

# Modelos de Teoría de Respuesta al Ítem para el Análisis del Clima Organizacional

## Item Response Theory Models for Organizational Climate Analysis

**Roberto Carlos Evaristo Broncano**

<https://orcid.org/0000-0001-9010-3641>

[roberto.evaristo@unmsm.edu.pe](mailto:roberto.evaristo@unmsm.edu.pe)

Universidad Nacional Mayor de San Marcos, Lima, Perú

RECIBIDO: 2/07/2024 - ACEPTADO: 17/07/2024 - PUBLICADO: 31/07/2024

### RESUMEN

La variable satisfacción con el clima organizacional (CO) es una variable latente que no puede medirse directamente, generalmente, se utilizan cuestionarios donde se definen un conjunto de preguntas pertinentes, en una escala Likert, para su medición. Su análisis forma parte de la teoría de la medición, Teoría Clásica de los Test (TCT) o de la Teoría de Respuesta al Ítem (TRI). El presente trabajo ha evidenciado la utilidad de los modelos TRI en el análisis del CO, en los enfoques de ítems dicotómicos y politómicos. En el caso de la TRI se aplicaron los modelos logísticos de un parámetro o Rasch (1PL), modelos logísticos de dos parámetros o Birnbaum (2PL) y el Modelo de Respuesta Graduada (MRG), la cual es el modelo pertinente para el análisis del CO porque se ajusta mejor a preguntas con escala Likert o politómicas.

**Palabras clave:** Clima organizacional, Escala Likert, Teoría Clásica de los Test, Teoría de Respuesta al Ítem, Modelo de Respuesta Graduada.

### ABSTRACT

The variable "satisfaction with organizational climate (OC)" is a latent variable that cannot be directly measured. Typically, questionnaires are used to define a set of relevant questions for its measurement, often using a Likert scale. Its analysis is part of measurement theory, specifically Classical Test Theory (CTT) or Item Response Theory (IRT). This study has demonstrated the usefulness of IRT models in analyzing OC for both dichotomous and polytomous items. For IRT, the one-parameter logistic model (1PL or Rasch), the two-parameter logistic model (2PL or Birnbaum), and the Graded Response Model (GRM) were applied. The GRM is particularly suitable for analyzing OC as it better fits Likert scale or polytomous items.

**Keywords:** Organizational Climate, Likert Scale, Classical Test Theory, Item Response Theory, Graded Response Model.

## I. INTRODUCCIÓN

El CO es una abstracción que representa la estructuración cognitiva de un todo (conjunto de eventos, políticas, prácticas, procedimientos, etc.) que perciben los colaboradores de la organización a partir de sus experiencias (Ehrhart et al., 2014). El CO es un factor determinante para la satisfacción laboral y es importante para la gestión adecuada de los recursos humanos. (Downey et al, 1975).

La satisfacción con el CO se considera una variable latente, ya que no puede medirse directamente pero se manifiesta a través de comportamientos, actitudes y percepciones de los colaboradores (MacCallum y Austin, 2000). Esta variable se puede medir a través de encuestas o entrevistas y es posible conocer la opinión de los colaboradores sobre la organización. Normalmente, se utilizan cuestionarios donde se realizan preguntas que miden el grado de satisfacción con el CO desde diferentes dimensiones como trabajo en equipo, comunicación interna, etc.

La medición es un proceso fundamental en las ciencias sociales, que se basa en la comprensión de las propiedades de objetos, personas o instituciones mas no de las entidades mismas (David e Ida, 2019), por ejemplo, la satisfacción con el CO puede considerarse una propiedad de los colaboradores ya que se evalúa su percepción de la organización. Sin embargo, siempre habrá cierto grado de error al medir las variables psicológicas, por lo que es necesario el uso de metodologías adecuadas.

Los modelos de variables latentes son herramientas estadísticas que posibilitan el análisis y la representación de la relación entre un grupo de variables que pueden ser observadas y otro grupo de variables que son latentes, con ello abordan el desafío de la medición en fenómenos psicológicos. Existe una clasificación de los modelos en función a los tipos de variables latentes y observables, en el caso de que ambas variables sean continuas será considerado como modelo de variable latente, entre las cuales se tiene a la TCT y a la TRI (Bartholomew et al., 2011).

La TCT es un enfoque valioso para comprender la confiabilidad de los tests y determinar como el error afecta a las puntuaciones observadas. Al postular que la puntuación observada es la suma de la puntuación verdadera y del error aleatorio facilita la medición, pero también tiene limitaciones como la dependencia de la puntuación verdadera con el test, la suposición de una relación lineal, las estadísticas de los ítems dependen de la muestra y presuponen

que el error de medición es constante en todas las puntuaciones (Shultz et al, 2020).

Por otro lado, la TRI ofrece un enfoque más avanzado, proporcionando estadísticas que son invariables a la población, lo que permite obtener información consistente sobre los parámetros de los ítems (dificultad y discriminación) y de los individuos (habilidad). Se dice que la TRI supera las limitaciones de la TCT proporcionando un nivel de análisis más robusto y detallado, sin embargo, Ellis y Mead (2002) recomiendan que los enfoques de TCT y TRI se combinen en la realización de un análisis de ítems.

El trabajo está compuesto de una introducción, luego, en el capítulo dos, se detalla la parte teórica de los modelos de la TCT y de la TRI; en el tercer capítulo, se exponen los objetivos; en el cuarto capítulo se describen las herramientas utilizadas y el método seguido; en el cuarto capítulo, se detallan los resultados de la aplicación de los modelos de TCT y TRI; en el quinto capítulo se expuso la discusión de los resultados y en el sexto capítulo se especificaron las conclusiones.

## II. FUNDAMENTOS TEÓRICOS

### 2.1. Modelos de la TCT

La TCT parte del supuesto general de que la puntuación total del colaborador  $i$  en la ECO ( $X_i$ ) es la suma del valor teórico ( $T_i$ ) con el error de medición del test ( $E_i$ ) (Bartolucci et al., 2015). Según Bandalos (2018), la TCT tiene tres propiedades principales: (i) el promedio de los errores en una población de encuestados es cero, (ii) no hay correlación entre las puntuaciones verdaderas y los errores en una población de encuestados, y (iii) los errores de dos pruebas distintas, o de la misma prueba en diferentes ocasiones, no están correlacionados.

**Índice de fiabilidad:** Es la medida en que las puntuaciones observadas reflejan las puntuaciones verdaderas en lugar de las de error. Esto es importante para los investigadores porque quieren saber si las diferencias observadas en las puntuaciones reflejan diferencias verdaderas o errores aleatorios (Bandalos, 2018). Se tienen métodos para medir el índice de fiabilidad como son: (i) la forma paralela que consiste en aplicar dos test con un mismo contenido a los mismos individuos, (ii) el test-retest que se refiere a la aplicación del mismo test a los mismos individuos en más de una ocasión y (iii) el análisis interno que consiste en medir las correlaciones

entre cada ítem y con ello evaluar la similaridad de los resultados.

El Alfa de Cronbach ( $\alpha_c$ ) es una medida del análisis interno y es considerado como el valor mínimo del índice de fiabilidad teórica (Cronbach, 1951). Si ( $\alpha_c$ ) es superior a 0.7 se considera una mayor fiabilidad del test.

$$\alpha_c = \left( \frac{k}{k-1} \right) \left( 1 - \frac{\sum_{j=1}^k S_j^2}{S_X^2} \right)$$

Donde  $k$  es el número de ítems del test,  $S_j^2$  es la varianza de las puntuaciones en el ítem  $j$  y  $S_X^2$  es la varianza de la puntuación total.

**Error estándar de medición ( $S_E$ ):** Indica la precisión de la puntuación empírica que obtuvo el individuo en el test con respecto a su puntuación verdadera (De Ayala, 2013).

$$S_E = S_X \sqrt{1 - \alpha_c}$$

donde  $S_X$  es la desviación de la puntuación total y  $\alpha_c$  es el Alfa de Cronbach.

**Dificultad del ítem:** Es el porcentaje de individuos que respondieron positivamente el ítem de un test (Crocker y Algina, 1986).

**Correlación Ítem-Test:** Es la discriminación del ítem y se mide mediante la correlación entre las puntuaciones del ítem con el total (McDonald, 1999).

## 2.2. Modelos de la TRI

La TRI fue desarrollada principalmente por Lord (1952), Rasch (1960) y Birnbaum (1968) y es un enfoque psicométrico que analiza los ítems de una prueba para obtener medidas invariantes entre individuos y test. Se considera una familia de modelos estadísticos donde la probabilidad de responder correctamente a un ítem se modela como una distribución condicional que depende de la habilidad

del individuo, la dificultad y discriminación del ítem (ver Tabla 1)

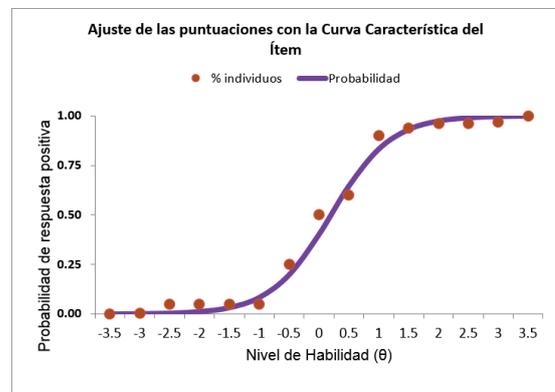
**Curva característica del ítem (CCI):** Es la función que representa la relación entre la probabilidad de responder correctamente un ítem y las características de los individuos y de los ítems, por lo tanto, cada ítem tiene una CCI.

En la Figura 1 se muestra un ejemplo de una CCI que sigue el mismo patrón de respuestas positivas en la medida que aumenta la habilidad.

**Tabla 1**  
Parámetros de la TRI.

Parámetro	Definición
Habilidad del individuo ( $\theta$ )	Es la posición que tiene el individuo en la escala continua de la variable latente estudiada.
Dificultad del ítem ( $\beta$ )	Es un punto en la escala de habilidad ( $\theta$ ), la cual sirve de referencia para identificar a individuos satisfechos con una probabilidad mayor al 0.5 (DeMars, 2010).
Discriminación del ítem ( $\alpha$ )	Indica que tanto cambia la probabilidad de respuesta correcta a medida que aumenta la habilidad (DeMars, 2010). Una mayor discriminación permite que el ítem diferencie entre individuos con diferentes niveles de habilidad, lo cual es deseable.

**Figura 1**  
CCI.



Por otro lado, en la TRI existen supuestos que respaldan el desarrollo metodológico de los modelos y de la estimación de los parámetros.

**Unidimensionalidad:** Se refiere a que las respuestas a los ítems son influenciadas por un factor dominante (Hattie, 1985), en ese sentido, los modelos de la TRI se basan en que las respuestas de los individuos dependen de un nivel latente de una variable. Para determinar la unidimensionalidad existen métodos como el análisis paralelo que, mediante simulaciones de monte carlo encuentran autovalores simulados que superan a los reales (Ledesma y Mora, 2007).

**Independencia Local:** Se refiere a que la respuesta del individuo a una pregunta no depende de las respuestas al resto de las preguntas del test. Entonces, la probabilidad de un patrón de respuestas a los ítems del test se determina con el producto de las probabilidades de respuesta a cada ítem (Hambleton et al, 1991). Es decir, sea  $X_i = (X_{i1}, X_{i2}, \dots, X_{ik})$  el patrón de respuestas a los  $k$  ítems del test, entonces

$$P(X_{i1}, X_{i2}, \dots, X_{ik}) = P(X_{i1}|\theta)P(X_{i2}|\theta) \dots P(X_{ik}|\theta)$$

No es necesario contrastar empíricamente el supuesto de independencia local (aunque se puede revisar a Santisteban (1990) para ello), ya que una adecuada especificación de la unidimensionalidad garantiza esta independencia (DeMars, 2010).

**Invarianza de Parámetros:** Asegura que las estimaciones de habilidad son independientes de la muestra de ítems e individuos utilizados. Este supuesto aplica con los parámetros de los ítems, ya que permanecen consistente entre diferentes grupos de individuos (DeMars, 2010).

A continuación, se detallan los dos modelos de la TRI, en ítems dicotómicos se considera al Modelo 2PL y para ítems politómicos se considera al MRG.

### 2.2.1. Modelo 2PL

Modelo propuesto por Birnbaum (1968), donde planteó que la probabilidad de respuesta correcta sigue una función Logit con dos parámetros para cada ítem  $j$ , dificultad ( $\beta_j$ ) y discriminación ( $\alpha_j$ ), y el parámetro de habilidad del individuo  $i$  ( $\theta_i$ ).

Sea  $X_{ij}$  que se caracteriza por ser una variable dicotómica (1 / 0).

$$\begin{aligned} X_{ij}|\theta_i, \beta_j, \alpha_j &\sim \text{Bernoulli}(P_{ij}) \\ P_{ij} = P(X_{ij}|\theta_i, \beta_j, \alpha_j) &= F(Z_{ij}) \\ Z_{ij} &= \alpha_j(\theta_i - \beta_j) \end{aligned}$$

Las probabilidades de respuesta correcta e incorrecta están dadas por:

$$\begin{aligned} P(X_{ij} = 1|\theta_i, \beta_j, \alpha_j) &= P_{ij} = \frac{1}{1 + \exp[-Z_{ij}]} \\ P(X_{ij} = 0|\theta_i, \beta_j, \alpha_j) &= Q_{ij} = \frac{1}{1 + \exp[Z_{ij}]} \end{aligned}$$

La Función de Información del Ítem (FII) indica la cantidad de información que el ítem proporciona sobre la habilidad de los individuos en la escala de medición (Natarajan, 2009). Los ítems con altas FII son aquellos que son más informativos para una amplia gama de habilidades latentes y pueden diferenciar mejor a los individuos con diferentes niveles de habilidad. Según Baker y Kim (2004), la FII para el ítem  $j$  es:

$$I_j(\theta) = \frac{[P_i'(\theta)]^2}{P_i(\theta)Q_i(\theta)} = \alpha_j^2 P_j(\theta)Q_j(\theta)$$

La Función de Información del Test (FIT) es la suma de la FII de los  $k$  ítems del test, considerando el supuesto de independencia local (Birnbaum, 1968). Los tests con alta FIT son aquellos que son más precisos para medir la habilidad latente de los individuos.

$$I(\theta) = \sum_{j=1}^k I_j(\theta) = \sum_{j=1}^k \alpha_j^2 P_j(\theta)Q_j(\theta)$$

### 2.2.2. El Modelo MRG

Fue propuesto por Samejima (1969) y representa una generalización del modelo 2PL. MRG permite modelar variables latentes que tienen variables manifiestas categóricas con al menos tres categorías, por lo que adicionalmente interviene un subíndice  $k$  que representa la categoría del ítem  $j$ . En ese sentido, las puntuaciones de un ítem ya no son  $x_j = (0, 1)$  como era en el modelo 2PL sino que se considera un vector de puntuaciones más amplio  $x_j = (0, 1, 2, \dots, m_j)$  donde  $m_j$  representa la puntuación más alta que se puede alcanzar en el ítem.

El MRG plantea que los ítems están caracterizados por una discriminación ( $\alpha_j$ ), así como en el modelo 2PL, y los  $m_j$  parámetros de umbral (*threshold*) entre las  $k_j$  categorías del ítem  $j$  ( $\beta_{jm}$ ) de tal forma que  $m_j + 1 = k_j$ , lo que difiere del modelo 2PL. Si una pregunta tiene  $k = 5$  opciones de respuesta entonces tendrá  $m = 4$  umbrales y cuatro curvas características de la categoría del ítem, conocida como Curva Característica Operativa (CCO).

Este umbral determina un punto en la escala de habilidad donde la probabilidad de elegir entre dos categorías adyacentes es igual, lo cual está relacionado conceptualmente con el parámetro de dificultad visto en el modelo 2PL (Embretson y Reise, 2000).

Según Baker y Kim (2004) y Tarazona (2015) el MRG es el siguiente:

$$X_{ij}|\theta_i, \beta_{jk}, \alpha_{jk} \sim \text{Categórica}(P_{ij1}, \dots, P_{ijk})$$

$$\sum_{k=1}^K P_{ijk} = 1$$

$$P_{ijk} = P(X_{ij} = k|\theta_i, \beta_{jk}, \alpha_j) = P_{ijk}^* - P_{ij(k+1)}^*$$

$$P_{ijk}^* = P(X_{ij} \geq k|\theta_i, \beta_j, \alpha_j) = F(Z_{ijk})$$

$$F(Z_{ijk}) = \frac{1}{1 + \exp[-Z_{ijk}]}$$

$$Z_{ijk} = \alpha_j(\theta_i - \beta_{jk})$$

$$-\infty = \beta_{j0} < \beta_{j1} < \dots < \beta_{j(k-1)} < \beta_{jk} = \infty$$

Del planteamiento anterior, se puede calcular  $P_{ijk}$  y  $Q_{ijk}$

$$P(X_{ijk} = 1|\theta_i, \beta_{jk}, \alpha_j) = P_{ijk} = \frac{1}{1 + \exp[-Z_{ijk}]}$$

$$P(X_{ijk} = 0|\theta_i, \beta_{jk}, \alpha_j) = Q_{ijk} = \frac{1}{1 + \exp[-Z_{ijk}]}$$

La FII de un MRG es la suma ponderada de la Función de Información de la Categoría de Respuesta del Ítem (FICRI),  $I_{jk}(\theta)$ , con la probabilidad de respuesta positiva a la categoría de respuesta del ítem  $I_{jk}(\theta)$ .

$$I_j(\theta) = \sum_{k=1}^m \left[ \frac{[P_{j,k-1}'(\theta) - P_{j,k}'(\theta)]^2}{P_{j,k-1}'(\theta) - P_{j,k}'(\theta)} \right]$$

### III. OBJETIVOS

#### 3.1. Objetivo principal

Evaluar la utilidad de los modelos de la Teoría de Respuesta al Ítem en el análisis del Clima Organizacional, a comparación de métodos tradicionales como la Teoría Clásica de los Test.

#### 3.2. Objetivos específicos

Analizar la satisfacción del clima organizacional con el enfoque de la Teoría Clásica de los Test, considerando ítems dicotómicos y politómicos.

Analizar la satisfacción del clima organizacional con el enfoque de la Teoría de Respuesta del ítem, considerando el modelo Rasch y Birnbaum para ítems dicotómicos y el modelo de respuesta graduada para ítems politómicos.

### IV. MATERIALES Y MÉTODOS

Para la aplicación de los modelos señalados se utilizó el software estadístico R a través del entorno colaborativo Google Colab.

**Google Colaborative**, donde se tiene disponible de manera libre Jupiter notebook que permite codificar en python o R a través de un servicio en la nube.

**Entorno R y librerías**, se utilizaron librerías como ggplot2, ltm, etc (ver Tabla 2):

**Tabla 2**  
Librerías usadas en el software R

Librería	Definición
ggplot2	Propuesto por varios autores como Hadley Wickham, entre otros en su publicación: "ggplot2: cree visualizaciones de datos elegantes utilizando la gramática de los gráficos". Es útil para crear gráficos. <a href="https://cran.r-project.org/web/packages/ggplot2/index.html">https://cran.r-project.org/web/packages/ggplot2/index.html</a>
DataExplorer	Es útil para realizar el análisis exploratorio de los datos. <a href="https://cran.r-project.org/web/packages/DataExplorer/readme/README.html">https://cran.r-project.org/web/packages/DataExplorer/readme/README.html</a>
Psych	Propuesto por William Revelle en su publicación "psych: Procedures for Psychological, Psychometric, and Personality Research". Es útil para realizar análisis exploratorio de los datos, mediciones de fiabilidad del test, entre otros. <a href="https://cran.r-project.org/web/packages/psych/psych.pdf">https://cran.r-project.org/web/packages/psych/psych.pdf</a>
psychometric	Propuesto por Thomas Fletcher en su publicación "psychometric: Applied Psychometric Theory". Es útil para realizar el análisis psicométrico y proceder con las estimaciones propias de las TCT. <a href="https://cran.r-project.org/web/packages/psychometric/index.html">https://cran.r-project.org/web/packages/psychometric/index.html</a>
LTM	Propuesto por Dimitris Rizopoulos en su publicación "ltm: Latent Trait Models under IRT". Es útil para la aplicación de los modelos TRI para ítems dicotómicos y politómicos. <a href="https://cran.r-project.org/web/packages/ltm/ltm.pdf">https://cran.r-project.org/web/packages/ltm/ltm.pdf</a>

En la Tabla 3 se brinda una aproximación de las definiciones de los parámetros de los ítems (dificultad y discriminación) y del individuo (habilidad) al contexto del CO, lo representa las bases conceptuales para utilizar la TRI en el análisis del CO.

**Tabla 3**  
*Aproximación de parámetros de la TRI en la ECO.*

Parámetro	Conceptualización
Habilidad	Es la percepción y evaluación global del colaborador del CO, representa la capacidad latente del colaborador para interpretar y responder a diferentes aspectos del CO.
Dificultad	Indica el grado de acuerdo o desacuerdo de los colaboradores con el ítem propuesto en la ECO, por ejemplo, para la afirmación "Siento que mis opiniones son valoradas en esta organización", un valor positivo de dificultad indica que la pregunta es crítica y requiere de un mayor nivel de satisfacción del colaborador para estar de acuerdo con la afirmación.
Discriminación	Indica la capacidad del ítem para diferenciar entre colaboradores satisfechos e insatisfechos, es decir, indica que tan bien un ítem distingue entre colaboradores con diferentes niveles de percepción con el CO.
FII	Indica la cantidad de información que proporciona una pregunta del test sobre el análisis de la percepción de los colaboradores en relación con el CO, por ejemplo, un ítem que pregunta sobre la satisfacción con el liderazgo podría tener una alta FII si es capaz de distinguir con precisión entre empleados que están altamente satisfechos con el liderazgo y aquellos que no lo están.
FIT	Indica la cantidad de información total proporciona el conjunto de preguntas sobre el CO en su conjunto, por ejemplo, una alta FIT implicaría que la encuesta es precisa para medir la percepción general de los empleados sobre el CO en toda la gama de aspectos evaluados.

El análisis del CO se abordó desde el enfoque tradicional de medición del test que es la TCT y desde un enfoque alternativo de medición que es la TRI. Ambos enfoques se desarrollaron considerando ítems dicotómicos y politómicos, dependiendo de ello se utilizó unos modelos específicos que son descritos en la Tabla 4.

**Tabla 4**  
*Uso de los modelos.*

Enfoque	Ítems	
	Dicotómicos	Politómicos
Tradicional	TCT con ítems dicotómicos	TCT con ítems politómicos
Alternativo	Modelo Logístico TRI con dos parámetros	Modelo TRI de respuesta Graduada

Respecto a las fases del análisis del CO, primero se recolectaron los datos, luego, se procedió con un análisis exploratorio de los datos recolectados realizando gráficos estadísticos (histogramas, barras y dispersión) y mediciones estadísticas univariadas y bivariadas (promedios, desviaciones y correlaciones de spearman). En tercer lugar, se evaluó la confiabilidad del test considerando como

indicador el Alfa de Cronbach y también se evaluó el supuesto de la unidimensionalidad relacionada con la TRI. En cuarto lugar, se realizó la aplicación del modelo de la TCT donde se determinaron la dificultad, discriminación y se midió el error estándar de medición, tanto para los ítems dicotómicos como politómicos. En quinto lugar, se aplicaron los modelos TRI dependiendo de los tipos de ítems, también se compararon los niveles de ajuste.

**Obtención de los datos:** Los datos fueron recolectados en una ECO de una entidad financiera. El cuestionario está compuesto de 15 dimensiones, pero, se seleccionó la dimensión de Remuneración que consta de seis preguntas (ver Tabla 5), las cuales están formuladas con la escala Likert de cinco categorías donde 1 indica que está totalmente en desacuerdo y 5 indica que está totalmente de acuerdo con la pregunta. En la Tabla 6 se muestran las cinco categorías consideradas en el cuestionario. Además, estas cinco categorías fueron dicotomizadas considerando de que las categorías "De acuerdo" y "Muy de acuerdo" son considerados como "Satisfecho" y el resto de categorías como "No satisfecho" que será útil para la aplicación de los modelos con respuesta dicotómica.

**Tabla 5**  
*Preguntas de la dimensión de remuneración.*

ID	Pregunta
REM_01	¿Considero que mi salario está en línea con el de otros puestos de igual responsabilidad dentro de la organización?
REM_02	¿Siento que mi compensación es adecuada considerando las responsabilidades asociadas con mi cargo?
REM_03	¿Mis condiciones salariales en la entidad son competitivas en comparación con roles similares en el mercado?
REM_04	¿Los beneficios adicionales que recibo en la organización son apropiados y equitativos?
REM_05	¿Tengo la percepción de que mi compensación refleja justamente las ganancias generadas por la organización?
REM_06	¿Creen las personas de esta organización que reciben un salario justo por el trabajo que desempeñan?

**Tabla 6**  
*Escala de las preguntas.*

Cat	Descripción	Grado
1	Muy en desacuerdo	No satisfecho
2	En desacuerdo	No satisfecho
3	Neutro	No satisfecho
4	De acuerdo	Satisfecho
5	Muy de acuerdo	Satisfecho

**Métricas de desempeño:** Para la comparación de los modelos de la TRI se utilizaron diferentes métricas como son las siguientes:

**FIT 3**( $-\leq \theta \leq 3$ ): Se mide el FIT considerando un rango de habilidad entre -3 y 3. Luego, a mayor FIT es mejor.

**SEM:** Es el error de medición obtenido por el modelo en un rango de habilidad entre -3 y 3. Luego, a menor SEM es mejor.

**Correlación entre la puntuación bruta y la habilidad estimada:** Se considera una medición informativa para ver que tan alineados están los resultados de los modelos de la TRI y de la TCT.

**El logaritmo de la verosimilitud (Log.lik):** Un mayor valor de Log.lik indica un mejor ajuste de los modelos a los datos.

**Criterios de Información de Akaike (AIC) y de Información Bayesiana (BIC):** Un menor valor del AIC o BIC indica un mejor ajuste del modelo a los datos,

**Prueba de razón de verosimilitud (P-valor):** Es una prueba de hipótesis estadística que parte de la H0: las restricciones impuestas en el modelo reducido son verdaderas.

**V. RESULTADOS**

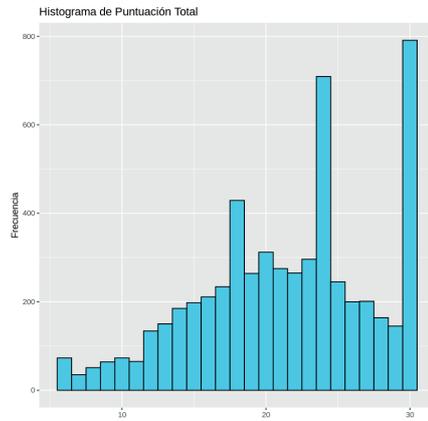
**5.1. Análisis exploratorio de los datos**

La puntuación total alcanzada en la dimensión *Remuneración* presenta una distribución trimodal en los valores 30, 24 y 18, los cuales hacen referencia a respuestas iguales en cada una de las seis preguntas con puntuaciones de 5, 4 y 3 respectivamente (ver Figura 2). En cada pregunta se observa una distribución asimétrica hacia puntuaciones altas por lo que se puede inferir un alto grado de satisfacción de los colaboradores respecto al CO. Además, la correlación de *spearman* entre cada pregunta y la puntuación total tiende a ser alta y positiva de al menos el 0.4, teniendo a las preguntas 4 y 5 con menores grados de correlación con el resto de las preguntas (ver Figura 3).

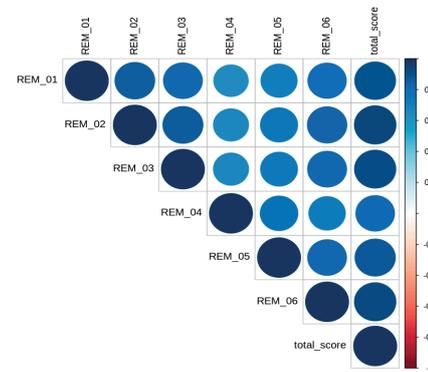
**Confiabilidad del test:** El indicador de consistencia interna Alfa de *Cronbach* alcanzó un nivel de 0.95, lo cual es evidencia de que las preguntas están altamente correlacionadas y miden de manera consistente la variable latente analizada.

**Verificación de la Unidimensionalidad:** Se evidenció la existencia de una sola componente capaz de explicar la variabilidad total del dataset conformado por las seis preguntas de la dimensión de remuneración (ver Figura 4).

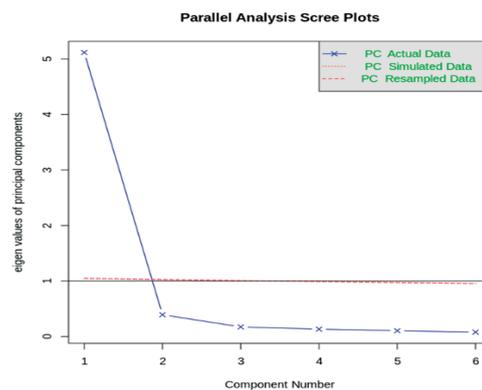
**Figura 2**  
Histograma de la puntuación total observada.



**Figura 3**  
Correlación de spearman entre ítems.



**Figura 4**  
Gráfico de análisis paralelo para verificar unidimensionalidad.



**5.2. Aplicación de la TCT**

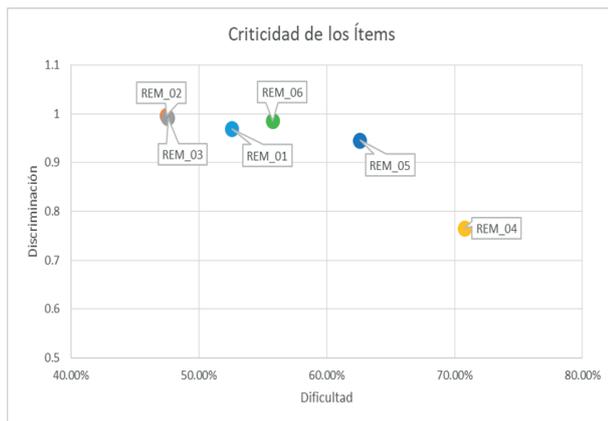
**Resultados en ítems dicotómicos:** Se evidenció que en los ítems 2 y 3 los colaboradores tienen el menor grado de satisfacción y a su vez la más alta discriminación, seguido por las preguntas 1, 6, 5 y

4 en cuanto a la dificultad y por los ítems 6, 1, 5 y 4 respecto a la discriminación (ver Tabla 7). Finalmente, la pregunta 4 parece ser un ítem con baja criticidad para el estudio del CO a comparación del resto de las preguntas (ver Figura 5). Otro punto a considerar, es que en las preguntas 1, 2, 3, 5 y 6 los niveles de discriminación son parecidos.

**Tabla 7**  
*Parámetros de la TCT para Ítems Dicotómicos.*

Ítem	Dificultad	Discriminación
REM_01	52.6%	0.968
REM_02	47.5%	0.996
REM_03	47.6%	0.992
REM_04	70.8%	0.765
REM_05	62.6%	0.945
REM_06	55.8%	0.985

**Figura 5**  
*Evaluación de la criticidad de los ítems con la TCT con ítems dicotómicos.*



El error estándar de medición (SEM por sus siglas en inglés) fue de 0.7, considerando un coeficiente de fiabilidad del 0,92 y una desviación de la puntuación total de 2,5. Teniendo en cuenta que la puntuación total puede fluctuar entre 0 y 6, se puede concluir que el SEM es relativamente bajo.

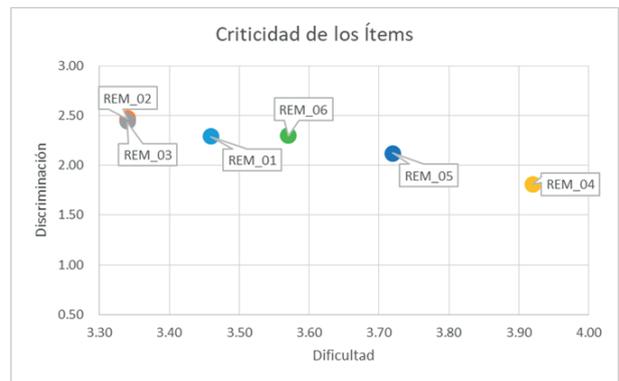
**Resultados en ítems politómicos:** Análogo al análisis con ítems dicotómicos, se evidenció que en las preguntas 2 y 3 los colaboradores tienen el menor grado de satisfacción y a su vez la más alta discriminación, seguido por los ítems 1, 6, 5 y 4 en cuanto a la dificultad y por los ítems 6, 1, 5 y 4 en discriminación (ver Tabla 8). Finalmente, el ítem 4 parece ser un ítem con baja criticidad para el estudio del CO a comparación del resto de las preguntas. Otro punto a considerar, es que en las

preguntas 1, 2, 3, 5 y 6 los niveles de discriminación son parecidos (ver Figura 6).

**Tabla 8**  
*Parámetros de la TCT para Ítems Politómicos.*

Ítem	Dificultad	Discriminación
REM_01	3.46	2.29
REM_02	3.34	2.47
REM_03	3.34	2.44
REM_04	3.92	1.81
REM_05	3.72	2.12
REM_06	3.57	2.30

**Figura 6**  
*Evaluación de la criticidad de los ítems con la TCT con ítems politómicos.*



El SEM fue de 1.35, considerando un coeficiente de fiabilidad del 0.95 y una desviación de la puntuación total de 6.1. Teniendo que la puntuación total puede fluctuar entre 0 y 30 se puede concluir que el SEM es relativamente bajo.

**5.2. Aplicación de la TRI**

La aplicación de la TRI se abordó desde el enfoque dicotómico con el modelo logístico de dos parámetros y desde el enfoque politómico con el modelo de respuesta graduada.

**Aplicación del modelo 2PL.** Se aplicaron los modelos logísticos de un parámetro o también conocido como Rasch donde sólo se considera la dificultad del ítem (Rasch1), también se aplicó un segundo modelo Rasch considerando como restricción un valor fijo de la discriminación del ítem (Rasch2) y finalmente se aplicó el modelo logístico de dos parámetros (2PL).

En la evaluación del ajuste de estos tres modelos propuestos se verificó que el modelo 2PL presentó un mejor ajuste, además, se evidenció que la

mejora es significativa al 5% a comparación de los modelos Rasch (ver Tabla 9). Adicionalmente, se evidenció de que el modelo 2PL tiene un nivel de información más alto que los otros modelos, por lo que su precisión es mejor con un SEM más bajo. Por otro lado, el grado de correlación entre la puntuación bruta y la habilidad estimada es alto con un nivel de 0.99.

**Tabla 9**  
Comparación del Ajuste de los Modelos 1PL y 2PL para Ítems dicotómicos.

Indicador	Rasch1	Rasch2	2PL
FIT	5.34	27.34	30.78
SEM	0.43	0.19	0.18
CORR	0.9999	0.9893	0.9873
log.Lik	-17,807.9	-13,914.9	-13,748.8
AIC	35,627.8	27,843.9	27,521.5
BIC	35,667.7	27,890.5	27,601.5
P-Valor		<0.001	<0.001

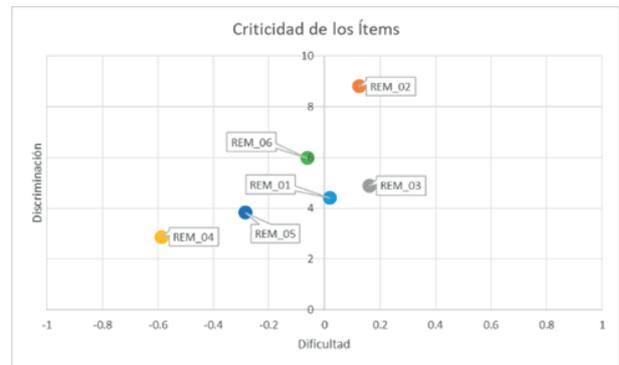
Respecto a las estimaciones de los parámetros de los ítems del modelo 2PL (ver Tabla 10), en el análisis de la dificultad se tuvo que las preguntas REM\_03 y REM\_02 son las que requiere mayor nivel de satisfacción del colaborador para estar de acuerdo con ellos, seguido por las preguntas REM\_01 y REM\_06 que tienen una dificultad regular y las preguntas REM\_05 y REM\_08 que no requieren mucho esfuerzo del colaborador para estar satisfecho. Respecto al análisis de la discriminación, la pregunta REM\_02 tiene el nivel más alto, seguido por las preguntas REM\_06, REM\_03 y REM\_01, mientras que las preguntas REM\_05 y REM\_04 tienen los niveles más bajos.

**Tabla 10**  
Parámetros de la TRI para Ítems dicotómicos.

Ítem	2PL	
	Dificultad	Discriminación
REM_01	0.0183	4.4096
REM_02	0.1248	8.8010
REM_03	0.1617	4.8906
REM_04	-0.5864	2.8655
REM_05	-0.2858	3.8264
REM_06	-0.0615	5.9854

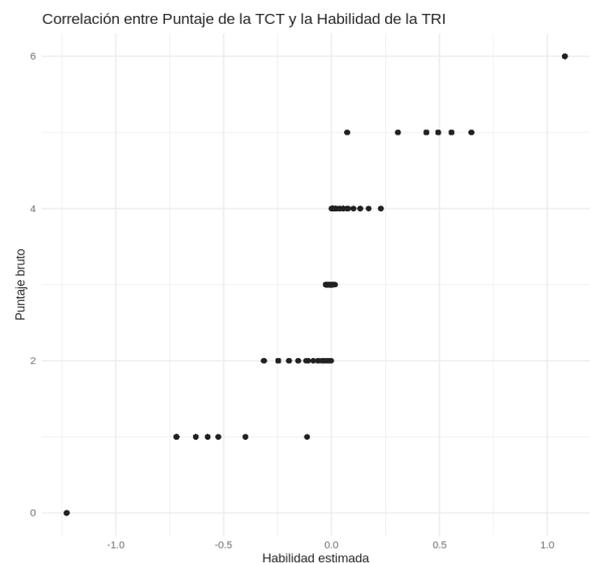
De acuerdo con la Figura 7, la pregunta REM\_02 es el atributo más importante de la dimensión de la remuneración que hay que tener en cuenta para mejorar el CO, mientras que la pregunta REM\_04 no es tan importante y las demás preguntas tienen niveles parecidos de dificultad y discriminación.

**Figura 7**  
Evaluación de la criticidad de los ítems con la TRI con ítems politómicos.



Adicionalmente, por el lado de la estimación de la habilidad con el modelo 2PL, se aprecia de que a pesar de ser diferentes metodologías entre la TRI y la TCT tienden a tener estimaciones parecidas. La utilidad del modelo 2PL es que es más granular las estimaciones, en cambio el modelo de TCT es discreto en cuanto a las puntuaciones (ver Figura 8).

**Figura 8**  
Correlación entre la Puntuación de la TCT y la habilidad de la TRI con ítems dicotómicos



Aplicación del MRG. Se aplicaron dos tipos de modelos MRG que se diferencian por el nivel de restricciones de los parámetros, el primer modelo MRG1 presenta restricciones en el parámetro de discriminación mientras que el modelo MRG2 es

el completo con las discriminaciones y dificultades diferenciadas por pregunta. En la evaluación del ajuste de estos tres modelos propuestos se verificó que el modelo MRG1 presentó un mejor ajuste con los indicadores más bajos (log.lik, AIC y BIC), además, se evidenció que estadísticamente a un 5% no hay diferencias en el ajuste de los dos modelos aplicados (ver Tabla 11). Además, se evidenció de que el modelo MRG1 tiene un nivel de información más alto que los otros modelos, por lo que su precisión es mayor con un SEM más bajo. Por otro lado, el grado de correlación entre la puntuación bruta y la habilidad estimada es alto con un nivel de 0.99.

**Tabla 11**

Comparación del Ajuste de los Modelos MRG para Ítems Politémicos.

Indicador	MRG1	MRG2
FIT ( $-3 \leq \theta \leq 3$ )	84.37	66.25
SEM	0.11	0.12
CORR	0.991195	0.9902875
log.Lik	-33,831.31	-34,412.42
AIC	67,712.63	68,884.84
BIC	67,879.13	69,084.65
P-Valor		1

Respecto a las estimaciones de los parámetros de los ítems del modelo MRG1 (ver Tabla 12), en el análisis de los umbrales de dificultad para las categorías de los ítems se encontró que estos están ordenados de forma creciente, lo que es consistente con el MRG y la discriminación es igual a 4.1 para todas las preguntas. Los valores de umbrales negativos sugieren que las primeras categorías de las preguntas son relativamente fáciles para la mayoría de los colaboradores. Los valores positivos en los umbrales superiores indican que se requiere un nivel más alto de satisfacción para alcanzar dichas categorías.

**Tabla 12**

Parámetros de la TRI para Ítems politémicos con MRG1.

Ítem	Umbral 1	Umbral 2	Umbral 3	Umbral 4
REM_01	-1.749	-1.069	-0.284	0.745
REM_02	-1.670	-0.934	-0.152	0.805
REM_03	-1.671	-0.931	-0.155	0.804
REM_04	-2.173	-1.551	-0.753	0.257
REM_05	-1.996	-1.345	-0.550	0.455
REM_06	-1.874	-1.222	-0.375	0.610

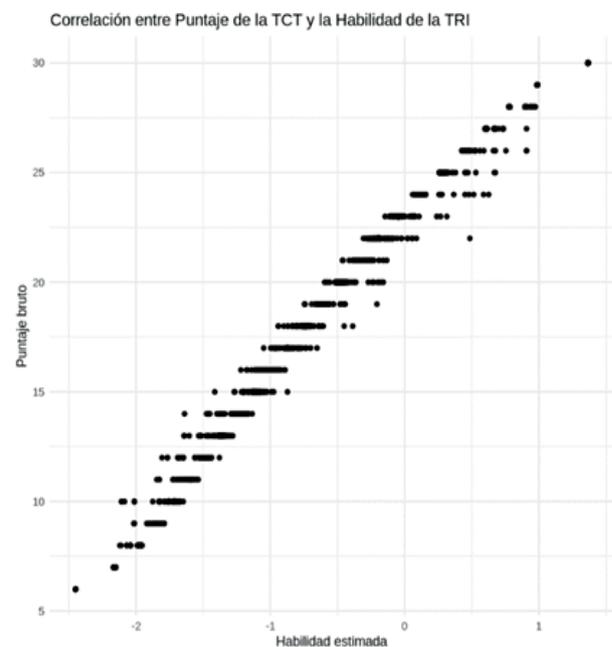
En el MRG1, la pregunta REM\_04 tuvo los menores niveles de umbrales en cada categoría a comparación del resto de preguntas. Por el contrario, las preguntas REM\_02 y REM\_03 alcanzaron mayores niveles de los umbrales, lo que sugiere la exigencia de una mayor satisfacción para responder a estas

preguntas. El resto de las preguntas tienen niveles promedios de los umbrales con un orden de dificultad de mayor a menor entre REM\_01, REM\_06 y REM\_05.

Al comparar la estimación de la habilidad con el modelo MRG1 es más granular las estimaciones en una escala continua de habilidad, en cambio el modelo de TCT es discreto en cuanto a las puntuaciones (ver Figura 9).

**Figura 9**

Correlación entre la Puntuación de la TCT y la habilidad de la TRI con ítems politémicos.



## VI. DISCUSIÓN

El análisis de la satisfacción con el CO en la dimensión de la remuneración se realizó utilizando modelos alternativos de la TRI, específicamente el modelo 2PL y el modelo MRG, los cuales permitieron tener una evaluación detallada de las preguntas y un mejor conocimiento de las percepciones de los colaboradores.

Los resultados del MRG determinaron la existencia de algunos factores críticos, también se determinó que la pregunta REM 04 no aportó en el análisis. Es importante precisar que la discriminación diferenciada por ítem no mejoró la precisión del modelo y resultó suficiente mantener un nivel de discriminación fijo de 4,101 lo cual es alto y sugiere que estos ítems son efectivos para distinguir entre colaboradores con diferentes niveles de satisfacción.

El modelo 2PL también mostró que la pregunta REM 04 no aporta en el análisis del CO, sin embargo, los ítems tenían buenos niveles de discriminación y diferentes niveles, los cuales permitieron realizar un ranking de las preguntas y evaluar su criticidad. Esto sugiere que aunque ambos modelos son efectivos para analizar la satisfacción con la dimensión de remuneración en la ECO, el MRG proporciona una visión más detallada debido a su capacidad para manejar respuestas politómicas.

Respecto a las diferencias encontradas con el modelo de la TCT, se ha encontrado que la estimación de la habilidad es más granular a lo largo de la escala continua de habilidad a comparación de la puntuación obtenida en la ECO, ello permite segmentar mejor a los colaboradores y aplicar mejores estrategias relacionadas con la gestión de los recursos humanos. En análisis de los ítems se han obtenido resultados parecidos, dependiendo del análisis en ítems dicotómicos o politómicos; sin embargo, en la TRI los resultados son modelos más sofisticados por su planteamiento probabilístico, que también mejora la medición de las variables latentes por su independencia del instrumento de medición.

Por otro lado, en el análisis de los ítems se obtuvo una tendencia de las respuestas a las categorías más altas de la escala Likert evidenciando un mayor grado de satisfacción de los colaboradores con el CO. Es posible que haya un efecto adicional en la CCI, hasta ahora se ha considerado la posición de CCI a través de la dificultad y también la pendiente de la CCI por la discriminación, sin embargo, se ha considerado que las curvas son simétricas lo cual no necesariamente se cumplen en la realidad, Tarazona (2015), Dodonov y Dodonova (2010), Bazán (2005) han propuesto modelos con funciones de enlace asimétrica a través de la inferencia bayesiana.

Además, existe la posibilidad de tener un mejor ajuste de los modelos considerando que los colaboradores se diferencian por las áreas donde trabajan, las posiciones, las edades y otras características. No se ha considerado el efecto de los segmentos de colaboradores e incluirlo podría mejorar el análisis del CO a través del uso de los modelos multinivel (Peterson et al, 2020) o los modelos con efectos mixtos (Hachuel et al, 2004) en el enfoque de la TRI.

## VII. CONCLUSIONES

La utilización de los modelos TRI como el modelo 2PL para ítems dicotómicos y el modelo MRG para ítems politómicos representó una mejora metodo-

lógica en el análisis de la satisfacción de los colaboradores con la dimensión de la remuneración de la ECO, proporcionando una comprensión más profunda y detallada de las preguntas formuladas.

El modelo de respuesta graduada demostró ser más detallado en la evaluación de la satisfacción con la remuneración debido a su capacidad para manejar múltiples categorías de respuesta, proporcionando una evaluación más precisa en comparación con el modelo 2PL. Además, en el modelo de respuesta graduada se obtuvo una estimación de la habilidad más granular a comparación de los modelos 2PL y del modelo de la TCT.

Los ítems con altos valores de discriminación son efectivos para diferenciar entre colaboradores con distintos niveles de satisfacción con la remuneración, permitiendo identificar áreas específicas de mejora. Los parámetros de dificultad variados en los ítems sugieren que la encuesta cubre una amplia gama de percepciones, desde muy satisfechos hasta menos satisfechos con la remuneración.

Las preguntas REM\_01, REM\_02 y REM\_03 representan aspectos específicos de la remuneración que necesitan atención y mejora, mientras que la pregunta REM\_04 podría retirarse en la próxima encuesta porque no agrega valor para determinar la satisfacción en la remuneración.

Si bien se ha alcanzado un buen entendimiento del CO, existe la posibilidad de mejorar la precisión de los modelos con planteamientos alternativos que se enfocan en estudios multiniveles o que toman en consideración la asimetría de las respuestas a lo largo de la escala *Likert*.

## VIII. RECOMENDACIÓN

**Modelos alternativos:** Se recomienda explorar otros modelos unidimensionales enfocados a ítems politómicos como es el modelo de respuesta nominal (Bock, 1972), modelo de crédito parcial (Masters, 1982), modelos de respuesta graduada generalizada (Samejima, 1969), modelos de respuesta categórica ordenada (Andrich, 1978) y modelos de respuesta fraccional (Muraki, 1992). También es posible evaluar los modelos multidimensionales cuando se evidencie que la variabilidad de los datos no sea explicada sólo por una variable latente (Adams et al, 1997; Reckase, 2009).

**Fuentes de datos:** Generalmente este tipo de modelos de la teoría de la medición se ha acotado a las pruebas psicométricas, sin embargo, con el presente trabajo se ha evidenciado que también es

plausible aplicarlo a otros tipos de estudios donde se utilicen instrumentos de medición.

**Gestión de los recursos humanos:** Con esta herramienta es posible determinar que preguntas de la encuesta son críticas para la satisfacción del clima organizacional, lo cual representa un insight importante para la gerencia de recursos humanos, porque puede acotar su gestión a ciertos factores y con ello mejorar el clima de la organización.

## REFERENCIAS

- [1] Adams, R., Wilson, M., & Wang, W. (1997). The multidimensional random coefficients multinomial logit model. *Applied psychological measurement*, 21(1), 1-23.
- [2] Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43, 561-573.
- [3] Baker, F. y Kim, S. (2004). *Item response theory: Parameter estimation techniques*. CRC press.
- [4] Bandalos, D. (2018). *Measurement theory and applications for the social sciences*. Guilford Publications.
- [5] Bartholomew, D., Knott, M. y Moustaki, I. (2011). *Latent variable models and factor analysis: A unified approach*. John Wiley & Sons.
- [6] Bartolucci, F., Bacci, S. y Gnaldi, M. (2015). *Statistical analysis of questionnaires: A unified approach based on R and Stata*. Chapman and Hall/CRC.
- [7] Bazán, J. (2005). *Uma familia de modelo de resposta ao ítem normal assimétrica*. Universidad de Sao Paulo, Sao Paulo.
- [8] Birnbaum, A. (1968) Some Latent Trait Models and Their Use in Inferring an Examinee's Ability. In: Lord, F.M. and Novick, M.R., Eds., *Statistical Theories of Mental Test Scores*, Addison-Wesley, Reading, 397-479.
- [9] Bock, R. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, 37(1), 29-51.
- [10] Crocker, L. y Algina, J. (1986). *Introduction to classical and modern test theory*. Holt, Rinehart and Winston, 6277 Sea Harbor Drive, Orlando, FL 32887.
- [11] Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.
- [12] David, A. y Ida, M. (2019). *A Course in Rasch Measurement Theory: Measuring in the Educational, Social and Health Sciences*.
- [13] De Ayala, R. (2013). *The theory and practice of item response theory*. Guilford Publications.
- [14] DeMars, C. (2010). *Item response theory*. Oxford University Press.
- [15] Dodonov, Y. y Dodonova, Y. (2010). Asymmetric item characteristic curve in item response theory: model construction and application. *Procedia Social and Behavioral Sciences*. doi:<https://doi.org/10.1016/j.sbspro.2010.07.330>
- [16] Downey, H., Hellriegel, D. y Slocum, J. (1975). Congruence between Individual Needs, Organizational Climate, Job Satisfaction and Performance. *Academy of Management Journal*, 18(1), 149-155 [DOI: 10.2307/255634].
- [17] Ellis, B., & Mead, A. (2002). Item analysis: Theory and practice using classical and modern test theory. In S. G. Rogelberg (Ed.), *Handbook of research methods in industrial and organizational psychology* (pp. 324-343). Malden, MA: Blackwell. A practical guide for using IRT and CTT in scale and item development, written by two measurement practitioners.
- [18] Embretson, S. y Reise, S. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum.
- [19] Ehrhart, M., Schneider, B., Macey, W. (2014). *Organizational Climate and Culture: An Introduction to Theory, Research, and Practice*. Routledge.
- [20] Hachuel, L., Boggio, G., Wojdyla, D. (2004). Modelos logit mixtos: una aplicación en el área de la salud Modelos logit mixtos: una aplicación en el área de la salud.
- [21] Hambleton, R., Swaminathan, H. y Rogers, H. J. (1991). *Fundamentals of item response theory* (Vol. 2). Sage.
- [22] Ledesma, D., y Mora, P. (2007). Determining the number of factors to retain in EFA: An easy-to-use computer program for carrying out parallel analysis. *Practica Assessment, Research & Evaluation*, 12(2), 1-11.

- [23] Lord, F. (1952). A Theory of Test Scores (Psychometric Monograph No. 7). Richmond, VA: Psychometric Corporation. Retrieved from <http://www.psychometrika.org/journal/online/MN07.pdf>
- [24] MacCallum, R. y Austin, J. (2000). Applications of structural equation modelling in psychological research. *Annual Review Psychology*, Vol. 51:201-226. <https://doi.org/10.1146/annurev.psych.51.1.201>
- [25] Masters, G. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47(2), 149-174.
- [26] McDonald, R. (1999). *Test theory: A unified treatment*. Mahwah, NJ.
- [27] Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied psychological measurement*, 16(2), 159-176.
- [28] Natarajan, V. (2009). Basic principles of IRT and application to practical testing & assessment. Retrieved from <http://www.merittracblog.blogspot.com>
- [29] Peterson, M., Arregle, J. y Martin, X. (2020). Multilevel models in international business research. *Research methods in international business*, 417-432.
- [30] Rasch, G. (1960). *Studies in mathematical psychology: I. probabilistic models for some intelligence and attainment tests*.
- [31] Reckase, M. (2009). Historical background for multidimensional item response theory (MIRT). *Multidimensional item response theory*, 57-77.
- [32] Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika monograph supplement*.
- [33] Shultz, K., Whitney, D. y Zickar, M. (2020). *Measurement theory in action: Case studies and exercises*. Routledge.
- [34] Tarazona, E. (2015). Modelos alternativos de respuesta graduada con aplicaciones en la calidad de servicios. Pontificia Universidad Católica del Perú. *parameters. Journal of Environmental Health Science and Engineering*, 12, 1-8.
- [35] Zhang, D., & Gong, Y. (2020). The comparison of LightGBM and XGBoost coupling factor analysis and prediagnosis of acute liver failure. *IEEE Access*, 8, 220990-221003.

**Financiamiento:**

Propia.

**Conflictos de interés:**

Los autores declaran no tener conflictos de interés.

**Contribuciones de autoría:**

Todos los autores participaron en las diferentes actividades para la elaboración del artículo.