M] Translated by Shi Liangfang Beijing: People's Education Press, 1994:68.

[7] Yu Xiulin Multiple Statistical Analysis and Programs [M] China Statistical Publishing House, 1993.

[8] WANG S, LI C, DING S, et al. Research and exploration of graduate education quality evaluation and monitoring system[J]. International Journal of Social Sciences in Universities, 2020, 3(3): 224-226.