

RELIGACIÓN

R E V I S T A

Evaluación de la calidad percibida en salud: análisis multivariante y segmentación en un centro privado

Assessment of Perceived Quality in Healthcare: Multivariate Analysis and Segmentation in a Private Center

Joseph Giovanni Villacreses Quevedo, Milena Jamileth Gualán Torres, Harry Vite Cevallos, Wiliam Tranquilino Medina Castillo

Resumen

Las evaluaciones de calidad percibida basadas en SERVQUAL suelen resumirse en medias que pueden ocultar heterogeneidad y problemas de calidad de respuesta. Este estudio transversal la evaluó en un centro privado de especialidades de Machala, Ecuador, mediante análisis multivariante y segmentación. Se analizaron 90 cuestionarios (30 ítems, cinco dimensiones) con coeficientes de fiabilidad, correlaciones de Spearman, componentes principales y K-means, validado con bootstrap, Ward y exclusión de respuestas invariantes. La consistencia interna fue elevada ($\alpha = \omega = 0.990$) y el primer componente explicó 91.7% de la varianza. La solución de tres estratos (19, 41 y 30 participantes) coincidió con la segmentación de la puntuación global (ARI = 1.000); al excluir las respuestas invariantes (32.2%), Silhouette favoreció dos estratos. El tiempo de espera fue el ítem peor valorado en el estrato crítico (media = 2.42). Los estratos representan niveles de una percepción global, sensibles a la calidad de respuesta.

Palabras clave: calidad percibida; satisfacción del paciente; SERVQUAL; análisis multivariante; análisis de conglomerados.

Joseph Giovanni Villacreses Quevedo

Universidad Técnica de Machala | Machala | Ecuador | jvillacreses_est@utmachala.edu.ec
<https://orcid.org/0000-0001-8859-8147>

Milena Jamileth Gualán Torres

Universidad Técnica de Machala | Machala | Ecuador | mgualan2@utmachala.edu.ec
<https://orcid.org/0009-0000-4732-9004>

Harry Vite Cevallos

Universidad Técnica de Machala | Machala | Ecuador | hvite@utmachala.edu.ec
<https://orcid.org/0000-0003-2056-7111>

Wiliam Tranquilino Medina Castillo

Universidad Técnica de Machala | Machala | Ecuador | wmedina@utmachala.edu.ec
<https://orcid.org/0000-0003-2609-6568>

<http://doi.org/10.46652/rgn.v11i52.1729>

ISSN 2477-9083

Vol. 11 No. 52, octubre-diciembre, 2026, e2601729

Quito, Ecuador

Enviado: junio 10, 2026

Aceptado: agosto 13, 2026

Publicado: octubre 15, 2026

Publicación Continua



Abstract

SERVQUAL-based assessments of perceived quality are usually summarized as means, which may conceal heterogeneity and response-quality problems. This cross-sectional study assessed perceived quality at a private specialty center in Machala, Ecuador, using multivariate analysis and segmentation. Ninety questionnaires (30 items, five dimensions) were analyzed with reliability coefficients, Spearman correlations, principal components, and K-means, validated with bootstrap, Ward's method, and exclusion of invariant responses. Internal consistency was high ($\alpha = \omega = 0.990$), and the first component explained 91.7% of the variance. The three-stratum solution (19, 41, and 30 participants) matched the segmentation based on the global score ($ARI = 1.000$); after excluding invariant responses (32.2%), the Silhouette index favored two strata. Waiting time was the lowest-rated item in the critical stratum (mean = 2.42). The strata represent levels of a global perception that are sensitive to response quality.

Keywords: perceived quality; patient satisfaction; SERVQUAL; multivariate analysis; cluster analysis.

Introducción

La calidad de la atención sanitaria es un constructo multidimensional (Cioffi et al., 2023) que, desde el modelo de Donabedian (1988), se analiza a partir de la estructura, los procesos y los resultados de la atención (Villa-Velásquez et al., 2023). Entre los resultados, la calidad percibida recoge la valoración que el paciente hace del servicio recibido y comprende tanto lo que se entrega (calidad técnica) como la forma en que se entrega (calidad funcional) (Grönroos, 1984). Esta percepción no equivale a la calidad clínica objetiva, pero la complementa: la experiencia del paciente se asocia de manera positiva y consistente con la seguridad clínica y la efectividad de la atención (Doyle et al., 2013). Su medición, no obstante, depende del contexto y de las propiedades del instrumento; una revisión sistemática en países de ingresos bajos y medios documentó una marcada heterogeneidad conceptual y psicométrica entre las herramientas disponibles (Banda et al., 2023).

Desde la gestión de operaciones de servicios, la atención ambulatoria es un proceso en el que el paciente participa en la producción del servicio y lo evalúa mientras transita por la admisión, la espera, la consulta y los trámites; su percepción depende, por tanto, del trato y de la interacción con cada etapa (Becerra-Canales et al., 2024; Torain et al., 2021). En consecuencia, las decisiones sobre capacidad, programación y flujo de pacientes condicionan la calidad percibida. El tiempo de espera es el ejemplo más estudiado: las esperas prolongadas se asocian con menor satisfacción y deterioran la valoración de otros atributos, incluida la confianza en el profesional (Anderson et al., 2007; Bleustein et al., 2014). Por ello, los enfoques de gestión de la calidad orientados a la mejora de procesos utilizan la percepción del paciente como criterio para priorizar intervenciones operativas y no solo como resultado descriptivo (Moraros et al., 2016; Naiker et al., 2018).

SERVQUAL es el marco más difundido para operacionalizar la calidad percibida. Parte del modelo de brechas entre expectativas y percepciones (Parasuraman et al., 1985) y organiza la evaluación en cinco dimensiones: elementos tangibles, fiabilidad, capacidad de respuesta, seguridad y empatía (Parasuraman et al., 1988). Se adaptó tempranamente a los servicios hospitalarios (Babakus y Mangold, 1992) y sigue utilizándose como herramienta de mejora (Jonkisz et al.,

2021). Dos precisiones son relevantes para este estudio. La primera es que las medidas basadas únicamente en percepciones constituyen una alternativa aceptada al cálculo de brechas (Cronin y Taylor, 1992; Ibrahim, 2020); cuando no se dispone de expectativas, el instrumento debe describirse como una escala de percepción basada en SERVQUAL, sin atribuirle brechas no observadas. La segunda es que la estructura de cinco dimensiones no siempre se reproduce empíricamente y con frecuencia emerge un factor general (Babakus y Boller, 1992), de modo que la dimensionalidad debe verificarse y no darse por supuesta.

En América Latina, la evidencia empírica muestra que la calidad percibida varía entre instituciones, usuarios y componentes del servicio. En hospitales públicos de Trujillo, la capacidad de respuesta y la seguridad tuvieron un papel relevante (Morales-Garrido et al., 2025); estudios con SERVQUAL en Quibdó (Colombia) y en establecimientos ecuatorianos documentaron diferencias entre dimensiones y su relación con la satisfacción (Duque Oliva et al., 2023; Eguez Morales et al., 2025; Suco-Guerrero y Zambrano-Santos, 2025); y, a escala nacional, la demora en la atención fue el problema señalado con mayor frecuencia por los usuarios de establecimientos peruanos (Murillo et al., 2019). Estos trabajos se apoyan mayoritariamente en promedios, que pueden ocultar subgrupos con experiencias desfavorables. La analítica de datos ofrece una extensión útil frente a esta limitación (Awrahman et al., 2022; Piovani y Bonovas, 2022; Subrahmanya et al., 2022): como sintetiza la Tabla 1, se ha combinado con SERVQUAL y con encuestas de satisfacción para clasificar y priorizar dimensiones (Altuntas et al., 2022; Lee et al., 2022; Rahim et al., 2021), segmentar usuarios (Nurmeksela et al., 2023; Okuda et al., 2021; Oloya-Salazar et al., 2026) y predecir la satisfacción (Datt et al., 2025; Liu et al., 2023; Simsekler et al., 2021; Tanantong et al., 2026).

Tabla 1. Antecedentes de integración de ciencia de datos y analítica avanzada en calidad de servicio y experiencia del paciente en salud.

Estudio	Contexto y datos	Analítica aplicada	Conocimiento adicional
Altuntas et al. (2022)	Caso de servicios de salud en hospital público; evaluación basada en SERVQUAL.	SERVQUAL + algoritmos de machine learning para clasificación y priorización.	Genera impacto relativo y prioridades de mejora de dimensiones; muestra que la analítica puede superar la lectura de medias o brechas.
Rahim et al. (2021)	Reseñas de Facebook de hospitales públicos de Malasia (2017-2019).	Clasificación supervisada de dimensiones SERVQUAL + regresión logística.	Convierte texto no estructurado en dimensiones de calidad y relaciona patrones con insatisfacción, facilitando monitoreo casi en tiempo real.
Lee et al. (2022)	392 reseñas válidas de Google Maps de un hospital docente regional en Taiwán.	Web scraping + análisis semántico de eWOM según SERVQUAL.	Identifica patrones positivos y negativos por dimensión; destaca respuesta y tiempos de espera como focos de gestión.
Nurmeksela et al. (2023)	1,810 pacientes de tres hospitales de cuidados agudos en Finlandia.	K-means sobre subescalas de satisfacción del paciente.	Revela cuatro grupos desde insatisfechos hasta satisfechos, mostrando heterogeneidad que los promedios agregados pueden ocultar.

Estudio	Contexto y datos	Analítica aplicada	Conocimiento adicional
Okuda et al. (2021)	Datos HCAHPS de 3,711 hospitales de Estados Unidos.	Análisis de conglomerados, análisis de correspondencias y escalamiento multidimensional ordinal.	Descubre estructura latente entre dominios de satisfacción y su relación con características hospitalarias; ejemplo de triangulación analítica.
Simsekler et al. (2021)	Encuesta de satisfacción de 411 pacientes en un hospital de Abu Dabi.	Random forest con validación cruzada e importancia de variables.	Jerarquiza determinantes de satisfacción en registro y consulta; identifica tiempos y comportamiento profesional como factores relevantes.
Liu et al. (2023)	Datos extraídos de una plataforma líder de salud en línea de China.	Comparación de 11 modelos de machine learning; LightGBM como mejor modelo.	Predice satisfacción e identifica determinantes, incluido el tiempo de espera, aportando conocimiento adicional para decisiones de servicio.
Datt et al. (2025)	Encuesta de calidad de servicios sanitarios basada en el modelo CIRMQUAL.	Validación factorial + múltiples modelos de machine learning.	Integra medición de calidad y predicción de satisfacción para orientar acciones correctivas en dimensiones clínicas, infraestructurales, relacionales y gerenciales.
Oloya-Salazar et al. (2026)	480 usuarios de instituciones públicas y privadas de salud en Trujillo, Perú.	SERVQUAL + análisis de conglomerados de dos etapas y análisis de satisfacción.	Identifica perfiles de usuarios y extiende la evaluación SERVQUAL más allá de resultados agregados, con utilidad para estrategias diferenciadas.
Tanantong et al. (2026)	Encuesta en un servicio ambulatorio de oftalmología de un hospital universitario de Tailandia.	Random forest, boosting, validación cruzada anidada y explicabilidad SHAP.	Predice satisfacción e interpreta los factores que impulsan las predicciones; ejemplifica ciencia de datos explicable para gestión de recursos.

Fuente: elaboración propia.

La segmentación, sin embargo, exige cautela. Cuando las dimensiones están muy correlacionadas, los conglomerados pueden representar niveles bajos, medios y altos de una misma evaluación global y no perfiles cualitativamente distintos. Además, los patrones de respuesta poco diferenciados (*straightlining*) son un indicador reconocido de calidad de respuesta que debe examinarse antes de interpretar la estructura de los datos (Kim et al., 2019; Meade y Craig, 2012). Por ello, la utilidad de estas técnicas depende de comprobar la estabilidad de los grupos y su valor añadido respecto de la puntuación global.

En Ecuador, generar evidencia institucional sobre la calidad percibida puede fortalecer la mejora continua en centros ambulatorios, donde los componentes clínicos y administrativos configuran el recorrido del usuario (Suco-Guerrero y Zambrano-Santos, 2025). ProsaludOro S.A., centro privado de especialidades de Machala, ofrece un escenario para examinarla con una escala de percepción basada en SERVQUAL. El objetivo de este estudio fue evaluar la calidad percibida por los pacientes de ProsaludOro S.A. y determinar si el análisis multivariante y la segmentación aportan información adicional a la estadística descriptiva para priorizar oportunidades de mejora. Se plantearon cuatro preguntas de investigación: (PI1) ¿cómo se distribuye la percepción en las cinco dimensiones?; (PI2) ¿existe una estructura multivariante dominante?; (PI3) ¿la segmentación

identifica estratos reproducibles y aporta información adicional respecto de la puntuación global?; y (PI4) ¿qué ítems concentran las principales oportunidades de mejora en el estrato de menor percepción? El estudio es exploratorio: no pretende generalizar sus resultados ni establecer relaciones causales.

Metodología

Diseño y contexto. Se realizó un estudio observacional, transversal y de alcance descriptivo-exploratorio en el Centro de Especialidades ProsaludOro S.A., institución privada de atención ambulatoria ubicada en Machala, provincia de El Oro, Ecuador. El diseño transversal es adecuado para caracterizar la percepción de los usuarios en un momento determinado y explorar la estructura de sus respuestas; no permite establecer relaciones causales ni evaluar cambios en el tiempo. Se optó por un enfoque analítico no supervisado porque la base no contiene expectativas ni una medida independiente de satisfacción global: sin ellas no es posible calcular brechas, y un modelo predictivo cuyo objetivo se derivará de los mismos 30 ítems incurriría en circularidad.

Participantes y muestreo. La población de referencia fueron los pacientes adultos atendidos en el centro. Se empleó un muestreo no probabilístico por conveniencia: el cuestionario se administró en línea (Google Forms) entre el 25 y el 29 de julio de 2026 y se analizaron los 90 cuestionarios completos recibidos, correspondientes a personas de 18 a 71 años. Al no disponerse de un marco muestral ni de un registro verificable de pacientes elegibles, no se estimó la tasa de respuesta ni se aplicó corrección por población finita; los resultados se interpretan a nivel de la muestra observada, sin atribuirles representatividad poblacional. El tamaño muestral es suficiente para los análisis exploratorios sobre las cinco puntuaciones dimensionales (18 observaciones por variable), pero limitado para modelos factoriales a nivel de ítem (Wolf et al., 2013).

Instrumento y evidencias de validez. La calidad percibida se midió con una adaptación local de la escala de percepciones de SERVQUAL (Parasuraman et al., 1988), integrada por 30 ítems distribuidos en cinco dimensiones de seis ítems: elementos tangibles (ítems 1-6), fiabilidad (7-12), capacidad de respuesta (13-18), seguridad (19-24) y empatía (25-30). Las respuestas se registraron en una escala Likert de cinco categorías (1 = totalmente en desacuerdo; 5 = totalmente de acuerdo). No se midieron expectativas, por lo que no se calculan brechas percepción-expectativa.

La validación siguió tres pasos. Primero, la validez de contenido se examinó mediante el juicio de cinco expertos del área de salud, quienes revisaron los ítems. La base disponible no contiene sus valoraciones individuales, por lo que no se calcularon índices formales de validez de contenido (Polit y Beck, 2006). Segundo, la consistencia interna se estimó con alfa de Cronbach (α) y omega total de McDonald (ω); este último se obtuvo con un modelo congénico unifactorial sobre ítems estandarizados, porque relaja el supuesto de tau-equivalencia de α (Hayes y Coutts, 2020). Como diagnóstico de redundancia se calculó la correlación interítem media, dado que una consistencia extremadamente alta puede reflejar ítems redundantes y no una mejor medición (Streiner, 2003). Tercero, la estructura interna se examinó con un análisis factorial confirmatorio

(AFC) a nivel de ítem, que comparó el modelo de cinco factores correlacionados con un modelo unifactorial. Se utilizó estimación de máxima verosimilitud con el paquete semopy (Igolkina y Meshcheryakov, 2020) y el ajuste se valoró con χ^2 , CFI, TLI y RMSEA según los criterios de Hu y Bentler (1999). Dado que la muestra ($n = 90$) es inferior a la recomendada para un modelo con 70 parámetros libres e ítems ordinales (Wolf et al., 2013), el AFC se interpreta como evidencia diagnóstica y no como validación de constructo.

Preparación y control de calidad de los datos. Las respuestas se exportaron a formato .xlsx. El control inicial verificó 90 marcas temporales distintas, ausencia de datos faltantes en los 30 ítems y correspondencia con las cinco categorías previstas; los valores de edad con texto adicional se normalizaron a formato numérico. Se detectó discordancia entre las preguntas «primera visita» y «número de atenciones» en 19 registros (21.1%), posiblemente por diferencias de interpretación; por ello ambas variables se describen por separado y el número de atenciones no se utilizó en las pruebas de asociación. Para cada participante se calcularon la media de los seis ítems de cada dimensión y una puntuación global como promedio de las cinco dimensiones. Como indicadores de calidad de respuesta se obtuvieron la proporción de respuestas favorables (categorías 4-5), el número de categorías distintas utilizadas y la varianza intraindividual de los 30 ítems; una varianza igual a cero se definió como patrón invariante (Kim et al., 2019; Meade y Craig, 2012). Estos cuestionarios no se consideraron inválidos de forma automática: se conservaron en el análisis principal y se excluyeron en un análisis de sensibilidad.

Análisis de datos. El análisis se organizó en cinco etapas vinculadas con las preguntas de investigación (Tabla 2). En la etapa descriptiva, las dimensiones y los ítems se resumieron con media, desviación estándar (DE), mediana, rango intercuartílico (RIC) e intervalo de confianza (IC) del 95%.

Tabla 2. Secuencia analítica: técnicas, propósito y criterios de interpretación.

Etapa	Técnicas	Propósito	Criterio de interpretación
1. Calidad de los datos y de la medida	Auditoría de respuestas; α y ω ; AFC diagnóstico	Verificar la calidad de las respuestas y de la medición	Patrón invariante: varianza intraindividual = 0. Ajuste aceptable: CFI y TLI ≥ 0.95 ; RMSEA ≤ 0.06
2. Descripción	Media, DE, mediana (RIC), IC 95%	Describir la percepción por dimensión (PI1)	Escala de 1 a 5; respuesta favorable: categorías 4-5
3. Estructura	Correlaciones de Spearman; KMO, MSA y Bartlett; PCA	Determinar si existe una estructura dominante (PI2)	Varianza explicada por el primer componente y cargas
4. Segmentación y robustez	K-means ($K = 2$ a 6); bootstrap; Ward; K-means sobre la puntuación global; exclusión de invariantes	Identificar estratos reproducibles y su valor añadido (PI3)	Silhouette y Calinski-Harabasz: mayor es mejor; Davies-Bouldin: menor es mejor; ARI: concordancia entre particiones

Etapa	Técnicas	Propósito	Criterio de interpretación
5. Priorización	Media y proporción de respuestas ≤ 3 por ítem en el estrato de menor percepción	Localizar oportunidades de mejora (PI4)	Ítems con menor media

Fuente: elaboración propia.

Nota. AFC = análisis factorial confirmatorio; DE = desviación estándar; RIC = rango intercuartílico; IC = intervalo de confianza; KMO = Kaiser-Meyer-Olkin; MSA = medida de adecuación muestral; PCA = análisis de componentes principales; ARI = índice de Rand ajustado; PI = pregunta de investigación.

Para examinar la estructura se calcularon correlaciones de Spearman entre las cinco puntuaciones dimensionales. La adecuación de la matriz se evaluó con el índice de Kaiser-Meyer-Olkin (KMO), las medidas individuales de adecuación muestral (MSA) y la prueba de esfericidad de Bartlett. A continuación, se aplicó un análisis de componentes principales (PCA) sobre las puntuaciones estandarizadas (Jolliffe y Cadima, 2016), con el fin de determinar si las dimensiones aportan información diferenciada o responden a una valoración general. El PCA se utilizó como herramienta exploratoria entre dimensiones agregadas y no como prueba de la estructura factorial de los 30 ítems.

Para la segmentación se eligió K-means porque no existe una etiqueta externa de satisfacción y porque permite explorar si los pacientes se agrupan según patrones similares en las cinco dimensiones. Se evaluaron soluciones con $K = 2$ a 6 (100 inicializaciones; semilla 42) mediante los índices Silhouette (Rousseeuw, 1987), Davies-Bouldin (Davies y Bouldin, 1979) y Calinski-Harabasz (Caliński y Harabasz, 1974). La selección de K consideró, además del mejor índice numérico, la parsimonia, el tamaño mínimo de los grupos y la utilidad interpretativa.

La robustez de la solución retenida se evaluó de cuatro maneras. Primero, mediante 500 réplicas bootstrap: en cada una se remuestrearon 90 filas con reemplazo, se ajustó K-means y las observaciones originales se asignaron a los centroides de la réplica; la concordancia con la partición de referencia se cuantificó con el índice de Rand ajustado (ARI; Hubert y Arabie, 1985; Liu et al., 2022). Segundo, mediante la comparación con un agrupamiento jerárquico de Ward. Tercero, mediante la comparación con un K-means aplicado únicamente a la puntuación global: una concordancia perfecta indicaría que los grupos son niveles de intensidad y no perfiles multidimensionales. Cuarto, repitiendo el análisis sin los cuestionarios invariantes.

Los estratos se ordenaron de menor a mayor puntuación global y se denominaron, con fines descriptivos, percepción crítica, favorable y muy favorable. Sus características se compararon con la prueba de Kruskal-Wallis (edad) y con χ^2 de Pearson o, cuando los conteos esperados fueron bajos, con una prueba de permutación Monte Carlo (5,000 réplicas; semilla 20260815). Se informó la V de Cramér como tamaño del efecto y se aplicó la corrección de Benjamini-Hochberg a las cinco comparaciones (Benjamini y Hochberg, 1995). Por último, en el estrato de menor percepción se calcularon, para cada ítem, la media y la proporción de respuestas neutrales o desfavorables ($\leq$

3); los ítems con menor media se interpretan como prioridades descriptivas de mejora y no como determinantes causales.

Software y consideraciones éticas. El análisis se implementó en Python (Google Colab) con pandas, NumPy, SciPy, scikit-learn, semopy y Matplotlib; los procedimientos estocásticos utilizaron semillas fijas. El cuestionario se aplicó de forma digital y la base analizada no contiene identificadores directos.

Resultados

Características de la muestra y calidad de las respuestas. Se analizaron 90 cuestionarios completos. La edad media fue de 35.4 años (DE = 10.6; rango: 18-71); 52.2% de los participantes eran mujeres, 56.7% tenía instrucción de tercer nivel, 57.8% acudía por primera vez y 72.2% recibió consulta médica (Tabla 3).

Tabla 3. Características sociodemográficas y de atención de los participantes (n = 90).

Variable	Categoría	n	%
Edad (años)	18-29	30	33.3
	30-39	37	41.1
	40-49	12	13.3
	50-59	10	11.1
	≥ 60	1	1.1
Sexo	Femenino	47	52.2
	Masculino	43	47.8
Nivel de instrucción	Secundaria	15	16.7
	Tercer nivel	51	56.7
	Cuarto nivel	24	26.7
Primera atención	Sí	52	57.8
	No	38	42.2
Número de atenciones	1 vez	47	52.2
	2 a 3 veces	22	24.4
	4 a 5 veces	7	7.8
	Más de 5 veces	14	15.6
Tipo de atención	Consulta médica	65	72.2
	Consulta especializada	10	11.1
	Exámenes o diagnóstico	3	3.3
	Emergencia	2	2.2
	Otro	10	11.1

Fuente: elaboración propia.

Nota. «Primera atención» y «número de atenciones» proceden de preguntas distintas y fueron discordantes en 19 registros (21.1%); se presentan por separado.

De las 2,700 respuestas emitidas a los 30 ítems, 2,100 (77.8%) correspondieron a las categorías 4 o 5. Veintinueve participantes (32.2%) respondieron de forma invariante a todos los ítems: 15

marcaron la categoría 4, 13 la categoría 5 y uno la categoría 1. Además, 47 participantes (52.2%) utilizaron exclusivamente las categorías 4 y 5, y 67 (74.4%) emplearon solo una o dos categorías a lo largo del cuestionario (Tabla 4).

Tabla 4. Indicadores de uso de la escala y de calidad de respuesta.

Indicador	n	%
Distribución de las respuestas (2,700 respuestas)		
1 = Totalmente en desacuerdo	54	2.0
2 = En desacuerdo	83	3.1
3 = Ni de acuerdo ni en desacuerdo	463	17.1
4 = De acuerdo	1,252	46.4
5 = Totalmente de acuerdo	848	31.4
Categorías distintas utilizadas por participante (n = 90)		
Una categoría (patrón invariante)	29	32.2
Dos categorías	38	42.2
Tres categorías	18	20.0
Cuatro categorías	5	5.6
Cinco categorías	0	0.0
Composición del patrón invariante (n = 29)		
Categoría 4 en los 30 ítems	15	51.7
Categoría 5 en los 30 ítems	13	44.8
Categoría 1 en los 30 ítems	1	3.4
Uso exclusivo de categorías favorables (n = 90)		
Participantes que solo utilizaron las categorías 4 y 5	47	52.2

Fuente: elaboración propia.

Nota. Patrón invariante: varianza intraindividual igual a cero en los 30 ítems.

Fiabilidad, estructura interna y percepción por dimensiones. La consistencia interna fue muy elevada: $\alpha = 0.990$ y $\omega = 0.990$ en la escala completa, y entre 0.947 y 0.975 en las dimensiones (Tabla 5). La correlación interítem media fue de 0.773 en el conjunto de los 30 ítems y osciló entre 0.750 y 0.864 dentro de las dimensiones. En el AFC, el modelo de cinco factores correlacionados no alcanzó los umbrales de ajuste, $\chi^2(395) = 1108.85$, $p < .001$; CFI = 0.849; TLI = 0.833; RMSEA = 0.142, aunque ajustó mejor que el modelo unifactorial, $\chi^2(405) = 1427.99$, $p < .001$; CFI = 0.783; TLI = 0.767; RMSEA = 0.168. En el modelo de cinco factores, las cargas estandarizadas oscilaron entre 0.80 y 0.97 y las correlaciones entre factores latentes, entre 0.87 y 0.96.

Las medias dimensionales se situaron entre 3.96 (capacidad de respuesta) y 4.08 (empatía), con intervalos de confianza superpuestos; la media global fue de 4.02 (DE = 0.78; IC 95%: 3.86-4.19). La proporción de respuestas favorables por dimensión varió entre 74.6% y 79.4%.

Tabla 5. Consistencia interna y estadísticos descriptivos de las dimensiones (n = 90).

Dimensión	α	ω	Media (DE)	IC 95%	Mediana (RIC)	Resp. 4-5 (%)
Tangibles	0.947	0.948	4.01 (0.82)	3.84-4.18	4.00 (3.83-4.62)	79.4
Fiabilidad	0.967	0.967	3.99 (0.84)	3.82-4.17	4.00 (3.50-4.83)	76.7
Capacidad de respuesta	0.966	0.967	3.96 (0.81)	3.79-4.13	4.00 (3.50-4.62)	74.6
Seguridad	0.974	0.975	4.06 (0.82)	3.89-4.24	4.00 (3.67-4.96)	79.1
Empatía	0.968	0.968	4.08 (0.82)	3.91-4.25	4.00 (3.83-5.00)	79.1
Escala global	0.990	0.990	4.02 (0.78)	3.86-4.19	4.00 (3.64-4.67)	77.8

Fuente: elaboración propia.

Nota. α = alfa de Cronbach; ω = omega total de McDonald; DE = desviación estándar; IC = intervalo de confianza; RIC = rango intercuartílico (percentiles 25-75); Resp. 4-5 = proporción de respuestas en las categorías 4 o 5. Cada dimensión consta de seis ítems.

Estructura multivariante. Las cinco dimensiones presentaron correlaciones de Spearman elevadas ($\rho = 0.858$ - 0.948 ; todas $p < .001$). La adecuación de la matriz fue alta (KMO = 0.884; MSA entre 0.853 y 0.935) y la prueba de esfericidad de Bartlett fue significativa, $\chi^2(10) = 689.65$, $p < .001$ (Tabla 6). En el PCA, el primer componente obtuvo un autovalor de 4.636 y explicó 91.7% de la varianza; el segundo explicó 3.0% (Figura 1). Las cargas en el primer componente fueron altas en las cinco dimensiones (0.947-0.981).

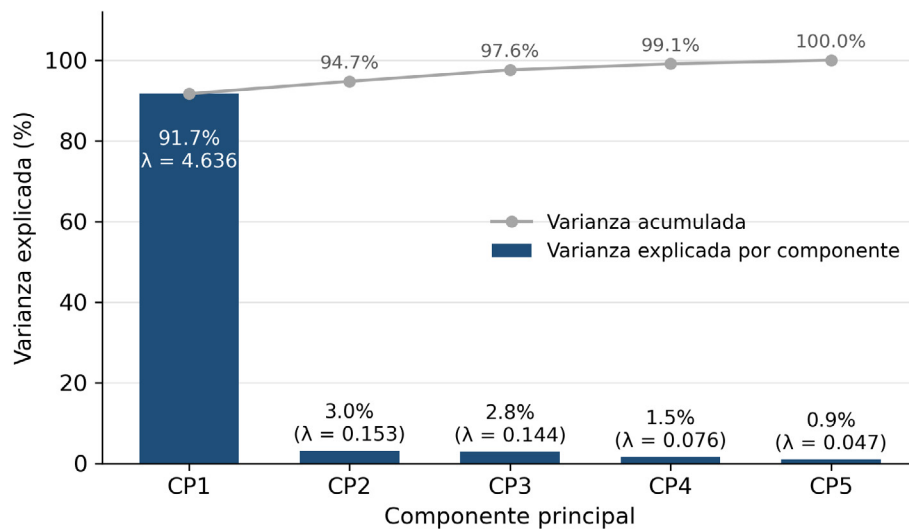
Tabla 6. Correlaciones de Spearman entre dimensiones, adecuación muestral y cargas en el primer componente principal.

Dimensión	1	2	3	4	MSA	Carga en CP1
1. Tangibles	—				0.935	0.947
2. Fiabilidad	0.918	—			0.853	0.981
3. Capacidad de respuesta	0.909	0.948	—		0.898	0.951
4. Seguridad	0.858	0.911	0.888	—	0.866	0.966
5. Empatía	0.860	0.874	0.897	0.923	0.878	0.970

Fuente: elaboración propia.

Nota. Coeficientes ρ de Spearman; todas las correlaciones fueron significativas ($p < .001$; $n = 90$). MSA = medida de adecuación muestral; CP1 = primer componente principal. KMO global = 0.884; prueba de Bartlett: $\chi^2(10) = 689.65$, $p < .001$. Estos índices evalúan la adecuación para la reducción dimensional y no constituyen evidencia de validez de constructo.

Figura 1. Varianza explicada por los componentes principales de las cinco puntuaciones dimensionales.



Fuente: elaboración propia a partir de las respuestas de los 90 participantes.

Nota. CP = componente principal; λ = autovalor.

Segmentación y robustez. Los índices internos favorecieron numéricamente $K = 4$ (Silhouette = 0.611; Davies-Bouldin = 0.511; Calinski-Harabasz = 193.4), pero esa solución aisló un grupo de dos respuestas extremas, con puntuaciones globales de 1.00 y 1.73. La solución $K = 3$ mantuvo índices adecuados (0.584, 0.578 y 153.0, respectivamente) y generó grupos de 19, 41 y 30 participantes, por lo que se retuvo como solución principal (Tabla 7; Figura 2).

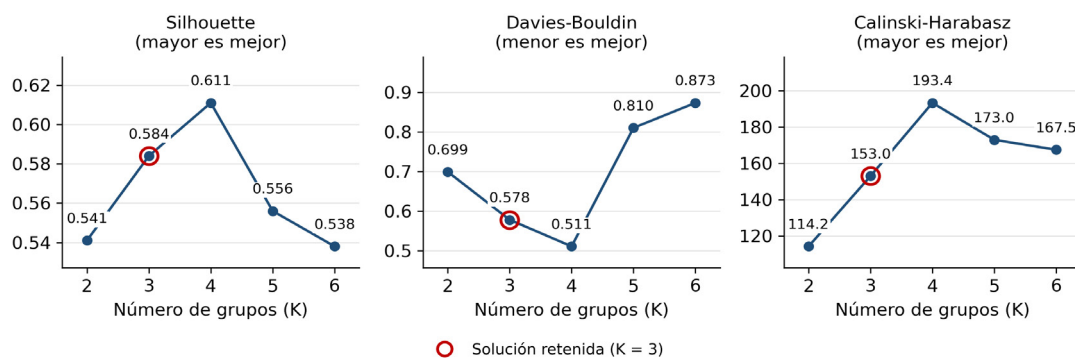
Tabla 7. Índices internos y tamaño de los grupos para $K = 2$ a 6.

K	Silhouette	Davies-Bouldin	Calinski-Harabasz	Tamaño de los grupos
2	0.541	0.699	114.2	24, 66
3	0.584	0.578	153.0	19, 41, 30
4	0.611	0.511	193.4	2, 18, 40, 30
5	0.556	0.810	173.0	2, 8, 15, 35, 30
6	0.538	0.873	167.5	2, 8, 12, 15, 20, 33

Fuente: elaboración propia.

Nota. Silhouette y Calinski-Harabasz: valores mayores indican mejor separación; Davies-Bouldin: valores menores indican mejor separación. En negrita, la solución retenida.

Figura 2. Índices internos de calidad del agrupamiento para K = 2 a 6.



Fuente: elaboración propia.

El estrato de percepción crítica agrupó a 19 participantes (21.1%), con medias inferiores a 3 en las cinco dimensiones; el de percepción favorable, a 41 (45.6%), con medias cercanas a 4; y el de percepción muy favorable, a 30 (33.3%), con medias entre 4.77 y 4.92 (Tabla 8). Los rangos de la puntuación global de los tres estratos no se superpusieron y los perfiles medios fueron paralelos entre dimensiones (Figura 3).

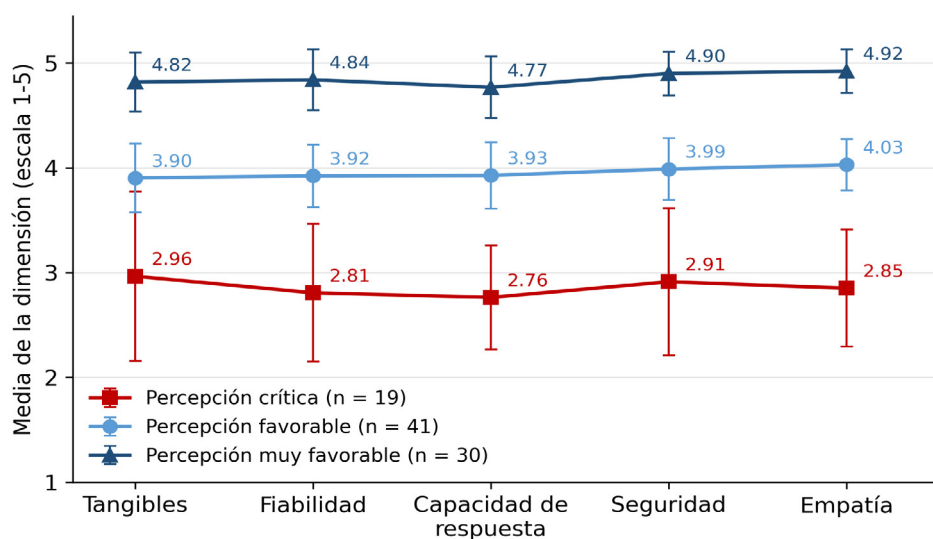
Tabla 8. Medias (DE) de las dimensiones según estrato de percepción (K-means, K = 3).

Dimensión	Percepción crítica (n = 19; 21.1%)	Percepción favorable (n = 41; 45.6%)	Percepción muy favorable (n = 30; 33.3%)
Tangibles	2.96 (0.81)	3.90 (0.33)	4.82 (0.28)
Fiabilidad	2.81 (0.66)	3.92 (0.30)	4.84 (0.29)
Capacidad de res- puesta	2.76 (0.50)	3.93 (0.32)	4.77 (0.30)
Seguridad	2.91 (0.70)	3.99 (0.29)	4.90 (0.21)
Empatía	2.85 (0.56)	4.03 (0.25)	4.92 (0.21)
Puntuación global	2.86 (0.58)	3.95 (0.19)	4.85 (0.19)
Rango de la pun- tuación global	1.00-3.37	3.43-4.37	4.43-5.00

Fuente: elaboración propia.

Nota. DE = desviación estándar. Los estratos se ordenaron de menor a mayor puntuación global.

Figura 3. Perfil medio de las cinco dimensiones según estrato de percepción.



Fuente: elaboración propia.

Nota. Las barras de error representan ± 1 desviación estándar.

En la muestra completa, la partición $K = 3$ fue estable ante el remuestreo (ARI bootstrap: media = 0.895; mediana = 0.968; percentiles 5-95: 0.585-1.000) y concordó con el método de Ward (ARI = 0.857). El K-means aplicado únicamente a la puntuación global reprodujo exactamente la misma partición (ARI = 1.000). Al excluir los 29 cuestionarios invariantes ($n = 61$), α descendió a 0.983, el primer componente explicó 86.7% de la varianza y el índice Silhouette fue mayor para $K = 2$ (0.512) que para $K = 3$ (0.474) (Tabla 9).

Tabla 9. Análisis de sensibilidad de la estructura y la segmentación según la calidad de respuesta.

Indicador	Muestra completa (n = 90)	Sin respuestas invariantes (n = 61)
Puntuación global, media (DE)	4.02 (0.78)	3.87 (0.73)
α global	0.990	0.983
ρ de Spearman entre dimensiones (mínimo-máximo)	0.858-0.948	0.793-0.904
Varianza explicada por el primer componente (%)	91.7	86.7
Silhouette, K = 2	0.541	0.512
Silhouette, K = 3	0.584	0.474
Silhouette, K = 4	0.611	0.482
Tamaño de los grupos, K = 2	24, 66	24, 37
Tamaño de los grupos, K = 3	19, 41, 30	18, 19, 24
ARI entre K-means multidimensional y K-means sobre la puntuación global, K = 3	1.000	1.000

Fuente: elaboración propia.

Nota. El análisis de sensibilidad excluyó los 29 cuestionarios con respuesta invariante en los 30 ítems. ARI = índice de Rand ajustado; DE = desviación estándar.

Caracterización de los estratos. No se observaron diferencias entre los estratos en edad, sexo, nivel de instrucción ni tipo de atención (Tabla 10). La condición de primera visita mostró una asociación nominal con el estrato, $\chi^2(2) = 8.33$, $p = .016$, V de Cramér = 0.304, que no se mantuvo tras la corrección de Benjamini-Hochberg ($q = .078$). Los cuestionarios invariantes se concentraron en los estratos favorable (15 de 41) y muy favorable (13 de 30).

Tabla 10. Características de los participantes según estrato de percepción.

Característica	Crítica (n = 19)	Favorable (n = 41)	Muy favorable (n = 30)	Prueba y tamaño del efecto	p	q
Edad, media (DE)	33.9 (9.1)	34.8 (10.5)	37.2 (11.7)	$H(2) = 0.89$	.640	.749
Mujeres, n (%)	11 (57.9)	23 (56.1)	13 (43.3)	$\chi^2(2) = 1.44$ $V = 0.127$	.486	.749
Instrucción de tercer o cuarto nivel, n (%)	17 (89.5)	34 (82.9)	24 (80.0)	$\chi^2(4) = 3.48$ $V = 0.139$	.499 ^a	.749
Primera visita, n (%)	13 (68.4)	17 (41.5)	22 (73.3)	$\chi^2(2) = 8.33$ $V = 0.304$	.016	.078
Consulta médica, n (%)	15 (78.9)	28 (68.3)	22 (73.3)	$\chi^2(8) = 5.54$ $V = 0.175$	.749 ^a	.749
Respuesta invariante, n (%)	1 (5.3)	15 (36.6)	13 (43.3)	—	—	—

Fuente: elaboración propia.

Nota. DE = desviación estándar; H = prueba de Kruskal-Wallis; V = V de Cramér; q = valor p ajustado por Benjamini-Hochberg para las cinco comparaciones. Las pruebas de nivel de instrucción (tres categorías) y tipo de atención (cinco categorías) se realizaron sobre la tabla completa de categorías. ^a Valor p obtenido por permutación Monte Carlo (5,000 réplicas). La respuesta invariante se presenta con fines descriptivos.

Priorización de ítems en el estrato de percepción crítica. En este estrato, el ítem con menor media fue «El tiempo de espera para recibir atención es adecuado» (media = 2.42; 100% de respuestas ≤ 3). Le siguieron el cumplimiento de los tiempos establecidos, el interés por solucionar los problemas del paciente y el interés por sus necesidades individuales (media = 2.63 en los tres casos), y la agilidad de los trámites (media = 2.68) (Tabla 11; Figura 4). Seis de los diez ítems con menor valoración pertenecen a las dimensiones de fiabilidad y capacidad de respuesta.

Tabla 11. Ítems con menor valoración en el estrato de percepción crítica.

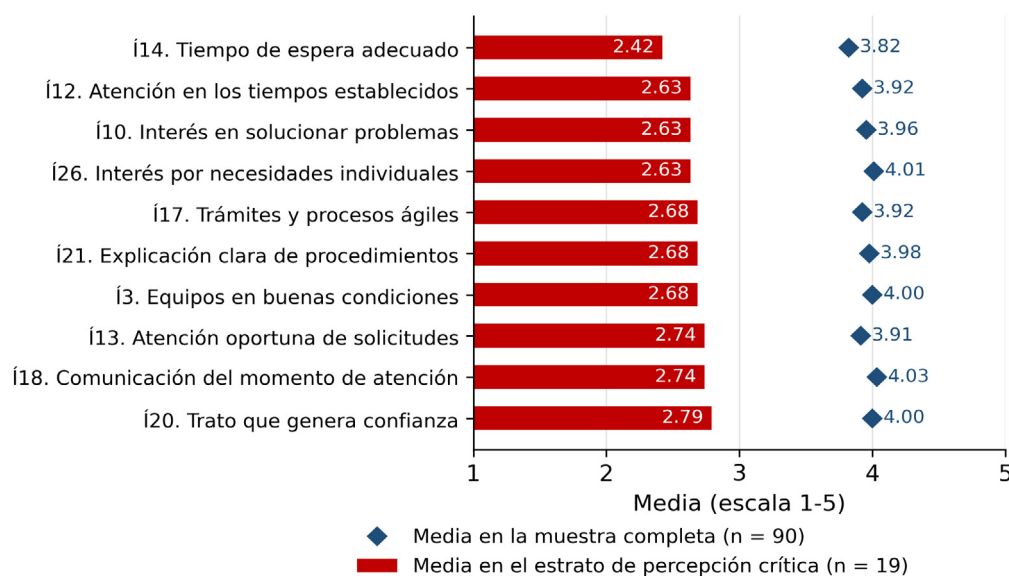
Ítem	Dimensión	Contenido	Total, media	Estrato crítico, media (DE)	Estrato crítico, % ≤ 3	Resto, media
14	Cap. de respuesta	El tiempo de espera para recibir atención es adecuado.	3.82	2.42 (0.69)	100.0	4.20
12	Fiabilidad	La institución brinda la atención ofrecida en los tiempos establecidos.	3.92	2.63 (0.68)	100.0	4.27
10	Fiabilidad	El personal demuestra interés en solucionar los problemas o necesidades del paciente.	3.96	2.63 (0.90)	89.5	4.31

Ítem	Dimensión	Contenido	Total, media	Estrato crítico, media (DE)	Estrato crítico, % ≤ 3	Resto, media
26	Empatía	El personal demuestra interés por las necesidades individuales de cada paciente.	4.01	2.63 (0.68)	100.0	4.38
17	Cap. de respuesta	Los trámites y procesos de atención se realizan de manera ágil.	3.92	2.68 (0.58)	100.0	4.25
21	Seguridad	El personal explica de manera clara los procedimientos realizados durante la atención.	3.98	2.68 (0.89)	89.5	4.32
3	Tangibles	Los equipos utilizados para la atención se encuentran en buenas condiciones.	4.00	2.68 (0.89)	84.2	4.35
13	Cap. de respuesta	El personal atiende oportunamente las solicitudes de los pacientes.	3.91	2.74 (0.56)	100.0	4.23
18	Cap. de respuesta	El personal comunica al paciente cuándo recibirá la atención o el servicio solicitado.	4.03	2.74 (0.65)	94.7	4.38
20	Seguridad	El trato recibido por parte del personal genera confianza y seguridad.	4.00	2.79 (0.63)	94.7	4.32

Fuente: elaboración propia.

Nota. Ítems ordenados de menor a mayor media en el estrato de percepción crítica (n = 19). Cap. = capacidad; DE = desviación estándar; Total = muestra completa (n = 90); % ≤ 3 = proporción de respuestas neutrales o desfavorables; Resto = estratos favorable y muy favorable (n = 71).

Figura 4. Ítems con menor valoración en el estrato de percepción crítica y su media en la muestra completa.



Fuente: elaboración propia.

Discusión

Los resultados responden a las cuatro preguntas planteadas. La percepción fue globalmente favorable y homogénea entre dimensiones (PI1); las respuestas estuvieron dominadas por un

componente general que concentró 91.7% de la varianza (PI2); la segmentación identificó tres estratos estables que, sin embargo, equivalen a niveles de la puntuación global y cuyo número depende del tratamiento de las respuestas invariantes (PI3); y las oportunidades de mejora del estrato crítico se concentraron en la oportunidad del servicio, en particular en el tiempo de espera (PI4).

La calidad de la medición condiciona la interpretación de estos hallazgos. Una consistencia interna de 0.990 y una correlación interítem media de 0.773 son compatibles con redundancia entre ítems y no deben asumirse como evidencia de mayor validez (Streiner, 2003). El AFC apunta en la misma dirección: el modelo de cinco factores no alcanzó los umbrales convencionales de ajuste (Hu y Bentler, 1999) y las correlaciones entre factores latentes (0.87-0.96) indican escasa validez discriminante, un problema descrito desde las primeras evaluaciones de SERVQUAL (Babakus y Boller, 1992). A ello se suma que 32.2% de los cuestionarios fue invariante. Estas respuestas pueden reflejar valoraciones genuinamente uniformes, pero también son compatibles con *straightlining*, aquiescencia, efecto halo o fatiga ante un cuestionario de 30 ítems (Kim et al., 2019; Meade y Craig, 2012; Podsakoff et al., 2003); el diseño no permite distinguir entre estas explicaciones. Cualquiera que sea su origen, este patrón incrementa la covarianza entre ítems y reduce la capacidad del análisis multivariante para detectar una estructura diferenciada: al excluirlo, la varianza del primer componente descendió de 91.7% a 86.7% y la solución preferida pasó de tres a dos estratos. Por tanto, los resultados multivariantes constituyen una descripción condicionada por la calidad de respuesta y no evidencia de una estructura poblacional.

El uso de conglomerados para estudiar la experiencia asistencial cuenta con antecedentes: en pacientes finlandeses se identificaron cuatro grupos de satisfacción no visibles en los promedios (Nurmeksela et al., 2023) y en Trujillo se combinó SERVQUAL con análisis de conglomerados para comparar servicios públicos y privados (Oloya-Salazar et al., 2026). A diferencia de esos trabajos, en este estudio la partición multidimensional coincidió exactamente con la obtenida a partir de la puntuación global, lo que indica que K-means discretizó un continuo en lugar de descubrir tipologías. La aportación del análisis es, por tanto, más modesta pero verificable: muestra que uno de cada cinco participantes se sitúa en un nivel de percepción claramente inferior, algo que la media institucional de 4.02 no revela, y documenta que esa estratificación no añade información dimensional. En muestras mayores, la segmentación por cuantiles o los modelos de mixtura finita evaluados con criterios de información permitirían contrastar si existen clases latentes.

Los hallazgos son coherentes con la evidencia empírica latinoamericana. La proporción de respuestas favorables (77.8%) se aproxima a la satisfacción informada en encuestas nacionales peruanas, entre 70% y 74% (Hernández-Vásquez et al., 2019; Murillo et al., 2019), aunque los indicadores no son directamente comparables. La concentración de las valoraciones más bajas en el tiempo de espera coincide con lo descrito en esas encuestas, donde la demora en la atención fue el problema más frecuente (Murillo et al., 2019), y la relevancia de la capacidad de respuesta también se observó en hospitales de Trujillo (Morales-Garrido et al., 2025) y en centros asistenciales de Quibdó (Duque Oliva et al., 2023). En el Centro de Salud Jipijapa, las dimensiones de calidad se

relacionaron de forma significativa con la satisfacción (Eguez Morales et al., 2025). La diferencia es que en ProsaludOro S.A., un centro privado con una valoración media alta, el problema no aparece en los promedios sino en un subgrupo, lo que respalda analizar la distribución de las respuestas y no solo su tendencia central.

Respecto de las intervenciones de mejora, la literatura de operaciones ofrece referencias para el problema identificado. Las estrategias eficaces para reducir las esperas en consulta externa se agrupan en la alineación de recursos con la demanda, la eficiencia operativa y el rediseño de procesos (Naiker et al., 2018). En un contexto próximo, la implementación del plan «Cero Colas» en cuatro hospitales públicos de Ica (Perú) se acompañó de niveles de satisfacción global superiores a 70% y de mejoras en el cumplimiento del horario de atención y en la información brindada (Becerra-Canales y Condori-Becerra, 2019). En cambio, la evidencia sobre intervenciones Lean en salud es inconsistente y no demuestra efectos sobre la satisfacción del paciente (Moraros et al., 2016), lo que aconseja acompañar cualquier rediseño con mediciones antes y después. Además, la espera no actúa de forma aislada: su efecto negativo se extiende a la valoración del profesional y de la información recibida (Bleustein et al., 2014) y pesa menos que el tiempo que el médico dedica al paciente (Anderson et al., 2007). Esto es coherente con que, en el estrato crítico, junto a la espera aparezcan ítems sobre el interés por las necesidades del paciente y la claridad de las explicaciones.

La asociación entre primera visita y estrato no se mantuvo tras controlar la tasa de falsos descubrimientos ($q = .078$), por lo que solo puede considerarse una hipótesis para estudios posteriores.

Implicaciones para la gestión. De los resultados se derivan tres implicaciones. Primero, el seguimiento debería combinar la media con indicadores de distribución, como la proporción de respuestas ≤ 3 y los cuantiles de la puntuación global, en lugar de institucionalizar etiquetas de segmento cuyo número no es robusto. Segundo, conviene separar las acciones sobre el proceso de medición (acortar el cuestionario, revisar ítems redundantes, incorporar controles de atención y vigilar la proporción de respuestas invariantes) de las acciones sobre el servicio. Tercero, estas últimas deberían dirigirse a los puntos de fricción del recorrido asistencial, tiempo de espera, cumplimiento de los tiempos ofrecidos y agilidad de los trámites y acompañarse de indicadores objetivos de proceso que permitan contrastar la percepción.

Limitaciones. Este estudio presenta limitaciones que acotan el alcance de sus conclusiones. La muestra es pequeña ($n = 90$), no probabilística, procede de un único centro y de cinco días de recolección, por lo que no permite generalizar ni validar externamente los estratos. El instrumento fue revisado por expertos, pero no se dispone de índices de validez de contenido, y el AFC, estimado con una muestra insuficiente y con máxima verosimilitud sobre ítems ordinales, no respaldó la estructura de cinco dimensiones; por ello, las puntuaciones dimensionales deben interpretarse con cautela y el PCA describe la relación entre dimensiones agregadas, no la estructura factorial de los ítems. Un tercio de los cuestionarios fue invariante, lo que limita la potencia del análisis multivariante y hace que el número de estratos dependa de su tratamiento. Por último, no se midieron expectativas ni una variable independiente de satisfacción, de modo que no fue posible

estimar brechas ni construir modelos predictivos. Los estudios futuros deberían ampliar los centros y los periodos, registrar indicadores objetivos de proceso, controlar la calidad de respuesta durante la recolección y realizar una validación psicométrica con estimadores para datos ordinales.

Conclusiones

En ProsaludOro S.A., la calidad percibida fue favorable (media global de 4.02 sobre 5), pero uno de cada cinco participantes se situó en un estrato con medias inferiores a 3 en las cinco dimensiones. La media institucional, por sí sola, no es un indicador suficiente para la gestión: debe acompañarse de la proporción de valoraciones neutrales o desfavorables.

Las cinco dimensiones de la escala no se comportaron como atributos diferenciados, sino como expresiones de una valoración general del servicio. En consecuencia, la segmentación aportó niveles de intensidad y no perfiles de pacientes, y la puntuación global bastó para reproducirla. La calidad de respuesta resultó ser parte del hallazgo y no un detalle técnico: un tercio de cuestionarios invariantes modificó el número de estratos preferido, por lo que toda evaluación de calidad percibida debería informar este indicador antes de interpretar análisis multivariantes.

La prioridad operativa del centro es la oportunidad del servicio: tiempo de espera, cumplimiento de los tiempos ofrecidos y agilidad de los trámites. En muestras pequeñas, el valor del análisis multivariante no reside en descubrir tipologías, sino en poner a prueba y, en este caso, descartar lecturas más complejas de las que los datos pueden sostener. Las líneas futuras son tres: validar una versión abreviada del instrumento en una muestra multicéntrica, incorporar una medida independiente de resultado (satisfacción global o intención de retorno) y evaluar, con indicadores objetivos de tiempo de espera, el efecto de las intervenciones de mejora.

Referencias

- Altuntas, S., Dereli, T., & Erdogan, Z. (2022). Evaluation of service quality using SERVQUAL scale and machine learning algorithms: A case study in healthcare. *Kybernetes*, 51(2), 846–875. <https://doi.org/10.1108/K-10-2020-0649>
- Anderson, R. T., Camacho, F. T., & Balkrishnan, R. (2007). Willing to wait?: The influence of patient wait time on satisfaction with primary care. *BMC Health Services Research*, 7. <https://doi.org/10.1186/1472-6963-7-31>
- Awrahman, B. J., Aziz Fatah, C., & Hamaamin, M. Y. (2022). A review of the role and challenges of big data in healthcare informatics and analytics. *Computational Intelligence and Neuroscience*, 2022. <https://doi.org/10.1155/2022/5317760>
- Babakus, E., & Boller, G. W. (1992). An empirical assessment of the SERVQUAL scale. *Journal of Business Research*, 24(3), 253–268. [https://doi.org/10.1016/0148-2963\(92\)90022-4](https://doi.org/10.1016/0148-2963(92)90022-4)
- Babakus, E., & Mangold, W. G. (1992). Adapting the SERVQUAL scale to hospital services: An empirical investigation. *Health Services Research*, 26(6), 767–786.

- Banda, S., Nkungula, N., Chiumia, I. K., Rylance, J., & Limbani, F. (2023). Tools for measuring client experiences and satisfaction with healthcare in low- and middle-income countries: A systematic review of measurement properties. *BMC Health Services Research*, 23. <https://doi.org/10.1186/s12913-023-09129-9>
- Becerra-Canales, B., Chávez-Espinoza, J. H., Soto-Cárdenas, A. J., Castillo-Romero, P. C., Chávez-Gálvez, O. E., & Puertas-Anicama, V. F. (2024). Calidad del trato al paciente y satisfacción con el servicio de atención primaria. *Ciencia y Enfermería*, 30. <https://doi.org/10.29393/CE30-27CTBV60027>
- Becerra-Canales, B., & Condori-Becerra, A. (2019). Satisfacción de usuarios en hospitales públicos: experiencia del plan «Cero Colas» en Ica, Perú. *Revista Peruana de Medicina Experimental y Salud Pública*, 36(4), 658–663. <https://doi.org/10.17843/rpmpesp.2019.364.4299>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1), 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Bleustein, C., Rothschild, D. B., Valen, A., Valatis, E., Schweitzer, L., & Jones, R. (2014). Wait times, patient satisfaction scores, and the perception of care. *American Journal of Managed Care*, 20(5), 393–400.
- Calinski, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics*, 3(1), 1–27. <https://doi.org/10.1080/03610927408827101>
- Cioffi, A., Ślęzak, D., Robakowska, M., Makanga, P. T., Aleaddini, F., & Cecannecchia, C. (2023). Editorial: Perspectives and opinions in health services: 2022. *Frontiers in Health Services*, 3. <https://doi.org/10.3389/frhs.2023.1295810>
- Cronin, J. J., & Taylor, S. A. (1992). Measuring service quality: A reexamination and extension. *Journal of Marketing*, 56(3), 55–68. <https://doi.org/10.1177/002224299205600304>
- Datt, M., Gupta, A., & Misra, S. K. (2025). Advances in management of healthcare service quality: A dual approach with model development and machine learning predictions. *Journal of Advances in Management Research*, 22(4), 675–701. <https://doi.org/10.1108/JAMR-07-2024-0251>
- Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2), 224–227. <https://doi.org/10.1109/TPAMI.1979.4766909>
- Donabedian, A. (1988). The quality of care: How can it be assessed? *JAMA*, 260(12), 1743–1748. <https://doi.org/10.1001/jama.1988.03410120089033>
- Doyle, C., Lennox, L., & Bell, D. (2013). A systematic review of evidence on the links between patient experience and clinical safety and effectiveness. *BMJ Open*, 3(1). <https://doi.org/10.1136/bmjopen-2012-001570>
- Duque Oliva, E. J., Palacios Palacios, D., & Córdoba Rodríguez, J. E. (2023). Evaluación de la calidad percibida del servicio en los centros asistenciales de la ciudad de Quidbó. *Revista EAN*, (94). <https://doi.org/10.21158/01208160.n94.2023.3752>

- Eguez Morales, J., Santistevan Villacreses, K., Sornoza-Parrales, D., & Parrales Poveda, M. L. (2025). Calidad del servicio y satisfacción del paciente: Modelo SERVQUAL en el Centro de Salud Jipijapa, Ecuador. *Revista de Ciencias Sociales*, 31(4), 447–459. <https://doi.org/10.31876/rcs.v31i4.44873>
- Grönroos, C. (1984). A service quality model and its marketing implications. *European Journal of Marketing*, 18(4), 36–44. <https://doi.org/10.1108/EUM0000000004784>
- Hayes, A. F., & Coutts, J. J. (2020). Use omega rather than Cronbach's alpha for estimating reliability. But... *Communication Methods and Measures*, 14(1), 1–24. <https://doi.org/10.1080/19312458.2020.1718629>
- Hernández-Vásquez, A., Rojas-Roque, C., Prado-Galbarro, F. J., & Bendezu-Quispe, G. (2019). Satisfacción del usuario externo con la atención en establecimientos del Ministerio de Salud de Perú y sus factores asociados. *Revista Peruana de Medicina Experimental y Salud Pública*, 36(4), 620–628. <https://doi.org/10.17843/rpmesp.2019.364.4515>
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification*, 2(1), 193–218. <https://doi.org/10.1007/BF01908075>
- Ibrahim, M. S. (2020). Validating service quality (SERVQUAL) in healthcare: Measuring patient satisfaction using their perceptions in Jordan. *Journal of Information & Knowledge Management*, 19(1). <https://doi.org/10.1142/S0219649220400213>
- Igolkina, A. A., & Meshcheryakov, G. (2020). semopy: A Python package for structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 27(6), 952–963. <https://doi.org/10.1080/10705511.2020.1704289>
- Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065). <https://doi.org/10.1098/rsta.2015.0202>
- Jonkisz, A., Karniej, P., & Krasowska, D. (2021). SERVQUAL method as an “old new” tool for improving the quality of medical services: A literature review. *International Journal of Environmental Research and Public Health*, 18(20). <https://doi.org/10.3390/ijerph182010758>
- Kim, Y., Dykema, J., Stevenson, J., Black, P., & Moberg, D. P. (2019). Straightlining: Overview of measurement, comparison of indicators, and effects in mail–web mixed-mode surveys. *Social Science Computer Review*, 37(2), 214–233. <https://doi.org/10.1177/0894439317752406>
- Lee, P.-C., Liang, L.-L., Huang, M.-H., & Huang, C.-Y. (2022). A comparative study of positive and negative electronic word-of-mouth on the SERVQUAL scale during the COVID-19 epidemic: Taking a regional teaching hospital in Taiwan as an example. *BMC Health Services Research*, 22. <https://doi.org/10.1186/s12913-022-08930-2>
- Liu, C., Li, Y., Fang, M., & Liu, F. (2023). Using machine learning to explore the determinants of service satisfaction with online healthcare platforms during the COVID-19 pandemic. *Service Business*, 17(2), 449–476. <https://doi.org/10.1007/s11628-023-00535-x>

- Liu, T., Yu, H., & Blair, R. H. (2022). Out-of-bag stability estimation for k-means clustering. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 15(6), 781–796. <https://doi.org/10.1002/sam.11593>
- Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, 17(3), 437–455. <https://doi.org/10.1037/a0028085>
- Morales-Garrido, A., Valderrama-Pazos, B., García-Carranza, J., Horna-Velásquez, A., Reyes-Anticona, W., Estela-Vargas, A., & Rojas-Villacorta, W. (2025). Quality and satisfaction in health care: A case study of two public hospitals in Trujillo, Peru. *International Journal of Environmental Research and Public Health*, 22(7). <https://doi.org/10.3390/ijerph22071119>
- Moraros, J., Lemstra, M., & Nwankwo, C. (2016). Lean interventions in healthcare: Do they actually work? A systematic literature review. *International Journal for Quality in Health Care*, 28(2), 150–165. <https://doi.org/10.1093/intqhc/mzv123>
- Murillo, J. P., Bellido-Boza, L., Huamani-Ñahuinlla, P., Garnica-Pinazo, G., Munares-García, O., & Del-Carmen, J. (2019). Satisfacción y tiempo de espera de usuarios de establecimientos de salud peruanos: análisis secundario de ENSUSALUD 2014-2016. *Anales de la Facultad de Medicina*, 80(3), 288–297. <https://doi.org/10.15381/anales.803.16846>
- Naiker, U., FitzGerald, G., Dulhunty, J. M., & Rosemann, M. (2018). Time to wait: A systematic review of strategies that affect out-patient waiting times. *Australian Health Review*, 42(3), 286–293. <https://doi.org/10.1071/AH16275>
- Nurmeksela, A., Kulmala, M., & Kvist, T. (2023). Patient satisfaction—Results of cluster analysis of Finnish patients. *BMC Health Services Research*, 23. <https://doi.org/10.1186/s12913-023-09625-y>
- Okuda, M., Yasuda, A., & Tsumoto, S. (2021). An approach to exploring associations between hospital structural measures and patient satisfaction by distance-based analysis. *BMC Health Services Research*, 21. <https://doi.org/10.1186/s12913-020-06050-3>
- Oloya-Salazar, P., Alayo-Ruiz, E., Vallejos-Salas, K., Ruiton-Castillo, M., Peña-López, J., Anicama-Ramirez, K., & Rojas-Villacorta, W. (2026). Public and private healthcare service quality in Trujillo, Peru: Evidence from a SERVQUAL survey. *Healthcare*, 14(6). <https://doi.org/10.3390/healthcare14060738>
- Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1985). A conceptual model of service quality and its implications for future research. *Journal of Marketing*, 49(4), 41–50. <https://doi.org/10.1177/002224298504900403>
- 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.
- Piovani, D., & Bonovas, S. (2022). Real world—big data analytics in healthcare. *International Journal of Environmental Research and Public Health*, 19(18). <https://doi.org/10.3390/ijerph191811677>
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5), 879–903. <https://doi.org/10.1037/0021-9010.88.5.879>

- Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Research in Nursing & Health*, 29(5), 489–497. <https://doi.org/10.1002/nur.20147>
- Rahim, A. I. A., Ibrahim, M. I., Musa, K. I., Chua, S.-L., & Yaacob, N. M. (2021). Patient satisfaction and hospital quality of care evaluation in Malaysia using SERVQUAL and Facebook. *Healthcare*, 9(10). <https://doi.org/10.3390/healthcare9101369>
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- Simsekler, M. C. E., Alhashmi, N. H., Azar, E., King, N., Luqman, R. A. M. A., & Al Mulla, A. (2021). Exploring drivers of patient satisfaction using a random forest algorithm. *BMC Medical Informatics and Decision Making*, 21. <https://doi.org/10.1186/s12911-021-01519-5>
- Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, 80(1), 99–103. https://doi.org/10.1207/S15327752JPA8001_18
- Subrahmanya, S. V. G., Shetty, D. K., Patil, V., Hameed, B. M. Z., Paul, R., Smriti, K., Naik, N., & Somani, B. (2022). The role of data science in healthcare advancements: Applications, benefits, and future prospects. *Irish Journal of Medical Science*, 191(4), 1473–1483. <https://doi.org/10.1007/s11845-021-02730-z>
- Suco-Guerrero, I. I., & Zambrano-Santos, R. O. (2025). Aplicación del modelo SERVQUAL para medir la satisfacción del usuario en la atención administrativa sanitaria. *Journal of Economic and Social Science Research*, 5(4), 1–17. <https://doi.org/10.55813/gaea/jessr/v5/n4/214>
- Tanantong, T., Pannakkong, W., Chemkomnerd, N., & Boonkwan, P. (2026). Interpretable SHAP-based machine learning framework for patient satisfaction prediction: A case study in Thammasat University Hospital. *BMC Medical Informatics and Decision Making*, 26. <https://doi.org/10.1186/s12911-026-03647-2>
- Torain, M. J., Bennett, G. G., Matsouaka, R. A., Olsen, M. K., Yang, H., Bolton, J. H., Johnson, K. S., & Svetkey, L. P. (2021). The patient's point of view: Characterizing patient-level factors associated with perceptions of health care. *Health Equity*, 5(1), 457–465. <https://doi.org/10.1089/heq.2021.0062>
- Villa-Velásquez, J., Araya, A.-X., Reynaldós-Grandón, K., Rivera-Rojas, F., & Valencia-Contrera, M. (2023). Evaluación del modelo de calidad de Donabedian en base a los criterios de Chinn y Kramer. *Horizonte de Enfermería*, 34(2), 203–215. https://doi.org/10.7764/Horiz_Enferm.34.2.203-215
- Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety. *Educational and Psychological Measurement*, 73(6), 913–934. <https://doi.org/10.1177/0013164413495237>

Autores

Joseph Giovanni Villacreses Quevedo. Analista de adquisiciones experto en Gestión de la Calidad, Economista Agropecuario y Magister en Gestión de la Calidad.

Milena Jamileth Gualán Torres. Licenciada en Administración de Empresas, Cajera Financiera y Back up Supervisor Operativo.

Harry Vite Cevallos. Docente investigador, Ingeniero Comercial, Magister en Estadística Aplicada

Wiliam Tranquilino Medina Castillo. Docente investigador, Ingeniero Comercial, Lic. En Contabilidad y Auditoría, Maestría en Auditoría y Contabilidad, Especialista en Tributación, Coordinador de la Carrera de Turismo de la Facultad de Ciencias Empresariales de la Universidad Técnica de Machala.

Declaración

Conflicto de interés

No tenemos ningún conflicto de interés que declarar.

Financiamiento

Sin ayuda financiera de partes externas a este artículo.

Nota

El artículo es original y no ha sido publicado previamente.