International Journal of Humanities, Arts and Social Sciences
volume 6 issue 1 pp. 10-24 doi: https://dx.doi.org/10.20469/ijhss.6.20002-1

The Construction of Meta-evaluation Indicators of Taiwan's University Program Evaluation: JCSEE Program Evaluation Standards as a Framework

Yi-Ning Chen
National Chengchi University, Taipei, Taiwan

Jia-Ling Kang *
National Chengchi University, Taipei, Taiwan

Chao-Yu Guo
National Chengchi University, Taipei, Taiwan

Abstract: Taiwan's university program evaluation has become an important trend. To ensure their quality, meta-evaluation is one of the important methods. The university program evaluation is a central issue in what concerns in Taiwan's university programs. In terms of meta-evaluation, the Joint Committee on Standards for Educational Evaluation (JCSEE) Program Evaluation Standards have wider applications but more general standard descriptions and more trivial related checklists. This study aims at the meta-evaluation indicators of Taiwan's university departmental self-evaluation based on the JCSEE Program Evaluation Standards. According to the weights of meta-evaluation's five major category standards of university program evaluation analyzed in this study, the most important is Propriety Standards, and the secondary important in order are Utility Standards, Feasibility Standards, Accuracy Standards and Evaluation Accountability Standards, different from the JCSEE's original order, which is worthy of follow-up studies. In this regard, it mainly amended and identified indicators via expert questionnaires and then used Fuzzy Delphi method questionnaires to integrate experts and scholars' opinions on indicator importance and their weights. The data analysis of 60 meta-evaluation indicators of Taiwan's university programs was constructed by referring to the JCSEE Program Evaluation Standards' 30 standards in five major categories. It was concluded that the importance of meta-evaluation category standards of university program evaluation in order is U6 Meaningful Processes and Products (Utility Standards), F3 Contextual Viability (Feasibility Standards), P1 Responsive and Inclusive Orientation (Propriety Standards), A3 Reliable Information (Accuracy Standards) and E1 Evaluation Documentation (Evaluation Accountability Standards). The JCSEE's standards were constructed, and conclusions and recommendations were provided to university departments for reference while carrying out self-evaluations.

Keywords: *University program evaluation, meta-evaluation, meta-evaluation indicators, fuzzy Delphi method*

Received: 19 August 2019; ***Accepted:*** 2 February 2020; ***Published:*** 28 February 2020

INTRODUCTION

Amidst the process of higher education's transition from traditional elitist education to mass education, the evaluation of schools is becoming more and more important due to their expansion in number, the shrinking in their expenditure and the international cooperation demand (Gall, Gall, & Borg, 2003; Kyriakides & Campbell, 2004; Perterson, 2000; Wudhikarn, 2016). The effective application of evaluations can not only ensure higher education quality but can also enhance educational performance (Schumacher & McMillan, 2006; Kyriakides & Campbell, 2004). It is very important in higher education. Higher education evaluation is also called program evaluation. However,

*Correspondence concerning this article should be addressed to Jia-Ling Kang, National Chengchi University, Taipei, Taiwan. E-mail: 106152513@mail2.nccu.tw

program evaluation is a necessary procedure and increasingly valued (Alford & Aharonian, 2011; Cousins, Goh, Elliott, & Bourgeois, 2014; Hou, Morse, & Wang, 2015). This has been the case with the development of evaluation of higher education in Taiwan: in recent times, Taiwan's universities have been gradually authorized to implement self-accreditation for university program evaluation (Chan & Chan, 2015; HEEACT, 2015; Nazari-Shirkouhi et al., 2020; Hou et al., 2015). Consequently, finding ways of ensuring the quality of university program evaluation is very important.

The meta-evaluation is an evaluation of an evaluation. It aims to ensure evaluation quality. The Program Evaluations Metaevaluation Checklist (Stufflebeam, 2012; Mertens & Wilson, 2018; Sturges & Howley, 2016) was explicitly designed around the JCSEE's (Yarbrough, Shulha, Hopson, & Caruthers, 2011) Program Evaluation Standards. The checklist is used to assess all 30 standards within the 5 attributes, utilizing a scale of 0-6 to measure an evaluation's level of compliance with these standards. The Program Evaluation Standards developed by the JCSEE include 30 standards in 5 major categories Utility Standards, Feasibility Standards, Propriety Standards, Accuracy Standards and Evaluation Accountability Standards and are generally recognized. Based on these standards, Stufflebeam further developed 150 summative meta-evaluation checkpoints (JCSEE, 2015; Stufflebeam, 2012; Mertens & Wilson, 2018; Sturges & Howley, 2016). Nevertheless, the Program Evaluation Standards contain more abstract descriptions, but there are too many Stufflebeam's checkpoints. In this study, all these standards used as a framework by referring to Stufflebeam's checkpoints to develop the meta-evaluation indicators suitable for the evaluation of Taiwan's program evaluation. This study canbe expected to benefit the establishment and quality enhancement of the evaluation system for the program evaluation.

LITERATURE REVIEW

Program evaluation is a professional task. The reason why Taiwan's evaluation needs a professional program evaluation system is to ensure the school quality of the university. In the process of evaluation, no matter evaluations' project instructors, organization operators or coordinators, they all need a considerable degree of specialized knowledge as well as an understanding of basic concepts, evaluation skills and evaluation objects for evaluation integration, and other related professional competences (Blair & Noel, 2014; Crawford, Aguayo, & Valle, 2017; Froh, 1991). External institutions usually play an important role of promoting program evaluations. Under the double demands of guaranteed quality and performance accountability, on the one hand, they reinforce universities' internal self-adjustment and management, and on the other hand, they also reduce the government's intervention in and control of universities, encouraging universities to establish a program evaluation mechanism (Ayob & Morell, 2016; Blair & Noel, 2014; Froh, 1991; Harvey & Williams, 2010; Wolff, 2010). The university program evaluations are the most important. Thus, the program evaluation is the core of university program evaluation, focusing on the department itself to understand the department's goal achievement levels, methods and to make self-improvements.

Taiwan's discipline evaluation, which has been under development since 1970, is mainly led by the Ministry of Education in Taiwan and authorized to related the academic institutions for implementation, but its discipline implementation schedules, frequencies, and influences are different. The first University Act was amended in 2005, mandating that universities should regularly carry out program evaluations of teaching, research, services, counselling and guidance, academic affairs, administration, and student participation. To promote the development of every university, the Ministry of Education will organize an assessment committee or educational commission organizations or professional accreditation bodies to carry out regular university program evaluations and make the results public (Misnistry of Education, 2015). In the same year, the Ministry of Education wrote a letter to universities and colleges in the country asking them to jointly donate funds for the establishment of the Higher Education Evaluation & Accreditation Council of Taiwan (HEEACT) for the overall planning and implementation of university program evaluations. Regarding university program evaluation, during the first cycle from 2006 to 2010, 3,119 departments in 76 universities were evaluated. During the process of evaluation, departments had to carry out program evaluations to provide the HEEACT as a reference while visiting and evaluating on-site. During the second cycle of university program evaluation from 2012 to 2016, to meet the demands of universities' autonomous management and featured development. The HEEACT and the Ministry of Education agreed that 34 universities could carry out the external evaluation of departments by themselves without undertaking an external evaluation from the HEEACT, thus allowing universities to take responsibility for department quality by themselves (HEEACT, 2015). In the future, the number of universities authorized to carry out the external evaluation of departments by themselves will be raised in Taiwan each

year. Under the principle that universities are authorized to carry out university program evaluations by themselves, the quality of university program evaluation is becoming increasingly important.

There are several university evaluation's problems. First, it's a vast number of universities. Second, the number of students has increased. Third, public sectors cut down expenses. Fourth, the proportion of government investment in student funding has been decreasing each year. Fifth, student's tuition is not high; students cannot meet the employment requirements. Sixth, it's the pressure from international university ratings. Finally, higher education institutions face great challenges and crises (Attuluri, 2019; Gatpandan & Ambat, 2017; Harvey & Williams, 2010; Horta, 2017; Nazari-Shirkouhi et al., 2020; Wolff, 2010). Meta-evaluation is the re-evaluation (the secondary evaluation) of evaluation (the primary evaluation)'s goals, systems, contents, processes, and results, establishing a point of reference and improvement for the primary evaluation in order to enhance evaluation quality. Meta-evaluation standards are principles that allow evaluators to judge values or estimate the strengths and weaknesses of primary evaluations (Harvey & Williams, 2010; Yarbrough et al., 2011; Kyriakides & Campbell, 2004). The Program Evaluation Standards developed by the JCSEE are sorted into 30 standards across five major categories Utility Standards, Feasibility Standards, Propriety Standards, Accuracy Standards and Evaluation Accountability Standards. Compared with the Personnel Evaluation Standards, which are sorted into 22 standards across four major categories, including Propriety Standards, Utility Standards, Feasibility Standards and Accuracy Standards, and which especially value evaluation propriety (JCSEE, 2015; Volden, 2018), the Program Evaluation Standards put greater value on Utility Standards and Feasibility Standards in program evaluation.

As the JCSEE Program Evaluation Standards are applied to Taiwan's departmental self-evaluations, each standard's connotation is more abstract and different from Taiwan's departmental evaluation contexts. In addition, the 150 checkpoints made by Stufflebeam (2012) referring to the Program Evaluation Standards are too many and not easy to be used. Therefore, the meta-evaluation indicators suitable for Taiwan's university program evaluation will be constructed in this study, basing on the JCSEE Program Evaluation Standards and referring to Stufflebeam's checkpoints, to provide university program evaluation reference.

Stufflebeam (2012) made a list of 150 checkpoints. For each of the 30 standards, the checklist includes 5 checkpoints drawn from the substance of the standard. For example, consider U1, Evaluator Credibility has five checkpoints (Engage evaluator whom the stakeholders trust; engage evaluators who are appropriately responsive to issues of gender, socioeconomic status, race, and language and cultural differences; assure that the evaluation plan responds to key stakeholders' concerns; attend appropriately to stakeholders' criticisms and suggestions; keep interested parties informed about the evaluation's progress). The Program Evaluation Standards developed by the JCSEE have become widely accepted and applied to evaluations across a diverse field of disciplines due to the systematic and comprehensive nature this method provides for quality evaluations. JCSEE Program Evaluation Standards have been widely used.

Therefore, in this article, the meta-evaluation indicators suitable for the evaluation of Taiwan's university program evaluation will be constructed in this study's purpose, basing on the JCSEE Program Evaluation Standards and referring to Stufflebeam's checkpoints, to provide university for program evaluation reference.

RESEARCH METHODOLOGY

In this study, the thirty standards across five major categories (Utility Standards, Feasibility Standards, Propriety Standards, Accuracy Standards and Evaluation Accountability Standards) of the JCSEE Program Evaluation Standards and 150 summative meta-evaluation checkpoints of Stufflebeam's program evaluation were used as a framework (JCSEE, 2015; Mertens & Wilson, 2018; Stufflebeam, 2012; Sturges & Howley, 2016) to construct meta-evaluation indicators for the evaluation of university program evaluations. Firstly, referring to the above-mentioned standards and Stufflebeam's checkpoints, concrete and readily understandable draft indicators were developed in a simplified way. If two experts modify the indicators or check for inappropriate, it will combine or adjust. All the indicators need to carry out the Propriety Standards analysis for the indicators by experts. Six evaluation-related experts and scholars were invited to carry out the Propriety Standards analysis for the indicators and make amendments in order to apply for University program evaluation. Supported by the experts' and scholars' knowledge and assistance in selection and amendment, 60 meta-evaluation indicators for the evaluation of university program evaluation were constructed.

Secondly, 11 important evaluation experts and scholars, including several former HEEACT executive directors and deans of office of evaluation and other meta-evaluation related experts and scholars, were invited to fill out an

expert questionnaire by Fuzzy Delphi Method. The Fuzzy Delphi Method to the selection process of system variables to increase the confidence of the model. Fuzzy Delphi Method is a method of constructing indicators, indicators of a relatively clear situation to seek consensus or process. The concept of combining fuzzy set theory and Delphi was proposed by (Murray, Pipino, & van Gigch, 1985) and named the Fuzzy Delphi Method. In this study, it used the Fuzzy Delphi Method.according to Noorderhaben (1995), applying the Fuzzy Delphi Method to group decisions can solve the fuzziness of the common understanding of expert opinions. Murry describes simplicity as the advantage of the Fuzzy Delphi Method. All expert opinions can be encompassed in one investigation. Hence, this method can permit a more effective selection of criteria. Opinions on the importance of indicators were received from experts, including the minimums and maximums of importance, and the values of most likely importance level, and shown by fuzzy values. The fuzzy values were shown via the defuzzifying steps of triangular fuzzy numbers and converted into actual numbers to represent the importance levels of the indicators.

The definition of triangular fuzzy number is as follows (Cheng-Ta, 2008):

The fuzzy number $\tilde{A}$ is a fuzzy set, and its membership function is $\mu\tilde{A}(X)R[0,1]$. As it meets the following three conditions, it is called a triangular fuzzy number, which is shown in Figure 1.

$\mu\tilde{A}(X)$ is iecewise Continuous.

$\mu\tilde{A}(X)$ is convex fuzzy subset.

$\mu\tilde{A}(X)$ is normality of a fuzzy subsetit means a real number (X0) that can make $\mu\tilde{A}(X0)=1$.

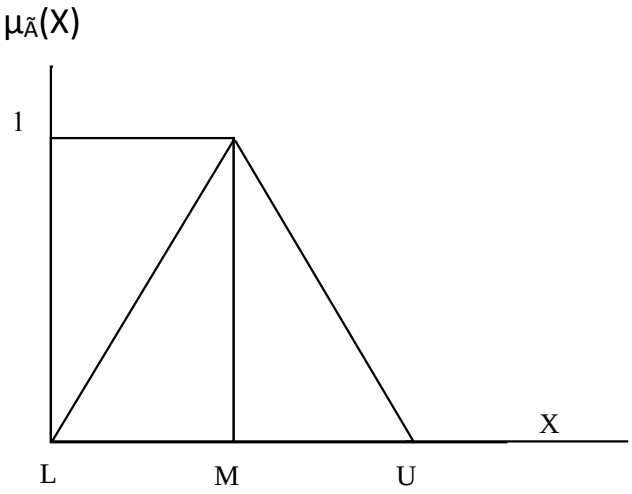

Figure 1 *Triangular fuzzy number*

In Figure 1, point L represents the minimum of experts' consensus, and the point U represents the maximum of experts' consensus. The membership degree of these two points is shown as 0 because of its extreme value. When designing the expert questionnaire by Fuzzy Delphi Method, the indicator importance level was designed ranging from 0 to 1 for the evaluation items, and experts were invited to evaluate the importance range while filling it out and checking the importance level of the most likely importance level, so as to collect their data. Cheng-Ta (2008) considered that, compared with other methods, the values calculated through geometric means are not affected by extreme values. Consequently, the researcher calculated the point M of each indicator by geometric means.

For triangular fuzzy numbers, the minimum of all the expert opinions is the consensus minimum fitness value L, and the maximum is the consensus minimum fitness value U. Of an indicator, "the most likely importance level", filled out by all expert opinions, is calculated into a geometric mean as the median M. And then, L-value and R-value are obtained by defuzzifying the above-mentioned three values L, M and U. Finally, a total value is obtained according to Chen and Hwang (1992) proposed defuzzification method. Their Right and Left Score of Importance Level refers to experts' and scholars' consensus on the importance of indicators. Defuzzification is the concept of hypothesizing the triangular fuzzy number A's maximum set the membership function $\mu\max(X)$ and minimum set the membership function $\mu\min(X)$, after obtaining L-value and R-value. Further, it is calculating the total membership value of indicators. Firstly, the triangular fuzzy number A=(L,M,U), which represents three coordinates, and these three coordinates can set up two fuzzy functions, which are $\mu\max(X)$ and $\mu\min(X)$, to separately intersect the triangular fuzzy number A's L-value and R-value. The R-value $\mu R(A)$ can be obtained from $\mu\max(X)$ and A's fuzzy function, and the L-value can be obtained

from μL(A) and A's fuzzy function. Via L-value and R-value, the total value of triangular fuzzy number A can be calculated, which can represent the crisp value of triangular fuzzy number A. Its calculation method is shown here (Cheng-Ta, 2008):

The crisp score of a fuzzy number A is calculated in the following way.

First, there are defined two functions μmax(x) and μmin(x):

$$\mu_{\max}(x) = \begin{cases} x, & 0 \le x \le 1 \\ 0, & \text{otherwise} \end{cases}$$
$$\mu_{\min}(x) = \begin{cases} 1-x, & 0 \le x \le 1 \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

Then, the left and the right scores of A and the two functions are defined as:

$$\mu_L(A) = \mathrm{Sup}_x\left[\mu_{\tilde{A}}(x) \cap \mu_{\min}(x)\right]$$
$$\mu_R(\tilde{A}) = \mathrm{Sup}_x\left[\mu_{\tilde{A}}(x) \cap \mu_{\max}(x)\right] \quad (2)$$

Here, Sup stands for the least upper bound. The total crisp score of the fuzzy number A is defined as:

$$\mu_T(\mathrm{A}) = \frac{[\mu_R(A) + 1 - \mu_L(A)]}{2} \quad (3)$$

Finally, it is to compare with the total value μt (A) of triangular fuzzy numbers of indicators. The membership function μ of a triangular fuzzy number is a simple linear function. The statistical moment solution (e.g., the first temporal moment (μt)) was then evaluated for each combination of the sampled parameter values. The higher the value is, the higher the experts consider the likely importance level.

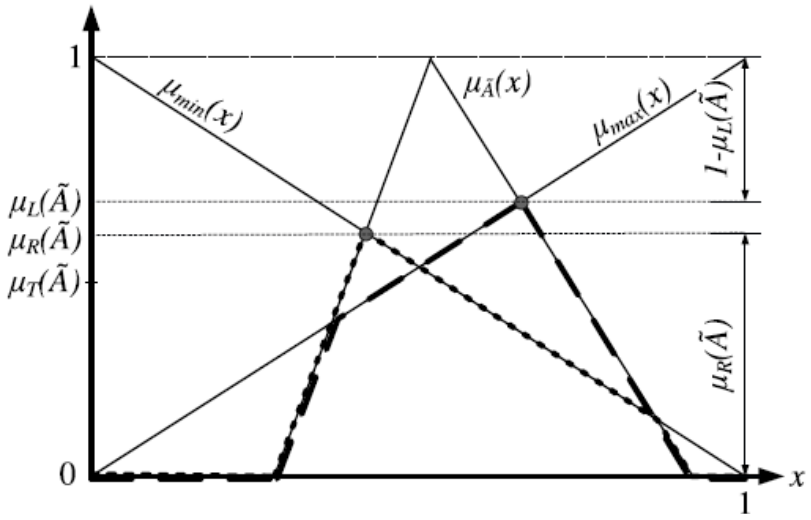

Figure 2 *Conversion of triangular fuzzy number to crisp value*

Fuzzy Delphi technique constructed on the index, through the triangular fuzzy number and the defuzzification method selected indicators, each indicator value of the triangular fuzzy number represents the consensus of members of the Fuzzy Delphi panel, and the threshold value is 0.6 to keep the indicators. Researchers should define a threshold value of meta-evaluation on their own to determine indicator reservation and deletion of the checkpoints. In this study, the threshold value of triangular fuzzy numbers is set as 0.6, on the basis of relevant studies (Chen & Hwang, 1992; Wei, 2018; Tang, Wei, & Gao, 2019). If an indicator's total value is lower than the threshold value, which will be determined as unfit and deleted, and the reserved indicators will be used for follow-up analysis.

RESULTS AND DISCUSSION

In this study, 60 indicators were developed by referring to the JCSEE's 30 standards and 150 checkpoints (see the checkpoints that indicators refer to in the table and compare them with the checkpoints developed by Stufflebeam

(2012), which were analyzed for fitness by six experts and scholars and amended by referring to experts' and scholars' opinions. After the fuzzy and defuzzification, indicator fitness (including fitness and fitness after amendment) is fully identified.

JCSEE's 30 standards and Stufflebeam's 150 checkpoints For each of the 30 standards, the checklist includes 5 checkpoints drawn from the substance of the standard. For example, consider U1 Evaluator Credibility. It has five checkpoints (Engage evaluator whom the stakeholders trust; engage evaluators who are appropriately responsive to issues of gender, socioeconomic status, race, and language and cultural differences; assure that the evaluation plan responds to key stakeholders' concerns; attend appropriately to stakeholders' criticisms and suggestions; keep interested parties informed about the evaluation's progress). Its indicators are U1-1 Evaluation committee members consider characteristics of program evaluation departments, and their professionalism and fairness can be trusted (refer to Stufflebeam's Checkpoint 1.2); U1-2 Evaluation committee members pay attention to opinions of stakeholders (e.g. program evaluation departments' teachers, students, administrators, alumni, etc.) and make stakeholders understand the evaluation process (refer to Stufflebeam's Checkpoint 3.4.5).

Besides, the Fuzzy Delphi method was carried out by the above-mentioned 60 indicators, which is showing the indicators' values, total values and orders of triangular fuzzy numbers from Table 1 to Table 5, respectively:

Table 1 *The Values and Total Values of Triangular Fuzzy Numbers of Meta-evaluation Indicators of University Program Evaluation (Utility Standards)*

Category/Indicator (Checkpoint Induction)	Triangular Fuzzy Numbers			Indicator μT	Standard μT	Category Order
	L	M	U			
U1.Evaluator Credibility						
U1-1 Evaluation committee members consider characteristics of program evaluation departments, and their professionalism and fairness are worth to be trusted. (refer to Checkpoint 1.2)	0.6	0.934	1.0	0.819	0.784	3
U1-2 Evaluation committee members pay attention to opinions of stakeholders (e.g., program evaluation departments' teachers, students, administrators, alumni, etc.) and make stakeholders understand the evaluation process. (refer to Checkpoint 3.4.5)	0.5	0.848	1.0	0.749		
U2.Attention to Stakeholders						
U2-1 Compared with the higher education evaluation authority's ideas, stakeholders' opinions and needs should more be valued in the evaluation process. (refer to Checkpoint 1.3)	0.4	0.829	1.0	0.717	0.676	7
U2-2 Stakeholders should be arranged to appropriately participate in university program evaluations. (refer to Checkpoint 2.4.5)	0.2	0.733	1.0	0.634		
U3.Negotiated Purposes						
U3-1 The purposes of university program evaluations should be negotiated with stakeholders and adjusted by needs if necessary. (refer to Checkpoint 1)	0.2	0.644	1.0	0.601	0.648	8
U3-2 Evaluation procedures should be recorded as appropriate, described in evaluation reports and summaries, and flexibly adjusted if necessary. (refer to Checkpoint 2.3.4)	0.4	0.786	1.0	0.695		

Table 1 *Continued...*

U4.Explicit Values						
U4-1 University program evaluations should have explicit values, including consideration of related laws, disciplines missions and departments goals. (refer to Checkpoint 2.3.4)	0.6	0.915	1.0	0.809	0.776	4
U4-2 Evaluation committee members should consider multiple values when interpreting evaluation findings and make appropriate alternative interpretations when values are different. (refer to Checkpoint 1.5)	0.5	0.836	1.0	0.742		
U5.Relevant Information						
U5-1 Information collected from university program evaluations can reflect the issues stakeholders value and flexibly address extra questions. (refer to Checkpoint 1.2.3)	0.6	0.875	1.0	0.788	0.799	2
U5-2 Collected information is sufficient to evaluate departmentsstrengths and values. (refer to Checkpoint 4.5)	0.6	0.915	1.0	0.809		
U6.Meaningful Processes and Products						
U6-1 University program evaluations can show important information and guide participants to discover or induce core issues, focusing evaluations on concrete questions. (refer to Checkpoint 1.2.3)	0.7	0.935	1.0	0.848	0.831	1
U6-2 University program evaluations should clearly interpret evaluation purposes, procedures and findings, provide summaries and make participants understand and agree on evaluation conclusions and recommendations. (refer to Checkpoint 4.5.6)	0.6	0.923	1.0	0.813		
U7.Timely and Appropriate Communicating and Reporting						
U7-1 During evaluations, temporary reports are provided as appropriate, and department members and students are communicated within time to understand their ideas. (refer to Checkpoint 3.4)	0.4	0.771	1.0	0.688	0.709	6
U7-2 A summarized report for final evaluation findings should be appropriately released to make the public understand. (refer to Checkpoint 1.2.5)	0.4	0.854	1.0	0.730		
U8.Concern for Consequences and Influence						
U8-1 During university program evaluations, stakeholders are communicated with via oral, written or interim reports and encouraged to use evaluation findings. (refer to Checkpoint 1.2.4)	0.5	0.857	1.0	0.753	0.758	5
U8-2 University program evaluation reports should be open, transparent and concrete, and workshops should be held if necessary to interpret and promote proper applications of evaluation findings. (refer to Checkpoint 3.5)	0.5	0.873	1.0	0.762		

As shown in Table 1, the Values and Total Values of Triangular Fuzzy Numbers of Meta-evaluation Indicators of University Program evaluation (Utility Standards), the total values of 16 indicators of 8 standards of the first category "Utility Standards" are all higher than 0.6, which are reserved. Of these standards, the most important is U6 Meaningful Processes and Products, which shows that university program evaluations should be able to guide stakeholders in paying attention to questions as well as understanding and agreeing with evaluation findings and conclusions, making program evaluations meaningful and reaching the requirements of Utility Standards. Of secondary important are U5 Relevant Information and U1 Evaluator Credibility, which show whether university program evaluations can reflect departments' strengths and weaknesses without a bias, due to committee members, which is evaluation committee members' professionalism.

Table 2 *The Values and Total Values of Triangular Fuzzy Numbers of Meta-evaluation Indicators of University Program Evaluation (Feasibility Standards)*

Category/Indicator (Checkpoint Induction)	Triangular Fuzzy Numbers			Indicator μT	Standard μT	Category Order
	L	M	U			
F1.Project Management						
F1-1 Evaluation projects should be effectively managed with professional manners and consistency and appropriate responses to participants concerns. (refer to Checkpoint 1.2.3)	0.6	0.853	1.0	0.776	0.758	2
F1-2 Implementation of university program evaluations can reduce interventions and respect privacy, but it should not ignore incompetence or fraud. (refer to Checkpoint 4.5)	0.5	0.829	1.0	0.739		
F2.Practical Procedures						
F2-1 Evaluation procedures and methods should be practical and able to simplify evaluation data needs to reduce burdens. (refer to Checkpoint 1.2)	0.4	0.816	1.0	0.710	0.750	3
F2-2 Arrangement of university program evaluation procedures is practical and feasible, which considers department operations and limitations and requests department staff to assist in evaluation. (refer to Checkpoint 3.4.5.6)	0.6	0.879	1.0	0.790		
F3.Contextual Viability						
F3-1 Evaluation reports should fully reflect department background contexts and avoid biases or misapplications. (refer to Checkpoint 1.3)	0.7	0.935	1.0	0.848	0.764	1
F3-2 Evaluation reports should integrate multiple perspectives and reflect the higher education authoritys decision needs. (refer to Checkpoint 2)	0.4	0.755	1.0	0.680		
F4.Resource Use						
F4-1 University program evaluations should be able to effectively use resources and propose feasible programs to foster department improvement or inspire department development. (refer to Checkpoint 1.2.3.4)	0.6	0.907	1.0	0.804	0.758	2
F4-2 In the university program evaluation process, department personnel impacts and administration burdens should be efficiently minimized as a principle. (refer to Checkpoint 5)	0.5	0.775	1.0	0.712		

As shown in Table 2, The Values and Total Values of Triangular Fuzzy Numbers of Meta-evaluation Indicators of University Program evaluation (Feasibility Standards), the total values of eight indicators of four standards of the second category "Feasibility Standards" are all higher than 0.6, which are reserved. Of these standards, the most important is F3 Contextual Viability, which shows whether university program evaluations can consider different department contexts, avoid biases, meet multiple perspectives and reflect the higher education authority's decision needs. To increase the feasibility of program evaluations is important.

Table 3 *The Values and Total Values of Triangular Fuzzy Numbers of Meta-evaluation Indicators of University Program Evaluation (Propriety Standards)*

Category/Indicator (Checkpoint Induction)	Triangular Fuzzy Numbers			Indicator μT	Standard μT	Category Order
	L	M	U			
P1.Responsive and Inclusive Orientation						
P1-1 Make stakeholders understand that program evaluations are service-oriented and that the key of outcome evaluations lies in students learning needs. (refer to Checkpoint 1.2)	0.6	0.889	1.0	0.795	0.832	1
P1-2 program evaluations can identify department strengths and weaknesses needed to be improved. (refer to Checkpoint 3.4)	0.7	0.972	1.0	0.868		
P2.Formal Agreements						
P2-1 An evaluation manual should be drawn up with explicit evaluation purposes, procedures, schedules, report formats, release methods, data confidentiality and evaluation resources. (refer to Checkpoint 1.2.3.4.5.6)	0.6	0.915	1.0	0.809	0.767	5
P2-2 The evaluation manual should be drawn up and negotiated with the department by considering the higher education authority and stakeholders needs and expectations. (refer to Standard P2 and Checkpoint 1.2.3.4.5.6)	0.5	0.802	1.0	0.725		
P3.Human Rights and Respect						
P3-1 University program evaluations have explicit use plans and make stakeholders understand. (refer to Checkpoint 1.2)	0.5	0.854	1.0	0.752	0.770	4
P3-2 University program evaluations can respect department characteristics and value privacy confidentiality to avoid harms. (refer to Checkpoint 3.4.5)	0.6	0.873	1.0	0.787		
P4.Clarity and Fairness						
P4-1 University program evaluation reports should clearly and fairly reflect departments strengths, weaknesses, and intended and unintended outcomes. (refer to Checkpoint 1.2.3.4)	0.6	0.915	1.0	0.809	0.774	3
P4-2 Overall judgments on departments should consider evaluation limitations and their effects, and claimed opinions on evaluation reports should also be appropriately handled. (refer to Checkpoint 5.6)	0.5	0.828	1.0	0.738		

Table 3 *Continued ...*

P5.Transparency and Disclosure						
P5-1 Written reports of University program evaluations should be transparent and open, appropriately showing evaluation findings and limitations. (refer to Checkpoint 1.5)	0.5	0.827	1.0	0.738	0.724	6
P5-2 If necessary, evaluation reports should show different perspectives and propose balanced conclusions and recommendations according to proofs. (refer to Checkpoint 2.3.4)	0.4	0.813	1.0	0.709		
P6.Conflicts of Interests						
P6-1 University program evaluations can hire professional and multiple evaluation committee members and draw up an avoidance norm to avoid conflicts of interest. (refer to Checkpoint 1.2.3)	0.6	0.897	1.0	0.799	0.792	2
P6-2 University program evaluations can release evaluation procedures, data and reports as appropriate and appropriately keep evaluation records for reference to reduce conflicts of interest. (refer to Checkpoint 4.5)	0.6	0.868	1.0	0.784		
P7.Fiscal Responsibility						
P7-1 University program evaluations can budget in advance and adjust flexibly for evaluation needs by approval. (refer to Checkpoint 1.2.3)	0.5	0.785	1.0	0.717	0.696	7
P7-2 University program evaluations can fully and accurately record budget incomes and expenses, personnel allocation and time, appropriately using evaluation resources. (refer to Checkpoint 4.5.6)	0.4	0.744	1.0	0.675		

As shown in Table 3, The Values and Total Values of Triangular Fuzzy Numbers of Meta-evaluation Indicators of University Program evaluation (Propriety Standards), the total values of 14 indicators of 7 standards of the third category "Propriety Standards" are all higher than 0.6, which are reserved. Of these standards, the most important is P1 Responsive and Inclusive Orientation, which shows that university program evaluations should be able to respond to students' learning needs with respect and tolerance so as to reach the Propriety Standards' requirements. Of secondary importance are P6 Conflicts of Interests and P4 Clarity and Fairness, which shows that for university program evaluations, fairness is still a factor highly valued in terms of fitness.

Table 4 *The Values and Total Values of Triangular Fuzzy Numbers of Meta-evaluation Indicators of University Program Evaluation (Accuracy Standards)*

Category/Indicator (Checkpoint Induction)	Triangular Fuzzy Numbers			Indicator μT	Standard μT	Category Order
	L	M	U			
A1.Justified Conclusions and Decisions						
A1-1 Universityprogram evaluations can focus on evaluation questions and have sufficient proofs to support seasonable conclusions. (refer to Checkpoint 1.2.4)	0.6	0.914	1.0	0.808	0.770	3

Table 4 *Continued...*

A1-2 Evaluation conclusions only reasonably interpret departments then operation products. (refer to Checkpoint 3.5.6)	0.5	0.816	1.0	0.732		
A2.Valid Information						
A2-1 Data collectors can be trained to use multiple methods to collect the valid information meeting the needs of evaluation purposes and items and appropriately record the data collection process. (refer to Checkpoint 2.3.4)	0.5	0.837	1.0	0.743	0.757	6
A2-2 How the obtained information be analyzed and interpreted can be recorded, and evaluation key questions can also be effectively responded to for comprehensive assessments. (refer to Checkpoint 1.5.6)	0.6	0.842	1.0	0.771		
A3.Reliable Information						
A3-1 Evaluation instrument reliabilities should be identified, and the factors that may influence them (e.g. department characteristics or data collection conditions) should also be assessed to ensure the data collection dependability. (refer to Checkpoint 1.3.5)	0.5	0.860	1.0	0.755	0.787	1
A3-2 Universityprogram evaluation committee members should take training, trying to ensure the consistency of their analyses and judgments. (refer to Checkpoint 2.4)	0.6	0.934	1.0	0.819		
A4.Explicit Program and Context Descriptions						
A4-1 Universityprogram evaluations can collect the data of how stakeholders describe actual department operation conditions and how they expect from departments. (refer to Checkpoint 1.3)	0.5	0.876	1.0	0.763	0.763	5
A4-2 Collect the data of how departments operate and how they function from different information sources, and analyze the reasons of how departments intended operations are different from actual operations. (refer to Checkpoint 2.4.5)	0.5	0.874	1.0	0.762		
A5.Information Management						
A5-1 Information from different sources can be collected by multiple methods, and criteria, methods and strategies for effective information selection can also be described. (refer to Checkpoint 1.2.3.4)	0.5	0.877	1.0	0.764	0.733	8
A5-2 Evaluation information should be stored systematically, and data collection instruments should also be attached to evaluation reports. (refer to Checkpoint 5.6)	0.4	0.797	1.0	0.701		
A6.Sound Designs and Analyses						
A6-1 Universityprogram evaluations should be appropriately designed to ensure the quality of evaluation information, including the consistency norm that multiple committee members should comply with during evaluations and scorings. (refer to Checkpoint 1.2.3.4)	0.6	0.879	1.0	0.790	0.774	2
A6-2Appropriate procedures can be selected by evaluation questions and data nature to obtain key information. (refer to Checkpoint 5.6)	0.5	0.864	1.0	0.757		

Table 4 *Continued...*

A7.Explicit Evaluation Reasoning						
A7-1 Universityprogram evaluations can adopt multiple methods to analyze data (including central tendency, variation and outlier) to ensure the dependability of evaluation products. (refer to Checkpoint 1.2.3)	0.5	0.849	1.0	0.749	0.735	7
A7-2 Data analyses and interpretations can be displayed by graphs and tables, developed into meaningful conclusions and recommendations, and recorded completely. (refer to Checkpoint 4.5)	0.4	0.836	1.0	0.721		
A8.Communication and Reporting						
A8-1 Evaluation reports should show stakeholdersperspectives and reasonable conclusions with appropriate communication. (refer to Checkpoint 2.3)	0.6	0.862	1.0	0.781	0.769	4
A8-2 Evaluation findings should be presented publicly with a safeguard and communication (e.g. claim procedures) to avoid biases or misconceptions. (refer to Checkpoint 1.4.5)	0.5	0.864	1.0	0.757		

As shown in Table 4, The Values and Total Values of Triangular Fuzzy Numbers of Meta-evaluation Indicators of University Program evaluation (Accuracy Standards), the total values of 16 indicators of 8 standards of the fourth category "Accuracy Standards" are all higher than 0.6, which are reserved. Of these standards, the most important is A3 Reliable Information, which shows that evaluation committee members' training and evaluation instruments' reliability can ensure the consistency of program evaluation analysis. Of secondary importance are A6 Sound Designs and Analyses, A1 Justified Conclusions and Decisions, which shows that evaluation design and planning at the beginning and conclusions and decisions at the end are relatively important during university program evaluations.

Table 5 *The Values and Total Values of Triangular Fuzzy Numbers of Meta-evaluation Indicators of University*

Category/Indicator (Checkpoint Induction)	Triangular Fuzzy Numbers			Indicator μT	Standard μT	Category Order
	L	M	U			
E1.Evaluation Documentation						
E1-1 The process of university program evaluation should be implemented by evaluation design and recorded (including the documents related to data collection and analysis) (refer to Checkpoint 1.2.3.4)	0.5	0.820	1.0	0.734	0.730	1
E1-2 Complete evaluation reports should show relevant recording documents (including the information of data collection, analysis and finding) (refer to Checkpoint 5)	0.5	0.802	1.0	0.725		
E2.Internal Meta- evaluation						
E2-1 Meta-evaluation standards can be appropriately used to examine the accountability of university program evaluations design, procedures, data collection and outcomes. (refer to Checkpoint 1)	0.5	0.795	1.0	0.722	0.724	2
E2-2Aspecially-assigned person can budget, record it and assess the university program evaluations processes and products. (refer to Checkpoint 2)	0.5	0.804	1.0	0.726		

Table 5 *Continued...*

E3.External Meta-evaluation						
E3-1 Appropriately and sufficiently budget for external meta-evaluations of university program evaluations. (refer to Checkpoint 1)	0.5	0.732	1.0	0.691	0.689	3
E3-2 Appropriately keep all program evaluation procedures, information and analysis records and support using evaluation standards for meta-evaluations to improve evaluation processes. (refer to Checkpoint 2)	0.4	0.767	1.0	0.686		

As shown in Table 5, the Values and Total Values of Triangular Fuzzy Numbers of Meta-evaluation Indicators of University Program evaluation (Evaluation Accountability Standards), the total values of 6 indicators of 3 standards of the fifth category "Evaluation Accountability Standards" are all higher than 0.6, which are reserved. Of these standards, the most important is E1 Evaluation Documentation, which shows that the record reports and relevant document reservation for reference during evaluations can ensure the quality of university program evaluations reaching accountability needs (including the information of data collection, analysis and findings).

In addition, it has data collection, analysis and findings as calculating the average total values of standards of five categories (Utility Standards, Feasibility Standards, Propriety Standards, Accuracy Standards and Evaluation Accountability Standards), the importance in order are Propriety Standards (0.765), Accuracy Standards (0.761), Feasibility Standards (0.757), Utility Standards (0.747) and Evaluation Accountability Standards (0.714) (see Table 6). The Weights of Meta-evaluation Category Standards of University Program evaluation). It is thus clear that in university program evaluations, Propriety Standards are especially valued.

Table 6 *The Weights of Meta-evaluation Category Standards of University Program Evaluation*

No.	Standards	Average Total Value μT	Order
1.	Utility Standards	0.747	4
2.	Feasibility Standards	0.757	3
3.	Propriety Standards	0.765	1
4.	Accuracy Standards	0.761	2
5.	Evaluation Accountability Standards	0.714	5

In Table 6, The Weights of Meta-evaluation Category Standards of University Program evaluation. Table 6 shows that in terms of Taiwan's university program evaluations, the importance of Propriety Standards should be the same order as that of Personnel Evaluation Standards, ranked top, showing that there is an expectation for university program evaluations to be proper, fair, legal, right and just. Of secondary important is Accuracy Standards, the expectation of an increase in the dependability and truthfulness of university program evaluations' representations, propositions and findings.

The importance of Utility Standards, which is top in the JCSEE Program Evaluation Standards, is comparatively low in the university program evaluations found in this study, which is a finding worthy of follow-up studies. In the JCSEE Program Evaluation Standards, the importance of "valuation Accountability Standards," the category newly added in the revised third edition of JCSEE Program Evaluation Standards, is comparatively not valued.

CONCLUSION AND IMPLICATIONS

In this study, meta-evaluation indicators for Taiwan's university program evaluation were constructed according to the JCSEE Program Evaluation Standards' 30 standards across five major categories Propriety Standards, Utility Standards, Feasibility Standards, Accuracy Standards and Evaluation Accountability Standards by referring to the 150

summative meta-evaluation checkpoints of Stufflebeam (2012) program evaluation. Sixty indicators were reserved and identified for fitness via experts and scholars' opinions and Fuzzy Delphi Method questionnaire analysis. In the future, when university departments carry out program evaluations, this indicator system can be cited during the examination of the quality of departments' program evaluations.

As the meta-evaluation indicators of university program evaluation constructed in this study were examined by the standards in five major categories respectively, the most valued standards in order are U6 Meaningful Processes and Products (Utility Standards), F3 Contextual Viability (Feasibility Standards), P1 Responsive and Inclusive Orientation (Propriety Standards), A3 Reliable Information (Accuracy Standards) and E1 Evaluation Documentation (Evaluation Accountability Standards). It can remind the university program evaluation of the standards that should be especially valued while carrying out a university program evaluation.

Finally, the comparison of the JCSEE Program Evaluation Standards' standards across five major categories (Utility Standards, Feasibility Standards, Propriety Standards, Accuracy Standards and Evaluation Accountability Standards) shows that experts and scholars consider Propriety Standards the most important and Accuracy Standards the second most important for university program evaluations, contrary to the JCSEE Program Evaluation Standards' own order of importance, a finding worthy of follow-up studies in the near future. This study canbe expected to benefit the establishment and quality enhancement of the evaluation system for the program evaluation.

REFERENCES

Alford, H., & Aharonian, A. (2011). *What is institutional research?* Retrieved from https://bit.ly/2Vo6U5T

Attuluri, S. (2019). Institutional changes in Mexico higher education: A reform to U.S accreditations. *International Journal of Humanities, Arts and Social Sciences*, *5*(1), 9-19. doi:https://dx.doi.org/10.20469/ijhss.5.10002-1

Ayob, A. H., & Morell, J. A. (2016). The historical path of evaluation as reflected in the content of evaluation and program planning. *Evaluation and Program Planning*, *58*, 20–27. doi:https://doi.org/10.1016/j.evalprogplan.2016.05.006

Blair, E., & Noel, K. V. (2014). Improving higher education practice through student evaluation systems: Is the student voice being heard? *Assessment & Evaluation in Higher Education*, *39*(7), 879–894. doi:https://doi.org/10.1080/02602938.2013.875984

Chan, S.-J., & Chan, Y. (2015). Higher education research community in Taiwan: An emerging field. *Higher Education Policy*, *28*(4), 459–475. doi:https://doi.org/10.1057/hep.2015.16

Chen, S. J., & Hwang, C. L. (1992). Fuzzy multiple attribute decision making methods. In *Fuzzy multiple attribute decision making.* Heidelberg, Germany: Springer.

Cheng-Ta, W. (2008). *Educational policy analysis: Concepts methods, and applications*. Taipei, Taiwan: Higher Education.

Cousins, J. B., Goh, S. C., Elliott, C. J., & Bourgeois, I. (2014). Framing the capacity to do and use evaluation. *New Directions for Evaluation*, *2014*(141), 7–23. doi:https://doi.org/10.1002/ev.20076

Crawford, E. R., Aguayo, D., & Valle, F. (2017). Counselors as leaders who advocate for undocumented students' education. *Journal of Research on Leadership Education*, *14*(2), 119–150. doi:https://doi.org/10.1177/1942775117739301

Froh, R. C. (1991). Using student persistence research to strengthen the quality of teaching and institutional process. In M. Nugent & J. L. Ratcliff (Eds.), *Measurement of quality in higher education.* Washington, DC,WA: Office of Educational Research and Improvement.

Gall, M. D., Gall, J. P., & Borg, W. R. (2003). *Educational research: An introduction*. Boston, MA: Allyn & Bacon.

Gatpandan, M. P., & Ambat, S. C. (2017). Implementing knowledge discovery in enhancing university student services portfolio management in higher education institutions. *Journal of Advanced Research in Social Sciences and Humanities*, *2*(4), 211-220. doi:https://doi.org/10.26500/jarssh-02-2017-0306

Harvey, L., & Williams, J. (2010). Fifteen years of quality in higher education. *Quality in Higher Education*, *16*(1), 3–36. doi:https://doi.org/10.1080/13538321003679457

HEEACT. (2015). *Ministry of educations pilot implementation project for certifying universities program evaluation outcome examination.* Retrieved from https://bit.ly/2ZabTrM

Horta, H. (2017). Higher-education researchers in Asia: The risks of insufficient contribution to international higher-education research. In *Researching higher education in Asia* (pp. 15–36). Gateway East, Singapore:

Springer.

Hou, A. Y.-C., Morse, R., & Wang, W. (2015). Recognition of academic qualifications in transnational higher education and challenges for recognizing a joint degree in Europe and Asia. *Studies in Higher Education*, *42*(7), 1211–1228. doi:https://doi.org/10.1080/03075079.2015.1085010

JCSEE. (2015). *Program evaluation standards statements.* Retrieved from https://bit.ly/2Bfbe0g

Kyriakides, L., & Campbell, R. (2004). School self-evaluation and school improvement: A critique of values and procedures. *Studies in Educational Evaluation*, *30*(1), 23–36. doi:https://doi.org/10.1016/s0191-491x(04)90002-8

Mertens, D. M., & Wilson, A. T. (2018). *Program evaluation theory and practice*. New York, NY: Guilford Publications.

Misnistry of Education. (2015). *Administrative goals and key tasks.* Retrieved from https://bit.ly/2ZbTzys

Murray, T. J., Pipino, L. L., & van Gigch, J. P. (1985). A pilot study of fuzzy set modification of delphi. *Human Systems Management*, *5*(1), 76–80. doi:https://doi.org/10.3233/hsm-1985-5111

Nazari-Shirkouhi, S., Mousakhani, S., Tavakoli, M., Dalvand, M. R., Šaparauskas, J., & Antuchevičienė, J. (2020). Importance-performance analysis based balanced scorecard for performance evaluation in higher education institutions: An integrated fuzzy approach. *Journal of Business Economics and Management*, *21*(3), 647–678. doi:https://doi.org/10.3846/jbem.2020.11940

Noorderhaben, N. (1995). *Strategic decision making*. Boston, MA: Addison Wesley.

Perterson, K. D. (2000). *Teacher evaluation*. New York, NY: Corwin Press Inc.

Schumacher, S., & McMillan, J. (2006). *Research in education evidencebased inquiry* (6th ed.). Boston, MA: Allyn & Bacon.

Stufflebeam, D. L. (2012). *Program evaluations summary meta evaluation checklist (based on the program evaluation standards).* Retrieved from https://bit.ly/3i6VJIe

Sturges, K. M., & Howley, C. (2016). Responsive meta-evaluation. *American Journal of Evaluation*, *38*(1), 126–137. doi:https://doi.org/10.1177/1098214016630405

Tang, X., Wei, G., & Gao, H. (2019). Models for multiple attribute decision making with interval-valued pythagorean fuzzy muirhead mean operators and their application to green suppliers selection. *Informatica*, *30*(1), 153–186. doi:https://doi.org/10.15388/informatica.2018.202

Volden, G. H. (2018). Public project success as seen in a broad perspective. *Evaluation and Program Planning*, *69*, 109–117. doi:https://doi.org/10.1016/j.evalprogplan.2018.04.008

Wei, G. (2018). TODIM method for picture fuzzy multiple attribute decision making. *Informatica*, *29*(3), 555–566. doi:https://doi.org/10.15388/informatica.2018.181

Wolff, R. A. (2010). Key trends for quality assurance in the us today. *Evaluation in Higher Education*, *4*(2), 55–87.

Wudhikarn, R. (2016). An efficient resource allocation in strategic management using a novel hybrid method. *Management Decision*, *54*(7), 1702–1731. doi:https://doi.org/10.1108/md-08-2015-0380

Yarbrough, D. B., Shulha, L. M., Hopson, R. K., & Caruthers, F. A. (2011). *The program evaluation standards: A guide for evaluators and evaluation users*. Thousand Oaks, CA: Sage.