6

DISEÑO Y VALIDACIÓN DE UNA PRUEBA PARA EVALUAR LA DISPOSICIÓN HACIA EL SERVICIO AL CLIENTE

Carlos Forero Aponte*, Mónica García Rubiano,
Martha Liliana Ramírez Garzón y Syndy Viviana
Sastoque

Universidad Católica de Colombia, Bogotá, Colombia.

Introducción

El servicio al cliente es crítico en el éxito de las organizaciones, ya que son los empleados quienes gestionan la relación entre estas y los consumidores de sus bienes y servicios; con esto, se debe tener en cuenta que la percepción que tienen los empleados es diferente a la que realizan los usuarios, por lo que se puede afirmar que la percepción de la calidad resulta de la comparación que hacen los clientes entre sus expectativas y la percepción de cómo se les ha prestado el servicio (Hirmukhe, 2012; Kheng, Mohamad, Ramayah & Mosahab, 2010). Esta percepción de calidad del servicio se ve afectada por el compromiso organizacional de los empleados, así como por su motivación y satisfacción laboral (García & Forero, 2014; Walsh & Deery, 2006).

Según Bettencourt, Gwinner y Meuter (2001), los trabajadores deben ser eficientes en la promoción de las organizaciones y sus servicios (fidelidad), eficaces en proporcionar al cliente un servicio (prestación de servicios), y propensos a involucrarse en el desarrollo de la orientación al servicio de la empresa (participación), con base en la comprensión de las diversas características de los clientes. Para estos autores, se deben tener en cuenta los antecedentes actitudinales, de personalidad y el conocimiento del cliente al momento de evaluar la disposición del empleado ante la prestación del servicio.

Acá, los antecedentes actitudinales hacen referencia a las actitudes que los empleados obtienen de acuerdo con el rendimiento en el trabajo y pueden servir como predictores de las conductas de ciudadanía organización (OCB), que está compuesta por dos aspectos: (a) la satisfacción laboral y (b) las POS, es decir, las creencias globales de los empleados.

*cforero@ucatolica.edu.co

Específicamente, la *satisfacción laboral* es definida como la evaluación afectiva general que realiza el empleado, tiene como finalidad alentar a los empleados a involucrarse en conductas orientadas al servicio, y tiene dos componentes: (a) los *afectivos intrínsecos*, referidos a la motivación del empleado para realizar tareas laborales de acuerdo con las fuentes que se originan desde el interior de la persona y que tienen un valor psicológico; y (b) los *afectivos extrínsecos*, que tienen que ver con la motivación del empleado para realizar tareas laborales tomando como fundamento los factores externos o del medio ambiente (Bateman et al., como se cita en Bettencourt et al., 2001).

Por otra parte, las *creencias globales*, o POS, son las valoraciones personales que el empleado genera frente al apoyo organizacional y la valoración organizacional que la empresa le otorga, y está compuesta por dos aspectos: (a) la percepción de las valoraciones que dan los empleados ante el apoyo que le otorga la organización o las creencias sobre los aportes personales (Eisenberger et al., como se cita en Bettencourt et al., 2001); y (b) la percepción de las creencias sobre los valores personales y las autovaloraciones del empleado de acuerdo con sus aportes a la organización y el trato que le da a los clientes.

De igual forma, los *antecedentes de personalidad* hacen referencia a los rasgos de personalidad, y pueden organizarse jerárquicamente en función de su grado de abstracción y en los aspectos relacionados con el servicio, específicamente, estos son la orientación de servicio, la empatía y la lealtad. Según la literatura, estos aspectos están estrechamente vinculados con el contexto de las conductas y probablemente conformados por la historia individual y la cultura (Bettencourt et al., 2001).

Adicionalmente, la *orientación al servicio* es la predisposición de un individuo para ofrecer un servicio superior mediante su capacidad para reflejar declaraciones positivas acerca de los productos y servicios de la empresa. En particular, esta variable tiene dos componentes: (a) la responsabilidad, entendida como la disposición del individuo a ofrecer un servicio y a asumir las consecuencias de este (Bettencourt et al., 2001); y (b) la cortesía, comprendida como el deseo genuino de satisfacer al cliente por medio de la amabilidad y la buena educación.

Otro aspecto relacionado es la empatía, la cual, según los mismos autores, se define como un rasgo de expansión para los empleados, debido a que les ayuda a percibir cómo experimentan los consumidores el servicio. Esta variable se divide en dos: (a) la empatía cognoscitiva, que hace referencia a la anticipación y el entendimiento de los puntos de vista de los clientes y las expectativas de estos hacia un producto o servicio –para posteriormente participar en comportamientos de prestación de servicios apropiados–; y (b) la empatía afectiva, que tiene que ver con la compasión para satisfacer las necesidades de los clientes a través de la prestación de un servicio superior.

Y otro antecedente de la personalidad es la lealtad, definida como el reflejo de la obediencia a la organización a través de la promoción de sus intereses y la imagen a los externos. Específicamente, esta variable tiene dos componentes: (a) la promoción efectiva; y (b) la disposición positiva de la firma (Bettencourt et al., 2001).

Por último, la tercera dimensión de la disposición a la prestación del servicio es el conocimiento del cliente, que se define como la adecuada comprensión de los tipos de clientes y las estrategias de interacción en combinación con una personalidad orientada al servicio. En particular, esta se encuentra compuesta por dos aspectos: (a) las experiencias enriquecedoras del servicio; y (b) las estrategias enriquecedoras del servicio.

Específicamente, las experiencias enriquecedoras del servicio tienen que ver con los conocimientos de los empleados sobre las características de los diferentes tipos de clientes (Sujan et al., como se cita en Bettencourt et al., 2001), ya que, debido a que los empleados interactúan con diferentes segmentos de clientes con diferentes expectativas y necesidades, es útil que consoliden la información sobre los tipos de clientes en categorías significativas para guiar sus conductas (Bettencourt et al., 2001). Así, las experiencias enriquecedoras tienen dos componentes: (a) los conocimientos de diferentes tipos de usuarios; y (b) el contexto con diferentes tipos de usuarios.

Como se ha encontrado en la literatura, las estrategias enriquecedoras del servicio hacen referencia a los conocimientos de los empleados sobre estrategias para hacer negocios frente a diferentes necesidades y situaciones de los clientes (Sujan et al., como se cita en Bettencourt et al., 2001); y se componen de dos aspectos: (a) el conocimiento de diferentes estrategias adecuadas a las necesidades de los usuarios; y (b) el empleo de múltiples estrategias con un mismo usuario.

Al respecto, al revisar diferentes investigaciones se encontró que las estrategias enriquecedoras del servicio hacían énfasis en la evaluación de la calidad, pero no en la disposición de los funcionarios de las organizaciones a la prestación del servicio (Greenland, Coshall & Combe, 2005; Parasuraman, Zeithaml & Berry, 1988, 1994; Pena, Santos da Silva, Rizzato & Melleiro, 2006; Snoj & Mumel, 2002; Subramony, Terry, Beehr & Merle, 2004; Susskind, Borchgrevink & Kacmar, 2003; Zeithaml, Berry & Parasuraman, 1996; Pascual, Pascual, Frías & Rosel, 2006; Varela, Rial & García, 2003; Varela, Prat, López & Rial, 2006).

Partiendo de estas consideraciones, y teniendo en cuenta la forma en que se evalúan las variables mencionadas, es necesario mencionar, además, algunos aspectos formales que tienen que ver con la construcción de pruebas, es decir, lo referente a un proceso sistemático y organizado mediante el cual se recopilan datos cualitativos o cuantitativos sobre rasgos de conocimientos, de actitudes o de habilidades, y que al no ser directamente observables se evalúan de acuerdo con las respuestas de los individuos ante pruebas objetivas (Mikulic, 2007).

La teoría de respuesta al ítem y el modelo de Rasch

Para el diseño de instrumentos es importante mencionar tres modelos matemáticos: la Teoría Clásica de los Test (TCT), la teoría de la generalizabilidad y la teoría de respuesta al ítem (TRI), esta última relevante para la presente investigación.

Específicamente, la TRI permite una visión amplia de la relación entre un rasgo y las características de los ítems (Hidalgo & French; 2016). Según Martínez (1996), esta no es una teoría, sino una serie de modelos bajo los cuales se asume que subyacente al rendimiento del examinado en un test se encuentra un rasgo o aptitud única, y que la relación entre el rasgo o aptitud y la respuesta del sujeto al ítem puede ser descrita por medio de una función o curva característica del ítem que establece las probabilidades de respuesta (Sijtsma & Junker, 2006).

Por otra parte, la TRI basa su funcionamiento en la probabilidad de que una persona responda a un ítem de cierta forma, según la posición del sujeto en un continuo con base en el cual se le está midiendo (Boscaïno, 2005). Así, las características de los ítems no dependen del grupo en que se calculen, las puntuaciones que describen el comportamiento de los sujetos examinados en el test no dependen de los contenidos particulares de este ni de su dificultad, los modelos están expresados a nivel del ítem –y no con respecto al test completo–, y la estrategia proporciona medidas de precisión para cada nivel de habilidad (Martínez, 1996).

De este modo, la TRI se plantea como un intento por formalizar las respuestas dadas por los examinados en ítems de pruebas educativas y psicológicas (Sijtsma & Junker, 2006), está relacionada con probabilidades de respuesta determinadas que son representadas a partir de modelos no lineales (logísticos), y su característica más relevante está en la posibilidad de comparar pruebas compuestas por distintos ítems y generar evidencias respecto a niveles diferentes de habilidad a pesar de haber resuelto adecuadamente la misma cantidad de ítems (Nunnally & Bernstein, 1995).

Adicionalmente, de acuerdo con Nunnally y Bernstein (1995), la TRI agrupa una serie de modelos diferenciados en cuanto a la cantidad de parámetros de medida, en donde, por ejemplo, se pueden encontrar modelos monoparamétricos, biparamétricos, y triparamétricos, los cuales, aunque son diferentes entre sí, también poseen aspectos comunes, ya que se trata de modelos estructurales que establecen una relación matemática formalizada entre las respuestas a un ítem concreto y el nivel o rasgo de aptitud del sujeto.

Asimismo, de acuerdo con Muñiz (1997), en los postulados básicos de la TRI se plantea que: (a) todo sujeto puede situarse mediante un punto en el espacio del rasgo, y los ítems pueden ser situados en función de la cantidad de rasgo que demanden para su adecuada ejecución; (b) el nivel de aptitud del individuo es preponderante

e independiente del test usado; (c) las unidades básicas del test son los ítems, y, por lo tanto, los modelos matemáticos se establecen a nivel del ítem y no del test; y (d) todos los modelos se establecen antes de puntuar el test, lo que permite transformar las respuestas en puntuaciones.

Específicamente, uno de los modelos más conocidos de la TRI es el de Rasch, el cual representa una aproximación a la medición en las ciencias sociales distinta al resto de modelos de la TRI. Este modelo puede ser visto desde tres perspectivas distintas: (a) se caracteriza por el tratamiento de las habilidades de la persona como efectos aleatorios, y la intención original de este tratamiento puede facilitar la máxima probabilidad de estimación marginal para los parámetros de los ítems; (b) considera los parámetros de las personas como variables aleatorias, lo que lleva a que las variables características de la persona sean incluidas como predictores dentro del modelo; y (c) es representado por un modelo multigrupo de la TRI, para el cual se asume que los individuos se agrupan de acuerdo a características comunes, tales como el grupo étnico o el nivel académico (Kamata & Cheong, 2007).

Según González (2008), el modelo de Rasch busca “describir la relación entre la probabilidad de una respuesta correcta ante un ítem y la diferencia entre la habilidad de quienes responden y la dificultad de este” (p. 7); y, en este sentido, el modelo tiende a ser dicotómico, ya que solo considera las diferencias en las dificultades de los ítems para producir medidas de habilidad en quien responde. Por tanto, desde el punto de vista de Wrigth y Stone (1998), lo que se busca en este modelo es establecer la probabilidad de que esta diferencia determine lo que sucede cuando una persona utiliza su habilidad frente a la dificultad del ítem.

Finalmente, cabe mencionar que desde este modelo se asumen dos supuestos fundamentales: (a) la unidimensionalidad, entendida como la tendencia del conjunto de ítems a definir un solo constructo; y (b) la independencia local, definida como la influencia de las habilidades sobre el desempeño de los individuos de forma independiente para cualquier par de reactivos (González, 2008). Teniendo esto en cuenta, el modelo de Rasch proporciona un marco eficiente con el cual se puede verificar y quizá mejorar el funcionamiento de las categorizaciones de las escalas de calificación (Linacre, 2002; 2006; 2013).

Construcción de pruebas: diseño basado en la evidencia

Una vez descritos los aspectos propios del enfoque matemático considerado en esta investigación, es conveniente hacer mención de algunos aspectos relevantes en el proceso específico de la construcción del instrumento de evaluación; en este caso, el enfoque denominado Diseño Centrado en la Evidencia.

Este enfoque es un modelo lógico y sistemático para la creación de pruebas, fundamentado en los principios del razonamiento probatorio (Mislevy, Steinberg & Almond, 2003), y que tiene como objetivo basar aspectos importantes del diseño de la prueba con su desarrollo y puntuación. Específicamente, este modelo posee dos funciones principales: la primera de ellas, razonar sobre el diseño de la evaluación; y, la segunda, realizar evaluaciones analíticas más avanzadas en comparación con otros modelos tradicionales (Razzouk, 2011).

Adicionalmente, el modelo se sustenta en tres premisas: (a) una evaluación debe construir todo el conocimiento importante en el dominio de interés y una comprensión de cómo se adquiere ese conocimiento y el uso que se le da; (b) la cadena de razonamiento de los participantes en relación a sus evaluaciones o inferencias debe basarse en los principios de evidencia de razonamiento; y (c) el propósito debe ser la fuerza impulsora detrás de las decisiones de diseño que reflejan las limitaciones, los recursos y las condiciones de uso (Mislevy, Almond & Lukas 2003). Y para cumplir con estas premisas, es necesario especificar las estructuras y fundamentos de apoyo para el argumento probatorio de la evaluación, en donde si el argumento es explícito, este facilita la examinación; así, la estructura de los argumentos abarca, entre otras cosas, las inferencias que se quieren hacer acerca del evaluado y lo observable (datos de rendimiento), además de la tarea situacional de rendimiento que provoca los resultados de los evaluados, y los fundamentos para vincular todo junto (Mislevy, Steinberg & Almond, 2003; Shute, Hansen & Almond, 2008).

Específicamente, según Mislevy et al. (2003) y Baxter y Mislevy (2005), el modelo divide el proceso de diseño, desarrollo y uso de pruebas en cinco actividades:

- a. *Análisis de dominio*: donde se recopila la información sobre el tema de interés que tendrá implicación en la evaluación y la forma en que esta se obtuvo y se comunica; particularmente, incluye el contenido, conceptos, terminología, herramientas y formas de representación que las personas tienen, y es la base sustantiva de los argumentos de evaluación.
- b. *Modelado del dominio*: tiene que ver con la organización de la información de análisis de dominio a lo largo de las líneas de argumentos de evaluación; así, el argumento tiene forma narrativa con descripciones coherentes de competencias de interés que evidencian las capacidades y las formas de organización de las situaciones.
- c. *Marco de evaluación conceptual*: el argumento se comienza a articular en patrones de diseño en términos de tipos de elementos y procesos necesarios para la evaluación; en esta se tienen en cuenta variables, esquemas de tareas y mecanismos de puntuación, y se genera una estructura que coordine aspectos sustantivos,

estadísticos y operativos. Acá también se toman decisiones de diseño, como el modelo estadístico, materiales y procedimientos para calificar el trabajo con el fin de generar una familia de tareas y procesos auxiliares.

- d. *Evaluación de la instrumentación*: que incluye la presentación de las tareas, la recolección y el análisis de las respuestas.
- e. *Evaluación de entrega*: donde se coordina las interacciones de los evaluados y las tareas.

Método

Tipo de estudio

El presente es un estudio empírico con metodología cuantitativa de tipo instrumental, ya que se realizó el desarrollo de una prueba, y dado que incluyó el diseño y el estudio de sus propiedades psicométricas (Montero & León, 2007). De acuerdo con Chow (2002), se trata de una investigación psicométrica y de metodología no experimental, ya que se busca evaluar las condiciones métricas de una prueba y las características técnicas de los ítems que la componen; y se considera no experimental, ya que no se produce manipulación ni se introducen nuevas variables en el ambiente de los participantes.

Participantes

El pilotaje se realizó con 100 funcionarios de empresas de diferentes sectores económicos, con edades entre los 18 a 65 años (35 % eran mujeres y 65 %, hombres). La validación se realizó con 366 trabajadores, con edades entre los 18 y 65 años (64 % mujeres, 34.7 % hombres; el 0.5 % no reportó sexo). Los participantes trabajaban en diferentes entidades públicas y privadas prestadoras de servicios en la ciudad de Bogotá; y proceden de estratos socioeconómicos entre 1 y 6. Respecto a la escolaridad, el 26.8 % eran bachilleres, 24 % técnicos, 43.2 % profesionales y 22 participantes no reportaron escolaridad. El 51.6 % eran solteros, 22.4 % casados, 3.8 % divorciados, 18.6 % en unión libre, 0.5 % viudos y 3 % no hicieron mención de su estado civil.

Instrumentos

Se trata de una prueba diseñada en escala tipo Likert para evaluar tres aspectos centrales: (a) *antecedentes actitudinales*, con dos componentes: satisfacción laboral y POS –creencias globales de los empleados–; (b) *antecedentes de personalidad*, con tres componentes: orientación al servicio, empatía y lealtad; y (c) y *conocimiento del cliente*, compuesto por: experiencias enriquecedoras del servicio y estrategias enriquecedoras del servicio. Para el pilotaje se emplearon 81 ítems, y en la validación se tuvieron en cuenta 60 ítems.

Procedimiento

Inicialmente, se realizó la contextualización teórica de la investigación, teniendo en cuenta aspectos teóricos del servicio al cliente, de la teoría de respuesta al ítem, del modelo de Rasch y del diseño centrado en la evidencia. Posteriormente, se definieron las variables, al igual que sus componentes, indicadores, índices e ítems, de acuerdo con lo propuesto desde el diseño centrado en la evidencia. En tercer lugar, los ítems fueron sometidos a diferentes procesos de evaluación para proceder con el proceso de pilotaje. Y, por último, se procedió a realizar el pilotaje y el análisis de resultados.

Consideraciones éticas

La investigación se implementó de acuerdo con procedimientos donde se garantizó la ética, el bienestar y la dignidad de los participantes, así como la garantía y veracidad de la información (Ley 1090 de 2006 del Ministerio de Protección Social). El consentimiento informado se redactó de acuerdo con la Resolución 8430 de 1993 del Ministerio de Salud; y, conforme a la misma resolución, el nivel de riesgo de la investigación fue el mínimo, puesto que no se realizó modificación o intervención de variables físicas, biológicas o psicológicas, ni se identificaron o trataron aspectos sensibles del comportamiento.

Resultados

El análisis de las características métricas de los ítems y de la prueba se hizo mediante el modelo de Rasch, con el software Winsteps 6.8, mientras que las condiciones sociodemográficas de los participantes se analizaron mediante estadística descriptiva, con el software IBM Statistics SPSS 20.0.

Para evaluar la unidimensionalidad de la prueba se tuvo en cuenta el criterio de Recase (1979), el cual considera aceptable un 20 % de varianza explicada para el primer factor en un análisis de componentes principales. En el pilotaje, los resultados arrojaron una varianza explicada de 35.9 %, mientras que la separación fue de 2.63 para los ítems, con una confiabilidad alfa de Cronbach de 0.94, y una separación de 2.77 para las personas, lo cual refleja niveles explicativos adecuados para este tipo de evaluación.

Los análisis para la toma de decisiones en el pilotaje se concentraron en el *Infit* y *Outfit* (estadígrafos de ajuste interno y externo, respectivamente). Por una parte, el valor del estadígrafo *Infit* se interpreta como una media cuadrática ponderada de residuales que es sensible a patrones de respuesta irregulares, que tiene como función capturar comportamientos de respuestas no esperadas a reactivos calibrados cerca del nivel de habilidad del sustentante. Y, por otra, el estadígrafo *Outfit* es sensible a valores extremos en los datos, es decir, capta las “anomalías” en patrones de respuesta a reactivos lejanos al nivel de la habilidad medida en los sujetos respondientes (González, 2008).

Para efectos del análisis y el establecimiento de la calidad de la prueba se consideraron niveles de exigencia de 0.7 a 1.3 tanto para el *Infit* como para el *Outfit*; y, a partir de estos

criterios, se encontró que los ítems 4, 8, 11, 17, 19, 24, 30, 38, 47, 51, 58, 59, 66, 67, 68, 71, 72, 76, 80 no cumplen con el criterio del *Outfit*, y que los ítems 37 y 60 no cumplen con ninguno de los dos, por lo cual fueron excluidos del proceso de validación.

En total, la validación se realizó para 60 ítems con 365 personas. Inicialmente se evaluó la unidimensionalidad de cada una de las variables (véase Tabla 25). Como se puede observar, tanto para cada una de las variables como para la prueba completa la varianza explicada en el primer factor supera el 20 % mencionado anteriormente, así pues, se trata de variables con comportamiento unidimensional tanto para las variables como para la prueba en sí misma. Sin embargo, para efectos de empleo de la prueba, el comportamiento total de la misma no es relevante.

Tabla 25. Evaluación de unidimensionalidad por variables a partir de la varianza explicada

Variable	<i>Eigenvalue</i>	Varianza explicada en el primer factor
Antecedentes actitudinales	6.2	25.8 %
Antecedentes de personalidad	12.9	31.5 %
Conocimiento del cliente	9.2	39.7 %
Creencias globales de los empleados	3.7	29.5 %
Empatía	7.1	39.2 %
Estrategias enriquecedoras del servicio	4.6	36.7 %
Experiencias enriquecedoras del servicio	6.2	50.8 %
Lealtad	3.9	30 %
Orientación al servicio	3.2	28.7 %
Satisfacción laboral	4.2	32 %
Prueba completa	23.6	28.3 %

Respecto a los aspectos de confiabilidad y capacidad explicativa de cada una de las variables y de la prueba completa, se encontró que en ambos casos se evidencia una tendencia a puntuaciones altas; las desviaciones inferiores a uno muestran que las variables y la prueba tienden a ser homogéneas; sin embargo, parece que los puntajes se ajustan a la población, ya que la diferencia de la media respecto al cero ideal es superior a una desviación estándar (véase Tabla 26).

Por otra parte, en cuanto al criterio de separación, las separaciones entre personas son pequeñas y se ven reflejadas en las puntuaciones de baja confiabilidad; sin embargo, las separaciones de los ítems son amplias e insinúan una explicación del comportamiento de los ítems respecto a las variables. Esta confiabilidad se ve reflejada en valores de alfa de Cronbach, donde la prueba completa presenta una confiabilidad de 88 %, mientras que las variables fluctúan entre 54 % y 81 %. Específicamente, la

variable satisfacción laboral muestra valores bajos que no se ajustan realmente a las necesidades de la evaluación; sin embargo, las decisiones centrales sobre la capacidad explicativa de los ítems se generan desde el comportamiento en el *Infit* y el *Outfit* de estos mismos en cada una de las variables, por lo que las decisiones respecto al funcionamiento de los ítems en los componentes de la prueba se toman a ese nivel.

Tabla 26. Confiabilidad, separación, promedios y desviaciones de cada una de las variables de la prueba de evaluación de servicio al cliente

Variable	Promedio personas	D. S. personas	Separación personas	Separación ítems	Confiabilidad personas	Confiabilidad ítems	Alfa de Cronbach
Antecedentes actitudinales	0.75	0.36	1.13	3.69	0.56	0.93	0.66
Antecedentes de personalidad	0.81	0.45	1.6	3.92	0.72	0.94	0.81
Conocimiento del cliente	0.82	0.43	1.17	5.83	0.58	0.97	0.68
Creencias globales de los empleados	0.89	0.57	1.00	4.11	0.50	0.94	0.61
Empatía	0.75	0.48	1.06	4.41	0.53	0.95	0.58
Estrategias enriquecedoras del servicio	0.90	0.53	0.84	5.02	0.42	0.96	0.54
Experiencias enriquecedoras del servicio	0.78	0.50	0.71	7.74	0.33	0.98	0.44
Lealtad	0.80	0.53	0.98	2.70	0.49	0.88	0.67
Orientación al servicio	1.14	0.69	0.69	2.48	0.32	0.86	0.54
Satisfacción laboral	0.72	0.37	0.66	0.90	3.57	0.93	0.40
Prueba completa	0.75	0.33	2.29	4.02	0.84	0.94	0.88

Por otra parte, se encontró que dos de los ítems de la variable antecedentes actitudinales exceden levemente los valores críticos (22 y 46) y, por tanto, deben ser excluidos de la prueba. Asimismo, los criterios que se asumieron corresponden a un rango entre 0.7 a 1.3, con valores ideales cercanos a 1 (Gothwal, 2009, citado por Jafari, Bagheri & Safe, 2012), por tanto, los valores que se salieron de ese rango deben ser excluidos de la prueba (Véase Tabla 27).

Específicamente, en la variable antecedentes de personalidad, los reactivos 32, 7, 9, 29, 33, 11, 10 y 15 muestran valores que indican la necesidad de replanteamiento; mientras que en la variable antecedentes actitudinales, el reactivo 39 muestra un comportamiento fuera del límite; en la variable creencias globales de los empleados no aparecen ítems que excedan los valores críticos; en la variable empatía, los reactivos 11, 32 y 33 exceden los valores críticos aceptables; en la variable estrategias enriquecedoras del servicio,

Tabla 27. Análisis por ítems de la prueba por variables e ítems totales

Dimensión	Ítem	Infit MNSQ	Outfit MNSQ	Dimensión	Ítem	Infit MNSQ	Outfit MNSQ	Dimensión	Ítem	Infit MNSQ	Outfit MNSQ
Antecedentes Actitudinales	22	1.08	1.42	Conocimiento del Cliente	39	1.88	2.29	Prueba Total	39	1.64	2
	46	1.19	1.4		19	1.06	1.25		22	1.07	2.03
	1	0.98	1.18		42	0.92	1.21		32	1.41	1.63
	44	1.17	1.3		21	0.99	0.88		44	1.23	1.51
	28	1.06	1.24		17	1.1	1.25		46	1.22	1.71
	3	0.94	0.72		59	0.89	0.8		1	1.02	1.6
	6	0.87	0.81		40	0.9	0.87		9	1.06	1.36
	48	1.04	1.05		60	1.04	1.16		7	0.99	1.14
	5	1.04	1.11		57	0.96	1.13		6	1.01	1.01
	45	1.02	1.17		20	1.01	1.03		19	1.01	1.13
	25	0.91	0.91		43	1.24	1.28		11	1	1.38
	4	0.92	0.94		41	0.85	0.77		28	1.04	1.31
	23	1.08	1.19		18	0.82	0.68		33	1.08	1.35
	27	0.96	1		58	0.8	0.75		3	0.97	0.95
	2	1.02	1.05	Creencias Globales de los Empleados	28	1.13	1.25		43	1.22	1.32
	24	0.96	1.09		6	0.87	0.93		27	1.13	1.37
	26	0.84	0.81		25	1.02	1		29	0.96	1.18
	47	0.93	0.92		48	1.14	1.29		51	0.94	0.95
Conocimiento del Cliente de la Prueba Sercliente	39	1.88	2.29		5	1.18	1.27		4	1.04	1.06
	19	1.06	1.25		26	0.84	0.8		35	1.02	1.27
	42	0.92	1.21		4	0.94	0.93		48	1.09	1.33
	21	0.99	0.88		27	0.96	0.98		8	0.96	1.23
	17	1.1	1.25		47	0.91	0.88		13	0.97	1.14
	59	0.89	0.8	Lealtad	56	0.87	1		5	1.07	1.21
	40	0.9	0.87		38	0.89	0.83		23	1.15	1.5
	60	1.04	1.16		15	1.13	1.33		15	1.03	1.15
	57	0.96	1.13		37	0.97	0.88		10	1.05	1.12
	20	1.01	1.03		55	0.99	1.05		53	0.99	1.18
	43	1.24	1.28		36	1.01	1.05		16	1.08	1.13
	41	0.85	0.77		54	0.89	0.86		60	1.02	1.2
	18	0.82	0.68		16	1.2	1.22		17	1.05	1.12
	58	0.8	0.75		14	1.08	1.1		47	1.05	1.16
Antecedentes de Personalidad	32	1.68	1.99	Orientación al Servicio	7	1.08	1.11		42	0.89	1.04
	7	1.02	1.51		51	0.92	0.71		31	0.95	1.07
	9	1.11	1.59		9	1.17	1.45		21	0.94	1
	29	0.98	1.59		8	0.97	0.78		59	0.94	0.94
	33	1.14	1.63		29	1.03	1.16		56	0.88	1.06
	11	1	1.46		49	0.86	0.7		37	0.97	0.94
	51	0.93	0.79		50	0.85	0.69		38	0.94	0.8
	8	0.97	1.12		30	1.13	1.03		49	0.88	0.72
	10	1.16	1.51	Satisfacción Laboral	1	0.99	1.22		25	0.9	0.88
	15	1.09	1.31		22	1	1.16		30	1	1.17
	35	1.02	1.09		3	0.86	0.64		24	1.03	1.08
	13	0.96	1.1		44	1.1	1.22		57	0.94	1.06
	38	0.94	0.96		46	1.09	1.12		34	0.94	1.1
	16	1.17	1.27		45	0.98	1.04		50	0.87	0.75
	53	1.01	1.17		2	0.99	0.99		40	0.92	0.88
	49	0.86	0.7		24	0.95	0.98		45	0.97	0.99
	56	0.86	1.02		23	0.97	0.99		55	0.93	1.01
	31	0.96	1.04						26	0.9	0.87
	50	0.87	0.79						2	1.01	1.12
	30	1.05	1.2						14	0.98	0.98
	55	0.95	1.04						20	0.94	0.98
	37	0.96	0.87						12	0.96	0.99
	14	1.04	1.06						54	0.89	0.98
	34	0.94	0.97						52	0.91	0.86
	54	0.9	0.99						36	0.89	0.92
	12	0.99	0.98						58	0.85	0.85
	36	0.9	0.99						41	0.87	0.8
	52	0.87	0.8						18	0.84	0.74

el ítem 43 excede los valores críticos; en la variable experiencias enriquecedoras del servicio, el ítem 39 debe ser replanteado; en la variable lealtad, el reactivo 15 excede los valores críticos; los ítems 9 y 50 de la variable orientación al servicio deben ser replanteados; y en la variable satisfacción laboral se requiere lo mismo con el ítem 3. Finalmente, para la prueba total, los ítems 1, 9, 11, 22, 23, 27, 32, 33, 39, 43, 44, 46 y 48 requieren replantearse; sin embargo, es recomendable hacer ajustes retirando los límites que indican mayor error para poder hacer los ajustes pertinentes.

Discusión

Los resultados muestran que la prueba muestra niveles globales de ajuste adecuados para el modelo de Rasch, particularmente en las características de unidimensionalidad y confiabilidad; sin embargo, variables como servicio al cliente podrían requerir ajustes importantes. A partir de los resultados, sería importante realizar nuevos análisis métricos retirando los ítems que generan mayor alteración en el comportamiento general de la prueba; y una evaluación a fondo de la estructura de esta sugiere la necesidad de hacer ajustes de texto en algunos ítems, lo que implicaría hacer un proceso de revalidación.

Por otra parte, los valores establecidos en el *Infit* y en el *Outfit* de la prueba muestran un ajuste adecuado en la prueba, ya que se demuestra que las variables de antecedentes actitudinales, conocimiento del cliente, creencias globales de los empleados, empatía, estrategias del servicio, experiencias del servicio, lealtad, orientación al servicio y satisfacción laboral se encuentran con el nivel de ajuste esperado, en donde los ítems están balanceados y son homogéneos –sin embargo, se puede mejorar la confiabilidad al hacer la revisión mencionada anteriormente–. Adicionalmente, y en concordancia con las anteriores, la variable de antecedentes de personalidad no muestra cambios en los factores internos o externos que afecten la resolución de la prueba de acuerdo con el nivel de habilidad de las personas, por lo que se demuestra un nivel de ajuste adecuado, con una confiabilidad alta, donde los ítems son balanceados y homogéneos, lo cual es lo ideal para la investigación.

Ahora bien, en cuanto a los ítems, se puede indicar que estos también presentan una condición general óptima y productiva para el proceso de medición según los resultados en todas las variables; sin embargo, se recomienda replantear los ítems mencionados para que se ajuste su métrica a las necesidades de la prueba.

Asimismo, se puede decir que la mayoría de los respondientes se encuentran ubicados por encima de la dificultad 0 (cero) para desarrollar actitudes positivas con relación al rendimiento en el trabajo, la orientación del servicio, la empatía, las estrategias de interacción con cada tipo de clientes en relación con el trabajo, las valoraciones

personales frente al apoyo organizacional, la percepción de cómo los consumidores están experimentando el servicio, las estrategias para hacer negocios con las diferentes necesidades y situaciones de los clientes, los conocimientos necesarios en relación con las características de los diferentes tipos de clientes, la obediencia a la organización a través de la promoción de los intereses, y la imagen de la misma; inclusive, también para la predisposición a ofrecer un servicio superior a través de la capacidad para reflejar las declaraciones positivas a los clientes acerca de los productos y servicios de la empresa, la satisfacción en el trabajo que tiene como finalidad alentar a los empleados a involucrarse en conductas orientadas al servicio, y, por último, la disposición al servicio al cliente.

En cuanto a lo anterior, se puede decir que se cumple con lo planteado en la investigación, debido a que se evalúan diversos factores orientados al servicio al cliente, teniendo en cuenta la habilidad de las personas para responder. Sin embargo, es importante considerar que hay algunos ítems que evalúan un mismo factor o característica, por lo que se recomienda hacer una revisión a los mismos y a cada variable.

Dentro de los resultados obtenidos para cada una de las variables, cabe destacar que la variable experiencias enriquecedoras del servicio posee un contenido de ítems adecuado debido a que todos los ítems que se encuentran dentro de esta evalúan diferentes factores.

Ahora bien, según lo propuesto en la investigación y los resultados obtenidos, puede decirse que se cumplieron los objetivos, ya que se logró diseñar la prueba desde la perspectiva del diseño basado en la evidencia y, a partir del modelo de Rasch de la TRI, evaluar la disposición al servicio al cliente. Esto debido a que se plantearon algunas variables operacionales y a partir de la información revisada de servicio al cliente, donde se tomó en cuenta la perspectiva de Bettencourt et al. (2001) para realizar la prueba encaminada hacia la evaluación de los empleados y no de los clientes.

De igual forma, se pudo observar de acuerdo con los resultados que es necesario hacer una revisión y ajuste de algunos ítems para que la prueba sea confiable, ya que en varias dimensiones se evidencia la necesidad de modificar o eliminar algunos ítems porque evalúan los mismos factores o no se encuentran entre los requerimientos establecidos en la investigación.

Adicionalmente, cabe mencionar que se logró realizar los análisis métricos de la prueba y de los ítems mediante el modelo de Rasch de la TRI por medio de los programas SPSS y Winsteps, con lo que se logró obtener un marco formal de especificaciones de los conocimientos y habilidades a medir (Mislevy, Almond & Lukas, 2003).

Por otra parte, también destaca el hecho de que se llevaron a cabo los cuatro procesos principales en el ciclo de evaluación de Mislevy (2000), ya que se creó un banco de ítems según las dimensiones, se realizó el monitoreo al estado de la prueba,

y se establecieron los códigos de clasificación y de información para equilibrar el formulario. Así mismo, se entregó la prueba al examinado para su ejecución –a fin de optimizar la selección de información que tomó como evidencia–, y se realizaron las modificaciones pertinentes para capturar los datos de entrada de usuarios.

A partir de lo anterior, se realizó el análisis de resultados para recolectar la evidencia sobre el conocimiento y las habilidades necesarias para que la persona responda la prueba, se clasificó la información según la relevancia, se ejecutaron las reglas de evidencia, se crearon resultados, y se dio retroalimentación con respecto a los hallazgos finales.

Finalmente, puede afirmarse que se logró la capacidad para establecer diferencias entre lo que se observa o se tiene como material informativo y lo que en realidad se quiere hacer y obtener con este (Mislevy, Behrens & Dicerbo, 2012). Sin embargo, de acuerdo con Bejar (2008), los puntos de corte para las pruebas que evalúan las habilidades de las personas se determinan a partir de distintos niveles de rendimiento, pero para obtener estos valores es necesario que se realicen análisis adicionales que, además, permitan identificar las diferencias de acuerdo a factores como el sexo, el nivel educativo o el cargo, entre otros aspectos significativos.

Referencias

- Baxter, G., & Mislevy, J. (2005). *The case for an integrated design framework for assessing Science inquiry*. CSE Technical Report 638. Los Angeles: The National Center for Research on Evaluation, Standards, Student Testing (CRESST), Center for Studies in Education, UCLA.
- Bejar, I. (2008). *Standard Setting: What is it? Why is it important?* Educational Testing Service 7. Recuperado de https://www.ets.org/Media/Research/pdf/RD_Connections7.pdf.
- Bettencourt, L., Gwinner, K., & Meuter, M. (2001). A Comparison of Attitude, Personality, and Knowledge Predictors of Service-Oriented Organizational Citizenship Behaviors. *Journal of Applied Psychology*, 86(1), 29-41.
- Boscaino, G. (2005). La qualità della didattica: la calibrazione dello strumento di misura con il modello di Rasch (Tesis para optar por el título de doctor). Universidad de Palermo, Palermo, Italia.
- Chow, S. L. (2002). Methods in psychological research. En *Methods in Psychological Research*, In *Encyclopedia of Life Support Systems (EOLSS)*. Oxford, U. K.: Eolss Publishers, Oxford, UK.
- García, M., & Forero, C. (2014). Motivación y satisfacción laboral como determinantes asociados al cambio organizacional. En Orjuela, J. (editores académicos), *Psicología De Las Organizaciones Y El Trabajo: Apuestas En Investigación* (pp. 327-350). Cali: Editorial Bonaventuriana.
- González, M. (2008). *El Análisis de Reactivos con el Modelo de Rasch*. Instituto Nacional para la evaluación de la educación. Universidad Sonora. México D.F.
- Greenland, S., Coshall, J., & Combe, I. (2005). Evaluating Service Quality and Consumer Satisfaction in Emerging Markets. *International Journal of Consumer Studies*, 30(6), 582-590. doi: <https://doi.org/10.1111/j.1470-6431.2005.00484.x>

- Hidalgo, M., & French, B. (2016). Una introducción didáctica a la Teoría de Respuesta al Ítem para comprender las construcciones a escalas. *Revista de Psicología Clínica con Niños y Adolescentes*, 3(2), 13-21. Recuperado de <http://www.revistapcna.com/sites/default/files/16-11.pdf>
- Hirmukhe, J. (2012). Measuring Internal Customers' Perception on service quality Using SERVQUAL in Administrative services. *International Journal of scientific and research publications*, 2(3), 1-6.
- Jafari, P., Bagheri, Z., & Safe, M. (2012). Item and Response-Category Functioning of The Persian Version of the KIDSCREEN-27: Rasch Partial Credit Model. *Health Quality of Life Outcomes*, 10, 127. doi: <https://doi.org/10.1186/1477-7525-10-127>
- Kamata, A., & Cheong, Y. (2007). Multilevel Rasch models. En M., Von Davier., & C., H. Carstensen. (Eds.), *Multivariate and mixture distribution Rasch models: Extensions and applications* (pp. 217-232). New York: Springer science and Business Media.
- Kheng, L., Mohamad, O., Ramayah, T., & Mosahab, R. (2010). The Impact of Service Quality on Costumer Loyalty: A Study of Banks in Penang, Malaysia. *International Journal of Marketing Studies*, 2(2), 57-66. doi: 10.5539/ijms.v2n2p57
- Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement*, 3(1), 85-106. doi: 10.1.1.424.2811
- Linacre, J. M. (2006). *WINSTEPS Rasch measurement computer program*. Chicago: Winsteps.com
- Linacre, J. M. (2013). *Winsteps® Rasch measurement computer program*. Beaverton, Oregon: Winsteps.com
- Martínez, R. (1996). *Psicometría: Teoría de los Tests Psicológicos y Educativos*. Madrid, España: Editorial Síntesis.
- Mikulic, I. (2007). *Construcción y adaptación de pruebas psicológicas* [Manuscrito inédito]. Universidad de Buenos Aires, Argentina.
- Ministerio de la Protección Social. (2006). Ley Numero 1090 de 2006. Ministerio de la Protección Social. Gobierno de Colombia.
- Ministerio de Salud (1993). Resolución No. 8430 de 1993. Ministerio de Salud. Gobierno de Colombia.
- Mislevy, J., Almond, G., & Lukas, F. (2003). A brief introduction to evidence-centered design. *ETS Research Report Series*. 20(2), 79-87. doi: <https://doi.org/10.1002/j.2333-8504.2003.tb01908.x>
- Mislevy, J., Steinberg, S., & Almond, G. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives*, 1(1), 3-62. doi: http://dx.doi.org/10.1207/S15366359MEA0101_02
- Mislevy, R. (2000). *A sample assessment using the Four Process Framework*. Princeton: Educational Testing Service.
- Mislevy, R., Almond, R., & Lukas, J. (2003). *A brief introduction to evidence-centered design*. Research report. Princeton: Educational Testing Service.
- Mislevy, R., Behrens, J., & Dicerbo, K. (2012). Design and discovery in educational assessment: Evidence-centered design, psychometrics, and educational data mining. *Journal of educational data mining*, 2(4), 11-48.

- Montero, I., & León, O. G. (2007). A guide for naming research studies in Psychology. *International Journal of Clinical and Health Psychology*, 7(3), 847-862.
- Muñiz, J. (1997). *Introducción a la teoría de respuesta a los ítems*. Madrid, España: Ediciones Pirámide.
- Nunnally, J., & Bernstein, I. (1995). *Teoría Psicométrica* (3ra. ed.). México: McGraw-Hill.
- Parasuraman, A., Valarie, A., Zeithaml, Leonard, L., & Berry. (1988). SERVQUAL: A Multiple-Item Scale for Measuring Consumer Perceptions of Service Quality. 64(1), 12-40.
- Parasuraman, A., Valarie, A., Zeithaml, Leonard, L., & Berry. (1994). Reassessment of Expectations as a Comparison Standard in Measuring Service Quality: Implications for further Research. *Journal of Marketing*, 58(1), 111-124. doi: 10.2307/1252255
- Pascual, M., Pascual, J., Frias, M., & Rosel, J. (2006). Calidad De Servicio En Supermercados: Una Propuesta De Medición. *Psicothema*, 18(3), 661-667.
- Pena, M., Santos da Silva, E., Rizatto, D., & Melleiro, M. (2013). The use of the quality model of Parasuraman, Zeithaml and Berry in health services. *Revista da Escola de Enfermagem da U S P*, 47(5), 1227-1232. doi: 10.1590/S0080-623420130000500030
- Razzouk, R. (2011). *Using Evidence-Centered Design for developing valid assessment of 21st century skills. Advancing education for 21st century success*. Maryland: Edvation.
- Reckase, M. (1979). Unifactor Latent Trait Models Applied to Multifactor Tests: Results and Implications. *Journal of Educational and Behavioral Statistics*, 4(3), 207-230. doi: doi.org/10.3102/10769986004003207
- Shute, J., Hansen, G., & Almond, G. (2008). You can't fatten a hog by weighing it, or can you? Evaluating an assessment for learning system called ACED. *International Journal of Artificial Intelligence and Education*, 18(4), 289-316.
- Sijtsma, K., & Junker, B. (2006). Item response theory: past performance, presents developments, and future expectations. *Behaviormetrika*, 33 (1), 75-102. doi: https://doi.org/10.2333/bhmk.33.75
- Snoj, B., & Mumel, D. (2002). The measurement of perceived differences in service quality-The case of health spas in Slovenia. *Journal of Vacation Marketing*, 8(4), 362-376. doi: doi/10.1177/135676670200800407
- Subramony, M., Terry, A., Beehr, & Merle J. (2004). Employee and Customer Perception of Service Quality in an Indian Firm. *Applied psychology. An international Review*, 53(2), 311-327. doi: https://doi.org/10.1111/j.1464-0597.2004.00173.x
- Susskind, A., Borchgrevink, C., & Kacmar, K. (2003). Customer Service Providers' Attitudes Relating to Customer Service and Customer Satisfaction in the Customer – Server exchange. *Journal of Applied Psychology*, 88(1), 179-187. doi: http://dx.doi.org/10.1037/0021-9010.88.1.179
- Varela, J., Rial, A., & García, E. (2003). Presentación De Una Escala De Satisfacción Con Los Servicios Sanitarios De Atención Primaria. *Psicothema*, 15(4), 656-661.
- Varela, J., Prat, R., Lopez, C., & Rial, A. (2006). Una nueva escala para la evaluación de la calidad de los servicios de hostelería. *Psicothema*, 18(1), 135-142.
- Walsh, J., & Deery, S. (2006). Refashioning Organizational Boundaries: Outsourcing Customer Service Work. *Journal of Management Studies*, 43(3), 557-582. doi: https://doi.org/10.1111/j.1467-6486.2006.00601.x
- Wright, B., & Stone, M. (1998). *Diseño De Mejores Pruebas Utilizando La Técnica De Rasch*. México: CENEVAL.
- Zeithmal, V., Berry, L., & Parasuraman, A. (1996). The Behavioral Consequences of Service Quality. *Journal of Marketing*, 60(2), 31-46.