

Comparación de las escalas de Likert y Vigesimal para la evaluación de satisfacción de atención en un hospital del Perú.

LLANOS ZAVALAGA Fernando*, ROSAS AGUIRRE Angel**, MENDOZA REQUENA Daniel**, CONTRERAS RÍOS Carlos***

SUMMARY

Objectives: To compare patient satisfaction using the Likert and twenty-point scales in patients that were attended in a Peruvian hospital. **Material and Methods:** In this cross-sectional study, we included 2445 patients of the “Instituto Nacional de Salud del Niño” during March 1998. Patients were asked about their satisfaction with a questionnaire that was performed through the two scales. **Results:** The two scales had weak association ($\rho=0.48$ $p<0.001$), and a determinant coefficient of 0.23. According the definition of the satisfaction, the percent of patients “not insatisfied” was higher in the vigesimal scale than the Likert scale (37.9% vs 21.4%, $\chi^2=160.8$ $p<0.001$). The twenty-point scale had 70.9% of sensitivity and 71.0% of specificity, with the Likert scale as a gold standard. **Conclusions:** Measurement of patient satisfaction is different according the scale utilized. The election of a scale must consider the objectives of investigation. (*Rev Med Hered* 2001; 12: 52-57).

KEYWORDS: Patient satisfaction, Likert scale, scale measurement.

RESUMEN

Objetivo: Comparar la medición de la satisfacción de pacientes atendidos en un hospital del Perú mediante el uso de las escalas Likert y Vigesimal. **Material y métodos:** Estudio transversal. 2445 pacientes atendidos en el Instituto de Salud del Niño fueron evaluados mediante una encuesta durante el mes de marzo de 1998, para medir satisfacción mediante dos escalas: Likert y Vigesimal. **Resultados:** Existe asociación significativa baja entre ambas escalas ($\rho=0.483$, $p<0.001$), con un coeficiente de determinación de 0.23. Según la definición de satisfacción, el porcentaje de usuarios “no insatisfechos” (37.9%) para la escala vigesimal fue mayor al comparar al de la escala Likert (21.4%) ($\chi^2=160.8$, $p<0.001$). Tomando como gold standard la escala de Likert para la medición de la satisfacción, la escala vigesimal tuvo una sensibilidad de 70.9% y una especificidad de 71.0%. **Conclusiones:** Los resultados obtenidos en satisfacción de la atención difieren según la escala usada para su medición. La elección de una determinada escala debe considerar los objetivos de la investigación. (*Rev Med Hered* 2001; 12: 52-57).

PALABRAS CLAVE: Satisfacción del usuario, escala de medición, Likert.

* Magister en Economía en Salud. Profesor Asociado de la Facultad de Salud Pública y Administración. Universidad Peruana Cayetano Heredia.

** Médico Cirujano. Unidad de Gestión. Facultad de Salud Pública y Administración. Universidad Peruana Cayetano Heredia.

*** Magister en Gestión de Programas Sociales. Facultad de Salud Pública y Administración. Universidad Peruana Cayetano Heredia Magister en Gestión de Programas Sociales. Facultad de Salud Pública y Administración. Universidad Peruana Cayetano Heredia.

INTRODUCCIÓN

En la actualidad existe un creciente interés por parte de las organizaciones de salud por evaluar la calidad de sus servicios. En este sentido, la satisfacción del paciente constituye uno de los más versátiles indicadores de calidad (1,2), por ser una medida de los resultados de la atención y un predictor de fidelización del cliente (3,4). Para medir la satisfacción del usuario se disponen de métodos cualitativos y cuantitativos. Dentro de estos últimos, los cuestionarios usando métodos de estimación directa, son instrumentos que permiten analizar cuantitativamente la percepción de los niveles de calidad por parte del usuario utilizando para ello diferentes tipos de escalas de evaluación.

La escala gráfica es uno de los tipos de escalas más antiguas y simples que existen (5), y dentro de éstas la de mayor representatividad es la "Escala Likert" (6) (Rensis Likert, 1932). Esta se ha convertido en una las técnicas psicométricas más populares para generar escalas, utilizada por economistas, sociólogos y psicólogos en el desarrollo de cuestionarios (7-10). El método es relativamente rápido y permite que el sujeto establezca un estimado cuantitativo de la magnitud de un atributo. La percepción del usuario se extiende sobre lo que él o ella opina a favor o en contra de la atención, en una escala de 3 a 9 puntos, que va desde desacuerdo hasta acuerdo.

En los estudios hechos en servicios de salud del Perú (11-15), se ha medido la calidad de atención desde la perspectiva del paciente utilizando la escala vigesimal, la cual se usa frecuentemente en la calificación en los diferentes niveles de educación en nuestro país. Se parte del supuesto que puntajes menores a 10 significan mal servicio o desaprobación de la atención, y puntajes mayores a 10 buen servicio o atención aceptable, sin embargo esta suposición no ha sido confirmada.

El presente estudio tiene por objetivo comparar dos escalas (Likert y Vigesimal) de medición del grado de satisfacción de pacientes atendidos en el Instituto de Salud del Niño de Lima, Perú durante Marzo de 1998. Este estudio será de valiosa información en el desarrollo de futuros cuestionarios para que investigadores en calidad del servicio, elijan la escala de medición que se acerque más a sus objetivos.

MATERIAL Y METODOS

Como parte de un estudio transversal que evaluó la calidad del servicio desde la perspectiva de los pacientes, en el Instituto de Salud del Niño de Lima, Perú, se utilizó

un cuestionario basado en la teoría Servqual. En dicho cuestionario se incluyeron dos preguntas para medir la satisfacción global del usuario, usando escalas de respuesta tipo Likert y Vigesimal. El cuestionario Servqual es uno de los más utilizados en la actualidad ya que permite medir la satisfacción del paciente evaluando el servicio recibido y cuantificando la diferencia entre el servicio percibido y las expectativas del paciente (15).

Se consideró como universo al total de consultas externas, altas de hospitalización, consultas de emergencia y servicios intermedios (laboratorio, rayos X, farmacia, etc.). Se determinó un tamaño muestral de 2000 encuestas. Las personas que contestaron el cuestionario fueron los padres o apoderados de los niños que recibieron atención en dichos servicios. Se realizó un conteo del movimiento de personas por hora del día, con lo cual se estableció el cronograma de encuestas a los pacientes, controlando el sesgo de aglomeración. Los encuestadores debidamente capacitados fueron ubicados estratégicamente en las puertas de salida del Instituto de Salud del Niño. Se excluyó del estudio aquellas encuestas, en las que no habían respondido a la vez las dos preguntas que correspondían a las dos escalas.

La escala de Likert usada consistió en una escala ordinal continua de 7 valores, en donde 1 correspondía a "muy malo", 2 a "malo", 4 a "regular", 6 a "bueno" y 7 "muy bueno". La escala Vigesimal es una escala comúnmente usada en la educación de nuestro país, conformada en realidad por 21 puntos (0-20). A nivel escolar, puntajes de 10 o menos significan desaprobación o inaceptable, el 0 pésimo o ausencia total de conocimientos y/o competencias y el 20 óptimo, conocimiento total y/o competencia insuperable de aquello que se está evaluando.

Los resultados fueron expresados en frecuencias absolutas y relativas. Se consideró como "no insatisfecho" cuando la atención recibida tuvo un valor de 7 (muy buena) en la escala de Likert (16) y valores de 18 a 20 en la escala vigesimal (15); y como "insatisfecho" valores diferentes de 7 en la escala de Likert y valores diferentes de 18 a 20 en la escala vigesimal. En el análisis estadístico se utilizaron el test de ji cuadrado (χ^2) y la correlación de Spearman para determinar asociación entre los resultados de la Escala de Likert y Vigesimal. En el procesamiento y análisis de los datos se utilizó el paquete estadístico SPSS 9.0 para Windows. A todos los pacientes se les solicitó un consentimiento informado, explicándoseles los objetivos del estudio en forma clara, directa y veraz.

RESULTADOS

Se incluyeron a 2445 pacientes, quienes respondieron las preguntas de ambas escalas. De ellos, usando la escala de Likert, 21.4% calificó la atención recibida con una puntuación de 7 (muy buena), 42.9% con una puntuaciones de 5 y 6, y sólo 7.6% con puntuaciones de 3, 2 y 1 (Tabla N° 1).

En las calificaciones obtenidas mediante la escala vigesimal se encontró que el mayor porcentaje (37.9%) dieron un puntaje de 18 o más a la atención recibida, el 35.2% entre 11 y 15 puntos y sólo un 3.3% la desaprobaron (menor ó igual a 10) (Tabla N° 2).

Se observó asociación significativa entre ambas mediante la correlación de Spearman ($\rho=0.483$, $p<0.001$) con un coeficiente de determinación de 0.23.

Se compararon los resultados de ambas escalas, encontrándose que de los pacientes que calificaron la atención en la escala de Likert como 1 (muy mala), el 50.0% dieron un puntaje entre 11 y 15, y un 16.7% entre 18 y 20 en la escala vigesimal, con un promedio de 13.3 ± 3.4 . Para la puntuación 4 (regular) de Likert, el 20.3% calificó con 18 a 20, el 54.3% con 11 a 15 y 0.6% con 0 a 5 en la escala vigesimal; con un promedio de 15.4 ± 2.4 . Así, para la puntuación 7 (muy buena) de Likert el mayor porcentaje (70.8%) correspondió a las puntuaciones entre 18 y 20 en la escala vigesimal, y 0,2% a puntuaciones entre 0 y 5 (Tabla N°3), con un promedio de 18.0 ± 2.0 .

Según la definición de grado de satisfacción, el porcentaje de usuarios "no insatisfechos" (37.9%) fue mayor al comparar con el porcentaje obtenido mediante

Tabla N°1. Calificación del servicio según la escala de Likert.

Escala de Likert (n=2445)	%
1 (Muy mala)	1.7
2 (Mala)	1.9
3	3.9
4 (Regular)	19.2
5	21.4
6 (Buena)	20.6
7 (Muy buena)	21.4

Tabla N°2. Calificación del servicio según la escala vigesimal.

Escala vigesimal (n=2445)	%
0 - 5	0.3
6 - 10	2.6
11 - 15	35.2
16 - 17	24.0
18 - 20	37.9

Tabla N°3. Calificaciones en la Escala de Likert y Vigesimal (%).

Puntaje	n	0 - 5 (n=8)	6 - 10 (n=63)	11 - 15 (n=861)	16 - 17 (n=586)	18 - 20 (n=927)	Calificación promedio
1	42	2.4	21.4	50.0	9.5	16.7	13.3 ± 3.4
2	47	4.3	10.6	59.6	10.6	14.9	13.9 ± 3.3
3	96	0.0	8.4	63.5	13.5	14.6	14.4 ± 2.8
4	713	0.6	3.2	54.3	21.6	20.3	15.4 ± 2.4
5	522	0.0	2.3	40.8	29.7	27.2	16.0 ± 2.2
6	503	0.0	0.4	17.5	34.0	48.1	17.2 ± 1.7
7	522	0.2	0.8	12.1	16.1	70.8	18.0 ± 2.0

Tabla N°4. Sensibilidad y especificidad de la escala Vigesimal usando como gold standard a la escala de Likert.

	SATISFECHO	INSATISFECHO	TOTAL
18 - 20	370	557	927
< 18	152	1366	1518
TOTAL	522	1923	2445

Especificidad: 70.9%
Sensibilidad: 71.0%

diseñadas para medir lo mismo: grado de satisfacción del usuario. Por tanto dicha relación es considerada débil (17), explicando sólo el 23% de la variabilidad de la Escala Vigesimal.

Ambos resultados sugieren un desfase entre las calificaciones obtenidas por ambas escalas y la percepción de los clientes sobre la calidad de atención, lo cual se hace más evidente al comparar los porcentajes de usuarios “no insatisfechos” ($p < 0.001$). Por tanto, el problema de elegir un determinado número de opciones es el efecto que pueda producir esta decisión en la confiabilidad de la escala y en la capacidad de discriminar (precisión) entre los grados de las percepciones de un ítem. Respecto a esto, podría pensarse que aquella escala con mayor cantidad de puntos (vigesimal) sería mejor que una con menor cantidad (Likert), sin embargo, ello la escala tipo Likert (21.4%) para la escala vigesimal ($\chi^2=160.8$, $p < 0.001$).

Tomando como gold standard la escala de Likert para la medición de la satisfacción del usuario, se obtuvo para la escala vigesimal una sensibilidad de 70.9% y una especificidad de 71.0% (Tabla N°4).

DISCUSIÓN

El interés por conocer cuál es el número óptimo de puntos que debe tener una escala que evalúe la satisfacción del usuario o cliente no es tema nuevo; y su importancia radica en lograr una adecuada medición de la calidad del servicio, que permita la toma de decisiones, con la finalidad de mejorar la calidad del servicio. No existe una escala perfecta en los instrumentos para todos los usos, ya que cada escala debe considerar los objetivos de la investigación y qué se harán con los resultados. Así, el investigador debe decidir si es necesario obtener información extra, que podría de alguna forma confundir la medición en la toma de decisiones. Algunos autores sostienen que la mejor escala debe ser fácilmente entendida por los respondientes, discriminar bien sus percepciones (precisión), ser fácil de interpretar y tener un mínimo sesgo de respuesta (2,4).

En este estudio se encontró que un 16.7% de los pacientes que calificaron con 1 (muy mala) la atención en la Escala de Likert, la calificaron excelente (entre 18 y 20) en la escala vigesimal, y el 50% le dieron un puntaje entre 11 y 15, lo cual es considerado como aceptable o regular. En los cuestionarios que calificaron con 7 (muy buena) en la escala Likert los mayores porcentajes se ajustaron más a lo esperado en la escala vigesimal con 70.9% para 18-20, siendo los porcentajes correspondientes a 0-5 menores a 0.5%. Se observó además, que para la calificación 1 (muy mala) en la escala

de Likert se obtuvo un promedio de 13.3, para la calificación 4 (regular) en la escala de Likert se obtuvo un promedio de 15.3 y para la calificación 7 “muy buena” se obtuvo un promedio de 18.0. Esto sugeriría que la escala de Likert (del 1 a 7) para medición de satisfacción de atención de pacientes, no corresponde verdaderamente a una escala del 0 al 20, sobretodo si tenemos en cuenta que las puntuaciones en la escala Likert tienen promedios de calificación en la escala vigesimal que fluctúan entre 13.3 y 18.0.

Al hacer el análisis de correlación de Spearman se encontró que la Escala de Likert tuvo relación lineal positiva con la Escala Vigesimal ($Rho=0.483$, $p < 0.001$); sin embargo se esperaba un coeficiente de relación más cercano a 1, debido a que ambas escalas fueron no es necesariamente cierto.

En la literatura sobre este tópico no se encuentra un consenso. Sin embargo, una pauta a seguir es que el número de puntos no debe ser menor que la habilidad del evaluador para discriminar el resultado ya que resultaría en una pérdida de la información. Chang (18) resumió la mayoría de opiniones, encontrando investigadores que sostienen que la confiabilidad es independiente del número de puntos de una escala, así como otros que plantean que dicha confiabilidad es maximizada usando una escala ya sea de 7, 5, 4 ó 3 puntos. Cohen (19) sostiene que por lo menos 3 puntos deben ser usados en las escalas. Bass y col.(18), en cambio, afirman que escalas con más de 9 puntos pueden ser efectivamente usadas por los respondientes y Conklin planteaba que un número límite razonable sería 10 a 15 puntos (20).

A pesar de todas estas opiniones, encontramos que la mayoría de instrumentos diseñados para medir satisfacción del usuario usan escalas validadas tipo Likert que tienen de 3 a 7 puntos. En nuestro país, la escala vigesimal ha sido utilizada para evaluar la satisfacción, de manera similar en que se usa para calificar a los estudiantes de nivel escolar y superior, no encontrándose evidencia de una validación previa. Así, considerando a la escala de Likert como gold standard, obtenemos que la escala vigesimal logra identificar un 70.9% de todos los usuarios “satisfechos” (sensibilidad) y 71.0% de todos los usuarios “insatisfechos” (especificidad), lo cual haría que la escala vigesimal sea poco válida para la medición de la satisfacción del usuario.

Por tanto, para explicar las elevadas puntuaciones en la escala vigesimal es conveniente tener en cuenta los sesgos de respuestas que pueden ocurrir en una escala que tiene mayor cantidad de puntos: abuso de puntos

extremos, aversión por puntos extremos, desviación a la derecha y diferente patrón de referencia de acuerdo a la cultura. El primero, afectaría tanto a la escala Likert como la vigesimal; sin embargo, en esta última tiene mayor influencia en nuestro estudio, ya que los evaluadores frecuentemente usaron las posiciones extremas de la escala reduciendo probablemente, la precisión de la medición. Una explicación para este sesgo estaría en relación con el bajo grado de instrucción y el pobre conocimiento del tema evaluado por parte del usuario. Así, Chang plantea que un factor a considerar en la elección de una escala es el conocimiento del respondiente sobre la materia evaluada. Una carencia de este conocimiento podría llevar a un abuso de los puntos extremos en las escalas largas, resultando en una menor confiabilidad al comparar con escalas cortas (18).

El segundo sesgo llamado sesgo de aversión por puntos extremos o sesgo de tendencia central, se refiere al rechazo de uso de categorías extremas, por la dificultad o temor de la gente para hacer juicios absolutos. Este efecto reduciría el rango de posibles respuestas resultando en pérdida de sensibilidad y confiabilidad. En nuestro estudio este sesgo, se dio de manera similar, por lo que no alcanzaría importancia en la comparación de ambas escalas.

El tercer sesgo, desviación a la derecha, es una situación que ocurre comúnmente en escalas que evalúan personal o estudiantes, sobretodo si estas escalas tienen un buen número de puntos. Consiste en la tendencia de algunos respondientes de ser incapaces de reportar sentimientos o comportamientos que ellos perciben como socialmente indeseables, o por el contrario incapaces de separar sus sentimientos de la materia evaluada, calificándola como positiva. El resultado sería un alto porcentaje de excelencia o satisfacción como se ve en mayor medida en la calificación de la escala vigesimal, siendo imposible detectar alguna distinción entre tantos grados de excelencia.

Finalmente, debemos señalar la importancia dada a los patrones de referencia que tienen los usuarios en los tópicos de estudio (18). Por tanto, en una escala con más opciones de respuesta, como la escala vigesimal podría introducirse un sesgo de respuesta. Estos patrones de referencia dependen de la cultura y concepción de la vida por la sociedad, y podemos encontrar que se manifiesta de dos formas en nuestro medio. En primer lugar, la escala vigesimal es usada frecuentemente para calificar el esfuerzo del estudiante de nivel escolar o superior para adquirir conocimientos, ya sea mediante exámenes o trabajos de investigación.

Es común observar que cuando el conocimiento es nulo o no se presenta ningún trabajo, no se le califica al estudiante con un cero, sino con un rango de notas suele fluctuar entre 5 y 10. Así, cuando uno presenta un examen o trabajo deficiente, es calificado con un 11 no merecido, siendo finalmente considerado "aprobado". Todo esto resultaría, en un no uso de la escala vigesimal en su integridad, convirtiéndose en una escala de 5 ó 10 a 20, más que de 0 a 20. Además se debe señalar que pocas veces los evaluadores califican a sus estudiantes con notas de excelencia (entre 19 y 20), lo que resultaría aún en un menor rango de calificaciones posibles de esta escala, correlacionando con los promedios obtenidos por las puntuaciones de la escala Likert que fluctuaron entre 13.3 y 18.0.

En segundo lugar, debe considerarse el tiempo en que fue realizada la encuesta al usuario y la última persona que tuvo contacto con dicho usuario, los cuales podrían influir positivamente en la satisfacción del usuario por la atención recibida. Así, una encuesta realizada justo al momento de terminada la atención y el hecho de ser atendidos finalmente por un médico, no tomando en cuenta las otras personas o procesos que componen la atención; haría poco probable, en nuestro medio, calificaciones menores de 10 en la escala vigesimal, así como puntuaciones menores de 4 en la escala de Likert.

Esperamos que el presente trabajo sirva de información para que investigadores en calidad de servicios elijan la escala apropiada para medir satisfacción del usuario, de acuerdo a sus necesidades.

Correspondencia:

Angel Martín Rosas Aguirre
Dirección: Calle 10, Mz I, Lt 1. La Calera de la Merced.
Surquillo, Lima, Perú.
Teléfono: 511-4818283
Correo electrónico: amrosasa@hotmail.com

REFERENCIAS BIBLIOGRAFICAS

1. Carr-Hill RA. The measurement of patient satisfaction. *Journal of Public Health Medicine* 1992; 14(3):236-349.
2. Wilkin D, Hallam L, Doggett MA. *Measures of Need and Outcome for Primary Health Care*. Oxford University Press, 1992.
3. Farache L. El Servicio: Fuente de Fidelidad. Investigación y Marketing-La revista de AEDEMO. 1999. Disponible en: <http://www.aedemo.es/diciembre97/5704.html>.
4. Hutchinson A, McColl E, Christie M, Ricalton C. *Health Outcome Measures in Primary and Out-Patient Care*. Hardwood Academic Publishers. 1996.
5. Streiner DL, Geoffrey RN. *Health Measurement Scales*.

- Oxford University Press, 1991.
6. Likert RA. A technique for development of attitude scales. *Archives of Psychology* 1932; 140:44-53.
 7. Cassus Becerra S, Fuentes Vásquez C, Valdés Valenzuela C: Actitudes de autocuidado de embarazadas con respecto a alimentación, accidentes y salud oral de sus futuros hijos. Tesis para obtener el grado de Profesional de Enfermería. 1998. Pontificia Universidad Católica de Chile. Santiago de Chile, Chile.
 8. Morales Ramírez M, Ocampo Andréyeva V, De la Mora L, Alvarado Calvillo R: Validez y confiabilidad del cuestionario clínico del síndrome depresivo. *Ginecología y Obstetricia de México* 1997; 65 (6): 11-15.
 9. Gadbury-Amyot CC, Williams KB, Krust-Bray K, Manne D, Collins P. Validity and reliability of the oral health-related quality of life instrument for dental hygiene. *J Dent Hyg* 1999; 73(3):126-34.
 10. Millard AD. Scaling the quality of clinical audit projects: a pilot study. *Int J Qual Health Care* 1999; 11 (3):241-9.
 11. Universidad Peruana Cayetano Heredia. Estudio de medición de la Calidad del Servicio a los Usuarios y Orientación al Mercado de los Ejecutivos del Instituto de Seguridad Social en los Centros Asistenciales de la Gerencia Departamental de Lima. Lima, Perú, 1998.
 12. Universidad Peruana Cayetano Heredia. Estudio de medición de la Calidad del Servicio a los usuarios del Instituto de Salud del Niño y orientación al mercado de los ejecutivos del ISN. Lima, Perú, 1998.
 13. Superintendencia de Entidades Prestadoras de Salud. Encuesta Metropolitana de Satisfacción del Usuario de Seguros de Salud. Lima, Perú, 1999.
 14. Universidad de Ingeniería. Estudio de la medición de la calidad de Macrorregiones Sur y Centro. Lima, Perú, 1998.
 15. Michelsen Consultores. Estudios en calidad en servicios de salud. 1996-1997. Lima, Perú, 1998.
 16. Blanchard K, Bowles S. *Raving Fans. A Revolutionary Approach to customer service.* William Morrow and Company Inc. New York. 1993.
 17. Dawson-Saunders B, Trapp R. *Bioestadística médica. Manual Moderno.* México, 1997.
 18. National Computer Systems-NCS. Response scales: How many points and what labels? Research notes. July 96. Disponible en: <http://www.ncs.com>.
 19. Cohen J, Cohen P. *Applied multiple regression/correlation analysis for the behavioral sciences (2nd end.).* Hillsdale, NJ: Erlbaum. 1983.
 20. Conklin ES. *The scale of value method for studies in genetic psychology.* University of Oregon Publications, Eugene. 1923.