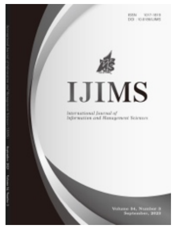
http://ijims.ms.tku.edu.tw/main.php

International Journal of
Information
and
Management
Sciences

International Journal of
Information and Management Sciences

37 (2026), 77-94. DOI:10.6186/IJIMS.202603_37(1).0005



A Data-Driven Performance Evaluation Matrix to Enhance EMI Course Quality

Shao-Yu Li[1], *Hung-Tso Lin*[2] *and Chun-Hung Yu*[3,4,*]
[1] Language Center, Chaoyang University of Technology, Taichung, Taiwan
[2] Department of Marketing and Distribution Management, National Chin-Yi University of Technology, Taichung, Taiwan
[3] Language Center, National Chin-Yi University of Technology, Taichung, Taiwan
[4] Language Teaching Center, National Chi Nan University, Nantou County, Taiwan

Keywords

EMI, Satisfaction Index, Importance Index, Performance Evaluation Matrix, Total Quality Management

Abstract

English as a Medium of Instruction (EMI) has gained prominence in higher education, yet ensuring teaching quality in non-English-speaking contexts remains a persistent challenge. This study develops a data-driven Performance Evaluation Matrix (PEM) to systematically assess and improve EMI course quality. Drawing on the service-operation perspective, instructors are treated as service providers and students as service recipients, and student feedback on instructional items is summarized through satisfaction and importance indices. Using statistical inference, the model estimates these indices and derives joint upper confidence limits as conservative decision aids that explicitly account for sampling variability when classifying instructional items. Within a Total Quality Management (TQM) logic of continuous improvement and matrix-based prioritization, the PEM partitions the satisfaction–importance space into four quadrants, highlights critical-to-quality items, and provides an improvement priority ranking under resource constraints. An empirical case study of an EMI course at a university in Taiwan illustrates how the model can support evidence-based decisions about course enhancement. While the findings offer initial empirical support for the proposed framework, they should be interpreted in light of the single-institution sample and the inherent subjectivity of student ratings. Overall, the PEM provides a transparent and statistically grounded tool for enhancing EMI teaching effectiveness and promoting continuous course improvement.

1. Introduction

*corresponding author. Email: chyu@ncnu.edu.tw

In pursuit of internationalization, many non-English-speaking countries have implemented bilingual education policies (Dafouz & Smit, 2016; Zhou et al., 2025; Nguyen et al., 2025; Li & Pei, 2024; Lin et al., 2024). These policies often promote English language instruction broadly, accompanied by standardized English proficiency assessments to encourage students to invest in English learning and improve their language skills (Green & Inoue, 2024; Wu & Lee, 2017). In addition to English language courses, a growing number of subject-specific courses are also delivered in English. These are commonly referred to as English as a Medium of Instruction (EMI) courses (Patterson et al., 2025).

EMI courses involve the teaching of disciplinary content through English rather than the native language. However, insufficient English proficiency among students can impede their understanding of subject matter, becoming a significant barrier to effective learning outcomes (Li & Pei, 2024; Kaiypova et al., 2025; Anghel et al., 2016). Beyond language proficiency, several other factors—such as the instructor's subject expertise, pedagogical approach, and the availability of learning-related resources—can also influence students' learning performance in EMI contexts (Nguyen et al., 2025; Lin et al., 2024; Kaiypova et al., 2025). Therefore, enhancing student learning outcomes in EMI courses remains a critical area of academic and pedagogical concern.

Some studies have conceptualized EMI courses as service operation systems (Lin et al., 2023; Chen et al., 2024). From this viewpoint, constructing a robust evaluation and analysis framework for service systems can aid instructors (as service providers) in understanding whether various instructional components meet students' expectations (as service recipients). It also allows for the identification and improvement of instructional dimensions that receive lower satisfaction, thereby enhancing students' overall satisfaction with the EMI learning experience (Liu et al., 2023; Yu et al., 2017).

Furthermore, prior research has indicated that the Performance Evaluation Matrix (PEM) is an effective and practical tool for assessing the performance of service operation systems (Chen & Yu, 2020; Chen et al., 2020; Chen et al., 2024; Li et al., 2021; Jeng et al., 2022; Kim et al., 2022). Considering this, the present study proposes the development of a performance evaluation and improvement of decision-making model for EMI courses based on the PEM framework. This model aims to support instructors in identifying underperforming instructional aspects and implementing targeted improvements to elevate students' overall satisfaction and learning effectiveness in EMI contexts. However, many existing EMI evaluation approaches still rely primarily on descriptive statistics or informal judgments, which may obscure item-level weaknesses and provide limited statistical guidance for improvement decisions. Building on the view of EMI as a service operation system, the present study therefore adopts a data-driven PEM to translate students' perceptions into explicit performance indicators and decision rules for systematic course evaluation and improvement.

Recent research has continued to refine the understanding of EMI course quality and learner performance. Studies such as Green and Inoue (2024), Li and Pei (2024), and Lin et al. (2024) highlight the need for more rigorous, data-driven approaches to evaluating language proficiency and academic achievement in EMI settings. Building on this work, more recent investigations from 2025 onwards (Kaiypova et al., 2025; Nguyen et al., 2025; Patterson et al., 2025; Zhou et al., 2025) further emphasize the importance of comprehensive performance indicators that capture both linguistic and instructional dimensions. The incorporation of this recent literature strengthens the study's scholarly credibility and underscores its relevance to current research discourse, while also clarifying how the proposed PEM framework aligns with and extends the latest developments in EMI evaluation.

To maintain generalizability, this study begins by assuming that the EMI course consists of q instructional items (Yu et al., 2018). A questionnaire can then be designed to collect data on students' perceived importance and satisfaction regarding these q instructional items. According to Lin et al. (2023) and Liu et al. (2023), this paper let random variable X_j represents the satisfaction of the *j*th service item, then X_j is distributed as $Beta(a_{Sj}, b_{Sj})$, where a_{Sj} and b_{Sj} are two parameters of this satisfaction distribution. Similarly, let random variable Y_j indicates the importance attached to the *j*th service item, then Y_j is distributed as $Beta(a_{Ij}, b_{Ij})$, where a_{Ij} and b_{Ij} are two parameters of this importance distribution. Thus, these two indices can be defined as follows:

$$\theta_{Sj} = \frac{a_{sj}}{a_{sj}+b_{sj}}, \text{ (satisfaction index of the } j\text{th service item)} \quad (1)$$

$$\theta_{Ij} = \frac{a_{Ij}}{a_{Ij}+b_{Ij}}, \text{ (importance index of the } j\text{th service item)} \quad (2)$$

where $j = 1,2,\cdots,q$. Both indices' values are between 0 and 1. When the value of the index is 0, it means complete dissatisfaction or lack of attention. When the value of the index is 1, it means complete satisfaction or attention.

Subsequently, a Performance Evaluation Matrix (PEM) for the EMI course is constructed by setting the satisfaction index as the horizontal axis and the importance index as the vertical axis. Based on the concept of continuous quality improvement in Total Quality Management (TQM), the PEM is divided into four evaluation quadrants using the average values of the observed satisfaction and importance indices as reference lines. It is important to note that both the satisfaction θ_{sj} and importance θ_{Ij} indices are unknown parameters. As pointed out by Chen et al. (2024), directly using point estimates to evaluate these indices may lead to misclassification due to sampling error. To address this concern, the present study derives joint upper confidence bounds for the satisfaction and importance indices across the q instructional items and uses these bounds as the evaluation coordinate points. Each instructional item is then classified into one of the four PEM quadrants based on the location of its coordinate point, and evaluation criteria are established following statistical inference principles. This approach enables decision-makers (i.e., instructors) to efficiently and reliably identify critical-to-quality (CTQ) instructional items that would yield the greatest benefit from targeted improvements. Furthermore, in consideration of resource constraints, the proposed evaluation model also provides a prioritization scheme, allowing instructors to rank the instructional items in need of improvement. This prioritization facilitates the optimal allocation of efforts and resources to enhance students' overall satisfaction with EMI courses.

The remainder of this paper is organized as follows. In Section 2, the derivation of the $1-\alpha$ confidence intervals for both the satisfaction index and the importance index is presented, and the corresponding upper confidence limits for these two indices are obtained. These upper limits are then utilized to construct evaluation coordinate points for each instructional item, reflecting the fact that the coordinates incorporate sampling variability through the use of upper confidence limits rather than the point estimates. In Section 3, based on the principle of continuous quality improvement in TQM, the PEM is divided into four evaluation quadrants using the point-estimated average values of the satisfaction and importance indices as reference lines, and decision rules are developed. Section 4 provides a real-world case study to demonstrate the practical application of the proposed evaluation and decision-making model.

2. The Joint Upper Confidence Limits of Satisfaction Index and Importance Index

As previously mentioned, since the satisfaction index is an unknown parameter, it must be estimated using sample data obtained from the surveyed students. Let the random variable X_{jh} denote the satisfaction score given by the hth respondent for the jth instructional item. The observed value of this random variable X_{jh} is represented as x_{jh}.

$$\begin{pmatrix} X_{11},\cdots,X_{1h},\cdots,X_{1n} \\ \vdots \quad \vdots \quad \vdots \\ X_{j1},\cdots,X_{jh},\cdots,X_{jn} \\ \vdots \quad \vdots \quad \vdots \\ X_{q1},\cdots,X_{qh},\cdots,X_{qn} \end{pmatrix} = \begin{pmatrix} x_{11},\cdots,x_{1h},\cdots,x_{1n} \\ \vdots \quad \vdots \quad \vdots \\ x_{j1},\cdots,x_{jh},\cdots,x_{jn} \\ \vdots \quad \vdots \quad \vdots \\ x_{q1},\cdots,x_{qh},\cdots,x_{qn} \end{pmatrix} \tag{3}$$

Then the estimator of jth satisfaction index is as follows:

$$\theta_{Sj}^{*} = \bar{X}_j = \frac{1}{n}\sum_{h=1}^{n} x_{jh} \tag{4}$$

The observed values of θ_{Sj}^{*} is

$$\theta_{Sj0}^{*} = \bar{x}_j = \frac{1}{n}\sum_{h=1}^{n} x_{jh} \tag{5}$$

where $j = 1,2,\cdots,q$. This study then calculates the average of all observed values for the satisfaction index as follows:

$$\theta_{S0} = \frac{1}{q}\sum_{j=1}^{q} \theta_{Sj0}^{*} \tag{6}$$

Similarly, the importance index is also an unknown parameter and must be estimated using sample data collected from the surveyed students. Let random variable Y_{jh} represent the importance of rating given by the hth respondent for the jth item. The set of observed variables Y_{jh} is denoted as y_{jh}.

$$\begin{pmatrix} Y_{11},\cdots,Y_{1h},\cdots,Y_{1n} \\ \vdots \quad \vdots \quad \vdots \\ Y_{j1},\cdots,Y_{jh},\cdots,Y_{jn} \\ \vdots \quad \vdots \quad \vdots \\ Y_{q1},\cdots,Y_{qh},\cdots,Y_{qn} \end{pmatrix} = \begin{pmatrix} y_{11},\cdots,y_{1h},\cdots,y_{1n} \\ \vdots \quad \vdots \quad \vdots \\ y_{j1},\cdots,y_{jh},\cdots,y_{jn} \\ \vdots \quad \vdots \quad \vdots \\ y_{q1},\cdots,y_{qh},\cdots,y_{qn} \end{pmatrix} \tag{7}$$

Then the estimator of jth importance index is as follows:

$$\theta_{Ij}^{*} = \bar{Y}_j = \frac{1}{n}\sum_{h=1}^{n} y_{jh} \tag{8}$$

The observed values of θ_{Ij}^{*} is

$$\theta_{Ij0}^{*} = \bar{y}_j = \frac{1}{n}\sum_{h=1}^{n} y_{jh} \tag{9}$$

where $j = 1,2,\cdots,q$. This paper then calculates the average of the observed values for the importance index across all respondents as follows:

$$\theta_{I0} = \frac{1}{q}\sum_{j=1}^{q} \theta_{Ij0}^{*} \tag{10}$$

Then, this paper derives the upper confidence limit of the satisfaction index and the upper confidence limit of the importance index. As noted by Chen et al. (2024), since $E[\theta sj^{*}] = \theta sj$, the unbiased estimator of the satisfaction index θsj is θsj^{*} . Let random variable Z_j be defined as follows:

$$Z_j = \frac{\theta_{Sj}^{*} - \theta_{Sj}}{S_{Xj}/\sqrt{n}} \tag{11}$$

where S_{Xj} is the sample standard deviation as follows:

$$S_{Xj} = \sqrt{\frac{1}{n-1}\sum_{h=1}^{n}\left(X_{jh} - \bar{X}_j\right)^2} \tag{12}$$

Based on the Central Limits Theorem (CLT) (Hogg & Tanis, 2021), random variable Z_j is approximately distributed as a standard normal distribution for large sample size n. Then, the $100(1-\alpha/2)\%$ upper confidence limit of satisfaction index θ_{Sj} as follows:

$$p\left(\theta_{Sj} \le \theta_{Sj}^{*} + z_{\alpha}\frac{S_{Xj}}{\sqrt{n}}\right) = 1 - \alpha/2 \tag{13}$$

Therefore, the $100(1-\alpha/2)\%$ upper confidence limit of satisfaction index is

$$U\theta_{Sj} = \theta_{Sj}^{*} + z_{\alpha/2}\frac{S_{Xj}}{\sqrt{n}} \tag{14}$$

where $z_{\alpha/2}$ is upper $\alpha/2$ quantile of standard normal distribution. Similarly, since $E\left[\theta_{Ij}^{*}\right] = \theta_{Ij}$, the unbiased estimators of importance index θ_{Ij} is θ_{Ij}^{*}. Let random variable Z'_j as below:

$$Z'_j = \frac{\theta_{Ij}^{*} - \theta_{Ij}}{S_{Yj}/\sqrt{n}} \tag{15}$$

where S_{Yj} is the sample standard deviation as follows:

$$S_{Yj} = \sqrt{\frac{1}{n-1}\sum_{h=1}^{n}\left(Y_{jh} - \bar{Y}_j\right)^2} \quad (16)$$

Based on the Central Limits Theorem (CLT), random variable Z'_j is approximately distributed as a standard normal distribution for large sample size n. Then, the $100(1-\alpha/2)\%$ upper confidence limit of importance index θ_{Ij} as follows:

$$p\left(\theta_{Ij} \le \theta_{Ij}^* + z_{\alpha/2}\frac{S_{Yj}}{\sqrt{n}}\right) = 1 - \alpha/2 \quad (17)$$

Therefore, the $100(1-\alpha/2)\%$ upper confidence limit of importance index is

$$U\theta_{Ij} = \theta_{Ij}^* + z_{\alpha/2}\frac{S_{Yj}}{\sqrt{n}} \quad (18)$$

To derive the $100(1-\alpha)\%$ joint upper confidence limits of $(\theta_{sj}, \theta_{Ij})$, this paper defines events E_{Sj} and E_{Ij} as follows:

$$E_{Sj} = \left\{\theta_{Sj} \le \theta_{Sj}^* + z_{\alpha/2}\frac{S_{Xj}}{\sqrt{n}}\right\} \quad (19)$$

$$E_{Ij} = \left\{\theta_{Ij} \le \theta_{Ij}^* + z_{\alpha/2}\frac{S_{Yj}}{\sqrt{n}}\right\} \quad (20)$$

Obviously, $p(E_{sj}) = p(E_{Ij}) = 1 - \alpha/2$ and $p(E_{sj}^C) = p(E_{Ij}^C) = \alpha/2$ 。According to Boole's inequality and DeMorgan's rules, we have $p(E_{Sj} \cap E_{Ij}) \ge 1 - p(E_{Sj}^C) - p(E_{Ij}^C) = \alpha/2$. Thus,

$$p\left\{\theta_{Sj} \le \theta_{Sj}^* + z_{\alpha/2}\frac{S_{Xj}}{\sqrt{n}}, \theta_{Ij} \le \theta_{Ij}^* + z_{\alpha/2}\frac{S_{Yj}}{\sqrt{n}}\right\} = 1 - \alpha \quad (21)$$

Therefore, $100(1-\alpha)\%$ joint upper confidence limits of $(\theta_{Sj}, \theta_{Ij})$ can be shown as:

$$JR = \left\{(\theta_{Sj}, \theta_{Ij}) \middle| \theta_{Sj} \le \theta_{Sj}^* + z_{\alpha/2}\frac{S_{Xj}}{\sqrt{n}}, \theta_{Ij} \le \theta_{Ij}^* + z_{\alpha/2}\frac{S_{Yj}}{\sqrt{n}}\right\} \quad (22)$$

Obviously, the evaluate point of jth service item is (x_j, y_j), where

$$x_j = U\theta_{Sj} = \theta_{Sj}^* + z_{\alpha/2}\frac{S_{Xj}}{\sqrt{n}} \quad (23)$$

$$y_j = U\theta_{Ij} = \theta_{Ij}^* + z_{\alpha/2}\frac{S_{Yj}}{\sqrt{n}} \quad (24)$$

3. Performance Evaluation Matrix of EMI: Evaluation and Improvement Decision Rules

This study uses the satisfaction index θ_{Sj} as the horizontal axis (x-axis) and the importance index θ_{Ij} as the vertical axis (y-axis) to construct a PEM for the EMI course, as shown in Figure 1. Based on the principle of continuous improvement in TQM, the PEM is divided into four quadrants by drawing a vertical line $x = \theta_{S0}$ and a horizontal line $y = \theta_{I0}$ at the average values of the satisfaction and importance indices, respectively, as illustrated in Figure.1. It should be noted that the coordinates plotted in the matrix are based on the estimated satisfaction and importance indices for each instructional item, whereas the upper confidence limits derived in Section 2 are used only in the subsequent hypothesis tests to determine whether an item meets or falls short of the benchmark values.

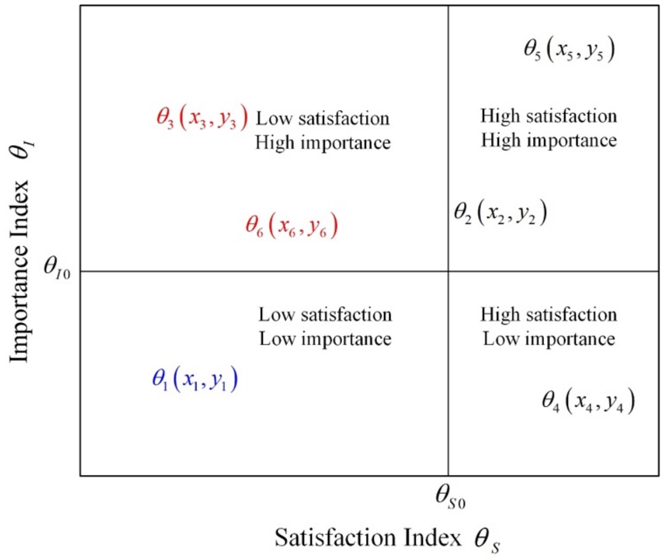


Figure 1 *Four Quadrants of Performance Evaluation Matrix*

Next, Q_I, Q_{II}, Q_{III}, and Q_{IV} denotes the following:

Quadrant 1: $Q_I = \{(x, y) | \theta_{S0} \leq x \leq 1, \theta_{I0} \leq y \leq 1\}$

Quadrant 2: $Q_{II} = \{(x, y) | 0 \leq x < \theta_{S0}, \theta_{I0} \leq y \leq 1\}$

Quadrant 3: $Q_{III} = \{(x, y) | 0 \leq x < \theta_{S0}, 0 \leq y < \theta_{I0}\}$

Quadrant 4: $Q_{IV} = \{(x, y) | \theta_{S0} \leq x \leq 1, 0 \leq y < \theta_{I0}\}$

Next, this study establishes the evaluation criteria for the PEM of the EMI course based on principles of statistical hypothesis testing. As previously mentioned, in line with the spirit of continuous improvement emphasized in TQM, the overall average of all satisfaction indices, denoted as θ_{S0}, is used as the benchmark for hypothesis testing of individual satisfaction indices. Then, the null and alternative hypotheses can be formulated as follows:

null hypothesis H_0: $\theta_{Sj} \geq \theta_{S0}$ versus alternative hypothesis H_1: $\theta_{Sj} < \theta_{S0}$

Next, this study conducts a statistical test using the upper confidence limit of the satisfaction index $U\theta_{Sj}$ for the *j*th instructional item. The statistical decision rule is formulated as follows:

1. When the upper confidence limit of the satisfaction index $U\theta_{Sj} \geq \theta_{S0}$, then do not reject the null hypothesis H_0 and conclude that $\theta_{Sj} \geq \theta_{S0}$ In this case, no improvement is deemed necessary for item *j*, and it can be maintained as is.
2. When the upper confidence limit $U\theta_{sj} < \theta_{S0}$, then reject the null hypothesis H_0 and conclude that $\theta_{Sj} < \theta_{S0}$. Consequently, item *j* should be considered for improvement to enhance the overall effectiveness of the EMI course.

After completing the statistical testing of the satisfaction index, the analysis proceeds with the statistical testing of the importance index. This step serves as a reference for prioritizing improvements when instructional resources are limited. Like the approach used for the satisfaction index, this study adopts θ_{I0} as the benchmark for hypothesis testing across all instructional items. Then, the null hypothesis and alternative hypothesis can be shown as follows:

Then, null hypothesis H'_0: $\theta_{Ij} \geq \theta_{I0}$ versus alternative hypothesis H'_1: $\theta_{Ij} < \theta_{I0}$

Next, the statistical test will be conducted using the upper confidence limit of the importance index $U\theta_{Ij}$. The corresponding hypothesis testing rules are outlined as follows:

3. When the upper confidence limit of the importance index $U\theta_{Ij} \geq \theta_{I0}$, then do not reject the null hypothesis H'_0, and conclude that $\theta_{Ij} \geq \theta_{I0}$. In this case, service item *j* is considered of high importance, and thus should be given a higher priority for improvement.
4. When the upper confidence limit of the importance index $U\theta_{Ij} < \theta_{I0}$, then reject the null hypothesis H'_0, and conclude that $\theta_{Ij} < \theta_{I0}$. In this case, the service item *j* is considered of lower importance, and therefore may be assigned a lower priority for improvement.

The following section of this study utilizes the evaluation coordinate point (x_j, y_j) for the *j*th instructional item, derived from Equations (Chen et al., 2024 ; Li et al., 2019), as the basis for constructing the evaluation rules of the PEM for EMI courses. These rules are established in accordance with the statistical testing criteria developed for the satisfaction index. This approach facilitates the identification of instructional items requiring improvement. Furthermore, by applying the statistical testing criteria for the importance index, this framework allows instructors to determine the priority order for improvements, particularly under conditions of limited resources.

1. When the evaluation coordinate point $(x_j, y_j) \in Q_I \cup Q_{IV}$ lies in the region where the upper confidence limit of the satisfaction index $U\theta_{Sj} \geq \theta_{S0}$, then according to Statistical Test Rule 1, do not reject the null hypothesis H_0, and conclude that the satisfaction level for instructional item *j* meets expectations. Therefore, instructional item *j* can be maintained as is and does not require improvement.
2. When the evaluation coordinate point $(x_j, y_j) \in Q_{II} \cup Q_{III}$ lies in the region where the upper confidence limit of the satisfaction index $U\theta_{sj} < \theta_{S0}$, then according to

Statistical Test Rule 2, reject the null hypothesis H_0 , and conclude that the satisfaction level for instructional item j does not meet expectations. Therefore, instructional item j requires improvement.

3. When the evaluation coordinate point $(x_j, y_j) \in Q_{II}$ lies in the region where the upper confidence limit of the importance index $U\theta_{Ij} \geq \theta_{I0}$, then according to Statistical Test Rule 3, do not reject the null hypothesis H'_0, and conclude that $\theta_{Ij} \geq \theta_{I0}$. Therefore, instructional item j should be assigned a high priority for improvement.

4. When the evaluation coordinate point $(x_j, y_j) \in Q_{III}$ lies in the region where the upper confidence limit of the importance index $U\theta_{Ij} < \theta_{I0}$, then according to Statistical Test Rule 4, reject the null hypothesis H'_0 and conclude that $\theta_{Ij} < \theta_{I0}$. Therefore, instructional item j should be assigned as a lower priority for improvement.

4. A Case Study

Over the past two decades, English as a Medium of Instruction (EMI) has witnessed rapid growth in the domain of higher education (Dafouz & Smit, 2016; Zhou et al., 2025). Universities in many countries have actively promoted EMI environments with the aim of fostering multicultural learning contexts, enhancing the global competitiveness of academic institutions, and attracting international students while increasing student mobility across borders (Nguyen et al., 2025; Li & Pei, 2024). In 2021, Taiwan's Ministry of Education launched the Bilingual Education for Students in College (BESIC) initiative, promoting EMI to "enhance students' English proficiency, expand English-taught courses, and strengthen the international competitiveness of higher education institutions." This policy aims to improve both faculty members' English teaching competencies and students' English abilities, while also fostering students' global outlook and intercultural integration skills in an English-medium learning environment (Lin et al., 2024).

Against this backdrop, this study uses an EMI course implemented at a university in Taiwan as an empirical case to illustrate the application process of the proposed performance evaluation model. Specifically, the study examines learners' perceived satisfaction and importance regarding EMI instruction, providing valuable insights for improving EMI course quality and enhancing student learning effectiveness.

To explore learners' perceptions of satisfaction and perceived importance in EMI (English as a Medium of Instruction) courses, this study adopts a questionnaire based on Yu et al.'s instrument for evaluating satisfaction with foreign language teaching (Yu et al., 2017), with modifications referencing the work of Yeh et al. (2005). Minor textual adaptations were made to tailor the questionnaire to the EMI course context, resulting in an EMI Course Learning Effectiveness Questionnaire. The instrument includes five key dimensions: (1) Teaching Preparation, (2) Teaching Attitude, (3) Teaching Capability, (4) Teaching Management, and (5) Coursework and Evaluation.

For each item, respondents are asked to evaluate both their level of satisfaction and the importance they assign to the item, using a scale from 0 to 10—where 0 indicates "extremely dissatisfied" (or "least important") and 10 indicates "extremely satisfied" (or "most important").

The participants of this study were students at the university level or above who had previously taken EMI (English as a Medium of Instruction) courses. A total of 329 questionnaires were distributed, with 272 returned, resulting in a response rate of 82.7%. After eliminating 9 invalid responses, 263 valid questionnaires were retained, yielding an effective response rate of 79.9%. Among the valid respondents, 65.31% were undergraduate students, 32.84% were master's students, and 1.85% were doctoral students. In terms of gender distribution, male and female participants accounted for 41.70% and 58.30%, respectively. Regarding academic disciplines, 81.74% were from business and management faculties, 17.11% from information and engineering, and 1.14% from humanities and social sciences. Because the sample is drawn from a single institution and is dominated by business and management majors, the empirical findings should be generalized to other EMI contexts (such as STEM or humanities programs) with caution.

Regarding reliability, the analysis conducted using SPSS revealed that the Cronbach's α values for each dimension were as follows: Dimension 1 reached 0.916; Dimension 2 reached 0.931; Dimension 3 reached 0.875; Dimension 4 reached 0.888; and Dimension 5 reached 0.800. The overall Cronbach'sαvalue was 0.971, indicating high internal consistency across the questionnaire items. According to DeVellis (2023), these results suggest that the reliability of the questionnaire scale is excellent and suitable for the purposes of this study. As for validity, the EMI questionnaire adopted in this study is based on the well-established instruments developed by Yu et al. (2017) and Yeh et al. (2005), both of which demonstrate strong content validity. Therefore, it can be concluded that the questionnaire used in this study meets the required standards of content validity.

Based on the analysis of the 263 valid questionnaires, the evaluation point of the *j*th service item (x_j, y_j), as shown in Table 1.

Table 1. *Summary of survey for EMI course*

Dimensions	Items	(x_j, y_j)	Quadrant	Improvement	Priority
Teaching preparation	. The course material that teachers have prepared is at an adequate level of complexity.	(0.869, 0.868)	4	N	
	. The quantity of the material is appropriate.	(0.876, 0.871)	4	N	
	. The content of the material helps to improve my English proficiency	(0.859, 0.867)	3	Y	Low
	. Teachers have prepared thoroughly for class	(0.897, 0.903)	1	N	
Teaching attitude	. Teachers emphasize conversation practice in English	(0.881, 0.881)	1	N	
	. Teachers value the opinions of students	(0.895, 0.897)	1	N	
	. Student-teacher interaction in class is strong	(0.891, 0.898)	1	N	

	. Teachers are happy to help students solve problems	(0.906, 0.914)	1	N	
	. Teachers treat all students fairly	(0.913, 0.909)	1	N	
Teaching capability	. Teachers speak English clearly	(0.880, 0.891)	1	N	
	. Teachers express themselves logically	(0.888, 0.894)	1	N	
Teaching management	. Teachers use a variety of teaching methods	(0.871, 0.878)	1	N	
	. Teachers interest me in learning English	(0.855, 0.866)	3	Y	Low
	. Teachers adequately control the pace of learning	(0.874,0.884)	1	N	
Coursework and evaluation	. The coursework or evaluation helps me improve my English proficiency.	(0.863, 0.872)	2	Y	High
	. The level of difficulty for assignments and exams is suitable.	(0.864, 0.885)	2	Y	High

Subsequently, using the satisfaction index θ_{Sj} as the horizontal axis (x-axis) and the importance index θ_{Ij} as the vertical axis (y-axis), a PEM for the EMI course is constructed, as shown in Figure. 2. In line with the TQM philosophy of continuous improvement, the matrix is divided into four evaluation quadrants by a vertical line at $x = \theta_{S0} = \frac{1}{16}\sum_{j=1}^{16} \theta^*_{Sj0} = 0.865$ and a horizontal line at $y = \theta_{I0} = \frac{1}{16}\sum_{j=1}^{16} \theta^*_{Ij0} = 0.871$, as illustrated in Figure. 2.

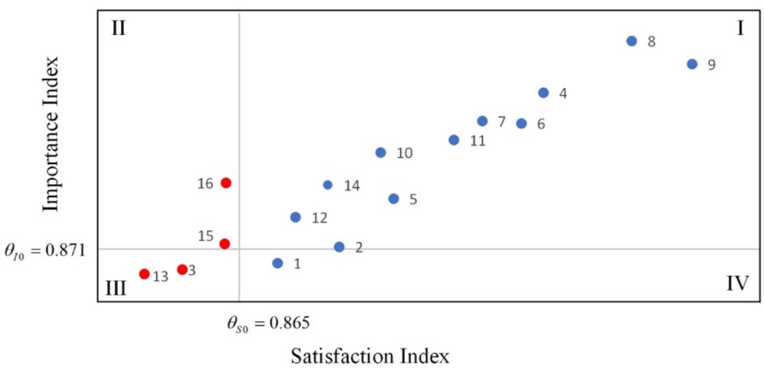


Figure 2 *Performance evaluation matrix for EMI Courses*

In this study, the four quadrants Q_I, Q_{II}, Q_{III}, and Q_{IV} are defined as follow:

Quadrant 1: $Q_I = \{(x, y)|0.865 \le x \le 1, 0.871 \le y \le 1\}$

Quadrant 2: $Q_{II} = \{(x, y)|0 \leq x < 0.865, 0.871 \leq y \leq 1\}$

Quadrant 3: $Q_{III} = \{(x, y)|0 \leq x < 0.865, 0 \leq y < 0.871\}$

Quadrant 4: $Q_{IV} = \{(x, y)|0.865 \leq x \leq 1, 0 \leq y < 0.871\}$

Next, based on the statistical testing rules, the evaluation rule for the PEM of EMI is established. As previously mentioned, in accordance with the spirit of continuous improvement emphasized in TQM, this case adopts the average value of all satisfaction indices $\theta_{S0} = 0.865$ as the testing benchmark for satisfaction indices.

Then, the null hypothesis and alternative hypothesis can be shown as follows:

null hypothesis $H_0: \theta_{Sj} \geq 0.865$ versus alternative hypothesis $H_1: \theta_{Sj} < 0.865$

Next, the statistical test is conducted using the upper confidence limit of the satisfaction index $U\theta_{Sj}$. Accordingly, the statistical testing rule for this case is defined as follows:

1. When the upper confidence limit of the satisfaction index is up to $U\theta_{Sj} \geq 0.865$, then do not reject null hypothesis H_0 and conclude that $\theta_{Sj} \geq 0.865$ and conclude that $\theta_{Sj} \geq 0.865$.
2. When the upper confidence limit of the satisfaction index is down to $U\theta_{Sj} < 0.865$, then reject null hypothesis H_0 and conclude that $\theta_{Sj} < 0.865$.

Then, the null hypothesis and alternative hypothesis can be shown as follows: Like the satisfaction index, this study adopts $\theta_{I0} = 0.871$ as the testing criterion for all importance indices in this case. Then, the null hypothesis and alternative hypothesis can be shown as follows:

null hypothesis $H'_0: \theta_{Ij} \geq 0.871$ versus alternative hypothesis $H'_1: \theta_{Ij} < 0.871$

Subsequently, based on the upper confidence limit of the importance index $U\theta_{Ij}$ for the *j*th item, the hypothesis testing rule for this case is established as follows:

1. When the upper confidence limit of the importance index $U\theta_{Ij} \geq 0.871$, then do not reject null hypothesis H'_0 and conclude that $\theta_{Ij} \geq 0.871$.
2. When the upper confidence limit of the importance index $U\theta_{Ij} < 0.871$, then reject null hypothesis H'_0 and conclude that $\theta_{Ij} < 0.871$.

Next, based on the testing rules and the PEM of EMI established in Section 3, the evaluation and decision-making outcomes of the above case are summarized as follows:

1. The evaluation coordinate points (x_1, y_1), (x_2, y_2), (x_4, y_4), (x_5, y_5), (x_6, y_6), (x_7, y_7), (x_8, y_8), (x_9, y_9), (x_{10}, y_{10}), (x_{11}, y_{11}), (x_{12}, y_{12}), $(x_{14}, y_{14}) \in Q_I \cup Q_{IV}$ indicate that the upper confidence limits of the satisfaction indices for teaching items 1, 2, 4, 5, 6, 7, 8, 9, 10, 11, 12, and 14 are all greater than 0.865. Therefore, no improvement is required for these items.
2. The evaluation coordinate points (x_3, y_3), (x_{13}, y_{13}), (x_{15}, y_{15}), $(x_{16}, y_{16}) \in Q_{II} \cup Q_{III}$ indicate that the upper confidence limits of the satisfaction indices for teaching items 3, 13, 15, and 16 are all less than 0.865. Therefore, these items require improvement.

3. The evaluation coordinate points (x_{15}, y_{15}), $(x_{16}, y_{16}) \in Q_{II}$ indicate that the upper confidence limits of the importance indices for teaching items 15 and 16 are greater than 0.871, suggesting that these items should be prioritized for improvement.
4. The evaluation coordinate points (x_3, y_3), $(x_{13}, y_{13}) \in Q_{III}$ indicate that the upper confidence limits of the importance indices for teaching items 3 and 13 are less than 0.871, suggesting that these items have a lower priority for improvement.

In TQM, continuous improvement is employed to enhance satisfaction, and the PDCA (Plan–Do–Check–Act) cycle serves as a core managerial tool for implementing such improvement processes (Deming, 1986; Oakland, 2014). Figure 2 and Table 1 summarize the questionnaire analysis results and key findings, fulfilling the "Check" function of the PDCA cycle. The subsequent "Act" phase involves taking appropriate actions according to the quadrant in which each teaching item falls. From a TQM perspective, the recommended managerial responses for items located in the four quadrants are as follows (Martilla & James, 1977; Slack et al., 2013):

Quadrant 1. High satisfaction and high importance — Maintain performance.

Teaching items 2, 4, 5, 6, 7, 8, 9, 10, 11, 12, and 14 fall into this quadrant and should be continuously maintained. If resources are limited, a portion of the resources allocated to Quadrants 3 or 4 may be redirected to this quadrant to support ongoing maintenance efforts.

Quadrant 2. Low satisfaction and high importance — Focus attention.

Teaching items 15 and 16 fall into this quadrant and require focused improvement. If resources are insufficient, some of those allocated to Quadrants 4 may be reallocated to this quadrant to support improvement actions.

Quadrant 3. Low satisfaction and low importance — Lower priority.

Teaching items 3 and 13 fall into this quadrant. Due to their low importance, improvements are of lower priority. If resources are limited, a portion of those allocated to this quadrant may be redirected to Quadrants 2 to support maintenance or improvement efforts.

Quadrant 4. High satisfaction and low importance — Potential overinvestment.

Teaching items 1 fall into this quadrant. Achieving high satisfaction on items of low importance suggests possible overuse of resources. When resources are constrained, it is advisable to moderately reduce investment in this quadrant and reallocate part of the resources to Quadrants 1 or 2 to support maintenance or improvement activities.

5. Conclusions

This study proposes a systematic performance evaluation model tailored for EMI courses, with a focus on data-driven instructional improvement. Conceptualizing EMI courses as service operation systems, the model assists in evaluating and improving student satisfaction across five key instructional dimensions: Teaching Preparation, Teaching Attitude, Teaching Capability, Teaching Management, and Coursework and Evaluation. Each EMI course is assumed to consist of q instructional items, for which student feedback is collected via a structured questionnaire measuring both satisfaction and importance. Based on these responses, the model estimates two core metrics—the Satisfaction Index and the Importance Index—using Beta distributions.

To visualize and analyze these metrics, a PEM is constructed, with the satisfaction index on the horizontal axis and the importance index on the vertical axis. Following the principles of TQM, the matrix is divided into four quadrants using the average values of the two indices as reference lines. To strengthen statistical rigor, the study further derives joint upper confidence limits for each item and conducts hypothesis testing to determine whether an instructional item meets the expected quality benchmarks.

A critical feature of this model is its prioritization mechanism, which enables the identification of critical-to-quality (CTQ) instructional items that receive low satisfaction ratings but are considered highly important by students. This mechanism provides practical guidance for instructors, particularly when resources are limited, by highlighting which items should be prioritized for improvement.

Overall, the findings provide initial empirical support for the proposed PEM-based evaluation model and illustrate its practical utility for analyzing EMI course quality. The case study demonstrates how satisfaction and importance indices can be summarized jointly and translated into concrete improvement priorities, helping instructors move from intuitive impressions to transparent, data-driven decisions.

The model offers a transparent, statistically grounded framework and a set of actionable decision rules for continuous quality enhancement in EMI instruction. It enables instructors and program administrators to identify CTQ instructional items, prioritize improvement efforts under resource constraints, and document the rationale for those decisions in a systematic way. Thus, the PEM framework functions both as an evaluative instrument that structures student feedback and as a strategic planning aid that can be integrated into broader course review or TQM-oriented improvement processes.

Future research could extend the PEM framework to other disciplinary and institutional contexts, including EMI courses in STEM and humanities programs, to examine its robustness across different learner profiles and teaching environments. It would also be valuable to incorporate external validation or sensitivity analyses—for example, by relating PEM-based classifications to independent measures of learning outcomes or by testing the impact of alternative benchmark choices.

References

Anghel, B., Cabrales, A., & Carro, J. M. (2016). Evaluating a bilingual education program in spain: The impact beyond foreign language learning. *Economic Inquiry, 54*(2), 1202-1223. https://doi.org/10.1111/ecin.12305

Chen, K. S., Hsieh, T. H., Chang, C. P., Yao, K. C., & Huang, T. H. (2024). Fuzzy decision-making and resource management model of performance evaluation indices. *Axioms, 13*(3), 198. https://doi.org/10.3390/axioms13030198

Chen, K. S., Lii, P. C., Li, S. Y., & Yu, C. M. (2020). Development and application of a performance evaluation matrix: A case study on exploring the items considered critical to quality. *Journal of Testing and Evaluation, 48*(5), 3468-3478. https://doi.org/10.1520/JTE20180166

Chen, K. S., & Yu, C. M. (2020). Fuzzy test model for performance evaluation matrix of service operating systems. *Computers and Industrial Engineering, 140*, 106240. https://doi.org/10.1016/j.cie.2019.106240

Chen, K. S., Yu, C. M., Yu, C. H., & Chen, Y. P. (2024). Performance management decision-making model: Case study on foreign language learning curriculums. *Information, 15*(8), 481. https://doi.org/10.3390/info15080481

Dafouz, E., & Smit, U. (2016). Towards a dynamic conceptual framework for english-medium education in multilingual university settings. *Applied Linguistics, 37*(3), 397-415. https://doi.org/10.1093/applin/amu034

Deming, W. E. (1986). Out of the crisis. MIT Press.

DeVellis, R. F. (2023). Scale development: Theory and applications, 5th ed.; Sage: Newbury Park, CA, USA.

Green, A., & Inoue, C. (2024). *Relating the BESTEP speaking test to the common European framework of reference for languages.* The Language Training & Testing Center, Taipei, Taiwan.

Hogg, R.V., & Tanis, E. A. (2021). *Probability and statistical inference*, 9th ed.; Prentice-Hall: New Jersey, USA.

Jeng, M. Y., Yeh, T. M., & Pai, F. Y. (2022). A performance evaluation matrix for measuring the life satisfaction of older adults using eHealth wearables. *Healthcare, 10*(4), 605. https://doi.org/10.3390/healthcare10040605

Kaiypova, F., Lee, H., Lo, Y. Y., & Lee, J. H. (2025). Effects of content and language integrated learning on secondary-level students' content learning: A meta-analysis. *System, 129*, 103580. https://doi.org/10.1016/j.system.2024.103580

Kim, J., Kim, E., & Hong, A. (2022). OTT streaming distribution strategies for dance performances in the post-covid-19 age: A modified importance-performance analysis. *International Journal of Environmental Research and Public Health, 19*(1), 327. https://doi.org/10.3390/ijerph19010327

Li, C.-H., Chang, H.-T., & Yu, C.-M. (2019).A statistical test method to establish performance evaluation matrix in service operations. *Journal of Testing and Evaluation, 47*(2), 1021–1035. https://doi.org/10.1520/JTE20170456

Li, M., & Pei, L. (2024). Exploring challenges in academic language-related skills of EFL learners in Chinese EMI settings. *Acta Psychologica, 247*, 104309. https://doi.org/10.1016/j.actpsy.2024.104309

Li, Y., Wang, L., & Li, F. (2021). A data-driven prediction approach for sports team performance and its application to National Basketball Association. *Omega, 98*, 102123. https://doi.org/10.1016/j.omega.2019.102123

Lin, C. H., Lin, I. H., Lin, S. Y., & Cheng, C. F. (2024). Students' learning experiences in bilingual Courses: The case of National Sun Yat-sen University. *Contemporary Educational Research Quarterly,* *32*(3), 35-70. https://doi.org/10.6151/CERQ.202409_32(3).0002

Lin, T. C., Chen, H. H., Chen, K. S., Chen, Y. P., & Chang, S. H. (2023). Decision-making model of performance evaluation matrix based on upper confidence limits. *Mathematics, 11*(16), 3499. https://doi.org/10.3390/math11163499

Liu, C. C., Yu, C. H., & Chen, K. S. (2023). Using statistical test method to establish a decision model of performance evaluation matrix. *Applied Sciences, 13*(8), 5139. https://doi.org/10.3390/app13085139

Martilla, J. A., & James, J. C. (1977). Importance–performance analysis. *Journal of Marketing, 41*(1), 77–79.

Nguyen, H. T., Trinh, L. Q., & Vo, T. D. (2025). Examining international student satisfaction: Insights from a medical English-medium instruction program in Vietnam. *Educational Process International Journal, 14*, e2025048. https://doi.org/10.22521/edupij.2025.14.48

Oakland, J. S. (2014). *Total quality management and operational excellence: Text with cases* (4th ed.). Routledge.

Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1988). SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing, 64*(1), 12–40.

Patterson, A. S., Brotherhood, T., & Thompson, G. (2025). Fun, money, and mom: Japanese EMI students' desire to study abroad. *System, 130*, 103619. https://doi.org/10.1016/j.system.2025.103619

Slack, N., Brandon-Jones, A., & Johnston, R. (2013). *Operations management* (7th ed.). Pearson.

Wu, J., & Lee, M. C. L. (2017). The relationships between test performance and students' perceptions of learning motivation, test value, and test anxiety in the context of the English benchmark requirement
for graduation in Taiwan's universities. *Language Testing in Asia, 7*, 9. https://doi.org/10.1186/s40468-017-0041-4

Yeh, L. C., Tung, C. C., Yang, S. Y., Chen, J. H., & Shiau, F. H. (2005). The development of the university teacher instructional evaluation scale. *Psychological Testing, 52*(1), 59-81.

Yu, C. M., Chang, H. T., & Chen, K. S. (2018). Developing a performance evaluation matrix to enhance the learner satisfaction of an e-learning system. T*otal Quality Management and Business Excellence, 29*(7-8), 727-745. https://doi.org/10.1080/14783363.2016.1233809

Yu, C. M., Chang, H. T., Hsu, S. Y. (2017). An assessment of quality and quantity for foreign language training course to enhance students' learning effectiveness. *International Journal of Information and Management Sciences, 28*(1), 53-66. https://doi.org/10.6186/IJIMS.2016.28.1.5

Zhou, S., Xu, J., Rose, H., & McKinley, J. (2025). Exploring the emotion-motivation nexus in English-medium higher education: A longitudinal mixed-methods study. *Learning and Individual Differences, 119*, 102628. https://doi.org/10.1016/j.lindif.2025.102628

Shao-Yu Li
Language Center, Chaoyang University of Technology, Taichung, Taiwan.
E-mail address: izzy@cyut.edu.tw
Major area(s): EMI pedagogy and quality assurance, ESP/EAP curriculum design, data-driven evaluation of educational service quality.

Hung-Tso Lin
Department of Marketing and Distribution Management, National Chin-Yi University of Technology, Taichung, Taiwan.
E-mail address: htl@ncut.edu.tw
Major area(s): Operations management, performance management, customer relationship management

Chun-Hung Yu
Language Center, National Chin-Yi University of Technology, Taichung, Taiwan.
Language Teaching Center, National Chi Nan University, Nantou County, Taiwan.
E-mail address: chyu@ncnu.edu.tw
Major area(s): Mandarin instruction, curriculum development, performance evaluation and English teaching effectiveness evaluation

(Received September 2025; accepted January 2026)