

Theory article

Multi-indicator comprehensive evaluation: reflection on methodology

Dong Qiu¹ and Tingyi Liu^{2,*}

¹ Jiangxi University of Finance and Economics, Nanchang, 330013, China

² Economics and Management School of Wuhan University, Wuhan, 430072, China

* **Correspondence:** Email: 2019101050047@whu.edu.cn.

Abstract: The number and field of researches on the application of Multi-Indicator Comprehensive Evaluation (MICE) are increasing. It is important to reflect on the understanding of the MICE method systematically and the issues implied behind it. This paper compares the core concepts and methodological elements of the three papers that systematically study the MICE method. It is found that the views of the three papers on the core issue are consistent and mutually supportive, but there are differences in the step division and sequence of the evaluation content. In addition, this paper considers the historical status of the MICE and holds that the key to solving the quality of weight lies in the “equivalent conversion” problem in the MICE. Taking the Human Development Index as an example, this paper illustrates the absoluteness of the “equivalent conversion” relationship. In addition, there are multiple processing methods for the MICE from the spatial dimension and multiple evaluation results accordingly, therefore, the results of the MICE need to be used carefully. Finally, based on the systematic summary and reflection of the MICE method, three suggestions are given for the application of the MICE method.

Keywords: Multi-Indicator Comprehensive Evaluation; methodological elements; statistical evaluation; equivalent conversion; spatial interpretation

JEL Codes: B4

1. Introduction

Multi-Indicator Comprehensive Evaluation (MICE) has been widely studied and applied. It is often referred to as “Composite indicators” or “Synthetic indicators”, and domestic scholars also refer to it as “Multi-Attribute Comprehensive Evaluation”, “Multi-Objective Comprehensive Evaluation” and “Multi-Variable Comprehensive Evaluation”, etc. In addition to the differences in name, there are

also differences in understanding. The MICE merge two or more data sources into a single measure, which is usually used to measure the performance of a country/entity in complex phenomena such as innovation, competitiveness and sustainable development. However, since the beginning of the application of the MICE method, there seems to be no unified definition to explain it. The European Commission's State-of-the-art Report on Current Methodologies and Practices for Composite Indicator Development argues that "composite indicators are based on sub-indicators that have no common meaningful unit of measurement and there is no obvious way of weighting these sub-indicators" (Saisana and Tarantola, 2002). Another definition provided in the OECD's Handbook on constructing composite indicators is that "composite indicators are formed when individual indicators are compiled into a single index on the basis of an underlying model" (Nardo et al., 2005). According to biologist Robert Rosen's book *Life Itself*, published in 1991, the complexity of complex systems refers to the causal impact of organization on the system as a whole (Rosen, 1991). In essence, composite indicators may reflect a "complex system" made up of numerous "components", making it easier to understand in full rather than reducing it back to its "spare parts" (Greco et al., 2019).

Despite the absence of a clear and uniform definition, comprehensive indicators have been applied and popularized in almost all research fields because of their simplicity and comparability. In 2006, Bandura introduced 165 leading composite indicators studied by international organizations, governments, research institutes and scholars, covering various aspects of national governance, competitiveness, environment, security, or other aspects (Bandura, 2005). Bandura also reviewed more than 400 official composite indicators in 2011, ranking or assessing national performance based on several economic, political, social or environmental indicators (Bandura, 2011). In addition, Salvatore Greco et al. searched for "composite indicators" in SCOPUS, showing results from 1997–2016. The growth in the literature over the last 20 years has been exponential, with no sign of a decline in the number of annual publications (Greco et al., 2019).

However, no method is perfect, including the MICE method. The advantages of comprehensive indicators are where they have been criticized. The simplicity of the results can lead to oversimplified policy conclusions, disguise severe failings in some dimensions, and even send misleading policy messages (OECD, 2008). The arbitrariness and subjectivity of its construction method are also controversial. Therefore, to better understand the complexity of the MICE method, it is necessary to systematically study its methodology and clarify the construction steps to ensure their transparency and soundness. In 1988, Dong Qiu presented a paper entitled "Systematic Analysis of Multi-Indicator Comprehensive Evaluation Methods" for the first National Youth Science Conference. In 1990, Dong Qiu submitted his doctoral dissertation of the same name, which was revised and published by China Statistics Press in 1991. This paper is the first systematic methodological study on the multi-indicator comprehensive evaluation of many cases. Su (2001) conducted a systematic methodological study on the multi-indicator comprehensive evaluation method. OECD (2008) developed a methodological manual for composite indicators, which formulates a standard guideline and identifies ten steps to guide users.

Based on the above three systematic studies on the MICE method, this paper reflects on its methodology by comparing their conceptual understanding and methodological elements (Sections 2 and 3). From the historical status of MICE, this paper objectively looks at the controversial point of the arbitrariness of weight setting (Section 4) and points out that the key of the dispute lies in the "equivalent conversion" of multi-index comprehensive evaluation (Section 5). Then this paper explains the MICE from the spatial dimension (Section 6) and discusses the nature of the MICE results

(Section 7). Finally, the main conclusions of this paper and the suggestions for applying the MICE method are presented (Section 8).

2. Comprehensive evaluation of multi-indicator of the broad views and narrow views

Dong Qiu identified this method as “the comprehensive evaluation of multi-indicator”, which had its specific connotation: “multi-indicator” means that evaluation uses multiple evaluation dimensions for the evaluated object, in order to achieve a comprehensive evaluation (Qiu, 1988). The name of “Comprehensive evaluation” is to make it clear that it is through the synthesis to reach the purpose of the sorting of evaluated object. The definition emphasizes the comprehensiveness and integrity of evaluation, and regards it as an important dimension for distinguish various indicators evaluation methods. In 1990, Dong Qiu summarized the concept, basic steps, basic variables and key issues, calculation properties, evaluation results and basic functions of multi-indicator comprehensive evaluation in his doctoral dissertation (Qiu, 1991).

This definition focuses on the evolution of the traditional evaluation to comprehensiveness and integrity, distinguishing MICE from physical indicator evaluation, comprehensive indicator (value indicator) evaluation and indicator system evaluation and so on. The “indicator synthesis” is considered to be the fourth method of statistical evaluation.

Dr. Wei-hua Su outlined the five basic elements of comprehensive evaluation of multi-indicator: evaluation subject, evaluation object, evaluation indicator system, evaluation model and evaluation results, to answer who will evaluate, whom to evaluate, what and how to evaluate (Su, 2001). Dr. Wei-hua Su believed that the evaluation model can be divided into seven methods, which are, qualitative comprehensive evaluation, quantitative comprehensive evaluation, equivalent evaluation method (utility function evaluation method), multivariate statistical method, fuzzy comprehensive evaluation, gray system comprehensive evaluation method, logistics optimization and decision-making method, intelligent evaluation. Correspondingly, the comprehensive evaluation results can be expressed as the value of quantitative evaluation, evaluation ranking and evaluation categories, namely “value evaluation”, “ranking evaluation” and “classification evaluation”.

In the methodology handbook of OECD, conventional methods and multivariate statistical analysis methods are used to construct composite indicators (OECD, 2008). In terms of content structure, the manual introduces more technical methods, but the discussion of ideology seems insufficient.

The core understanding of the multi-indicator comprehensive evaluation methods is usually consistent, and the difference lies mainly in its extension. Here only taking the difference in method understanding between Dr. Weihua Su and Dong Qiu as an example to illustrate.

First, Dr. Weihua Su has a broader understanding of the role of multi-indicator comprehensive evaluation (Su, 2001), while Dong Qiu narrowly restricts it to sorting, and does not regard fixed classification and fixed distance as the exclusive feature of multi-indicator comprehensive evaluation.

Second, in terms of the nature of the indicators, Dr. Weihua Su proposed that the comprehensive evaluation of multi-indicator can be included in the category of the comprehensive indicator method, while Dong Qiu emphasized the distinction between comprehensive indicators (value indicators) and indicator composite (composite Indicators).

Third, Dr. Weihua Su believed that a misunderstanding in evaluation practice is that sorting is the ultimate goal, and the comprehensive evaluation results can be used as a new variable for in-depth

statistical analysis. While Dong Qiu believes that comprehensive evaluation results should not be over-interpreted. Second, the re-application of comprehensive evaluation results should be more cautious.

Fourth, as mentioned above, Dr. Weihua Su summarized eight comprehensive rating methods. In 1990, Dong Qiu only summarized three types of comprehensive evaluation methods, which are the second, the third and the fourth method of the above eight methods. This is partly due to the research time, but it does not rule out the different understanding of the scope of the evaluation methods. Dong Qiu summarized the calculation nature of comprehensive evaluation of multi-indicator as “weighted average of the relative number of statistics”, which is only a narrow understanding, and some methods are difficult to generalize with this nature.

3. Understanding the method elements from the evaluation steps of MICE

For the discussion of the methodology of multi-indicator comprehensive evaluation, it is necessary to summarize the evaluation steps. Here we compare three generalizations of Dong Qiu, Weihua Su, and the OECD composite indicator expert group to show the method elements of multi-indicator comprehensive evaluation. Dong Qiu summarized the seven basic steps of MICE in 1990:

1. Select evaluation indicators and establish evaluation index system
2. Select dimensionless and composite formula
3. Determine the relevant thresholds and parameters of the indicator
4. Determine the indicator weight
5. Dimensionless, that is, the actual value of the indicator is converted into the evaluation value of the indicator
6. Weighted average, that is, to synthesize the evaluation value of each indicator to obtain a comprehensive evaluation value
7. Sort the evaluated objects according to the comprehensive evaluation value.

Dr. Weihua Su summarized basic process of comprehensive rating:

1. Determine the purpose of the evaluation
2. Establish the evaluation indicator system
3. Determine the indicator weight
4. Select the evaluation model or methods
5. Implement a comprehensive evaluation
6. Evaluate and test the results
7. Analyze and re-apply the evaluation results

The OECD Composite Research Expert Group presents ten steps in the handbook on the composite indicator construction methods and user guide:

1. Develop a theoretical framework
2. Select variables
3. Reckon missing data
4. Multivariate analysis
5. Normalize data
6. Weight and aggregate
7. Robustness and sensitivity analysis
8. Data reduction
9. Correlation analysis between the other variables

10. Presentation and dissemination

See Table1, Table 2, Table 3.

Table 1. Preprocessing step for multi-indicator comprehensive evaluation.

	Dong Qiu	Weihua Su	OECD
Determining evaluation purpose		(1)	
Development of theoretical framework			(1)
Establishing indicator system	(1)	(2)	
Selecting evaluation indicator	(1)	(2)	(2)
Reckoning missing data			(3)
Multivariate analysis			(4)
Selecting dimensionless and composition formula	(2)	(4)	
Determining indicator threshold and parameter	(3)		

Table 2. Step of core phase of multi-indicator comprehensive evaluation.

	Dong Qiu	Weihua Su	OECD
Determining the indicator weight	(4)	(3)	(6)
Dimensionless	(5)	(5)	(5)
Composition	(6)	(5)	(6)

Table 3. Step of post-processing stage of multi-indicator comprehensive evaluation.

	Dong Qiu	Weihua Su	OECD
Evaluation, inspection evaluation results		(6)	
Robustness analysis, sensitivity analysis			(7)
Data reduction			(8)
With the correlation analysis between the other indicators			(9)
Sorting	(7)		
Expression of evaluation results and release			(10)
Reapplication of evaluation results		(7)	

By comparison, it can be found that: first, the comprehensive evaluation of multi-indicator can be roughly divided into three stages, it is respectively the pretreatment stage of the evaluation, the core stage and the post-processing stage. In general, especially in the core stage, each step is recognized. The difference only lies in the order and step division of the comment content.

If the dimensionless and composition is regarded as a key step in the MICE method, then it is not appropriate to incorporate the multi-index comprehensive evaluation method without the rough evaluation method of these two steps, such as mandatory scoring method, which should not be included in MICE method, that is to say, the MICE method should be properly defined, and the range should not be too wide. The presence or absence of core steps should be a sign to determine whether it is the MICE method.

Second, selecting evaluation indicator is essentially the first time to determine indicator weights. The indicator that is not selected means that its weight is zero, and the weight of the selected indicator is yet to be determined.

However, the weight of a certain indicator is zero, which does not necessarily mean that it is not important for evaluation, but it is often because it is unmeasurable or difficult to measure. From the perspective of calculation feasibility, its measurement has to be abandoned. From the outside, it seems that giving up this indicator means that it is irrelevant to the measurement, and it is often misunderstood. This is one of many “measurement traps”, which requires special attention.

It is actually a second time to determine the weight at the core stage of the evaluation, according to certain principles to determine the specific value of selected indicator weight, that is the size of the weight. This is actually the second time the weight has been determined.

Third, we should pay attention to the relevance of the selected evaluation indicators. If the evaluation indicators A and B are completely related, and the information contained in the two indicators overlaps, it is not necessary to include both at the time of synthesis. If the two are completely unrelated, it means that at least one of the indicators is irrelevant to the object being evaluated, which cannot enter the evaluation indicator system at the same time. Therefore, the evaluation indicators in MICE are often between completely correlated and completely uncorrelated, that is, partial correlated (OECD, 2008). And the issues that have to be noticed evolve into: how to grasp the indicator, and how does different treatment affect the results of comprehensive evaluation?

Fourth, the content of the post-processing stage is slightly different. Dong Qiu only proposed the idea of result testing in 1990, but did not include it in the basic steps of evaluation. Dr. Weihua Su paid more attention to the inspection of the evaluation results and lists it as a separate step. The OECD pays more attention to the evaluation of the post-test, which needs three steps to carry out separately. At the same time, they also pay attention to the expression and release of the results, specifically listed as an independent step. On the whole, due to the abstractness of the comprehensive evaluation results, it is necessary to conduct multi-directional inspections, so that the comprehensive evaluation results have more practical guiding significance.

In addition to the examination of evaluation results, Dr. Weihua Su also emphasizes the reapplication of evaluation results, which corresponds to his broad understanding of evaluation methods. Qiu Dong’s question is that there is one biggest difference between the multi-index comprehensive evaluation results and the value indicators in economic operation, that is, the generation mechanism of the indicator data is different. The two types of data are then put into the measurement model to calculate the new results. How reliable is it? It is worthy of further consideration.

Fifth, the indicator system here is only for the selection of composite indicator. Whether it can be selected depends on whether it is helpful for the comprehensive evaluation. It belongs to the indicator preprocessing in the multi-indicator comprehensive evaluation. However, directly using the indicator system to evaluate things is different from this kind of preprocessing. It is a special evaluation method that focuses on the comprehensiveness of the evaluation and discards the integrity. It should be noted that the difference between the two in evaluation thinking.

Dr. Weihua Su believed that the thought and practice of comprehensive evaluation has existed since ancient times. From the pre-Qin “eight observations and six experience” to the Qing Dynasty “four grids and eight law”, they are various methods to understand people, and he used them as examples of comprehensive evaluation. Dong Qiu believes that “eight observations and six experiences” belongs to the indicator system of evaluation. Although the final decision has to be made, but there is no synthetic treatment in it, and it cannot be regarded as a comprehensive evaluation of multi-indicator.

4. Three comprehensive evaluations of statistical evaluation

To achieve a systemic grasp of multi-index comprehensive evaluation, researchers should not only compare and study all the methods in comprehensive evaluation practice horizontally, but also compare it with other statistical evaluation methods vertically. It should be the original intention of systematic research to make a thorough research, horizontally and vertically.

In 1990, Qiu Dong proposed that statistical evaluation can be divided into four methods: physical indicator, value indicator, indicator system, and multi-indicator comprehensive evaluation. In this way, our understanding of things tends to be increasingly comprehensive, holistic, and integrated. Comprehensiveness and integrity have always been the goals pursued by humans in statistical evaluation.

Comparing the four evaluation methods of statistical indicators, the value indicators are generally better than the physical indicator in terms of comprehensiveness and integrity. The level of integrity is the strongest in the comprehensive evaluation of multi-indicator, but it is more abstract. The indicator system covers the broadest information, but the comprehensiveness and integrity are compromised.

Therefore, Dong Qiu believes that, corresponding to the above four statistical evaluation methods, so far, there are three comprehensive evaluation tools: physical indicator, value indicator and composite indicator, without exception.

Even physical indicators have a certain degree of comprehensiveness. For example, for the earliest wealth indicator-total grain, the difference in calories between different grains are not considered, adding up only by output to get a comprehensive indicator. If we understand people in a broad way as some matter, the amount of time is the most comprehensive physical indicator (real indicator). Time can be added up, and its comprehensiveness is even no less than some value indicators. Now, people are paying more attention to indicators, and some international organizations specialize in developing and engaging in time accounting methods. Of course, there are still many limitations on the comprehensiveness of physical indicators.

Currency is one of the greatest inventions of mankind, and value indicator is a by-product of this invention. Value indicator has greatly improved the comprehensiveness of statistical evaluation and compensated for the shortcomings of physical indicator. They are an outstanding contribution of economic statisticians to mankind. The value indicator uses price as the same measurement factor for different evaluation factors, so as to solve the problem of indicator additivity that must be solved in the comprehensive evaluation to a certain extent.

However, not all things being evaluated have a price, or can be measured by price, which has created new restrictions on the comprehensiveness of value evaluation. Some economists advocate estimating a value for things without a market price, which is to use “price” to the extreme, such as the design of the “total social capital” indicator. There are also many economists who are not optimistic about this valuation. They advocate seeking new solutions.

Beginning in the 1860s, as developed countries paid more attention to social issues, multi-indicator comprehensive evaluation was gradually carried out. Different from the value indicator seeking and using the same measurement factor, the comprehensive indicator is a reverse operation: since it is impossible to find a general measurement of the same factor, then simply remove the dimensions of all evaluation indicators, and try to solve the additivity problem in the calculation of the composite indicator at one time. The presence or absence of the same measurement factor is the key difference between the value indicator evaluation and the multi-indicator comprehensive evaluation.

Although the three comprehensive evaluation tools are not mutually replaceable, their primary and secondary fluctuations are closely related to the changes of the national accounting paradigm. Under the political arithmetic paradigm (the Petty paradigm), physical indicators and value indicators were used in parallel, and then gradually became a supplementary position in evaluation. Under the modern national accounting paradigm (Kuznets-Stone paradigm), value indicators have achieved a dominant position in statistical evaluation, and the popularity of SNA is the best proof of this status.

From the beginning of social indicators movement, people began to pay attention to non-economic statistical evaluation, and non-comprehensive issue of value indicators began to emerge. To improve the comprehensiveness of the evaluation, it is necessary to add non-value indicators to the original evaluation, and the resulting problem is that the evaluation loses the possibility of integrity. The composite indicator is an exploration that people want to ensure comprehensiveness and achieve integrity.

In 2009, the measurement research report published by the Commission on the Measurement of Economic Performance and Social Progress, was hosted by Joseph E. Stiglitz, Amartya Sen, and Jean Paul Fitoussi. A series of issues about the existing value evaluation methods had been raised. They also put forward that the focus of measurement should shift from economy to people's life and welfare, which inevitably involves the historical position of comprehensive evaluation of multi-indicator. In fact, it still raises the issue of evaluation paradigm reform change.

There have always been two completely different opinions on whether the evaluation indicators can be synthesized into a total quantity indicator. The "gross faction" is based on realizing the temporal and spatial ordering of different things and supports the synthesis of indicators. The "non-gross faction" holds that it is impossible to achieve a comprehensive and holistic indicator, people can only stop at the evaluation of indicator system. It is arbitrary to put evaluation indicators into a comprehensive evaluation. The main objection lies in the arbitrariness of weighting in comprehensive evaluation.

In fact, even in the value indicator, the price as a weight has the phenomenon of distorting the total amount. However, in the multi-indicator comprehensive evaluation, it is often necessary to specifically generate weights. It seems that the objectivity of weights is not so strong, and people have doubts about the reliability of weights. However, the key issue of whether or not this change in statistical indicator evaluation can be achieved is not the quality of weight determination, but more importantly, the "equivalent conversion" problem in the comprehensive evaluation of multiple indicators.

5. The "equivalent conversion" of multi-indicator comprehensive evaluation

From the core steps of multi-indicator comprehensive evaluation, the implicit "equivalent conversion" problem can be found. This article uses the Human Development Index (HDI) as an example to illustrate this point.

1. Example of HDI

As we all know, HDI is composed of three aspects: GNI per capita, life expectancy per capita and educational level. It can be seen from its synthetic formula that the change of HDI can be the result of the combination of the three indicators. For example, if only one of the three constituent indicators changes, and the other two remain unchanged, which can lead to a change in the total indicator. Or, two of the three constituent indicators change, one remains constant, it can also change the total index. Of course, the most common thing is that all three indicators have changed, but the range of change is different, forming different combinations of indicator changes.

In order to introduce the concept of “equivalent conversion” concisely, we focus on the situation where only one constituent indicator changes.

Every 1% increase in HDI means a change in one of its PPs. It may be caused by a certain amount of per capita GNI increase, or it may be caused by a certain increase in the average life expectancy, or it may be caused by a certain increase in the level of education (Kagan, 2009).

To illustrate the problem more vividly, we assign three “certain amounts” to specific values. For example, we assume that every 3% increase in per capita GNI will increase HDI by 1%, and every half-year increase in average life expectancy will increase HDI by 1%, and every 2% increase in education level will increase HDI by 1%. Of course, it is the same to use other specific values to explain the problem.

Having each constituent indicators change a certain amount, can achieve an increase in HDI by 1%, which means that in terms of the growth of HDI, changes in different constituent indicators result in same effect. According to the hypothetical value, per capita GNI increased by 3%, which is the same as an increase in average life expectancy by half a year, and an increase in education level by 2%. The three contributions to the growth of HDI are the same, indicating that it doesn’t matter what the path to growth is. This is the “equivalent conversion” issue proposed in this article.

2. The general expression of “equivalent conversion”

General expression:

$$CI = f(x, y, z) \quad (1)$$

$$\Delta CI = f(\Delta x, \Delta y, \Delta z) \quad (2)$$

$$\Delta CI = f(\Delta x) \text{ when } \Delta y = 0, \Delta z = 0 \quad (3)$$

$$\Delta CI = f(\Delta y) \text{ when } \Delta x = 0, \Delta z = 0 \quad (4)$$

$$\Delta CI = f(\Delta z) \text{ when } \Delta x = 0, \Delta y = 0 \quad (5)$$

From the perspective of equal contribution of ΔCI :

$$f(\Delta x) = f(\Delta y) = f(\Delta z) \quad (6)$$

It is clear that this “equivalent conversion” relationship will also exist if there are only two constituent indicators that having a change or three constituent indicators all having a change. Such as when:

$$\Delta CI = a\Delta x + b\Delta y + c\Delta z \quad (7)$$

$$\Delta CI = l\Delta x + m\Delta y + n\Delta z \quad (8)$$

From the equal contribution to ΔCI , there can be

$$a\Delta x + b\Delta y + c\Delta z = l\Delta x + m\Delta y + n\Delta z \quad (9)$$

Where a, b, c, l, m, n are a certain amount of variation coefficients of constituent indicators.

3. The necessity of “mathematics additivity” and its expansion

The implicit question here is: Why does a 3% increase in per capita GNI equal to an increase in average life expectancy by half a year? Why can it be equivalent to a 2% increase in education level?

In general, why a certain increase in per capita GNI is equivalent to a certain increase in average life expectancy or education level? More generally, why is a combination of different constituent indicators equivalent to a certain amount of change? What is the socio-economic significance of the establishment of this “equivalent conversion” relationship? This problem is related to the issue of empowering different constituent indicators, but from a fundamental point of view, it is actually the “additivity problem” or “integrability problem” of each constituent indicator in the socio-economic sense.

Of course, in order to make the constituent indicators synthesizable, each constituent indicator is processed dimensionless in the synthetic indicator structure. However, dimensionless processing can only solve the problem of additivity and integrability in the mathematical sense, but it does not automatically guarantee the additivity and integrability in the socio-economic sense. We know that when the model is abstractly reduced to economic concrete, on-site factors must be added, which will inevitably reduce the effective space of pure mathematical models, which means that the problems of additivity and integrability in the socio-economic sense may still exist.

Additive and integrable in mathematics, but not necessarily additive and integrable in the socio-economic sense. Therefore, we cannot give up the discussion on additivity and integrability in the socio-economic sense just because mathematical processing can be done. Of course, if mathematics is non-additive and non-integrable, let alone additivity and integrable in the socio-economic sense. If it is considered that the synthesis process completely solves the problem of additivity, it is to confuse mathematical additivity with socioeconomic additivity, or there is an implicit assumption in the evaluation: mathematical additivity equals additivity in the socioeconomic sense.

In the socio-economic sense, how to determine the “equivalent conversion” relationship between evaluation indicators, using the input perspective? Or the output perspective? Or a process perspective? Or even a comprehensive perspective? If a comprehensive perspective is adopted, how to synthesize it? It is still a big problem that needs to be discussed in depth, but it is also a problem that has been ignored by many people engaged in multi-indicator comprehensive evaluation.

4. The “equivalent conversion” and “compensability”

In the OECD synthetic indicator manual, the concept of “equivalent conversion” emphasized in this article is not discussed, and the closest concept to it is the so-called “compensatory” problem.

Compensability is the possibility of offsetting a deficit in some dimension with an outstanding performance in another (OECD, 2008). For example, in the case of the HDI, the previous aggregation method was arithmetic mean, which allowed a low value of “life expectancy at birth” to be offset by a high value of “gross national income per capita”. This aggregation method is defined as compensable. After 2010, the geometric mean method was used to realize that all three dimensions are equally important and that there is no possibility of complete substitution, which is defined as non-compensable (UNDP, 2010). In a linear aggregation, the compensability is constant, while with geometric aggregations compensability is lower for the composite indicators with low values (OECD, 2008). Compensability is usually closely related to the concept of imbalance, that is, the disequilibrium between the indicators used to construct the composite index. All dimensions desired in a composite indicator may contribute to a comprehensive understanding of complex phenomena, so all dimensions must be balanced in a non-compensatory or partially compensatory approach (Mazziotta and Pareto, 2020). Intuitively, the levels of three dimensions of HDI for A country or region are (0.8,0.8,0.8), and the levels for B country or region are (0.95,0.8,0.65). Although the value of HDI for A is the same as that for B, the value of inequality-adjusted HDI for A must be higher than that for B. After 2010, an inequality adjustment was applied to the HDI.

The result of the composite indicator is essentially an average. The SSF report also points out this shortcoming, where averages may disguise structural changes in the composite indicator. The aggregation approach ignores correlations among dimensions and does not reflect state distributions within economies. Even if the actual structure changes, as long as the average value of the composite indicator is unchanged, the conclusion of the comprehensive evaluation will remain unchanged, which means that the “ergodic property” of the composite result does not exist (Qiu and Li, 2021). The various spatio-temporal states experienced by the evaluated object cannot be represented by the composite result, which is only one of many possible outcomes. Aggregation is the recognition of part of states as the whole states of the evaluated object. In other words, the same conclusion can be reached as long as the average value of the composite indicator is the same, even if the distribution structure of multiple components is different.

5. The “equivalent conversion” and “compensability” in this paper had three comments (Chang, 2011)

First, as far as the relationship between the two is concerned, “equivalent conversion” is a “standard concept”, while compensation is only a “secondary concept”. This is because only the existence of the “equivalent conversion” relationship between indicators can produce the so-called “compensatory”. In other words, if “equivalent conversion” and “compensation” are a generalization of the same phenomenon, then “equivalent conversion” can reveal the essence of the problem better than “compensation”.

Second, the OECD expert group proposed a certain solution (NMC approach) to reduce “compensation”, which is to abandon the fixed distance and fixed proportion information of the evaluation indicator, and only use the sequential information among them. It should be noted that this price does not guarantee that compensation will be removed, as in the case of the Borda Rule.

However, my question about this kind of treatment is: Why is it preferable to abandon the useful information of indicators in exchange for compensatory elimination? Or, must the cost of discarding fixed distance and fixed proportion information be less than the benefit of eliminating “compensatory”? How to prove which of the two treatments is better? For example, in the three countries A, B, and C, the average life expectancy is 71, 70, and 50 years. If only the sequencing information is used, the huge difference between B and C in this indicator will be concealed. The comprehensive evaluation result may be greatly distorted.

Third, as mentioned above, as long as the comprehensive evaluation of multiple indicators is carried out, there will be an “equivalent conversion” relationship, and the existence of the “equivalent conversion” relationship is absolute. As the research of OECD evaluation experts shows, the “compensation” can only be relatively reduced.

6. Spatial interpretation of multi-indicator comprehensive evaluation

If the evaluation object is conceived as the point in the multidimensional space, comprehensive evaluation of multi-indicator is used to evaluate them, then the composite indicator is the point sorting in the multidimensional space.

If these points meet the transitivity, then the sorting is very simple. If A is better than B in any respects, then A must be ahead of B. The difficulty is that these points in a multidimensional space often do not meet the transitivity. For example, A is better than B in one-dimensional and two-dimensional space, but in the three-dimensional space, A is worse than B, at this time, A and B which is the first?

If the evaluated objects are conceived as points in a multi-dimensional space, and multi-indicator comprehensive evaluation uses a composite indicator to evaluate them, then the composite indicator is to rank the points in the multi-dimensional space.

If these points satisfy transitivity, then the sorting is very simple. If A is better than B in any respect, then A must precede B. The difficulty is that these points in a multidimensional space often do not satisfy transitivity. For example, A is better than B in the first and second dimensions, but inferior to B in the third dimension. At this time, which is A and B first?

From this perspective, it can be summarized as a sorting of comprehensive evaluation of multi-indicator, which may not meet the transitivity of the multidimensional spatial point.

Intuitively, ordering is the order and distance of points on a straight line. Therefore, the composite indicator is constructed to be linear: the points in a multidimensional space are mapped to a line.

This linear processing at least includes two steps: the first multidimensional space on the point to a plane, and then mapped to a straight line from the plane. This two mapping contains a variety of possibilities, that is, a variety of uncertainties.

Suppose we start this process from a three-dimensional space, then the construction process will face a series of questions: First, how to define a point in a multi-dimensional space? Which three-dimensional space to choose? What is the basis for selection? Second, from the three-dimensional space to the two-dimensional plane, which dimension is to be compressed? How to choose the mapping angle? Obviously, the front and back, left and right, height, etc. of the mapping supervision will all affect the result of dimensionality reduction. Furthermore, from a two-dimensional plane to a one-dimensional straight line, the left and right changes of the mapping angle will not affect the dimensional order of the evaluated points, but will affect the distance between them. Therefore, the evaluation value is not suitable for further fixed ratio processing.

Since comprehensive evaluation of multi-indicator may have a variety of ways, and correspondingly a variety of evaluation results, then the re-application of the results should be more cautious.

How to choose among the many possibilities? The synthetic mathematical model itself cannot tell us the answer. Only from the socio-economic meaning of the things being evaluated can a relatively reliable evaluation result be obtained. This highlights the importance of grasping the socio-economic significance in the design of evaluation methods (Su, 2012). OECD experts particularly emphasized the theoretical framework for comprehensive evaluation of multiple indicators and listed it as an independent step. The author agrees with this approach, because the synthesis technology can only solve the problem of whether the data can be synthesized, and the qualitative understanding of the things being evaluated, that is, the theoretical framework of the comprehensive evaluation, determines the quality of the synthesized results, and determines the quality of the synthesized results. Therefore, it is more important.

However, no matter how much we attach importance to the construction of the evaluation theoretical framework, in practice, we can only rely on not so perfect theories, and the imperfection of the theory itself will in turn affect the quality of the comprehensive evaluation results. This is an inherent contradiction. We can only gradually approach the correct understanding of things in the evaluation practice.

7. The significance of the results of comprehensive evaluation of multi-indicator

1. The special discussions of the results of comprehensive evaluation of multi-indicator

It is very important to know the results of comprehensive evaluation of multi-indicator, which is directly related to the effectiveness of composition.

Dong Qiu made a special discussion on the homogeneity, abstraction, relativity and subjectivity of the results of comprehensive evaluation of multi-indicator in his Doctoral Dissertation in 1990. Dong Qiu believed that the quality of things is multilevel, comprehensive evaluation of multi-indicator from the function of indicator to protect the homogeneity of the evaluation things (Qiu, 2012). Any statistical evaluation indicator is relative and abstract. The difference lies in the degree. Therefore, the application of synthetic indicators cannot be denied, and the advantages and disadvantages of synthetic indicators cannot be judged from the standpoint of value indicators. Similarly, any statistical indicator evaluation will involve subjective judgment, and the method of reflecting subjective judgment should not be confused with subjective arbitrariness.

Dr. Wei-hua Su put forward his own point of view on the subjectivity and objectivity of the comprehensive evaluation results: a comprehensive evaluation conclusion is a point of view, not the fact itself. Every point of view must have its own angle and position, so there must be subjective factors in the comprehensive evaluation results. However, the conclusion of the comprehensive evaluation is a kind of “fact-based viewpoint” (by Dr. Su Weihua), and at the same time it has objectivity.

2. The enlightenment of scientific development research on comprehensive evaluation of multi-indicator

Jerome Kagan, a famous American development psychologist, pointed out “social scientists” are plagued by the continuous digital scale of the concept of all inventions, which forces them to bring together very different phenomena. The benefits of this approach are that one can estimate the value of a certain method, a certain standard deviation and apply relevant statistical techniques to estimate the probability that the observed result is not an accidental event. The disadvantage is that different phenomena are often grouped together, such as IQ, insecure attachment, GDP, etc., which are examples of concepts composed of things with different origins. Biologists would never average the status of one’s gastrointestinal, respiratory, reproductive and cardiovascular systems, in order to get a continuous indicator that called “health status”. (quoted from “Three Cultures: 21st Century Natural Sciences, Social Sciences and Humanities”, Gezhi Publishing House, 2011 edition).

Mr. Kagan’s point of view tells us that methods always have their advantages and disadvantages. We can’t just look at one and ignore the others. The key to evaluating effectiveness lies in the comparison of the two, which is more important.

Chinese physicists Sunny Y. Auyang said (Auyang, 1998), “An intentional philosophy believes that all knowledge can be deduced, and micro reductionism is a part of this philosophical system. When talking about combination, micro-reduction theory assumes that once we know the laws and concepts of system components, in principle, all we need are mathematical skills and large computers, so that we can know everything that these components make up, no matter how complicated it is. The concepts and theories of the system as a whole can be reduced, that is, they are dispensable in principle, because they are only the logical results and definitions of component concepts. The emergence characteristics need to be described by the system concept, so they should be excluded from science.”

For small and simple systems, this bottom-up reduction method (down-up) is very successful. When this method is commonly used, micro-reduction theory strategically assumes that a large system is simply composed of many identical small systems. The composition of the system can be dealt with by the same theoretical framework and methods. Like in the slogan “the whole is the sum of parts”, it assumes that there is no interaction and irrelevance between the parts.

The interaction between the components and their relations makes the whole greater than the sum of the parts, therefore, a greater whole is not a greater sum. They form a structure, emerge diversity,

create complexity, so that the combination becomes important. Micro-reduction theory believes that the influence of interaction can be explained by adding “and relationship” to its slogan. There is no way to add up without considering “relationships”. This easy addition is a self-deception, which makes many disciplines unconvincing, including the largest branch of physics. It is very difficult to theoretically deal with the structure formation of a large combined system with many interactions, and it brings a whole new situation to science.

Sunny Y. Auyang’s discussion is more profound. She pointed out the possible space for cognition, effective or not, and the limitations of our “micro reductionism”.

3. Three suggestions on the application of the comprehensive evaluation methods of multi-indicator
First, hold the middle ground and be critical.

For the comprehensive evaluation methods of multi-indicator, neither can it blindly follow it because of its popularity, nor can it be deterred by the existence of traps. It may be more appropriate to adopt a moderate attitude. We must work hard to apply but not abuse or use indiscriminately, “learning and using” instead of “rigidly applying”. According to Norbert Wiener, to maintain a critical scientific attitude, even if it is only the application of comprehensive evaluation methods for empirical analysis, we should also pay attention to exploring its methodological gains and losses.

Second, we don’t pursue the best, but we pursue the better.

The single economics research only pursues the limited goal, does not expect to obtain the optimal solutions, and it is important to obtain relatively good results. In the existing analysis, it has been better, it has been improved, and it is closer to economic reality. The most common issue in comprehensive evaluation of multi-indicator is the preference for comprehensiveness, and we often ignore whether the overall evaluation is feasible or not. In fact, we should focus on considering whether it is good to replace it.

The third is the honest reporting method and the process of opening up in good faith.

In empirical study, one should try to avoid evaluation traps. When reporting the results, honestly explain the evaluation methods to the readers. The results are all phased. The shortcomings of this research should also be clearly stated, so as to prevent future generations from making detours and to facilitate others to discover other traps.

Careful use of multi-index comprehensive evaluation results, efforts to open up research, and use of one’s own research experience as a public product are the true scientific attitudes that real intellectuals should adopt.

Acknowledgments

This paper is a phased achievement of the Major Program of National Philosophy and Social Science Foundation of China (Grant No. 18ZDA123), Innovation Team of Philosophy and Social Sciences in Henan Colleges and Universities (2017-CXTD-07), Major Projects in Basic Research of Philosophy and Social Sciences in Henan Colleges and Universities (2019-JCZD-002), and Projects of the National Social Science Foundation of China (2021-ATJ-003).

Conflict of interest

All authors declare no conflicts of interest in this paper.

References

- Bandura R (2005) Measuring country performance and state behavior: A survey of composite indices. Technical report, Office of Development Studies, United Nations Development Programme (UNDP), New York.
- Bandura R (2011) Composite indicators and rankings: Inventory 2011. Technical report, Office of Development Studies, United Nations Development Programme (UNDP), New York.
- Greco S, Ishizaka A, Tasiou M, et al. (2019) On the Methodological Framework of Composite Indices: A Review of the Issues of Weighting, Aggregation, and Robustness. *Soc Indic Res* 141: 61–94.
- Mazziotta M, Pareto A (2020) Composite Indices Construction: The Performance Interval Approach. *Soc Indic Res*.
- Nardo M, Saisana M, Saltelli A, et al. (2005) Handbook on constructing composite indicators. OECD Statistics Working Papers. Available from: https://www.oecd-ilibrary.org/economics/handbook-on-constructing-composite-indicators_533411815016.
- Qiu D, Li D (2021) Comments on the “SSF Report” from the perspective of economic statistics. *Green Finance* 3: 403–463.
- Rosen R (1991) *Life itself: A comprehensive inquiry into the nature, origin, and fabrication of life*. New York: Columbia University Press.
- Saisana M, Tarantola S (2002) State-of-the-art Report on Current Methodologies and Practices for Composite Indicator Development. In Joint Research Centre. Italy: European Commission.
- United Nations Development Programme (UNDP) (2010) Human Development Report 2010: The Real Wealth of Nations-Pathways to Human Development. New York.
- Qiu D (1991) *The system of the comprehensive evaluation method of multi-indicator*. China Statistical Publishing Press.
- Qiu D (1988) The system of the multi-indicator comprehensive evaluation. *Res Financ Econ Iss* 09: 49–55.
- Su WH (2001) The research on theory and method of research on theory and method. Chinese Price Publishing House.
- OECD (2008) Handbook on Constructing Composite Indicators-methodology and Use Guide.
- Auyang SY (1998) Foundations of Complex-system Theories. In: *Economics, Evolutionary Biology, and Statistical Physics*.
- Kagan J (2009) *The Three Cultures: Natural Sciences, Social Sciences and the Humanities in the 21st Century*. New York: Cambridge University Press.
- Chang CL (2011) *The relativism and beyond on the philosophy of science*. Shandong University Press.
- Su WH (2012) The review and understanding of the comprehensive evaluation technology of multi-indicator and application research in China. *Stat Res* 250: 98–107.
- Qiu D (2012) The Boundary Antinomy of Macro-measurement and It’s Significance. *Stat Res* 250: 83–90.



AIMS Press

© 2022 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)