SYNTIA – Systems and Technologies in Artificial Intelligence Applications.
Volume 1 Issue 1 (2025). Page: 1-8.

Decision Support System for Patient Satisfaction with Healthcare Services Using the Preference Selection Index (PSI)

Fein Dwi Muhammad[1*], Rahma Laila[2], Prakasit Poonwong[3]

[1] *Informatics Engineering Study Program, Faculty of Engineering, Universitas Tadulako, Palu, Indonesia.*
[2] *Department of Information Technology, Faculty of Engineering, Universitas Tadulako, Palu, Indonesia.*
[3] *Boromarajonani College of Nursing, Chainat, Praboromrajchanok Institute, Thailand.*

**Corresponding author: feininformatika17@gmail.com*

Abstract: Patient-satisfaction surveys at community health centers are still tallied by hand. This study built a web-based decision support system at Puskesmas Tawaeli, Palu, that ranks patients by satisfaction with the Preference Selection Index (PSI) from a ten-item questionnaire on a four-level scale, and evaluated it three ways. A black-box suite of 13 functions gave 11 results as expected and two mismatches (profile, logout) that the test sheet had recorded as passed. A recomputation of the PSI steps on the published five-patient example gave weights between 0.094 and 0.104 and a ranking that follows the mean score; the manual calculation in the source thesis contains arithmetic errors and negative weights, and the intermediate values the system displayed cannot arise from the PSI definition. A User Experience Questionnaire completed by 20 respondents rated all six scales above the "excellent" benchmark threshold (means 1.88–2.46), although the confidence intervals of three scales cross it and perspicuity has low reliability (α = 0.49). The system is usable, but its PSI module needs correction and retesting before its rankings are used.

Keywords : decision support system; patient satisfaction; Preference Selection Index; user experience questionnaire; community health center.

Received: November 01, 2025 | Revised: November 28, 2025 | Accepted: December 15, 2025 | Published: December 31, 2025

1. Introduction

Patient satisfaction is a standard indicator of health-service quality because it reflects whether care met what patients expected [1]. In Indonesia the community health center (Puskesmas) is the first point of contact for curative, preventive, promotive, and rehabilitative services, and each center is expected to collect feedback from the people it serves. At the center studied here, Puskesmas Tawaeli in Palu, the questionnaires are filled in on paper and summarized by hand, which is slow and leaves no record that can be re-examined. A decision support system (DSS) that stores the responses and computes a ranking is a natural remedy, since decision support has long been used to turn structured data into a comparison that a manager can act on [2].

The Preference Selection Index (PSI) is a multi-criteria decision-making method that needs no weights supplied by the decision maker. Each criterion receives a weight derived from how much its normalized values vary across the alternatives, and the alternatives are then ranked by a weighted sum [3]. It has been applied to design decisions in production systems [4] and, in Indonesian studies, to selecting tutors and scoring COVID-19 scan results [5], [6]. Most of them report the ranking but not how the ranking was checked, which matters because PSI contains a normalization step and a subtraction ($\Omega_j = 1 - \varphi_j$) that are easy to implement wrongly, and normalization choices are known to change the outcome of multi-criteria methods [7].

SYNTIA – Systems and Technologies in Artificial Intelligence Applications is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License

We therefore treat the system as an object of evaluation, not only of construction. We ask four questions. (RQ1) Do the functions of the web system behave as specified? (RQ2) Can the PSI steps be reproduced independently, and do the intermediate values the system displays agree with the PSI definition? (RQ3) What does PSI add to a plain average of the questionnaire scores? (RQ4) How do users rate the user experience of the system, and how reliable is that rating? User satisfaction with an information system is a construct of its own, studied since the work on end-user computing satisfaction [8]. The contributions are a working PSI-based satisfaction system for a primary-care setting, a corrected and fully worked PSI computation on the data of the original study, a record of where the original manual and system computations depart from the method, and a six-scale user-experience profile with confidence intervals and reliability coefficients. Section 2 describes the setting and methods, Section 3 reports and discusses the results, and Section 4 concludes.

2. Method

2.1. Setting, participants, and data

The study was carried out at Puskesmas Tawaeli, Palu, Central Sulawesi, from August to December 2023, with 20 patients as respondents. Satisfaction was recorded with a ten-item questionnaire shown on the web form (Fig. 2a); each item is answered on a four-level scale, Very Good (SB), Good (B), Less Good (KB), and Very Poor (SKB), stored as 4, 3, 2, and 1 (Fig. 2b). The ten items correspond to the criteria P1–P10 of the decision matrix and cover service conformity, procedure, speed, cost, product, staff competence, staff behavior, facilities, complaint handling, and safety at the center. Patient names are replaced by codes in this article. All 20 respondents gave informed consent before answering, and Puskesmas Tawaeli permitted the collection and use of the data for this study. The ten items adapt the nine elements of the national guideline for community satisfaction surveys of public service units (Permenpan-RB No. 14 of 2017) and add one item on safety.

2.2. System development

The system was developed as a PHP web application with the Waterfall model [9]: requirements analysis by interviews and observation, design with data-flow, use-case, activity, class, sequence, and entity-relationship diagrams, coding, testing, and operation. Fig. 1 shows the PSI computation implemented in the system and the three evaluations applied to it. Administrators manage patients, criteria, and scores; respondents fill in the questionnaire; the ranking module computes the PSI value of every patient.

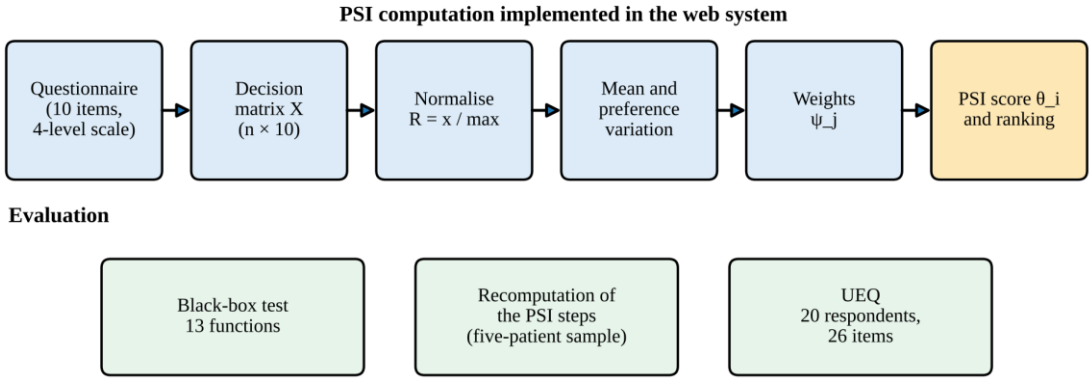


Figure 1. PSI computation implemented in the system (top) and the three evaluations reported in this article (bottom).

2.3. Preference Selection Index

Let $X = [x_{ij}]$ be the decision matrix of n alternatives (patients) and m = 10 criteria. All ten criteria are satisfaction ratings, so higher is better and every criterion is of the benefit type. The PSI procedure is [4]:

$$R_{ij} = \frac{x_{ij}}{\max_i x_{ij}}, \qquad \bar{N}_j = \frac{1}{n}\sum_{i=1}^{n} R_{ij} \tag{1}$$

$$\varphi_j = \sum_{i=1}^{n}\left(R_{ij} - \bar{N}_j\right)^2, \qquad \Omega_j = 1 - \varphi_j \tag{2}$$

$$\psi_j = \frac{\Omega_j}{\sum_{j=1}^{m} \Omega_j}, \qquad \theta_i = \sum_{j=1}^{m} R_{ij}\ \psi_j \tag{3}$$

R_{ij} is the normalized value, $\bar{N}_j$ the mean of criterion j, φ_j its preference variation, Ω_j the deviation, ψ_j the weight, and θ_i the PSI value; the alternative with the largest θ_i ranks first. Because $R_{ij} \leq 1$ for benefit criteria, $\bar{N}_j$ cannot exceed 1, and the weights are positive whenever $\varphi_j < 1$ for all j. We use these two properties as checks in Section 3.3.

2.4. Evaluation design

Black-box test. Each of the 13 buttons and menus of the administrator interface was exercised and the observed result was compared literally with the expected result [10]. *PSI recomputation.* The only complete decision matrix printed in the source thesis is a five-patient example, so we recomputed Eq. (1)–(3) on it in Python and compared the outcome with the thesis calculation and with the intermediate values shown in the screenshots of the system. Ranking agreement is measured with Spearman's rank correlation. *User experience.* The User Experience Questionnaire (UEQ), with 26 semantic-differential items on a seven-point scale, was completed by 20 respondents guided by the researcher [11]. Responses were transformed to −3…+3 using the published item polarity, averaged into six scales, and summarized with the respondent-level mean, standard deviation, a 95% t-interval, and Cronbach's α [12]; scale means were compared with the published benchmark thresholds [13], [14]. The UEQ respondents are the same 20 patients as in the PSI sample, so one group rated both the service and the system.

3. Results and Discussion

3.1. Implemented system

The administrator logs in, maintains the patient list, the criteria, and the criterion values, and opens the PSI page, which shows the decision matrix, the normalized matrix, the mean row, the preference-variation table, the deviation row, the weight row, the PSI value of each patient, and the resulting ranking, color-coded by satisfaction level. Fig. 2 shows two of the input screens. The questionnaire form lists the ten criteria with the four answer options, and the criteria-value page maps SB, B, KB, and SKB to 4, 3, 2, and 1.

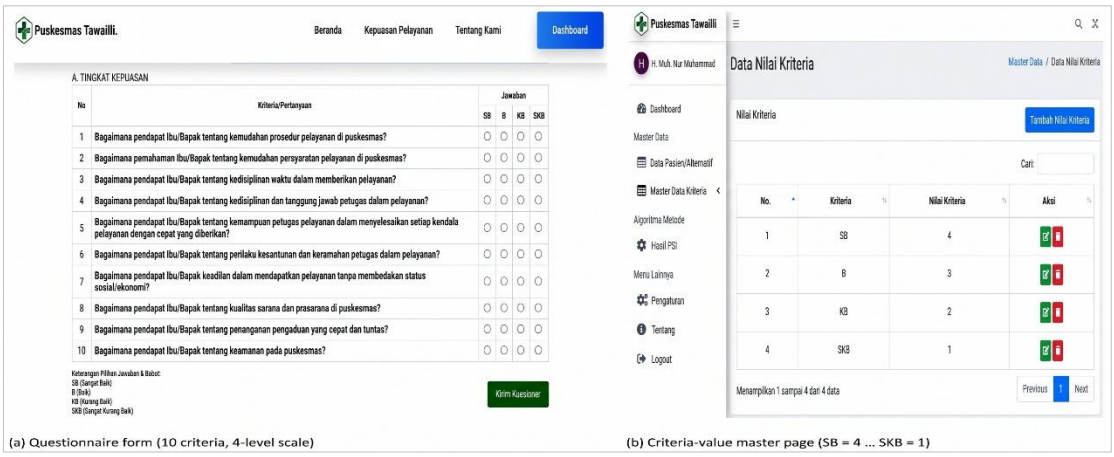


Figure 2. Input screens of the system: (a) questionnaire form; (b) criteria-value page.

3.2. Black-box test (RQ1)

Table 1 lists the 13 functions tested. For 11 of them the observed behavior equals the expected behavior. For two, the observed result differs from the expected result written in the same row, yet the test sheet records both as successful: the profile button returned to the previous page instead of showing account

information, and the logout button opened a "select category" page instead of the login page. The latter page does not belong to this application and suggests code carried over from another project. We count these as two open defects, so the functional pass rate is 11/13 (85%) under a literal reading of the test sheet, not 13/13. Both defects remain open in the version we reviewed, and a retest of the two buttons after correction is the first follow-up step (Section 4).

Table 1. Black-box test of the administrator interface. "Match" compares the expected and the observed column; the test sheet recorded all 13 rows as successful.

No	Function	Expected	Observed	Match
1	Login button	Dashboard if credentials are valid	Dashboard if credentials are valid	Yes
2	Patient-data menu	Patient identity page	Patient identity page	Yes
3	Add patient	Add-patient page	Add-patient page	Yes
4	Criteria-data menu	Criteria page	Criteria page	Yes
5	Add criterion	Add-criterion page	Add-criterion page	Yes
6	Scoring menu	Scoring page	Scoring page	Yes
7	Ranking process	Ranking-result page	Ranking-result page	Yes
8	Edit	Edit page	Edit page	Yes
9	Delete	Data deleted	Data deleted	Yes
10	Save	Data saved	Data saved	Yes
11	Add sub-criterion	Add-sub-criterion page	Add-sub-criterion page	Yes
12	Profile button	Account-information page	Returns to previous page	No
13	Logout button	Login page	"Select category" page	No

3.3. Recomputation of the PSI steps (RQ2)

Table 2 gives the five-patient decision matrix of the thesis, the recomputed PSI values, and the values printed in the thesis. The recomputed weights lie between 0.094 and 0.104 (Fig. 3) and sum to 1. The ranking by recomputed θ is A1, A3, A5, A4, A2. The ranking printed in the thesis is the reverse in its upper half: A2, the patient with the lowest scores on nine of ten criteria, is ranked first, and the Spearman correlation between the two rankings is −0.30 (n = 5, not significant but clearly not agreement).

Table 2. Five-patient decision matrix (P1–P10), mean score, recomputed PSI value θ (Eq. 3) with rank

Patient	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	Mean	θ	Rank	θ thesis	Rank thesis
A1	4	3	4	3	3	3	2	3	4	4	3.30	0.967	1	4.191	4
A2	2	3	2	3	2	3	2	2	2	2	2.30	0.706	5	9.599	1
A3	3	3	3	3	3	3	3	3	3	3	3.00	0.903	2	6.582	2
A4	2	2	2	2	3	3	3	2	3	3	2.50	0.753	4	3.304	5
A5	3	3	3	3	2	2	2	3	3	3	2.70	0.802	3	6.235	3

A1 = Mimi, A2 = Gunawan, A3 = Iksan, A4 = Belfin, A5 = Indri Astuti in the thesis. Mean is the plain average of P1–P10.

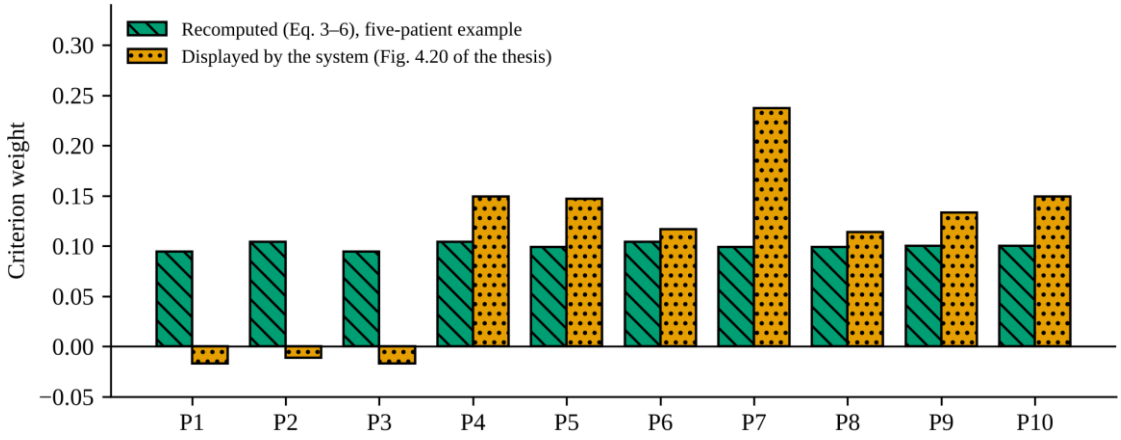


Figure 3. Criterion weights recomputed with Eq. (1)–(3) on the five-patient example, and the weights displayed by the system.

Three departures explain the difference. First, the thesis normalizes P4 as x / min (3/2 = 1.5), a cost-type formula, although all criteria are ratings of the benefit type, so R values above 1 appear. Second, several column totals in the thesis do not equal the sum of the entries printed above them: P4 is given as 3.668 although its entries sum to 7.0, and P5 as 6.0 although they sum to 4.333. Third, the resulting Ω_j sum to −0.284, so dividing by this sum produces negative weights (for example −2.908 for P1) and inverts the ranking. The thesis calculation therefore does not follow Eq. (1)–(3).

The screenshots of the system show the same symptoms. The mean row of the 20-patient run has values of 1.28–1.40 for P5–P10, which cannot be the mean of normalized benefit values bounded by 1, the deviation row sums to −34.15, and the weight row begins with −0.017, −0.011, and −0.017 for P1–P3 (Fig. 3). The thesis states that the manual and the system results are equal, but the two computations share these departures, so their agreement shows that the system reproduces the manual procedure, not that either implements PSI correctly. This answers RQ2: the PSI module as documented cannot be accepted until the normalization and the mean row are corrected and the 20-patient run is repeated. We could not inspect the source code of the ranking page, and the thesis does not print the 20-patient decision matrix, so Table 2 and Fig. 3 keep the five-patient example.

3.4. What PSI adds to a plain average (RQ3)

On the five-patient example the correct PSI weights vary by a factor of 1.10 (0.0945 to 0.1043). With nearly equal weights, θ is close to a rescaled mean of the ratings, and the PSI ranking coincides exactly with the ranking by mean score (Spearman $\rho = 1.00$, Table 2). Equal weights are the expected result when all criteria use the same four-level scale and have similar dispersion. PSI would give different answers only where some criteria separate patients much more than others. For a satisfaction survey with a common scale, a plain average would produce the same ranking for this sample; the benefit of PSI is that its weights are derived from the data instead of being set by the center, which is of value mainly when criteria differ in dispersion. A further design point is that the alternatives are patients, so the ranking orders patients by how satisfied they are. A manager who wants to know which service aspect to improve would instead read the criterion means or the PSI weights, and the system would be more useful if it displayed them.

3.5. User experience (RQ4)

Table 3 and Fig. 4 give the six UEQ scales. All scale means exceed the 0.8 threshold for a positive evaluation and the "excellent" threshold of the general benchmark [13], [14]: 2.46 for attractiveness, 2.30 for efficiency, 2.25 for dependability, 2.15 for perspicuity, 2.10 for stimulation, and 1.88 for novelty. The mean over all 26 items, a simple UX index [15], is 2.21, and only 9 of 520 item responses (1.7%) are negative. The ranking of the scales agrees with the thesis (attractiveness highest, novelty lowest), and our recomputation of every scale mean from the raw responses of the thesis reproduces its values exactly.

Table 3. UEQ scales for 20 respondents: mean, standard deviation of respondent-level scale means, 95% confidence interval, Cronbach's α, and benchmark category of the point estimate [14].

Scale	Items	Mean	SD	95% CI	α	Benchmark
Attractiveness	6	2.46	0.58	2.19–2.73	0.82	Excellent (>1.84)
Perspicuity	4	2.15	0.63	1.86–2.44	0.49	Excellent (>2.00)
Efficiency	4	2.30	0.67	1.99–2.61	0.77	Excellent (>1.88)
Dependability	4	2.25	0.79	1.88–2.62	0.79	Excellent (>1.70)
Stimulation	4	2.10	0.83	1.71–2.49	0.77	Excellent (>1.70)
Novelty	4	1.88	0.96	1.43–2.32	0.60	Excellent (>1.60)

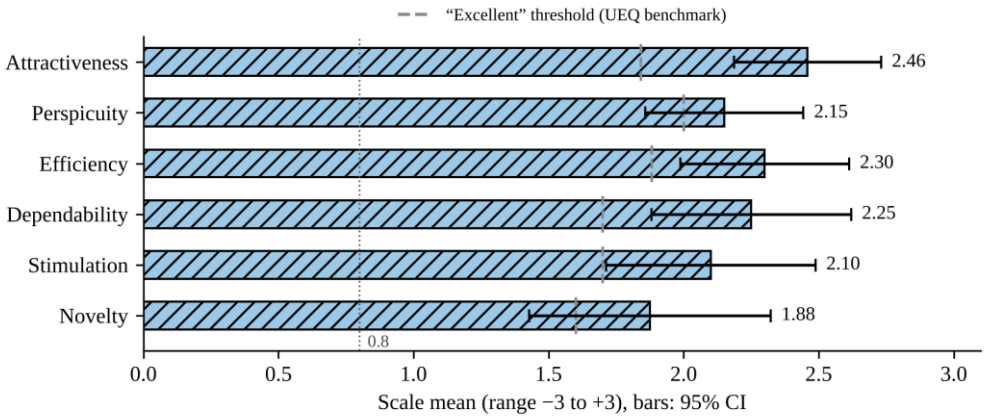


Figure 4. UEQ scale means with 95% confidence intervals. Dashed marks: lower bound of the "excellent" category in the general benchmark [14].

Three qualifications apply. First, with 20 respondents the intervals are wide (±0.27 to ±0.45): those of perspicuity (1.86–2.44), novelty (1.43–2.32), and stimulation (1.71–2.49) extend below their "excellent" thresholds, so the safe statement is that all scales are at least "good" and probably "excellent", not that all are excellent. Second, reliability is acceptable for attractiveness, efficiency, dependability, and stimulation ($\alpha = 0.77$–0.82) but low for perspicuity (0.49) and borderline for novelty (0.60); the handbook warns that α computed on 20 to 40 respondents is unstable and that item means should then be read directly [14]. The lowest item means are for "conservative–innovative" (1.55), "unpredictable–predictable" (1.70), and "fast–slow" (1.85), the highest for "not understandable–understandable" and "unfriendly–friendly" (both 2.65). Third, benchmark comparison uses products of other kinds, and the respondents were guided by the researcher, which can raise ratings. The result is best read as a favorable first impression, not as proof of superiority.

3.6. Limitations

Only the five-patient example could be recomputed, because the 20-patient decision matrix is not printed in the thesis, and the system was not available to us, so statements on the system rest on its screenshots. The black-box suite checks navigation and data handling but not the correctness of the computed values, which is why it did not detect the PSI defects; a numeric test with known inputs and outputs would have. The UEQ sample is small and was guided, there is no task-based usability test and no comparison with the paper process, and the study covers one center. Satisfaction values are ordinal ratings treated as interval data, as is common in practice but debated [16].

4. Conclusion

The web system gives Puskesmas Tawaeli a stored, rankable record of patient satisfaction, and 20 respondents rated its user experience above the 0.8 threshold on all six UEQ scales, with point estimates in the "excellent" range (1.88–2.46), although small-sample intervals and low reliability on perspicuity temper the claim. The evaluation also found that the PSI module is not yet correct: the thesis calculation uses a cost-type normalization for a benefit criterion, contains arithmetic errors, and yields negative weights, and the displayed intermediate values show the same symptoms. A correct computation on the five-patient example gives nearly equal weights and a ranking identical to the mean score, so PSI adds little for a single-scale survey. Two interface functions (profile, logout) behave differently from their specification.

Before the system is used for decisions, the authors should correct the normalization and mean computation, add a numeric test with known answers, rerun the 20-patient case, fix the two interface defects, and display criterion-level results so that managers can see which service aspects need attention. A

follow-up study should compare PSI with an equal-weight average and another weighting method on the full data, and should collect UEQ responses from patients without guidance.

AI Disclosure Statement

The authors used Generative AI and AI-Assisted Technologies throughout the preparation of this manuscript: to draft and revise portions of the text, to write and run the scripts behind the recomputation of the PSI steps and the UEQ analysis, to generate the figures, and to format the document from the authors' own research data. The authors verified the scripts and their outputs, reviewed the resulting text, cross-checked every reference against its original source, and accept full responsibility for the content of this publication.

Author Contributions Statement

Fein Dwi Muhammad: Conceptualization, Methodology, Software, Data Curation, Formal Analysis, Visualization, Writing – Original Draft.
Rahma Laila: Data Curation, Investigation, Validation, Software, Writing – Review and Editing.
Prakasit Poonwong: Supervision, Methodology, Resources, Writing – Review and Editing.

References

[1] A. Donabedian, "The quality of care: How can it be assessed?," *JAMA*, vol. 260, no. 12, pp. 1743–1748, 1988, doi: 10.1001/jama.1988.03410120089033.

[2] J. P. Shim, M. Warkentin, J. F. Courtney, D. J. Power, R. Sharda, and C. Carlsson, "Past, present, and future of decision support technology," *Decis. Support Syst.*, vol. 33, no. 2, pp. 111–126, 2002, doi: 10.1016/S0167-9236(01)00139-7.

[3] K. Maniya and M. G. Bhatt, "A selection of material using a novel type decision-making method: Preference selection index method," *Mater. Des.*, vol. 31, no. 4, pp. 1785–1789, 2010.

[4] R. Attri and S. Grover, "Application of preference selection index method for decision making over the design stage of production system life cycle," *J. King Saud Univ. – Eng. Sci.*, vol. 27, no. 2, pp. 207–216, 2015, doi: 10.1016/j.jksues.2013.06.003.

[5] S. Laia, F. Sonata, and S. Yakub, "Kelayakan hasil scanner data pasien Covid 19 menggunakan metode PSI (Preference Selection Index)" (in Indonesian), *J. Sist. Inf. Triguna Dharma (JURSI TGD)*, vol. 1, no. 3, p. 182, 2022, doi: 10.53513/jursi.v1i3.5217.

[6] F. Fauziyah, F. Nugroho, and E. Buulolo, "Sistem pendukung keputusan pemilihan instruktur bimbingan belajar menggunakan metode PSI (Preference Selection Index)" (in Indonesian), *Build. Informatics, Technol. Sci. (BITS)*, vol. 4, no. 3, pp. 1447–1455, 2022, doi: 10.47065/bits.v4i3.2695.

[7] N. Vafaei, R. A. Ribeiro, and L. M. Camarinha-Matos, "Data normalisation techniques in decision making: Case study with TOPSIS method," *Int. J. Inf. Decis. Sci.*, vol. 10, no. 1, 2018, doi: 10.1504/IJIDS.2018.090667.

[8] W. J. Doll and G. Torkzadeh, "The measurement of end-user computing satisfaction," *MIS Q.*, vol. 12, no. 2, pp. 259–274, 1988, doi: 10.2307/248851.

[9] W. W. Royce, "Managing the development of large software systems: Concepts and techniques," in *Proc. 9th Int. Conf. Softw. Eng. (ICSE)*, 1987, pp. 328–338.

[10] S. Nidhra, "Black box and white box testing techniques – A literature review," *Int. J. Embedded Syst. Appl.*, vol. 2, no. 2, pp. 29–50, 2012, doi: 10.5121/ijesa.2012.2204.

[11] B. Laugwitz, T. Held, and M. Schrepp, "Construction and evaluation of a user experience questionnaire," in *HCI and Usability for Education and Work*, Lecture Notes in Computer Science. Berlin, Germany: Springer, 2008, pp. 63–76, doi: 10.1007/978-3-540-89350-9_6.

[12] L. J. Cronbach, "Coefficient alpha and the internal structure of tests," *Psychometrika*, vol. 16, no. 3, pp. 297–334, 1951, doi: 10.1007/BF02310555.

[13] M. Schrepp, A. Hinderks, and J. Thomaschewski, "Construction of a benchmark for the user experience questionnaire (UEQ)," *Int. J. Interact. Multimed. Artif. Intell.*, vol. 4, no. 4, pp. 40–44, 2017, doi: 10.9781/ijimai.2017.445.

[14] M. Schrepp, "User experience questionnaire handbook," ueq-online.org. [Online]. Available: https://www.ueq-online.org/Material/Handbook.pdf (accessed Oct. 13, 2026).

[15] A. Hinderks, M. Schrepp, F. J. Domínguez Mayo, M. J. Escalona, and J. Thomaschewski, "Developing a UX KPI based on the user experience questionnaire," *Comput. Stand. Interfaces*, vol. 65, pp. 38–44, 2019, doi: 10.1016/j.csi.2019.01.007.

[16] G. M. Sullivan and A. R. Artino, "Analyzing and interpreting data from Likert-type scales," *J. Grad. Med. Educ.*, vol. 5, no. 4, pp. 541–542, 2013, doi: 10.4300/JGME-5-4-18.