

Bogotá D.C  
Enero 2026  
ISSN: 2590 - 4663  
Publicación Trimestral

  
Educación

  
Icfes

# SABER al detalle

Edición 20

¿Qué son los ítems politómicos y cómo se analizan en el Icfes?

---



## **Saber Al Detalle: ¿Qué son los ítems politómicos y cómo se analizan en el Icfes?**

Subdirección de Estadísticas

Presidente de la República  
**Gustavo Francisco Petro Urrego**

Ministro de Educación Nacional  
**José Daniel Rojas Medellín**

Directora General Icfes  
**Elizabeth Blandón Bermúdez**

Director Técnico de Evaluación  
**Gustavo Andrés Monsalve Londoño**

Subdirector de Diseño de Instrumentos  
**Heider Martínez Mena**

Subdirector de Estadísticas  
**Cristian Fabián Montaña Rincón**

Subdirectora de Análisis y Divulgación  
**Alejandra Neira Aroca**

Elaboración del documento  
**Juan Sebastián Argüello Chiquiza**  
**Luisa Fernanda Castañeda Ballesteros**

Revisado por  
**Mónica Liliana Manrique Galindo**

Diseño y diagramación  
**Linda Nathaly Sarmiento Olaya**

Bogotá D.C., enero de 2026

Todos los derechos de autor reservados ©.

# ¿Qué son los ítems politómicos y cómo se analizan en el Icfes?

## Introducción

En la evaluación educativa, el diseño de los ítems es una etapa fundamental para garantizar la confiabilidad, la validez y la utilidad de una prueba. Tradicionalmente, se han utilizado ítems dicotómicos para evaluar el rendimiento o las habilidades, dado que clasifican las respuestas como correctas e incorrectas en función del nivel de conocimiento requerido (Asún & Zúñiga, 2008). No obstante, este formato puede presentar limitaciones cuando se pretende evaluar constructos complejos o de naturaleza continua, tales como actitudes o competencias comportamentales, como en las pruebas de ejecución típica (Cronbach, 1960) o en la evaluación por parte de jueces expertos por medio de rúbricas.

Frente a estas limitaciones, los ítems politómicos se consolidan como una alternativa metodológica adecuada. A diferencia de los formatos dicotómicos, presentan tres o más categorías de respuesta que representan distintos niveles de intensidad del constructo evaluado, lo que los hace especialmente útiles para evaluar rasgos de personalidad, donde no existen respuestas correctas o incorrectas sino distintos grados del atributo evaluado (Asún & Zúñiga, 2008). En otras palabras, la selección de una determinada opción indica un mayor o menor grado del rasgo latente evaluado, y cada categoría tiene asociado un puntaje diferente (De Ayala, 2022). Esta capacidad para captar variaciones resulta esencial para evaluar constructos que se manifiestan a lo

largo de un continuo, como las habilidades socioemocionales o el clima escolar.

Los ítems politómicos han ganado especial relevancia en pruebas estandarizadas a gran escala, tanto en el contexto colombiano como en el internacional. En la prueba de pensamiento creativo de PISA 2022, por ejemplo, se aplicaron ítems de naturaleza politómica (OECD, 2024). Asimismo, podría ser útil ampliar su implementación en los exámenes de Estado en Colombia con el fin de evaluar nuevas competencias mediante formatos que van más allá de la opción múltiple tradicional.

A continuación, se presentan las características de los ítems politómicos, así como algunas consideraciones sobre los modelos de Teoría de Respuesta al Ítem (TRI), los cuales son utilizados para analizar pruebas que emplean este tipo de ítems.

## 1. Características de los ítems politómicos

Como ya se mencionó anteriormente, los ítems politómicos se caracterizan por ofrecer más de dos alternativas de respuesta, las cuales representan distintos niveles del atributo o constructo evaluado (López Pina, 2005). Estas categorías pueden presentarse de forma ordinal o nominal.

En la práctica, se usan principalmente los ítems politómicos ordinales porque sus

opciones de respuesta siguen un orden fácil de entender. Esto quiere decir que las respuestas avanzan de un nivel menor a uno mayor (o al revés) de forma progresiva. Por ejemplo, se puede pasar de “Nunca” a “Siempre”, o de “Muy en desacuerdo” a “Muy de acuerdo”. Gracias a este orden, es posible medir la frecuencia o la intensidad de una actitud o un comportamiento.

Las categorías ordinales presentan una estructura jerárquica, como ocurre en las escalas tipo Likert (Likert, 1932). Este formato se ha consolidado como el más utilizado en la evaluación de actitudes, intereses y rasgos de personalidad, al permitir que las respuestas se distribuyan a lo largo de un continuo. Habitualmente, estas escalas emplean entre tres y siete categorías, aunque existen variantes con un mayor número de opciones, generalmente impares (López Pina, 2005).

Además de las escalas tipo ordinal, se usa el formato de respuesta nominal, en el que las opciones no presentan un ordenamiento específico (*a priori*), sino que resultan mutuamente excluyentes (Rivera Espejo, 2019). En este formato, cada opción de respuesta resulta conceptualmente diferente, sin que exista una relación de mayor o menor intensidad entre ellas. A diferencia de las escalas ordinales, donde las opciones de respuesta avanzan progresivamente de un nivel a otro, los ítems nominales simplemente ofrecen alternativas distintas sin relación jerárquica aparente.

Los ítems politómicos requieren un análisis psicométrico complejo basado en modelos TRI, los cuales permiten estimar parámetros como la dificultad y la discriminación para cada categoría de respuesta, facilitando la interpretación de cómo los individuos y los

ítems se distribuyen a lo largo del continuo del constructo medido y, en consecuencia, mejorando la precisión en la medición de constructos psicológicos y educativos (Attorresi et al., 2009).

Según Asún y Zúñiga (2008), los estudios que han aplicado modelos politómicos a la medición de actitudes o personalidad se han centrado en cuatro áreas principales:

- » La evaluación del ajuste de modelos politómicos de la TRI a pruebas basadas en la Teoría Clásica de Tests (TCT), obteniendo resultados mixtos sobre con instrumentos no cognitivos.
- » La definición de un número óptimo de alternativas de respuesta cuando se usan escalas arbitrarias como los ítems tipo likert, evidenciando que, en ocasiones, tres o cuatro opciones son suficientes para obtener estimaciones precisas.
- » Las ventajas de estos ítems para el diseño de pruebas adaptativas eficientes, la medición del funcionamiento diferencial de los ítems (DIF) y la equiparación de puntuaciones.
- » La comprobación de nuevos modelos para análisis de ítems politómicos, tanto no paramétricos como aquellos adaptados a escalas tipo likert.

## 1.1 Consideraciones para definir el número de categorías de respuesta

Definir el número de categorías de respuesta en los ítems politómicos constituye una de las

decisiones más relevantes en el proceso de construcción de pruebas, ya que obliga a los diseñadores a considerar diversos factores; entre ellos, la complejidad del constructo que se pretende medir, las capacidades cognitivas y las características de la población objetivo, y las limitaciones prácticas del contexto en el que se aplicará el instrumento. Estas variables relacionan de manera directa con la calidad de las inferencias que pueden realizarse a partir de los puntajes obtenidos.

En relación con esta decisión, no debe asumirse de forma automática que un mayor número de categorías de respuesta conlleva mejores mediciones. De hecho, ante una gran cantidad de categorías, puede haber varias que se usen en proporciones muy bajas (por ejemplo, <5% de las respuestas). Además, esto conlleva a mayor esfuerzo o demora en la respuesta del evaluado, dado que debe gastar tiempo en revisar muchas opciones y entender cada categoría.

En consecuencia, resulta importante que la selección del formato de respuesta sea coherente con el tipo de instrumento, las características de la población, los objetivos de evaluación y el marco teórico que sustenta la prueba. De no considerarse estos aspectos, existe el riesgo de adoptar formatos de respuesta inadecuados para la población evaluada o inconsistentes con el constructo que se busca medir, lo que puede afectar la interpretación de los resultados (Lara, 2013).

No obstante, en la literatura se observa que pocos autores argumentan la elección de una cantidad determinada de categorías de respuesta (por ejemplo, cuatro frente a siete opciones) o la selección de determinados cuantificadores lingüísticos (como “nunca” o “raramente” frente a “muy poco” o “poco”).

Aún menos habitual es la realización de estudios empíricos destinados a evaluar si estas decisiones tienen implicaciones sobre propiedades psicométricas clave, tales como la validez, la confiabilidad o la sensibilidad de las escalas empleadas (Lozano et al., 2008; Abal et al., 2017).

## 2. Análisis psicométrico para ítems politómicos

Según Vásquez Gutiérrez (2022), para ítems ordinales, se utilizan modelos como el modelo de respuesta graduada, el modelo de respuesta graduada modificado, el modelo de crédito parcial y el modelo de crédito parcial generalizado; por su parte, para ítems nominales, se aplican modelos como el modelo de respuesta nominal, el modelo nominal modificado y el modelo de elección múltiple. La selección de un modelo adecuado depende de la naturaleza del constructo evaluado y el número de categorías de respuesta, y requiere una evaluación cuidadosa tanto de los supuestos teóricos de cada modelo como de la evidencia empírica disponible.

En este documento, se presentan, de manera general, algunos modelos Rasch (como el de escala rating, desarrollado por David Andrich, y el modelo de crédito parcial, desarrollado por Geoffrey Masters), no Rasch (como el modelo de crédito parcial generalizado, desarrollado por Eiji Muraki, y el modelo de respuesta graduada, desarrollado por Fumiko Samejima) y el modelo de respuesta nominal (desarrollado por Darrell Bock), siendo estos los que le Icfes ha utilizado históricamente para sus análisis psicométricos. Después, se abordan algunos aspectos que determinan ampliamente la elección de uno u otro

modelo para diferentes tipos de usos y pruebas, según lo expuesto por De Ayala (2022). Sin embargo, a modo de introducción, se describen algunas características compartidas entre modelos dicotómicos y politómicos.

## 2.1 Características de modelos TRI compartidas entre modelos dicotómicos y politómicos

En los boletines número 1 (Icfes, 2018) y número 8 (Icfes, 2020) de Saber al Detalle, se destacan algunos elementos conceptuales fundamentales para entender los modelos de la TRI para ítems dicotómicos, como la ecuación o definición funcional del modelo, la curva característica del ítem y los supuestos básicos de estos modelos de rasgo latente. Estos 3 aspectos son mantenidos en los modelos TRI para ítems politómicos, como se expone a continuación:

### a. Definición funcional del modelo

Cada modelo de TRI define la probabilidad de marcar una determinada respuesta dependiendo del nivel de habilidad de la persona evaluada y dependiendo de los tipos de ítem, como se expone a continuación: En para los ítems dicotomizados (como las preguntas de opción múltiple con única respuesta correcta) esta ecuación funcional corresponde a la probabilidad de marcar la respuesta correcta sobre las incorrectas. Por el contrario, en otros formatos de ítems se estima la probabilidad de marcar la opción “Sí” sobre la opción “No” o la presencia de algún comportamiento o característica en el evaluado sobre la ausencia del mismo. En ciertos casos, la elección de la opción puntuable es arbitraria, dependiendo de los fines evaluativos y las particularidades

técnicas del constructo, dado que se podría puntuar “NO” en lugar de “Sí”.

En la Ecuación 1 se presenta el modelo logístico de dos parámetros (2PL), en esta definición funcional se estima la probabilidad condicionada de acertar un ítem determinado, según el nivel de habilidad de la persona y la dificultad y discriminación<sup>1</sup> del ítem respondido. El modelo 2PL es de común uso para las preguntas dicotómicas en los formatos mencionados con anterioridad.

Del mismo modo, los modelos TRI politómicos se basan en el cálculo de la probabilidad de seleccionar o responder cada ítem con cada una de las 3 o más opciones de respuesta puntuables posibles, según el nivel de habilidad del evaluado y los demás parámetros que considere el modelo. En la Ecuación 2, se define el cálculo de la probabilidad de marcar una opción determinada dentro del modelo de respuesta graduada<sup>2</sup>, el cual se considera como una extensión del modelo 2PL para ítems politómicos.

---

<sup>1</sup> Cuando en el modelo 2PL el parámetro de discriminación se fija en 1 para todos los ítems, el modelo se reduce al 1PL. Sin embargo, esta reducción no implica equivalencia estricta con el modelo de Rasch, cuya formulación incluye una constante de escala.

<sup>2</sup> Cuando se habla del modelo de respuesta graduada, se hace alusión al modelo clásico homogéneo, en el cual se estima un solo parámetro de discriminación por ítem. En el apartado de “Aspectos que determinan la elección de modelos politómicos” se profundiza en esta característica del modelo.

- **Ecuación 1. Función de probabilidad del Modelo 2PL**

$$P(U_{ij} = 1 | \theta_j, a_i, b_i) = \frac{e^{a_i(\theta_j - b_i)}}{1 + e^{a_i(\theta_j - b_i)}}$$

- **Ecuación 2. Función de probabilidad del Modelo de Respuesta Graduada**

$$P(U_{ij} \geq k | \theta_j, a_i, b_{ik}) = \frac{e^{a_i(\theta_j - b_{ik})}}{1 + e^{a_i(\theta_j - b_{ik})}}$$

Donde:

- »  $\theta_j$  es el nivel de la habilidad (rasgo latente) de la persona “j”.
- »  $U_{ij}$  es la respuesta al ítem “i” de la persona “j”.
- »  $k$  es el puntaje de la categoría seleccionada o respondida por la persona.
- »  $b_i$  es el parámetro de dificultad del ítem “i”.
- »  $b_{ik}$  es el parámetro de ubicación del límite de la categoría con puntaje “k” y la subsiguiente categoría “k + 1” para el ítem “i”.
- »  $a_i$  es el parámetro de discriminación del ítem “i”.
- »  $e$  es la constante de Euler.

Nótese que, para el modelo dicotómico 2PL, la función se presenta como la probabilidad de puntuar uno (1), lo que significa que se modela la probabilidad de acertar o marcar la opción deseable sobre las que no lo son, la cual se cuantificara con el valor cero (0);

en contraste, para el modelo politómico de respuesta graduada, su ecuación funcional cuantifica la probabilidad de obtener el puntaje correspondiente a una determinada categoría o uno mayor.

El parámetro de ubicación del límite de la categoría del modelo de respuesta graduada (en inglés, category boundary location, denominado así por De Ayala (2022)), es el nivel de habilidad en el cual es igualmente probable marcar una determinada categoría (k) o su inmediatamente superior (k + 1); por esta razón, se considera que el parámetro de dificultad del modelo 2PL es una expresión del parámetro de ubicación del límite, ya que la dificultad del ítem sería el nivel de habilidad donde se espera que sea igualmente probable acertar (k + 1 = 1) o fallar (k = 0).

Una consecuencia de esta definición de probabilidad para modelos politómicos es que se tendrá una cantidad de ecuaciones funcionales igual al número de categorías de respuesta menos 1; por ejemplo, para un ítem con tres categorías de respuesta puntuables con uno (1), dos (2) y tres (3) puntos, se tendrán dos ecuaciones funcionales diferentes: para evaluados que responden la categoría que puntúa 1 con parámetro de ubicación del límite de responder 2 o más puntos y, para evaluados que responden la categoría que puntúa 2 con parámetro de ubicación del límite de responder 3 puntos. Esta característica puede observarse de mejor manera con las curvas características de los ítems que se verán en el siguiente apartado.

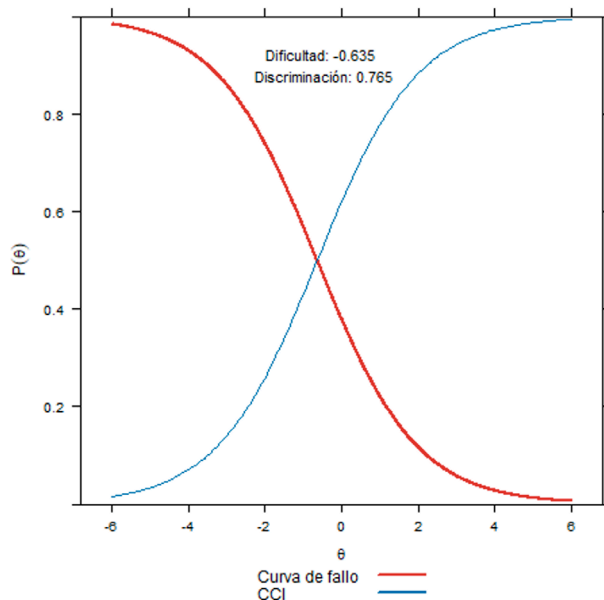
### b. Curvas características de los ítems

La curva característica de los ítem (CCIs) es la forma gráfica de representar la función (o funciones) de probabilidad de la respuesta

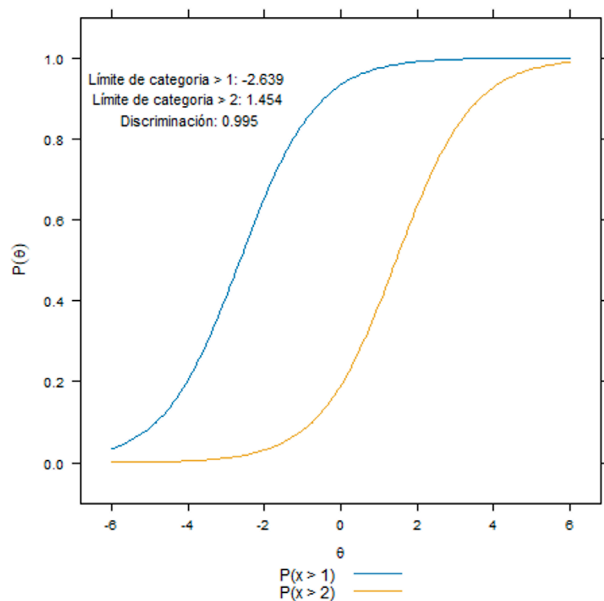
a cada ítem. Por norma general, para los modelos dicotómicos, una CCI representa la función de acierto y pocas veces se ve representada gráficamente la función de probabilidad de fallar, siendo esta última matemáticamente complementaria a la función de probabilidad de acierto. La ilustración 1 muestra la CCI junto con la curva de probabilidad de fallo de un ítem dicotómico.

Para los modelos politómicos, y en particular para el modelo graduado, existen dos representaciones gráficas de las funciones de probabilidad del modelo. La primera representación, denominada curvas de límites de categorías (en inglés, *category boundary curves*) por De Ayala (2022), representa las funciones de probabilidad de los límites de categorías. Por lo que cada ítem, según lo expuesto en el apartado anterior, tendrá una cantidad de curvas de límites de categorías igual al número de categorías puntuables menos uno. Por su parte, el segundo tipo de gráfico es análogo a las CCI de los ítems dicotómicos y se denomina gráfico de función de respuesta por opción (en inglés, *option response function*). Esta representación, derivada matemáticamente de las funciones de límites de categoría, proporciona información sobre la probabilidad de marcar cada una de las opciones del ítem. Aunque estos nombres son propios del modelo graduado, todos los modelos politómicos generan gráficas análogas de ambos tipos para representar el comportamiento de cada ítem. En las ilustraciones 2 y 3, se muestra un ejemplo de estas dos gráficas para un mismo ítem de tres opciones:

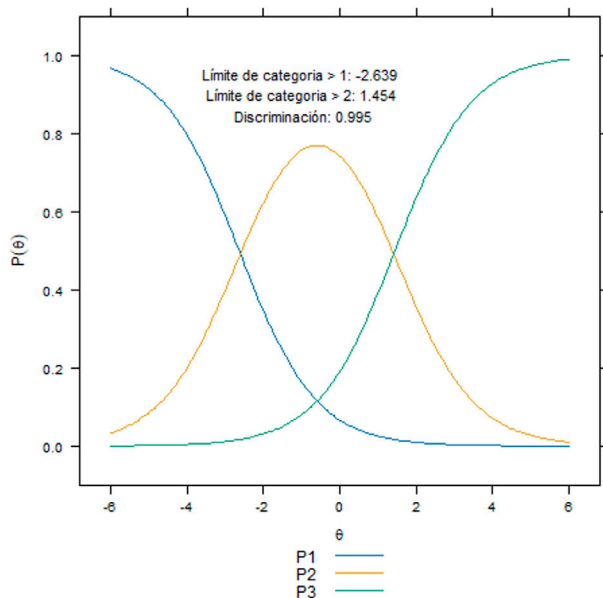
• **Ilustración 1. Curvas Características de un ítem dicotómico 2PL**



• **Ilustración 2. Curvas de límites de categorías de un ítem con tres opciones de respuesta**



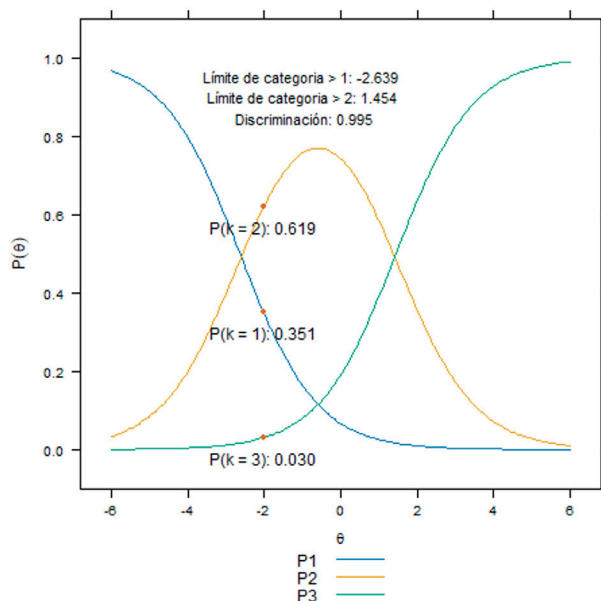
• **Ilustración 3. Curvas de función de respuesta por opción de un ítem con tres opciones de respuesta**



Al comparar las ilustraciones 1 y 3, se puede apreciar que el gráfico de funciones de respuesta por opción del modelo graduado y la CCI de un ítem dicotómico, graficada junto con la curva de fallo, comparten la propiedad en la cual en cualquier punto del eje  $X$  (denominado " $\theta$ ", que representa el nivel de habilidad del evaluado en escala logit), la suma de los valores de probabilidad (eje  $Y$ ) de las diferentes curvas es igual a uno. Esta propiedad puede resultar trivial en ítems dicotómicos, ya que, como se mencionó en el apartado anterior, la función de probabilidad de fallo es el complemento de la CCI. Sin embargo, en los modelos politómicos, al tratarse de modelos de TRI probabilísticos, indica que cualquier evaluado, independientemente de su nivel de habilidad, tiene una probabilidad (así sea ínfima) de responder cualquiera de las opciones del ítem. Para ejemplificar, en la ilustración 4, se señalan con puntos rojos las probabilidades

de seleccionar cada una de las tres opciones para el ítem de las ilustraciones 2 y 3, para un evaluado con una habilidad cercana a  $-2$  ( $\theta \approx -2.020$ ). Se puede observar que estas probabilidades suman 1 y que, con cualquier evaluado con este nivel de habilidad tiene solo un 3% de probabilidad de seleccionar la opción de mayor puntaje (P3).

• **Ilustración 4. Probabilidad de seleccionar cada una de las opciones de un ítem dado que un evaluado cuente con un nivel de habilidad, en logits, cercano a  $-2$**



**c. Supuestos básicos de los modelos TRI**

Este es el aspecto en el que los modelos politómicos presentan mayores similitudes con los dicotómicos. Los supuestos básicos de los modelos TRI son la dimensionalidad, la independencia local y la invarianza de ítems y evaluados, los cuales se revisan, a continuación, de manera general:

- » **Dimensionalidad:** Hay diversos tipos de modelos que asumen

cantidades diferentes de habilidades o rasgos latentes determinantes en las respuestas de los ítems. Existen modelos politómicos que asumen que las probabilidades de respuesta dependen de una habilidad única (modelos unidimensionales, que son utilizados por el Icfes) y otros que asumen que dependen de varias (modelos multidimensionales).

- » **Independencia Local:** este supuesto se refiere a que la probabilidad de respuesta para un ítem en particular debe ser independiente otros factores diferentes a la habilidad del evaluado. En este sentido, la presencia de ítems que afecten o sesguen los procesos de respuestas e otros ítems dentro de la prueba, o la deseabilidad social y otros sesgos de autorreporte, entre otros, son factores para tener en cuenta a la hora de analizar pruebas de ejecución típica.
- » **Invarianza entre ítems y personas:** para garantizar las cualidades psicométricas de los instrumentos de medida, es necesario que los parámetros de los ítems no varíen significativamente entre subpoblaciones (p. ej., género) ni entre versiones de la prueba (p. ej., para dos aplicaciones distintas del mismo instrumento). Cuando hay cambios sustanciales en las propiedades métricas de un ítem se conoce como funcionamiento diferencial del ítem (DIF, por sus siglas en inglés), por lo que los. Algunos procedimientos para la detección del DIF permiten abordar este supuesto (para más información, ver el Boletín número 15 de Saber al Detalle (Icfes, 2024)).

## 2.2 Aspectos a considerar en la elección de un modelo politómico

Los aspectos por considerar se enmarcan en tres categorías: lo concerniente a la cantidad de parámetros utilizados por el modelo para estimar sus funciones de probabilidad, la relación entre estos parámetros al interior de cada ítem y, por último, las suposiciones que hace el modelo sobre el contenido del ítem. A continuación, se detalla cada categoría:

### 2.2.1 Cantidad de parámetros estimados

Al igual que para los modelos TRI dicotómicos, según los intereses de los usuarios y desarrolladores de una prueba, es posible escoger un mayor o menor número de parámetros a expensas de mayores necesidades de recursos computacionales. Los parámetros obligatorios son los de discriminación y dificultad, esta última representada por los parámetros de ubicación del límite de las categorías, que se denominan, en inglés, con los términos *Threshold* o *Step*, según el modelo.

Cuando los desarrollares de una prueba no consideran necesario contar con información sobre la discriminación de los ítems, el parámetro de discriminación se fija en un valor constante para todos los ítems; específicamente, cuando el valor de la discriminación se fija en uno (1), se tiene un modelo de Rasch; en el caso contrario, cuando se estima la discriminación para cada ítem, se habla de un modelo no Rasch.

En otros modelos, se consideran parámetros adicionales, dadas las características de los ítems; por ejemplo, en el Modelo de Múltiples Facetas de Rasch, utilizado para generar

puntajes en pruebas calificadas por jueces, se estima un parámetro de severidad para cada juez interviniente en el proceso de puntuación del desempeño de un evaluado, que intenta controlar el efecto de jueces laxos y estrictos. Este modelo, en particular, es usado para las pruebas de Comunicación escrita y arquitectura de los exámenes Saber Pro y Saber TyT, y las preguntas abiertas de los exámenes Saber 3°, 5°, 7° y 9°, aplicadas por el Icfes. La calificación que hace el Icfes de este tipo de preguntas se describe de manera más completa en el Boletín 7 de Saber al Detalle (Icfes, 2020).

Por otro lado, el modelo de Escala de *rating*, además de estimar los parámetros de ubicación del límite de las categorías (*Thresholds*, en este modelo), estima un parámetro general del nivel de dificultad del ítem, denominado parámetro de ubicación (*Location*), el cual interactúa con los *Thresholds* para determinar las probabilidades de respuesta, en cada nivel de habilidad, para cada opción.

### 2.2.2 Relación entre los parámetros estimados

Según el modelo utilizado, se hacen ciertas asunciones sobre las relaciones que se tienen entre los diferentes parámetros de ubicación del límite de las categorías y con el parámetro de discriminación.

En cuanto a la discriminación y su relación con los demás parámetros, se tiene la restricción de la discriminación constante, lo cual todas las categorías de respuesta tienen el mismo nivel de discriminación y, por tanto, solo es necesario estimar (o fijar en el caso de los modelos Rasch) un parámetro general

de discriminación. Esta asunción es adoptada por los modelos de escala *rating* y crédito parcial, en el caso de los modelos Rasch, y por los modelos de respuesta graduada y de crédito parcial generalizado, en los modelos no Rasch. El modelo de respuesta nominal no asume tal condición, sino que estima un nivel de discriminación diferente para cada una de las categorías que conforman cada ítem y, además, derivado de esta asunción de discriminación variable, estima un parámetro de pseudodificultad para cada categoría, denominado parámetro de tendencia (en inglés, *Propensity*), que se aproxima a los parámetros de ubicación del límite de los demás modelos.

Sobre las relaciones entre los parámetros de ubicación del límite de las categorías, está la restricción de adyacencia, presente en los modelos de respuesta graduada y de escala *rating*, según el cual las categorías de respuesta están natural y correctamente ordenadas, forzando a que se respete ese ordenamiento en los parámetros, esto es, que el parámetro de ubicación del límite de entre las categorías que puntúan 1 y 2 siempre sea menor que el estimado entre las categorías 2 y 3, y así sucesivamente. Esta restricción y teniendo en cuenta que la discriminación se fija para todos los ítems del test, el modelo de escala *rating* estimará los mismos *Thresholds* para todos los ítems de una prueba, razón por la cual el parámetro de ubicación anteriormente descrito cobra sentido para “desplazar”, por medio de transformaciones, a estos parámetros de ubicación del límite a lo largo de los niveles de habilidad que sean evaluables, según la dificultad de cada ítem y manteniendo la equidistancia original estimada que resulta por tener la discriminación constante.

En el otro lado, se encuentran los modelos de crédito parcial, crédito parcial generalizado y de respuesta nominal, que estiman los parámetros de ubicación del límite para cada categoría, independientemente de la relación de orden *a priori* que pueda tener con las demás categorías de respuesta del ítem.

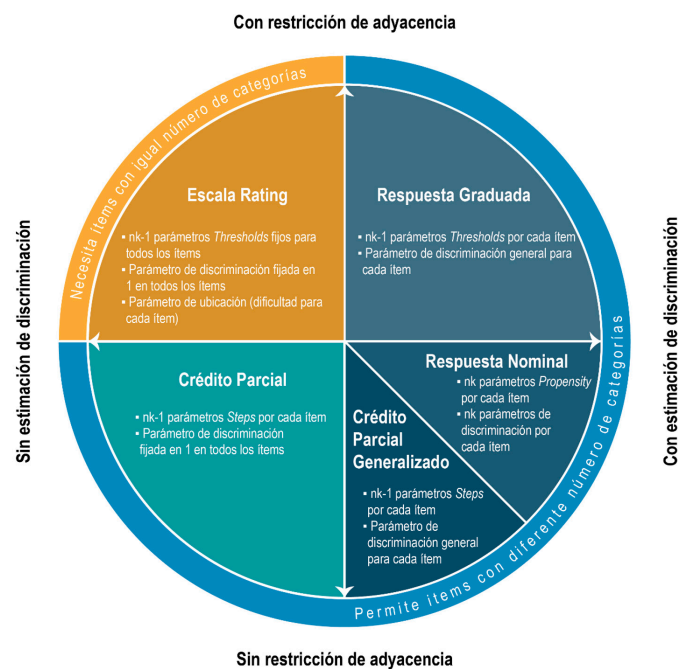
### 2.2.3 Suposiciones sobre contenido del ítem

El modelo de escala de rating se fundamenta en el supuesto de que la escala de valoración es la misma para los diferentes ítems, es decir que este modelo solo puede ser estimado cuando todos los ítems tienen el mismo número de categorías y su escala es comparable. Los ítems Likert son los formatos de ítem ideales para este modelo, ya que se suelen utilizar escalas de graduación del mismo tipo; por ejemplo, escalas de acuerdo o frecuencia con el mismo número de

opciones para los diferentes ítems. Esta suposición no es asumida por los demás modelos de interés, permitiendo estimar propiedades psicométricas para conjuntos de ítems con diferentes cantidades de puntuaciones, escalas de graduación o, en el caso del modelo de respuesta nominal, con categorías independientes sin un ordenamiento aparente.

Para finalizar, en la Ilustración 5, se resumen los aspectos determinantes para la elección de modelos politómicos. Dentro de cada uno de los cuadrantes de la figura, se lista la cantidad de parámetros a estimar o fijar para cada modelo. Se puede observar que el modelo escala rating es el que exige menos recursos computacionales, mientras que el modelo de respuesta nominal es el que más recursos necesita. Este efecto sobre los recursos computacionales se puede apreciar mayormente en evaluaciones a gran escala o con gran número de evaluados.

• Ilustración 5. Diagrama de propiedades de algunos modelos TRI politómicos



nk = número de categorías puntuables del ítem

Steps/Thresholds = parámetros de límite de ubicación de categorías de respuesta

### 3. Uso de estos modelos politómicos en el Icfes

Debido al aumento en la capacidad computacional y en la disponibilidad de nuevos desarrollos tecnológicos y de software, el Icfes ha migrado paulatinamente de los modelos Rasch a los modelos no Rasch aquí descritos, utilizando desde 2016 un modelo mixto en la estimación del Índice de Nivel Socioeconómico (INSE) y el Nivel Socioeconómico (NSE), en el que cada uno de los ítems puede ser analizado a través de diferentes tipos de modelos. Este es el caso del Cuestionario Socioeconómico de los evaluados de todos los exámenes de Estado (Icfes, 2019), así como es el caso del cuestionario de Clima Escolar para los estudiantes de Saber 11, aplicado desde 2023.

Estas prácticas son coherentes con los estándares metodológicos de evaluaciones internacionales de referencia como PISA 2022, que también empleó un enfoque mixto que combina el modelo logístico de dos parámetros (2PL) para ítems dicotómicos y el Modelo de Crédito Parcial Generalizado para ítems politómicos de las diferentes pruebas aplicadas (OECD, 2024).

Por su parte, para los exámenes Saber 3°, 5°, 7° y 9°, se han utilizado modelos de respuesta graduada para generar puntajes del componente no cognitivo, integrado por el cuestionario Socioeconómico (INSE-NSE), los cuestionarios para docentes, el cuestionario para directivos docentes, el cuestionario de Habilidades Socioemocionales y el Cuestionario de Factores Asociados al Desempeño Escolar, así como para la prueba de Acciones y Actitudes Ciudadanas.

En la misma línea, en algunos otros proyectos de evaluación de la educación básica, con pruebas similares, como la medición de trayectorias escolares y educativas o pruebas CRESE, también se ha hecho uso de estos modelos para generar puntajes. Por último, el modelo de respuesta nominal ha sido empleado para determinar graduaciones en ítems con escalas sin un aparente ordenamiento natural.

### Bibliografía

- Attorresi, H. F., Lozzia, G. S., Abal, F. J. P., Galibert, M. S., & Aguerri, M. E. (2009). Teoría de Respuesta al Ítem. Conceptos básicos y aplicaciones para la medición de constructos psicológicos. *Revista Argentina de Clínica Psicológica*, 18(2), 179-188.
- Asún, R., & Zúñiga, C. (2008). Ventajas de los Modelos Politómicos de Teoría de Respuesta al Ítem en la Medición de Actitudes Sociales. El Análisis de un Caso. *Psyche (Santiago)*, 17(2), 103-115.
- Abal, F. J. P., Auné, S. E., Lozzia, G. S., & Attorresi, H. F. (2017). Funcionamiento de la categoría central en ítems de Confianza para la Matemática. *Evaluar*, 17(2), 18-31.
- Cronbach, L. J. (1960). *Essentials of Psychological Testing* (2nd ed.). New York: Harper & Row.
- De Ayala, R. J. (2022). *The Theory and Practice of Item Response Theory* (2nd ed.). The Guilford Press.
- Instituto Colombiano para la Evaluación de la Educación – Icfes. (2018). ¿Cómo se generan los puntajes en las pruebas saber del Icfes? *Saber al Detalle. Edición 01*.

Disponible en: <https://www.icfes.gov.co/publicaciones-icfes/documentos-tecnicos-y-metodologicos/saber-al-detalle/>

Instituto Colombiano para la Evaluación de la Educación – Icfes. (2019). ¿Cómo se construye el Índice de Nivel Socioeconómico (INSE) en el contexto de las pruebas Saber? *Saber al Detalle. Edición 04*. Disponible en: <https://www.icfes.gov.co/publicaciones-icfes/documentos-tecnicos-y-metodologicos/saber-al-detalle/>

Instituto Colombiano para la Evaluación de la Educación – Icfes. (2020). ¿Cómo se califican las preguntas de respuesta abierta en el contexto de las pruebas Saber? *Saber al Detalle. Edición 07*. Disponible en: <https://www.icfes.gov.co/publicaciones-icfes/documentos-tecnicos-y-metodologicos/saber-al-detalle/>

Instituto Colombiano para la Evaluación de la Educación – Icfes. (2021). ¿Cuáles son los modelos de calificación de las pruebas Saber? *Saber al Detalle. Edición 08*. Disponible en: <https://www.icfes.gov.co/publicaciones-icfes/documentos-tecnicos-y-metodologicos/saber-al-detalle/>

Instituto Colombiano para la Evaluación de la Educación – Icfes. (2025). ¿Qué es el Funcionamiento Diferencial de los Ítems y cómo se detecta en las pruebas que aplica el Icfes? *Saber al Detalle. Edición 15*. Disponible en: <https://www.icfes.gov.co/publicaciones-icfes/documentos-tecnicos-y-metodologicos/saber-al-detalle/>

Lara, S. A. D. (2013). ¿Ítems Politómicos o Dicotómicos? Un estudio empírico con una escala unidimensional. *Revista Argentina de Ciencias del Comportamiento*, 5(3), 30-37.

Likert, R. A. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 22(140), 1-55.

López Pina, J. A. (2005). Ítems politómicos vs. dicotómicos: Un estudio metodológico. *Anales de Psicología*, 21(2), 339-344.

Lozano, L. M., García-Cueto, E., & Muñiz, J. (2008). Effect of the number of response categories on the reliability and validity of rating scales. *Methodology*, 4(2), 73-79.

OECD (2024), PISA 2022 Technical Report, PISA, OECD Publishing, Paris, [https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/03/pisa-2022-technical-report\\_599753f0/01820d6d-en.pdf](https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/03/pisa-2022-technical-report_599753f0/01820d6d-en.pdf)

Rivera Espejo, J. M. (2019). *El modelo de respuesta nominal: Aplicación a datos educacionales* [Tesis de maestría, Pontificia Universidad Católica del Perú]. PUCP-Institucional. <http://hdl.handle.net/20.500.12404/14600>

Vásquez Gutiérrez, L. (2022). *Exploración de la idoneidad de las alternativas de respuesta en ítems de elección única de pruebas estandarizadas* [Tesis de maestría profesional, Universidad de Costa Rica]. Repositorio Kérwa. <https://www.kerwa.ucr.ac.cr/server/api/core/bitstreams/b38eb513-d600-4c0d-b6ef-ab6bb0ca951b/content>