

Article

Using Statistical Test Method to Establish a Decision Model of Performance Evaluation Matrix

Chin-Chia Liu ¹, Chun-Hung Yu ^{1,2} and Kuen-Suan Chen ^{3,4,5,*}¹ Department of Industrial Education and Technology, National Changhua University of Education, Changhua 500208, Taiwan² Language Center, National Chin-Yi University of Technology, Taichung 411030, Taiwan³ Department of Industrial Engineering and Management, National Chin-Yi University of Technology, Taichung 411030, Taiwan⁴ Department of Business Administration, Chaoyang University of Technology, Taichung 413310, Taiwan⁵ Department of Business Administration, Asia University, Taichung 413305, Taiwan

* Correspondence: kschen@ncut.edu.tw

Abstract: Many studies have pointed out that the Performance Evaluation Matrix (PEM) is a convenient and useful tool for the evaluation, analysis, and improvement of service operating systems. All service items of the operating system can collect customer satisfaction and importance through questionnaires and then convert them into satisfaction indices and importance indices to establish PEM and its evaluation rules. Since the indices have unknown parameters, if the evaluation is performed directly by the point estimates of the indices, there will be a risk of misjudgment due to sampling error. In addition, most of the studies only determine the critical-to-quality (CTQ) that needs to be improved, and do not discuss the treatment rules in the case of limited resources nor perform the confirmation after improvement. Therefore, to address similar research gaps, this paper proposed the unbiased estimators of these two indices and determined the critical-to-quality (CTQ) service items which need to be improved through the one-tailed statistical hypothesis test by building a PEM method of the satisfaction index. In addition, through the one-tailed statistical hypothesis test method of the importance index, the improvement priority of service items was determined under the condition of limited resources. Confirmation of the effect on improvement is an important step in management. Thus, this paper adopted a statistical two-tailed hypothesis test to verify whether the satisfaction of all the CTQ service items that need to be improved was enhanced. Since the method proposed in this paper was established through statistical hypothesis tests, the risk of misjudgment due to sampling error could be reduced. Obviously, reducing the misjudgment risk is the advantage of the method in this paper. Based on the precondition, utilizing the model in this study may assist the industries to determine CTQ rapidly, implement the most efficient improvement under the condition of limited resources and also confirm the improvement effect at the same time. Finally, a case study of computer-assisted language learning system (CALL System) was used to illustrate a way to apply the model proposed in this paper.

Keywords: performance evaluation matrix; satisfaction index; importance index; unbiased estimator; statistical hypothesis test



Citation: Liu, C.-C.; Yu, C.-H.; Chen, K.-S. Using Statistical Test Method to Establish a Decision Model of Performance Evaluation Matrix. *Appl. Sci.* **2023**, *13*, 5139. <https://doi.org/10.3390/app13085139>

Academic Editor: Radu Godina

Received: 4 March 2023

Revised: 14 April 2023

Accepted: 17 April 2023

Published: 20 April 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In the face of global warming and the impact of the COVID-19 pandemic in recent years, human coexistence with the natural environment has become a relatively important issue. The awareness of sustainable development that enterprises and environmental protection have in common has also become the major concern for governments and enterprises in various countries. According to many studies, as the Internet of Things (IoT) has gradually become popular and mature, various smart APP platforms have also sprung up [1–4]. Not only can efficiency be leveled up, but the impact on the environment

can also be reduced. For example, by means of the digital learning model of the teaching APP, students can eliminate the inconvenience of going out as well as achieve the effect of learning at home. By doing so, not only can the amount of traffic jams be reduced, but the time of going out can also be saved, so that the effect of energy saving and carbon reduction can be achieved. Based on the above scenario, constructing a complete evaluation and improvement model of the digital learning effectiveness of the teaching APP will help attract more users to achieve digital learning effect by means of the teaching APP in order to achieve the above-mentioned effect of alleviating the congestion problem, saving energy, and reducing carbon emissions.

Some methods such as Quality Function Deployment (QFD) and Balanced Scorecard (BSC) are employed to weigh performance evaluations or product designs, and some apply Fuzzy inference system (FIS) to increase the competitive capability and customer satisfaction [5–7]. According to numerous studies, the Performance Evaluation Matrix (PEM) is a convenient and useful evaluation tool for the service operating systems. Thus, plenty of dissertations have engaged in the research on PEM [8–11]. In order not to lose generality, this paper assumes that there are q key service items of digital learning provided by the teaching APP, so q questions are designed to investigate the importance and satisfaction of the q service items for learners. In addition, many studies assert that since Beta assigns values between 0 and 1, it represents satisfaction ranging from 0, which is completely dissatisfied, to 1, which is 100% satisfied. Similarly, the importance ranges from 0, which means total ignorance, to 1, which refers to great importance. Consequently, satisfaction and importance are regarded as random variables subject to the Beta distribution [12,13]. Hung, Huang, and Chen [14] proposed a standardized and significant importance index and satisfaction index based on the two parameters assigned by Beta distribution. If random variable X_i represents the satisfaction of the i th service item, then X_i is distributed as the Beta distribution denoted by $X_i \sim \text{Beta}(a_i^S, b_i^S)$, where a_i^S and b_i^S are two parameters. Similarly, if random variable Y_i indicates the importance attached to the i th service item, then Y_i is distributed as the Beta distribution denoted by $Y_i \sim \text{Beta}(a_i^I, b_i^I)$, where a_i^I and b_i^I are two parameters. Thus, the satisfaction index and the importance index can be expressed, as follows, for satisfaction:

$$\theta_i^S = \frac{a_i^S}{a_i^S + b_i^S} (\text{satisfaction index}), \quad (1)$$

$$\theta_i^I = \frac{a_i^I}{a_i^I + b_i^I} (\text{importance index}). \quad (2)$$

According to the abovementioned data, the values of the two indices are between 0 and 1. In this paper, we set the satisfaction index as the x -axis and set the importance index as the y -axis to form the performance evaluation matrix. Since the indices have unknown parameters, if the evaluation is performed directly by the point estimates of the indices, there will be a risk of misjudgment due to sampling error [15–17]. In addition, most studies just determine the critical-to-quality (CTQ) that needs to be improved, without discussing the processing rules in resource-limited settings, and without the confirmation of the improvements. In order to make up for such research gaps, this paper adopts the one-tailed statistical hypothesis testing method of satisfaction index, determines the CTQ service items that need to be improved, and then defines their improvement order. Then, this paper uses a statistical two-tailed hypothesis test to verify whether the satisfaction of all CTQ service items that need to be improved is improved. First, this paper detects the unbiased estimators of the satisfaction index and the importance index for all service items. However, according to numerous studies, if the PEM locations of the point estimates are directly used to determine whether the service items need to be improved, then the risk of misjudgment may be caused by sampling error [18,19]. Therefore, this paper uses the average value θ_0^S of all satisfaction indices as the testing standard in accordance with the spirit of continuous improvement on the comprehensive quality management and

employs the statistical one-tailed test to determine the critical value C_0^S . At the same time, the average value θ_0^I of all importance indices is adopted as the testing standard, and the statistical one-tailed test is employed to determine the critical value. Next, this paper establishes the evaluation blocks and evaluation rules of PEM based on the critical value derived from the statistical test and judges whether the service item needs to be improved according to the location of the evaluation coordinate point $(x_i, y_i) = (\hat{\theta}_i^S, \hat{\theta}_i^I)$ of each service item, where $\hat{\theta}_i^S$ is the estimator of θ_i^S and $\hat{\theta}_i^I$ is the estimator of θ_i^I . Meanwhile, according to the locations of the evaluation coordinate points of the service items that need to be improved in the evaluation quadrant of PEM, the priority of improvement is determined under the condition of limited resources. Because this method is established by statistical tests, the risk of misjudgment caused by sampling error can be lowered. After the task of improvement is completed, the effectiveness of the improvement must be confirmed. Therefore, this paper then adopts the statistical two-tailed test to verify whether the satisfaction of all service items that need to be improved is leveled up. Finally, a case study of the computer-assisted language learning system (CALL System) is used in this paper to illustrate the application of the abovementioned statistical test rules to define the key service items that need to be improved and to determine the priority of improvement in the situation of limited resources.

The remainder of this paper is arranged as follows. In Section 2, we propose the unbiased estimators and the statistical hypothesis test method for the satisfaction index and the importance index, respectively. According to the statistical test results, we procure the service items that need to be improved as well as determine the order of improvement for all service items. In Section 3, we define the evaluation block of the Performance Evaluation Matrix based on the two critical values derived from the statistical test conducted in Section 2; at the same time, we also define evaluation rules. In Section 4, we adopt a statistical two-tailed test to verify the improvement effects on all key service items that need to be improved. In Section 5, we design a case study of the computer-assisted language learning system (CALL System), illustrating the application of the method proposed in this paper. In Section 6, we offer conclusions.

2. Unbiased Estimators and Statistical Hypothesis Test

As mentioned earlier, since the indices have unknown parameters, they must be estimated by sample data of the interviewed customers. Due to the application of the Central Limit Theorem in this paper to carry out the statistical test, the sample size must be sufficiently large. The sample size of the cases applied to the method in this study are usually larger than 100. They are much larger than the minimum sample size required in the Central Limit Theorem (the minimum sample size for requirements is 30). In this section, we propose the unbiased estimators and the statistical hypothesis test method for the satisfaction index and the importance index, respectively.

2.1. Satisfaction Index

First, we assume that the satisfaction sample data of the customer's i th service item is $(X_{i,1}, \dots, X_{i,j}, \dots, X_{i,n})$, $i = 1, 2, \dots, q$. Then, the estimator of the satisfaction index for service item i is expressed as follows:

$$\hat{\theta}_i^S = \frac{1}{n} \times \sum_{j=1}^n X_{i,j}. \quad (3)$$

The expected value of $\hat{\theta}_i^S$ is

$$\begin{aligned} E[\hat{\theta}_i^S] &= \frac{1}{n} \times \sum_{j=1}^n E[X_{i,j}] \\ &= \frac{1}{n} \times \sum_{j=1}^n \int_0^1 x \frac{\Gamma(a_i^S + b_i^S)}{\Gamma(a_i^S) \Gamma(b_i^S)} x^{a_i^S} (1-x)^{b_i^S} dx \\ &= \frac{1}{n} \times \sum_{j=1}^n \left(\frac{a_i^S}{a_i^S + b_i^S} \right) \\ &= \theta_i^S. \end{aligned} \quad (4)$$

Thus, $\hat{\theta}_i^S$ is an unbiased estimator of θ_i^S .

$$\theta_S = \frac{1}{q} \sum_{i=1}^q \hat{\theta}_i^S. \quad (5)$$

Then, based on the concept of total quality management, when the satisfaction index value of service item i is greater than or equal to the average value θ_S , it is fine to maintain the current situation. When the satisfaction index value of service item i is smaller than the average value θ_S , then the service item i needs to be improved [20,21]. Based on the stated above, the hypotheses of the statistical test for satisfaction index i are as follows:

$$\text{null hypothesis } H_0 : \theta_i^S \geq \theta_S; \quad (6)$$

$$\text{alternative hypothesis } H_1 : \theta_i^S < \theta_S. \quad (7)$$

We assume that the test statistic is the unbiased estimator $\hat{\theta}_i^S$ with significance level α ; then, the critical region for satisfaction index i is

$$CR_{Xi} = \left\{ \hat{\theta}_i^S \mid \hat{\theta}_i^S < C_{X0i} \right\}. \quad (8)$$

Thus, we have

$$p\left(\hat{\theta}_i^S < C_{X0i} \mid \theta_i^S = \theta_S\right) = \alpha. \quad (9)$$

Equivalently,

$$p\left(Z_{Xi} < \frac{\sqrt{n}(C_{X0i} - \theta_S)}{S_{Xi}}\right) = \alpha, \quad (10)$$

where

$$S_{Xi} = \sqrt{\frac{1}{n-1} \times \sum_{j=1}^n (X_{i,j} - \bar{X}_i)^2} \quad (11)$$

and

$$Z_{Xi} = \frac{\sqrt{n}(\theta_i^S - \theta_S)}{S_{Xi}} \quad (12)$$

are approximate values, and they are distributed as standard normal distributions when sample size n is large. Then, we have

$$\frac{\sqrt{n}(C_{X0i} - \theta_S)}{S_{Xi}} = -z_\alpha,$$

where the critical value C_{X0i} is displayed below:

$$C_{X0i} = \theta_S - z_\alpha \frac{S_{Xi}}{\sqrt{n}}. \quad (13)$$

2.2. Importance Index

We assume the importance sample data of the customer's i th service item is $(Y_{i,1}, \dots, Y_{i,j}, \dots, Y_{i,n})$, $i = 1, 2, \dots, q$; then, the estimator of importance index for service item i is expressed as follows:

$$\hat{\theta}_i^I = \frac{1}{n} \times \sum_{j=1}^n Y_{i,j}. \quad (14)$$

The expected value of $\hat{\theta}_i^I$ is

$$\begin{aligned}
 E[\hat{\theta}_i^I] &= \frac{1}{n} \times \sum_{j=1}^n E[Y_{i,j}] \\
 &= \frac{1}{n} \times \sum_{j=1}^n \int_0^1 y \frac{\Gamma(a_i^I + b_i^I)}{\Gamma(a_i^I) \Gamma(b_i^I)} y^{a_i^I} (1-y)^{b_i^I} dy \\
 &= \frac{1}{n} \times \sum_{i=1}^n \frac{\Gamma(a_i^I + b_i^I) \Gamma(a_i^I + 1)}{\Gamma(a_i^I + b_i^I + 1)} \times \int_0^1 \text{Beta}(a_i^I + 1, b_i^I) dy \\
 &= \frac{1}{n} \times \sum_{i=1}^n \left(\frac{a_i^I}{a_i^I + b_i^I} \right) \\
 &= \theta_i^I.
 \end{aligned} \tag{15}$$

Thus, $\hat{\theta}_i^I$ is an unbiased estimator of θ_i^I .

$$\theta_I = \frac{1}{q} \sum_{i=1}^q \hat{\theta}_i^I. \tag{16}$$

Then, when the importance index value of service item i is greater than or equal to the average value θ_I , the priority of improvement is high. When the importance index value of service item i is less than the average value θ_I , the priority of improvement is low. Based on the above statements, the hypotheses of the statistical test for importance index i are defined as follows:

$$\text{null hypothesis } H_0 : \theta_i^I \geq \theta_I; \tag{17}$$

$$\text{alternative hypothesis } H_1 : \theta_i^I < \theta_I. \tag{18}$$

We assume the test statistic is the unbiased estimator $\hat{\theta}_i^I$ with significance level α ; then, the critical region for importance index i is

$$CR_{Yi} = \left\{ \hat{\theta}_i^I \mid \hat{\theta}_i^I < C_{Y0i} \right\}. \tag{19}$$

Thus, we have

$$p\left(\hat{\theta}_i^I < C_{Y0i} \mid \theta_i^I = \theta_I\right) = \alpha. \tag{20}$$

Equivalently,

$$p\left(Z_{Yi} < \frac{\sqrt{n}(C_{Y0i} - \theta_I)}{S_{Yi}}\right) = \alpha, \tag{21}$$

where

$$S_{Yi} = \sqrt{\frac{1}{n-1} \times \sum_{j=1}^n (Y_{i,j} - \bar{Y}_i)^2} \tag{22}$$

and

$$Z_{Yi} = \frac{\sqrt{n}(\hat{\theta}_i^I - \theta_I)}{S_{Yi}} \tag{23}$$

are approximate values, and they are distributed as standard normal distributions when sample size n is large. Then,

$$C_{Y0i} = \theta_I - z_\alpha \frac{S_{Yi}}{\sqrt{n}}. \tag{24}$$

3. Performance Evaluation Matrix

As mentioned earlier, the performance evaluation matrix is widely used to evaluate and improve the performance levels of q service items for various service systems. In this paper, the satisfaction index is set as x -axis and the importance index is set as y -axis to form the performance evaluation matrix. Then, the horizontal line $y = C_{Y0}$ and the vertical

line $x = C_{X0}$ are used to divide the performance evaluation matrix into four quadrants, where $C_{X0} = \text{Min}\{C_{X01}, C_{X02}, \dots, C_{X0q}\}$ and $C_{Y0} = \text{Min}\{C_{Y01}, C_{Y02}, \dots, C_{Y0q}\}$. These four quadrants can be defined as follows:

$$\text{Quadrant 1: } Q_1 = \{(x, y) | C_{X0} \leq x \leq 1, C_{Y0} \leq y \leq 1\}, \quad (25)$$

$$\text{Quadrant 2: } Q_2 = \{(x, y) | 0 \leq x < C_{X0}, C_{Y0} \leq y \leq 1\}, \quad (26)$$

$$\text{Quadrant 3: } Q_3 = \{(x, y) | 0 \leq x < C_{X0}, 0 \leq y < C_{Y0}\}, \quad (27)$$

$$\text{Quadrant 4: } Q_4 = \{(x, y) | C_{X0} \leq x \leq 1, 0 \leq y < C_{Y0}\}. \quad (28)$$

Then, the performance evaluation matrix and four quadrants can be displayed in Figure 1 as follows:

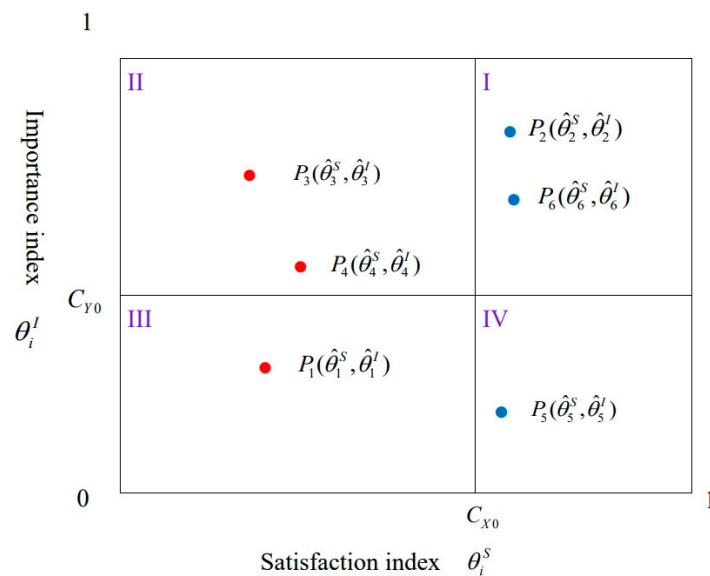


Figure 1. Performance evaluation matrix and four quadrants.

Obviously, when the evaluation coordinate points fall into Quadrant 2 and Quadrant 3, it means that the satisfaction of the service items is lower than the average satisfaction, so that the service items definitely need to be improved. Since the evaluation coordinate points falling into Quadrant 2 are more important than the ones falling into Quadrant 3, the priority of their improvement is higher. In addition, when the evaluation coordinate points fall into Quadrant 1 and Quadrant 4, it means that the satisfaction of the service items is higher than the average satisfaction, so that there is no need to make any improvement, and the current state can remain unchanged. Based on the above concept, we can define the evaluation coordinates for the i th service item as follows:

$$(x_i, y_i) = (\hat{\theta}_i^S, \hat{\theta}_i^I). \quad (29)$$

Then, the evaluation rules and improvement ranking rules are established as follows:

- (1) If $(x_i, y_i) \in Q_1 \cup Q_4$, then service item i does not need to make any improvement.
- (2) If $(x_i, y_i) \in Q_2 \cup Q_3$, then service item i needs to be improved.
- (3) If $(x_i, y_i) \in Q_2$, then the service item i is ranked the first priority of improvement.
- (4) If $(x_i, y_i) \in Q_3$, then the service item i is ranked the second priority of improvement.

4. Statistical Two-Tailed Test

As noted above, confirming the effect upon the improvement is an important step in management. Therefore, we use a statistical two-tailed test to judge whether the satisfaction of all the CTQ service items needs to be enhanced in this section. We let set A be the set formed by the service items that need to be enhanced; then,

$$A = \{i | (x_i, y_i) \in Q_2 \cap Q_3\}. \quad (30)$$

θ_{bi}^S is the value of satisfaction index before improvement, θ_{ai}^S is the value of satisfaction index after improvement, and $i \in A$. If the satisfaction sample data are $(X'_{i,1}, \dots, X'_{i,j}, \dots, X'_{i,n})$ for $i \in A$, then the estimator of θ_{ai}^S can be shown as follows:

$$\hat{\theta}_{ai}^S = \frac{1}{n} \times \sum_{j=1}^n X'_{i,j}. \quad (31)$$

Similarly, the estimator of θ_{bi}^S for $i \in A$ as follows:

$$\hat{\theta}_{bi}^S = \frac{1}{n} \times \sum_{j=1}^n X_{i,j}. \quad (32)$$

Then,

$$Z_{hi} = \frac{\sqrt{n}(\hat{\theta}_{hi}^S - \theta_{hi}^S)}{S_{Xhi}}, \quad h = a, b \text{ and } i \in A, \quad (33)$$

where

$$S_{Xai} = \sqrt{\frac{1}{n-1} \times \sum_{j=1}^n (X'_{i,j} - \bar{X}'_i)^2} \text{ and } S_{Xbi} = \sqrt{\frac{1}{n-1} \times \sum_{j=1}^n (X_{i,j} - \bar{X}_i)^2}. \quad (34)$$

Then, Z_{hi} is an approximate value, and it is distributed as a standard normal distribution when sample size n is large. Thus,

$$\begin{aligned} 1 - \alpha &= p(-Z_{\alpha/2} \leq Z_{hi} \leq Z_{\alpha/2}) \\ &= p\left(-Z_{\alpha/2} \leq \frac{\sqrt{n}(\hat{\theta}_{hi}^S - \theta_{hi}^S)}{S_{Xhi}} \leq Z_{\alpha/2}\right) \\ &= p\left(\hat{\theta}_{hi}^S - Z_{\alpha/2} \frac{S_{Xhi}}{\sqrt{n}} \leq \theta_{hi}^S \leq \hat{\theta}_{hi}^S + Z_{\alpha/2} \frac{S_{Xhi}}{\sqrt{n}}\right). \end{aligned} \quad (35)$$

Then, the $100(1 - \alpha)\%$ confidence interval of θ_{bi}^S and θ_{ai}^S can be demonstrated as follows:

$$\text{before improvement : } [L\theta_{bi}, U\theta_{bi}] = \left[\hat{\theta}_{bi}^S - Z_{\alpha/2} \frac{S_{Xbi}}{\sqrt{n}}, \hat{\theta}_{bi}^S + Z_{\alpha/2} \frac{S_{Xbi}}{\sqrt{n}}\right]; \quad (36)$$

$$\text{after improvement : } [L\theta_{ai}, U\theta_{ai}] = \left[\hat{\theta}_{ai}^S - Z_{\alpha/2} \frac{S_{Xai}}{\sqrt{n}}, \hat{\theta}_{ai}^S + Z_{\alpha/2} \frac{S_{Xai}}{\sqrt{n}}\right]. \quad (37)$$

Obviously, when the value of θ_{ai}^S is bigger than that of θ_{bi}^S , it indicates that the improvement is effective. When θ_{ai}^S is equal to θ_{bi}^S , it means that the improvement has no effect. When the value of θ_{ai}^S is smaller than that of θ_{bi}^S , improvement does not occur; in fact, it is made worse. Accordingly, the statistical hypothesis test for validating the effectiveness of improvement is displayed as follows:

$$\text{null hypothesis } H_0 : \theta_{bi}^S = \theta_{ai}^S; \quad (38)$$

$$\text{alternative hypothesis } H_1 : \theta_{bi}^S \neq \theta_{ai}^S. \quad (39)$$

Based on Equations (36) and (37), we perform the two-tailed statistical test using two confidence intervals. The test rules are defined as follows:

- (1) If $U\theta_{bi} < L\theta_{ai}$, then the improvement of service item i works well.
- (2) If $[L\theta_{bi}, U\theta_{bi}] \cap [L\theta_{ai}, U\theta_{ai}] \neq \emptyset$, then service item i does not improve significantly.
- (3) If $U\theta_{ai} < L\theta_{bi}$, then it means that improvement does not occur; in fact, it is made worse.

Based on the rules of the two-tailed statistical test, we can assist the system administrator of the teaching APP to verify the improvement effect upon all the service items that need to be improved and determine whether the improvements should continue so as to ensure the system quality of the teaching APP.

5. Case Study

As mentioned above, with the performance evaluation matrix, we can help the system administrator of the teaching APP quickly figure out the key service items that need to be improved and implement some improvements. In the field of English digital learning, many scholars have been involved in the research on the computer-assisted language learning system (CALL system) to help students use the CALL system to improve their foreign language [22–26]. The web-based e-learning system (WELS) satisfaction questionnaire designed by Shee and Wang [27], including four dimensions—1. Learner Interface, 2. Learning Community, 3. System Content, and 4. Personalization and Contains—and 13 questions can be said to be simple and complete. Therefore, it has been adopted by plenty of studies [28–31], for example, on the perceptions of students' e-learning experiences in English for specific purpose (ESP) classes [28], examining and categorizing proper conditions for mathematics teaching as a real-life utilization with a multi-criteria decision analysis and fuzzy-decision-making trial and evaluation laboratory (MCDA and F-DEMATEL) method [29], and examining 98 articles to confirm the satisfaction as a critical aspect of student success in online education [30]. This paper also used this questionnaire to collect users' satisfaction and importance and adopted the Call system as a case study to illustrate the model proposed in this study.

This questionnaire survey recruited students from a university in Taiwan as a sample. A total of 398 questionnaires were sent out, and 355 questionnaires were received (the response rate was 89.2%), of which 11 were invalid and 344 were effective (the effective response rate was 86.4%). The following is a data analysis based on the received 344 valid questionnaires ($N = 344$).

First, according to Equations (3) and (14), we calculated the estimator ($\hat{\theta}_i^S$) of the satisfaction index and the estimator ($\hat{\theta}_i^I$) of the importance index of each question, respectively. At the same time, according to Equation (5), we also calculated the average of the estimators of all satisfaction indices, $\theta_S = 0.52$. Then, the hypotheses of the statistical test for satisfaction index i are displayed as follows:

$$\text{null hypothesis } H_0 : \theta_i^S \geq 0.52; \quad (40)$$

$$\text{alternative hypothesis } H_1 : \theta_i^S < 0.52. \quad (41)$$

Subsequently, based on Equation (13) and $\theta_S = 0.52$, we calculated the critical value (C_{X0i}) of the danger zone of the satisfaction index of each question and filled it in Table 1.

Similarly, according to Equation (16), the average of estimators of all the importance indices, $\theta_I = 0.70$, was calculated. The hypotheses of the statistical test for importance index i were defined as follows:

$$\text{null hypothesis } H_0 : \theta_i^I \geq 0.70; \quad (42)$$

$$\text{alternative hypothesis } H_1 : \theta_i^I < 0.70. \quad (43)$$

Table 1. Satisfaction and Importance survey for CALL System.

Dimensions	Items	$\hat{\theta}_i^S$	$\hat{\theta}_i^I$	C_{X0i}	C_{Y0i}	Quadrant
Learner Interface	1. Ease of use	0.65	0.77	0.49	0.69	Q1
	2. User-friendliness	0.70	0.77	0.50	0.69	Q1
	3. Ease of understanding	0.68	0.76	0.50	0.69	Q1
	4. Operational stability	0.59	0.75	0.49	0.69	Q1
Learning Community	5. Ease of discussion with other learners	0.41	0.77	0.49	0.69	Q2 **
	6. Ease of discussion with teachers	0.39	0.77	0.49	0.69	Q2 **
	7. Ease of accessing shared data	0.37	0.63	0.49	0.68	Q3 *
	8. Ease of exchanging learning with the others	0.34	0.57	0.49	0.68	Q3 *
System Content	9. Up-to-date content	0.52	0.67	0.49	0.68	Q4
	10. Sufficient content	0.52	0.67	0.49	0.68	Q4
	11. Useful content	0.53	0.66	0.49	0.68	Q4
Personalization	12. Capability of controlling learning progress	0.51	0.65	0.49	0.68	Q4
	13. Capability of recording learning performance	0.51	0.67	0.49	0.68	Q1

Note: * indicates the need for improvement; ** indicates the need for the top priority of improvement.

Next, according to Equation (24) and $\theta_I = 0.70$, we determined the critical value (C_{Y0i}) of the danger zone of the importance index of each question and filled it in Table 1.

According to Table 1, we used horizontal line $y = C_{Y0}$ and vertical line $x = C_{X0}$ to divide the performance evaluation matrix into four quadrants, where $C_{X0} = \text{Min}\{C_{X01}, C_{X02}, \dots, C_{X0q}\} = 0.49$ and $C_{Y0} = \text{Min}\{C_{Y01}, C_{Y02}, \dots, C_{Y0q}\} = 0.68$. Next, according to the evaluation coordinate points of $(x_i, y_i) = (\hat{\theta}_i^S, \hat{\theta}_i^I)$ for all service items in Table 1, we plotted them in Figure 2, as shown below.

**Figure 2.** PEM of CALL System.

According to Figure 2, the evaluation coordinate points of these 13 service items fall into the four evaluation quadrants as follows:

Quadrant 1: items 1–4;

Quadrant 2: item 5, item 6;

Quadrant 3: item 7, item 8;

Quadrant 4: items 9–13.

According to the evaluation rules, service items 5–8 need to improve. Based on the improvement ranking rules, since evaluation coordinates $(x_5, y_5) \in Q_2$ and $(x_6, y_6) \in Q_2$, service item 5 and service item 6 were ranked to be of the first priority of improvement. Similarly, evaluation coordinates $(x_7, y_7) \in Q_3$ and $(x_8, y_8) \in Q_3$, then service item 7 and service item 8 were ranked to be of the second priority of improvement.

Next, managers would be able to refer to the recommendations of these rules to carry out improvement procedures. After the improvement concludes, a test of the improvement can be carried out. Based on Equations (38) and (39), the assumptions of its statistical test are as follows:

$$\text{null hypothesis } H_0 : \theta_{bi}^S = \theta_{ai}^S; \quad (44)$$

$$\text{alternative hypothesis } H_1 : \theta_{bi}^S \neq \theta_{ai}^S. \quad (45)$$

Based on Equations (36) and (37), the 95% confidence intervals with sample size $n = 258$ of the satisfaction indices of the four service items that need to implement improvements are calculated.

1. Service Item 5

$$\text{before improvement : } [L\theta_{b5}, U\theta_{b5}] = \left[0.41 - 1.96 \frac{0.28}{\sqrt{344}}, 0.41 + 1.96 \frac{0.28}{\sqrt{344}} \right] = [0.38, 0.44]; \quad (46)$$

$$\text{after improvement : } [L\theta_{a5}, U\theta_{a5}] = \left[0.48 - 1.96 \frac{0.27}{\sqrt{258}}, 0.48 + 1.96 \frac{0.27}{\sqrt{258}} \right] = [0.45, 0.51]. \quad (47)$$

2. Service Item 6

$$\text{before improvement : } [L\theta_{b6}, U\theta_{b6}] = \left[0.39 - 1.96 \frac{0.29}{\sqrt{344}}, 0.39 + 1.96 \frac{0.29}{\sqrt{344}} \right] = [0.36, 0.42]; \quad (48)$$

$$\text{after improvement : } [L\theta_{a6}, U\theta_{a6}] = \left[0.47 - 1.96 \frac{0.28}{\sqrt{258}}, 0.47 + 1.96 \frac{0.28}{\sqrt{258}} \right] = [0.43, 0.50]. \quad (49)$$

3. Service Item 7

$$\text{before improvement : } [L\theta_{b7}, U\theta_{b7}] = \left[0.37 - 1.96 \frac{0.29}{\sqrt{344}}, 0.37 + 1.96 \frac{0.29}{\sqrt{344}} \right] = [0.34, 0.40]; \quad (50)$$

$$\text{after improvement : } [L\theta_{a7}, U\theta_{a7}] = \left[0.42 - 1.96 \frac{0.27}{\sqrt{258}}, 0.42 + 1.96 \frac{0.27}{\sqrt{258}} \right] = [0.39, 0.46]. \quad (51)$$

4. Service Item 8

$$\text{before improvement : } [L\theta_{b8}, U\theta_{b8}] = \left[0.34 - 1.96 \frac{0.28}{\sqrt{344}}, 0.34 + 1.96 \frac{0.28}{\sqrt{344}} \right] = [0.31, 0.37]; \quad (52)$$

$$\text{after improvement : } [L\theta_{a8}, U\theta_{a8}] = \left[0.27 - 1.96 \frac{0.24}{\sqrt{258}}, 0.27 + 1.96 \frac{0.24}{\sqrt{258}} \right] = [0.24, 0.30]. \quad (53)$$

In summary, the aforementioned test rules are listed as follows:

- (1) For service items 5 and 6, their upper limits before improvement are smaller than their lower limits after improvement ($U\theta_{b5} = 0.44 < L\theta_{a5} = 0.45$; $U\theta_{b6} = 0.42 < L\theta_{a6} = 0.43$), indicating that the improvements of service items 5 and 6 are effective.
- (2) For service item 7, $[L\theta_{b7}, U\theta_{b7}] = [0.34, 0.40]$ and $[L\theta_{a7}, U\theta_{a7}] = [0.39, 0.46]$, then $[L\theta_{b7}, U\theta_{b7}] \cap [L\theta_{a7}, U\theta_{a7}] \neq \phi$. This shows that the improvement effect of service item 7 is not significant, so continuous improvement is necessary.
- (3) For service item 8, $U\theta_{a8} = 0.30 < L\theta_{b8} = 0.31$; it means that improvement does not occur; in fact, it is made worse. It also indicates that the improvement strategy needs to be re-examined so as to make effective improvements.

6. Conclusions

According to the viewpoints of Hung, Huang, and Chen [14], we proposed standardized and significant indices of importance and satisfaction on the premise that satisfaction and importance were assigned by Beta in this paper, and the values of these two indices were both between 0 and 1. Next, we set the satisfaction index as x -axis and the importance index as y -axis to form a performance evaluation matrix. In order not to lose generality, we assumed that the digital learning service items provided by the teaching APP in this paper had a total of q service items, so there were q satisfaction indices and q importance indices

in total. First, in this paper, we determined the unbiased estimators of the satisfaction index and the importance index for all service items since the indices had unknown parameters. Second, we used the average value θ_0^S of all satisfaction indices as the testing standard in accordance with the spirit of continuous improvement on the comprehensive product management as well as adopted the statistical one-tailed test to determine the critical value C_0^S . Third, we established testing rules to facilitate the management to procure the service items that need to be improved. At the same time, the average value θ_0^I of all the importance indices was adopted as the testing standard, and the ranking of all service items that need to improve was determined by the statistical one-tailed test. Next, the above two statistical testing rules were applied as the basis of defining the PEM evaluation blocks and establishing the PEM evaluation rules. Obviously, since the method in this paper was established via statistical tests, the risk of wrong judgment caused by sampling error can be diminished. After the task of improvement was completed, we employed the statistical two-tailed test to verify whether the satisfaction of all key service items that need to make improvements was leveled up. Finally, a case study of the computer-assisted language learning system (CALL System) was used to explain the application of the abovementioned statistical testing rules.

As noted above, most studies just identify critical qualities (CTQs) that need improvement, without discussing processing rules in resource-limited settings, and without confirmation of improvement. The method of this paper just makes up for the gap in similar research. This paper adopts the one-tailed statistical hypothesis testing method of satisfaction index, which can help the industry to detect the CTQ service items that need to be improved and determine the order of improvement with a low risk of misjudgment. Finally, the two-tailed statistical hypothesis testing method is used to assist the industry to confirm its improvement effect. In addition, the digital learning effect evaluation and improvement model of the teaching APP constructed in this paper will contribute to attracting more users to apply the teaching APP to achieve the effectiveness of digital learning. The APP enables more students to avoid the inconvenience of going out and obtain good learning outcomes at home. Not only can it alleviate the problem of traffic congestion, but it can also achieve the effect of energy saving and carbon reduction by decreasing the need of going out.

Clearly, this paper refers to the literature, closely examines the needs of the customers, and assumes that the distribution of customer satisfaction and importance lies with Beta distribution since Beta distribution belongs to continuous probability distribution; however, data would have to be collected from the highest score of 100 and the lowest score of 0 and then divided by 100 to produce a value between 0 and 1. In addition, the statistical test must be conducted by the Central Limit Theorem. These are the limitations of this study. Therefore, future research can focus on the data collection method of the Beta distribution, and explore the fitting quantity of sample size in the application of the Central Limit Theorem.

Author Contributions: Conceptualization, C.-C.L. and K.-S.C.; methodology, C.-C.L. and K.-S.C.; software, C.-H.Y.; validation, C.-H.Y.; formal analysis, C.-C.L. and K.-S.C.; resources, C.-H.Y.; data curation, C.-H.Y.; writing—original draft preparation, C.-C.L., C.-H.Y. and K.-S.C.; writing—review and editing, C.-C.L. and K.-S.C. visualization, C.-H.Y.; supervision, K.-S.C.; project administration, C.-H.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data underlying the results are available as part of the article and no additional source data are required.

Conflicts of Interest: The authors declare no conflict of interest.

Nomenclature

q	the number of service items
X_i	the satisfaction of the i th service item
Y_i	the importance attached to the i th service item
a_i^S, b_i^S	parameter of Beta distribution for satisfaction
a_i^I, b_i^I	parameter of Beta distribution for importance
θ_i^S	satisfaction index
$\hat{\theta}_i^S$	the estimator of θ_i^S .
θ_i^I	importance index
$\hat{\theta}_i^I$	the estimator of θ_i^I
θ_0^S	the average value of all satisfaction indices
C_0^S	critical value
θ_I	the average value of $\hat{\theta}_i^I$
$(x_i, y_i) = (\hat{\theta}_i^S, \hat{\theta}_i^I)$	evaluation coordinate point
n	sample size
$(X_{i,1}, \dots, X_{i,j}, \dots, X_{i,n})$	the satisfaction sample data of the customer's i th service item
$E[\hat{\theta}_i^S] = \theta_i^S$	the expected value of $\hat{\theta}_i^S$
$\hat{\theta}_i^S$	an unbiased estimator of θ_i^S
θ_S	the average value of $\hat{\theta}_i^S$
α	significance level
CR_{Xi}	the critical region for satisfaction index i
S_{Xi}	the standard deviation of satisfaction index i
Z_{Xi}	the standard score of satisfaction index i
C_{X0i}	the critical value of satisfaction index i
$(Y_{i,1}, \dots, Y_{i,j}, \dots, Y_{i,n})$	the importance sample data of the customer's i th service item
$\hat{\theta}_i^I$	an unbiased estimator of importance index i
$E[\hat{\theta}_i^I] = \theta_i^I$	the expected value of $\hat{\theta}_i^I$
θ_I	the average value of $\hat{\theta}_i^I$
CR_{Yi}	the critical region for importance index i
S_{Yi}	the standard deviation of importance index i
Z_{Yi}	the standard score of importance index i
C_{Y0i}	the critical value of importance index i
H_0	null hypothesis
H_1	alternative hypothesis
$y = C_{Y0}$	the value at which the horizontal line intersects the y -axis
$x = C_{X0}$	the value at which the vertical line intersects the x -axis
C_{X0}	$\text{Min}\{C_{X01}, C_{X02}, \dots, C_{X0q}\}$
C_{Y0}	$\text{Min}\{C_{Y01}, C_{Y02}, \dots, C_{Y0q}\}$
Q_1	Quadrant 1
Q_2	Quadrant 2
Q_3	Quadrant 3
Q_4	Quadrant 4
A	the set formed by the service items that need to be enhanced
θ_{bi}^S	the value of satisfaction index before improvement
θ_{ai}^S	the value of satisfaction index after improvement
$(X'_{i,1}, \dots, X'_{i,j}, \dots, X'_{i,n})$	the satisfaction sample data is for $i \in A$
θ_{ai}^S	the average value of satisfaction index after improvement
θ_{bi}^S	the average value of satisfaction index before improvement
Z_{hi}	the standard score of satisfaction index i after improvement ($h = a$) or before improvement ($h = b$)
S_{Xai}	the standard deviation of satisfaction index i after improvement
S_{Xbi}	the standard deviation of satisfaction index i before improvement
$[L\theta_{bi}, U\theta_{bi}]$	the $100(1 - \alpha)\%$ confidence interval of θ_{bi}^S
$[L\theta_{ai}, U\theta_{ai}]$	the $100(1 - \alpha)\%$ confidence interval of θ_{ai}^S

References

1. Ahmad, S.; Wong, K.Y.; Tseng, M.-L.; Wong, W.P. Sustainable product design and development: A review of tools, applications and research prospects. *Resour. Conserv. Recycl.* **2018**, *132*, 49–61. [\[CrossRef\]](#)
2. Amindoust, A.; Ahmed, S.; Saghafinia, A.; Bahreininejad, A. Sustainable supplier selection: A ranking model based on fuzzy inference system. *Appl. Soft Comput.* **2012**, *12*, 1668–1677. [\[CrossRef\]](#)
3. Awasthi, A.; Govindan, K.; Gold, S. Multi-tier sustainable global supplier selection using a fuzzy AHP-VIKOR based approach. *Int. J. Prod. Econ.* **2018**, *195*, 106–117. [\[CrossRef\]](#)
4. Meshram, C.; Ibrahim, R.W.; Deng, L.; Shende, S.W.; Meshram, S.G.; Barve, S.K. A robust smart card and remote user password-based authentication protocol using extended chaotic maps. *Soft Comput.* **2021**, *25*, 10037–10051. [\[CrossRef\]](#)
5. Yazdi, A.K.; Hanne, T.; Gómez, J.C.O. Evaluating the performance of colombian banks by hybrid multicriteria decision making methods. *J. Bus. Econ. Manag.* **2020**, *21*, 1707–1730. [\[CrossRef\]](#)
6. Angtuaco, D.S.; Barria, N.M.A.; Lee, J.M.C.; Tangsac, J.C.; Chiu, A.S.F.; Mutuc, J.E. A redesign of the toothpaste tube using green QFD II for improved usability and sustainability. *J. Clean. Prod.* **2023**, *393*, 136279. [\[CrossRef\]](#)
7. Sarfaraz, A.H.; Yazdi, A.K.; Wanke, P.; Nezhad, E.A.; Hosseini, R.S. A novel hierarchical fuzzy inference system for supplier selection and performance improvement in the oil & gas industry. *J. Decis. Syst.* **2022**. [\[CrossRef\]](#)
8. Chen, K.-S.; Yu, C.-M. Fuzzy test model for performance evaluation matrix of service operating systems. *Comput. Ind. Eng.* **2020**, *140*, 106240. [\[CrossRef\]](#)
9. Martínez-Caro, E.; Cegarra-Navarro, J.G.; Cepeda-Carrión, G. An application of the performance-evaluation model for e-learning quality in higher education. *Total. Qual. Manag. Bus.* **2015**, *26*, 632–647. [\[CrossRef\]](#)
10. Wong, R.C.P.; Szeto, W.Y. An alternative methodology for evaluating the service quality of urban taxis. *Transp. Policy* **2018**, *69*, 132–140. [\[CrossRef\]](#)
11. Wu, J.; Wang, Y.; Zhang, R.; Cai, J. An Approach to Discovering Product/Service Improvement Priorities: Using Dynamic Importance-Performance Analysis. *Sustainability* **2018**, *10*, 3564. [\[CrossRef\]](#)
12. Cezar, A.; Ögüt, H. Analyzing conversion rates in online hotel booking: The role of customer reviews, recommendations and rank order in search listings. *Int. J. Contemp. Hosp. Manag.* **2016**, *28*, 286–304. [\[CrossRef\]](#)
13. Gómez-Déniz, E.; Pérez-Rodríguez, J.V.; Boza-Chirino, J. Modelling tourist expenditure at origin and destination. *Tour. Econ.* **2020**, *26*, 437–460. [\[CrossRef\]](#)
14. Hung, Y.H.; Huang, M.L.; Chen, K.-S. Service quality evaluation by service quality performance matrix. *Total. Qual. Manag. Bus.* **2003**, *14*, 79–89. [\[CrossRef\]](#)
15. Wang, C.-H.; Chen, K.-S. New process yield index of asymmetric tolerances for bootstrap method and six sigma approach. *Int. J. Prod. Econ.* **2020**, *219*, 216–223. [\[CrossRef\]](#)
16. Yu, K.T.; Chen, K.S. Testing and analyzing capability performance for products with multiple characteristics. *Int. J. Prod. Res.* **2016**, *54*, 6633–6643. [\[CrossRef\]](#)
17. Chen, K.-S. Fuzzy testing of operating performance index based on confidence intervals. *Ann. Oper. Res.* **2022**, *311*, 19–33. [\[CrossRef\]](#)
18. Cheng, S.W. Practical implementation of testing process capability indices. *Qual. Eng.* **1994**, *7*, 239–259. [\[CrossRef\]](#)
19. Xu, Y.; Zhang, X.; Meng, P. A Novel Intelligent Deep Learning-Based Uncertainty-Guided Network Training in Market Price. *IEEE Trans. Ind. Inform.* **2022**, *18*, 5705–5711. [\[CrossRef\]](#)
20. Yang, C. Establishment and applications of the integrated model of service quality measurement. *Manag. Serv. Qual.* **2003**, *13*, 310–324. [\[CrossRef\]](#)
21. Yu, C.-M.; Chang, H.-T.; Chen, K.-S. Developing a performance evaluation matrix to enhance the learner satisfaction of an e-learning system. *Total. Qual. Manag. Bus. Excel.* **2018**, *29*, 727–745. [\[CrossRef\]](#)
22. Li, R. Research trends of blended language learning: A bibliometric synthesis of SSCI-indexed journal articles during 2000–2019. *ReCALL* **2022**, *34*, 309–326. [\[CrossRef\]](#)
23. Bobunova, A.; Sergeeva, M.; Notina, E. Integrating Computer-Assisted Language Learning into ESL Classroom: Formation of Moral and Aesthetic Values. *Int. J. Inf. Educ. Technol.* **2021**, *11*, 24–29. [\[CrossRef\]](#)
24. Warschauer, M.; Healey, D. Computers and language learning: An overview. *Lang. Teach.* **1998**, *31*, 57–71. [\[CrossRef\]](#)
25. Leszczyński, P.; Charuta, A.; Łaziuk, B.; Gałazkowski, R.; Wejnarski, A.; Roszak, M.; Kołodziejczak, B. Multimedia and interactivity in distance learning of resuscitation guidelines: A randomised controlled trial. *Interact. Learn. Environ.* **2018**, *26*, 151–162. [\[CrossRef\]](#)
26. Zhang, Y.; MacWhinney, B. The role of novelty stimuli in second language acquisition: Evidence from the optimized training by the Pinyin Tutor at TalkBank. *Smart Learn. Environ.* **2023**, *10*, 3. [\[CrossRef\]](#)
27. Shee, D.Y.; Wang, Y.-S. Multi-criteria evaluation of the web-based e-learning system: A methodology based on learner satisfaction and its applications. *Comput. Educ.* **2008**, *50*, 894–905. [\[CrossRef\]](#)
28. Gaffas, Z.M. Students' perceptions of e-learning ESP course in virtual and blended learning modes. *Educ. Inf. Technol.* **2023**. [\[CrossRef\]](#)
29. Martin, F.; Bolliger, D.U. Developing an online learner satisfaction framework in higher education through a systematic review of research. *Int. J. Educ. Technol. High. Educ.* **2022**, *19*, 1–21. [\[CrossRef\]](#)

30. Qing, J.; Cheng, G.; Ni, X.-Q.; Yang, Y.; Zhang, W.; Li, Z. Implementation of an interactive virtual microscope laboratory system in teaching oral histopathology. *Sci. Rep.* **2022**, *12*, 5492. [[CrossRef](#)]
31. Souza, F.V.; Motoki, F.Y.S.; Mainardes, E.W.; Azzari, V. Public Corporate e-Learning: Antecedents and Results. *Public Organ. Rev.* **2021**, *22*, 1139–1156. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.