



# Simulación computacional de predicciones de rendimiento en el contexto de pruebas estandarizadas

**Diana Chacón-Camacho**\* 

**Vanessa Smith-Castro**† 

**Alejandra Alfaro-Barquero**‡ 

**Mauricio Molina-Delgado**§ 

Recibido: 30 de junio de 2023  
Evaluado: 8 de marzo de 2024  
Publicado: 1 de enero de 2025

## Resumen

Cuando una persona se enfrenta a una prueba realiza predicciones acerca de su rendimiento, estas estimaciones se conocen como juicios metacognitivos. Los juicios metacognitivos pueden diferir del rendimiento real por factores como la ansiedad, el género, entre otros. Frecuentemente la diferencia entre el juicio metacognitivo y el resultado real genera un efecto negativo que a largo plazo puede comprometer el autoconcepto de la persona. Para estimar la probabilidad de acertar un ítem en una prueba estandarizada se utiliza la ecuación de Rash que depende de dos variables, la habilidad del sujeto resolutor en el constructo y la dificultad del ítem. La hipótesis del presente estudio es que los humanos realizamos un proceso análogo al descrito por la ecuación de Rash para establecer los juicios metacognitivos. El presente estudio tiene como objetivo proponer un modelo computacional de la relación entre el desempeño de ítems en pruebas estandarizadas de matemática y los juicios metacognitivos de la predicción del rendimiento. Se realizaron múltiples simulaciones computacionales en las cuales se asume la ecuación de Rash para generar estimaciones de las posibilidades de los sujetos de acertar un ítem alterando mediante errores aditivos los parámetros de habilidad del sujeto y dificultad del ítem. Se compararon estas simulaciones con juicios metacognitivos emitidos por estudiantes en pruebas anteriores y se encontraron dos simulaciones con resultados similares a la de los datos reales.

## Palabras clave

Juicios; cognición; modelo de simulación

\* dianalucia.chacon@ucr.ac.cr

† vanessa.smith@ucr.ac.cr

‡ alealfaro@itcr.ac.cr

§ mauricio.molina.delgado@gmail.com

# Computer Simulation of Performance Predictions in the Context of Standardized Tests

## Abstract

When a person faces a test, they make predictions about their performance; these estimates are known as metacognitive judgments. Metacognitive judgments can differ from actual performance due to factors such as anxiety, gender, and others. Frequently, the difference between metacognitive judgment and actual results generates a negative effect that, in the long term, can compromise a person's self-concept. To estimate the probability of correctly answering an item in a standardized test, the Rasch equation is used, which depends on two variables: the subject's ability in the construct and the difficulty of the item. The hypothesis of this study is that humans perform a process analogous to that described by the Rasch equation to make metacognitive judgments. The aim of this study is to propose a computational model of the relationship between item performance in standardized math tests and metacognitive judgments of performance prediction. Multiple computer simulations were conducted, assuming the Rasch equation to generate estimates of the likelihood of subjects correctly answering an item by altering the ability parameters of the subject and the difficulty of the item through additive errors. These simulations were compared with metacognitive judgments made by students in previous tests, and two simulations were found to have results similar to the real data.

## Keywords

metacognitive judgments; Rasch equation; computational modeling

## Simulação computacional de previsões de desempenho no contexto de testes padronizados

## Resumo

Quando uma pessoa enfrenta um teste, ela faz previsões sobre seu desempenho; essas estimativas são conhecidas como julgamentos metacognitivos. Os julgamentos metacognitivos podem diferir do desempenho real devido a fatores como ansiedade, gênero, entre outros. Frequentemente, a diferença entre o julgamento metacognitivo e os resultados reais gera um efeito negativo que, a longo prazo, pode comprometer o autoconceito da pessoa. Para estimar a probabilidade de acertar um item em um teste padronizado, é utilizada a equação de Rasch, que depende de duas variáveis: a habilidade do sujeito no construto e a dificuldade do item. A hipótese deste estudo é que os humanos realizam um processo análogo ao descrito pela equação de Rasch para fazer julgamentos metacognitivos. O objetivo deste estudo é propor um modelo computacional da relação entre o desempenho em itens de testes padronizados de matemática e os julgamentos metacognitivos de previsão de desempenho. Foram realizadas múltiplas simulações computacionais nas quais se assume a equação de Rasch para gerar estimativas das probabilidades dos sujeitos acertarem um item, alterando os parâmetros de habilidade do sujeito e a dificuldade do item por meio de erros aditivos. Essas simulações foram comparadas com julgamentos metacognitivos emitidos por estudantes em testes anteriores, e duas simulações apresentaram resultados semelhantes aos dados reais.

## Palavras-chave

julgamentos metacognitivos; equação de Rasch; modelagem computacional

## Para citar este artículo:

Chacón-Camacho, D., Smith-Castro, V., Alfaro-Barquero, A. y Molina-Delgado, M. (2025). Simulación computacional de predicciones de rendimiento en el contexto de pruebas estandarizadas. *Revista Colombiana de Educación*, (94), e19808. <https://doi.org/10.17227/rce.num94-19808>

## Introducción

Entender el proceso cognitivo que se lleva a cabo en la solución de ítems es de suma importancia en la actualidad, debido al rol protagónico que cumplen las pruebas estandarizadas en los sistemas de educación y los modelos de evaluación. Las pruebas estandarizadas como PISA (Programme for International Student Assessment), bachillerato internacional y exámenes de admisión a las universidades son consideradas un referente de la calidad de la educación (Consejo Nacional de Rectores [Conare], 2017); por tanto, es necesario comprender los procesos cognitivos involucrados en la solución de estas pruebas.

Uno de los procesos metacognitivos que las personas efectúan es el correspondiente a estimaciones sobre la posibilidad de resolver con éxito los ítems (Alfaro, 2014; Macbeth *et al.*, 2009; Valenzuela, 2018), estas estimaciones se denominan juicios metacognitivos. Este cálculo puede presentar sesgos (errores sistemáticos) con respecto al desempeño real de las personas, lo que significa que las estimaciones realizadas no están ajustadas con respecto al desempeño real.

La modelación computacional ha sido empleada por distintas personas investigadoras para analizar fenómenos cognitivos (Curşeu *et al.*, 2014; Kaber y Kim, 2011; Thilakarathne y Treur, 2015; Vancouver y Purl, 2017). Estos modelos son especialmente útiles porque permiten, primero, contrastar los resultados del propio modelo con los datos empíricos; segundo, realizar una gran cantidad de iteraciones de los procesos, y tercero (Summerfield y Yeung, 2014), modelar la estructura y la fluidez de la toma de decisiones, entre ellas la solución de tareas. Sin embargo, se ha utilizado menos en el campo de la metacognición (Kepecs y Mainen, 2012).

El presente estudio tiene como objetivo proponer un modelo computacional de la relación entre los juicios metacognitivos de la predicción del rendimiento (juicios que las personas establecen sobre su desempeño cognitivo) y el desempeño en ítems de pruebas estandarizadas, a partir de la idea de que los seres humanos establecen una especie de cálculo probabilístico informal para evaluar su rendimiento. De ser esto cierto, los seres humanos emplearían información sobre sus habilidades y sobre la dificultad de los ítems de pruebas que podría tener sesgos de diversos tipos; sin embargo, serían capaces de realizar un proceso análogo al empleado en el cálculo de probabilidades para estimar sus posibilidades de tener éxito. Este enfoque concuerda con lo propuesto por Kepecs y Mainen (2012), para

quienes los juicios de confianza, lo mismo que las predicciones de resultados en pruebas estandarizadas, no corresponden exclusivamente a un proceso de alto nivel propio del ser humano, sino a formas de razonamiento probabilístico que responden a algoritmos relativamente simples para ser implementados en el sistema nervioso.

## Juicios metacognitivos

El concepto de metacognición, desarrollado por Flavell (1976), posee dos aspectos esenciales: el primero refiere al conocimiento de los procesos y productos cognitivos de uno mismo, y el segundo, al análisis, regulación y organización de ese conocimiento de manera consciente.

En el trabajo de Flavell (1976) se presentan algunos ejemplos de la metacognición como conocimiento de los procesos y productos cognitivos, cuando una persona infante aprende que la información que posee es insuficiente para justificar los juicios sobre la confianza en otra persona. O, bien, cuando comprende que existen algunas tareas cognitivas más difíciles que otras.

Wilson (1997) destaca las siguientes tres funciones de la metacognición:

1. Conocer: dónde se encuentra en el proceso de aprendizaje, las estrategias personales y lo que necesita ser hecho.
2. Evaluar: los juicios sobre la propia capacidad y las limitaciones del pensamiento.
3. Regular: el propio pensamiento, usando sus habilidades de control.

Basándose en las dos funciones de la cognición descritas anteriormente por Flavell (1976), Efkliides (2006) presenta un esquema de tres facetas de la metacognición y sus respectivas manifestaciones. Con respecto a la función de monitoreo, Efkliides considera las facetas de conocimiento metacognitivo (MK, por su sigla en inglés) y experiencias metacognitivas (ME, por su sigla en inglés), mientras que para la función de control establece la faceta de habilidades metacognitivas (MS, por su sigla en inglés).

El conocimiento metacognitivo hace referencia a los conocimientos, ideas, creencias y teorías de sí mismo, y de su conocimiento acerca de las tareas, objetivos, acciones, estrategias, funciones vinculadas con su cognición, etc. Los conocimientos metacognitivos permiten comprender la cognición propia y la de los otros (Efkliides, 2006).

Para Efklides (2006), las habilidades metacognitivas son acciones específicas que una persona realiza con el fin de controlar la cognición. Como ejemplos de las habilidades metacognitivas están la orientación y monitoreo de la comprensión de la tarea, planificar los pasos a seguir para resolver una tarea, verificación y regulación del procesamiento cognitivo y evaluación del resultado del proceso cognitivo.

Las experiencias metacognitivas se dividen en dos: los sentimientos metacognitivos y los juicios o estimaciones metacognitivas (Efklides, 2006). Ambas manifestaciones cognitivas provienen del monitoreo del procesamiento de tareas o del resultado de estas. Como ejemplo de los sentimientos que las personas perciben al enfrentarse a una tarea metacognitiva están: familiaridad, dificultad, saber, confianza y satisfacción. Los juicios o estimaciones pueden referir a la calidad del aprendizaje, de la información o del desempeño en una tarea.

Efklides y Tsiora (2002) resaltan que los sujetos establecen sentimientos y juicios sobre la cognición; además señalan que estas experiencias provienen del autocontrol y monitoreo del procesamiento cognitivo, es decir, las ME son juicios y reacciones en el dominio cognitivo.

Los juicios, para Efklides y Petkaki (2005), se alimentan del aprendizaje alcanzado, el esfuerzo cognitivo y algunos otros factores, a su vez sirven para tomar control del procesamiento cognitivo de una tarea y la respuesta que se tendrá ante esta. Además, para los autores, las experiencias metacognitivas pueden tener lugar antes, durante o después del procesamiento de una tarea.

Allueva (2002) considera también la habilidad de la predicción dentro de los procesos metacognitivos. Esta consiste en prever si se es capaz de resolver un problema, lo cual incluye predecir si se tienen los conocimientos necesarios para dicha resolución.

Valenzuela (2018) destaca dos tipos de juicios que las personas pueden hacer: los primeros son aquellos sobre su aprendizaje (JOL, por su sigla en inglés), y los segundos son las predicciones sobre su rendimiento (POP, por su sigla en inglés). Los JOL se basan en el rendimiento actual respecto a lo bien o mal que se ha aprendido algo, de acuerdo con el objetivo de aprendizaje deseado. Los POP son estimaciones a futuro sobre el desempeño que se tendrá en una tarea.

En este estudio son de interés los juicios metacognitivos que las personas realizan para predecir el desempeño que se tiene en la solución de un conjunto de ítems en el contexto de una prueba.

En resumen, una parte importante de la metacognición es la estimación que los sujetos hacen en su desempeño, de ahí el interés por indagar sobre

estas predicciones. Además, es necesario conocer acerca del origen y las características de los sesgos metacognitivos, los cuales representan diferencias sistemáticas entre el juicio metacognitivo y la realidad de los resultados.

## Sesgos metacognitivos

Como ya se mencionó, cuando se resuelven tareas existe una serie de factores que influyen en los juicios metacognitivos, y que provocan sesgos. Un sesgo es la diferencia entre la esperanza del estimador de un valor con el valor real, lo cual representa un error que no es aleatorio, sino que se repite de manera sistemática y que puede tener distintos orígenes, entre ellos, la naturaleza misma de los sujetos (Castro, 2019).

Según Budiño y Sánchez (2019), los sesgos que tienen origen en los sujetos pueden ocurrir porque estos presentan una tendencia a lanzar conclusiones con información limitada, tomando en cuenta contenidos presentes, pero obviando datos ausentes.

Macbeth *et al.* (2009), para el caso de problemas de lógica modal, mencionan un modelo matemático para un proceso psicológico denominado calibración, propuesto por Oskamp en 1965. El fenómeno de la calibración, también denominado confianza, consiste en una relación entre el éxito objetivo y el éxito subjetivo de una persona en la resolución de tareas. El éxito objetivo está relacionado con el éxito observable, mientras que el subjetivo hace referencia a la creencia o estimación personal sobre el propio desempeño.

La variable de calibración es una diferencia entre el juicio metacognitivo y el desempeño real. Esta variable de calibración representa entonces los sesgos metacognitivos y resultaría positiva en los casos en los que se daría el sesgo de sobreconfianza, esto es, cuando el éxito subjetivo es mayor que el éxito objetivo. La calibración será negativa en el caso en el cual el éxito subjetivo sea menor que el objetivo, de modo que en este caso se presentaría el sesgo de subconfianza. Finalmente, la calibración sería nula, en los casos en que el éxito objetivo y el subjetivo sean iguales, es decir, cuando exista ausencia de sesgos.

Macbeth *et al.* (2009) encuentran una reducción en el monitoreo metacognitivo atribuible a una ilusión de simplicidad que generan las proposiciones lógicas, de modo que la coincidencia de operadores en ambos términos de una proposición modal provoca una disminución en la calibración. Por ejemplo, en la proposición compuesta “p tiene que ser q”, ambas proposiciones carecen de la conectiva negación; sin embargo, en la

proposición “no p tiene que ser q”, únicamente el primer término contiene una negación. Los hallazgos del estudio indican que en el primer caso los participantes tendieron a producir un sesgo en la calibración, cuya dirección apunta a la sobreconfianza, lo cual presumiblemente se debe al citado efecto de simplicidad.

Se ha encontrado que la resolución de problemas matemáticos genera un sesgo de sobreconfianza tanto en estudiantes de tecnologías como de humanidades (Macbeth, 2009). Camerer y Lovallo (1999) no encontraron evidencia de que un grupo de participantes en un juego que simulaba inversiones en el mercado moderara su tendencia a la sobreconfianza después de obtener repetidamente resultados negativos. En cambio, cuando se dan intervenciones tendientes a mejorar las estrategias sobre el monitoreo metacognitivo sí se han observado cambios importantes en los sesgos de sobreconfianza y subconfianza (Macbeth y Razumiejczyk, 2008).

Se conoce también un efecto de las tareas denominado difícil-fácil, en el cual, en una tarea difícil se presenta un sesgo de sobreconfianza y en una tarea fácil el sesgo de subconfianza (Macbeth *et al.*, 2009).

Por su parte, Corrales (2015), al analizar la resolución de problemas matemáticos, encontró que los hombres tendían más a realizar sesgos de sobreconfianza, pero esta sobreconfianza no significó un mejor rendimiento en la resolución de los problemas en comparación con las mujeres.

Alfaro (2014) analiza la relación entre la habilidad matemática, la autoeficacia y la dificultad del ítem en la calibración o precisión metacognitiva para una prueba de matemáticas, mediante un análisis de ecuaciones estructurales. El estudio muestra que la precisión metacognitiva aumenta en proporción directa con la autopercepción y la destreza en matemática. Además, al igual que en el estudio de Macbeth (2009), se encontró que la precisión metacognitiva se relaciona de manera inversa con la dificultad de la tarea, dándose una tendencia a la sobreconfianza en el caso de tareas difíciles. Finalmente, no se encontró una relación directa entre el rendimiento académico y la precisión metacognitiva.

Un factor que parece generar sesgos en los juicios metacognitivos es el estado de ánimo (Efklides y Petkaki, 2005), pues cuando este es positivo se tendería a una visión más favorable de la situación y una sobreestimación de las posibilidades de que se logre un buen resultado. Por su parte, un estado de ánimo negativo implicaría un análisis más cuidadoso y analítico de las posibilidades de obtener un resultado favorable.

En síntesis, los sesgos metacognitivos parecen estar asociados a distintas variables como el grado de dificultad de un ítem, la forma lógica de una proposición, estados de ánimo, género, etc.

### Modelación de juicios metacognitivos

Kepecs y Mainen (2012) proponen que los juicios metacognitivos de confianza corresponden a procesos de bajo nivel que podrían ser atribuidos a animales no humanos. Mientras que varios paradigmas de investigación han sido empleados para estudiar la metacognición animal, estos autores postulan que el uso de modelos computacionales constituye una herramienta para entender los juicios de confianza en diversas especies. Bastaría con encontrar algoritmos lo suficientemente sencillos como para ser implementados en un sistema nervioso dado y comparar el rendimiento de simulaciones con resultados empíricos con animales reales.

Los autores defienden que los enfoques computacionales permiten vincular las variables internas que no pueden observarse, con variables externas que, por el contrario, sí serían observadas de manera cuantitativa. Estos enfoques facilitan la creación de modelos “probados, refutados y mejorados iterativamente” (Kepecs y Mainen, 2012, p. 1322) y que permiten avanzar el debate científico.

### Modelos de la confianza y la toma de decisiones

Kepecs y Mainen (2012) definen la confianza desde un punto de vista teórico como “una estimación por parte del responsable de la toma de decisiones sobre la probabilidad de que una decisión tomada sea correcta” (p. 1328). Ahora bien, entendiendo la confianza en las decisiones como un cálculo de probabilidad, esta se puede estimar utilizando modelos bayesianos.

La confianza en la decisión de un evento  $s$  con base en la información de dos eventos  $A$  y  $B$  está dada en el modelo de Kepecs y Mainen (2012) por la ecuación (1):

$$d(s) = \log \frac{P(s|A)}{P(s|B)} + \log \frac{P(A)}{P(B)} \quad (1)$$

En la ecuación (1), se tiene que la confianza  $d$  en una decisión  $s$  está en función de la probabilidad de  $s$ , si se ha dado antes el evento  $A$  y de la probabilidad de  $s$  si se ha dado antes el evento  $B$ , además de la probabilidad de los eventos  $A$  y  $B$  respectivamente.

Este modelo permite obtener una definición matemática de confianza que facilita la posibilidad de desarrollar simulaciones y programas computacionales sobre este constructo, así como su influencia en la toma de decisiones. Los autores señalan que, aunque la confianza se puede calcular de manera algorítmica, esto no explica totalmente la toma de decisión. Sin embargo, Kepecs y Mainen (2012) indican que en las investigaciones de Insabato *et al.* (2010) y Rolls *et al.* (2010) se han creado redes neuronales para poder dar cuenta de la toma de decisiones basadas en la confianza.

El modelo de Kepecs y Mainen (2012) no se va a utilizar en la presente investigación, pues no es posible determinar los eventos A y B, ya que estos corresponden a estados internos no observables del sujeto. Sin embargo, consideramos que la estrategia seguida en este caso está en línea con el espíritu de este modelo, por cuanto puede ser interpretada desde una perspectiva bayesiana. Mientras que estos autores se centran en un problema de naturaleza discreta donde se debe decidir entre dos opciones, el caso que analizamos aquí es de naturaleza continua, donde los sujetos cuentan con distribuciones a priori que representan sus creencias sobre sus habilidades y revisan sus juicios metacognitivos con base en sus análisis de las tareas a las que se enfrentan, evaluación inicial de la dificultad que igualmente puede ser interpretada como una distribución de probabilidad.

Por otra parte, el modelo de Rasch representa un mecanismo natural que podría ser utilizado para determinar la probabilidad de acertar un ítem en una prueba estandarizada. Este modelo establece la probabilidad de responder correctamente a un ítem en función del nivel de dificultad del ítem y de la habilidad de la persona que responde el ítem en el constructo correspondiente (Montero y Rojas, 2012).

Dicho modelo de Rasch (Prieto y Delgado, 2003) define la probabilidad de acertar una respuesta mediante la ecuación (2) donde  $\theta_i$  representa la habilidad de la persona en el constructo mientras  $\beta_j$  corresponde a la dificultad del ítem.

$$P_{ij} = \frac{e^{\theta_i - \beta_j}}{1 + e^{\theta_i - \beta_j}} \quad (2)$$

El modelo de Rasch representa una manera adecuada de modelar la probabilidad de que una persona responda correctamente un ítem. Podemos pensar que los seres humanos hacemos un proceso similar para evaluar las probabilidades de éxito en la solución de ítems, es decir, nuestros cálculos de dichas probabilidades se podrían fundamentar en procedimientos análogos al modelo de Rasch. Esto equivale a pensar que nuestro modelo

para estimar el éxito en la resolución de una tarea efectivamente se basa en una representación del modo en que resolvemos la tarea.

Así, suponemos aquí que los seres humanos realizan un procesamiento análogo al descrito por el modelo de Rasch para establecer una especie de cálculo subjetivo de la probabilidad de resolverlo correctamente.

El presente estudio pretende simular niveles de dificultad de ítems y niveles de habilidad, y a partir de ellos generar juicios metacognitivos de confianza en la solución de ítems, así como predicciones sobre la cantidad de ítems correctos. Para ello, se aplicó un algoritmo que utiliza la ecuación del modelo de Rash agregando una serie de posibles sesgos.

Se aplicaron dos simulaciones distintas. En la simulación 1 se emplearon valores aleatorios de los parámetros de dificultad de los ítems y habilidad de los sujetos, así como de los valores de sesgos metacognitivos correspondientes a ambos parámetros. Posteriormente, los resultados de una nueva simulación (simulación 2) fueron contrastados con datos reales obtenidos en un estudio previo donde se realizó la aplicación de una prueba de matemáticas (Alfaro, 2014).

## Simulación 1

### Modelo de la simulación

La tarea modelada corresponde a un paradigma de investigación en metacognición en el cual las personas participantes deben resolver una serie de problemas y reportar qué tanta confianza tienen en que su resolución para cada uno de ellos sea correcta. Usualmente, los sujetos deben reportar estos juicios de confianza antes de intentar resolver los problemas (prospectivos) y después de resolverlos (retrospectivos). A diferencia de la tarea real en caso de la simulación, para cada ítem simulado se consideró un único juicio en lugar de los juicios prospectivos y retrospectivos. Otra diferencia fue que en lugar de trabajar una escala ordinal para los juicios de confianza se empleó un valor real entre 0 y 1. Esta decisión se tomó por el hecho de que diferentes investigaciones empleaban escalas distintas.

Según nuestro modelo, en el proceso de evaluación de la confianza los seres humanos usaríamos un sistema análogo al del cálculo probabilístico, de modo que bajo los reportes ordinales (por ejemplo, totalmente seguro, muy seguro, etc.) deberían subyacer dichos valores de probabilidad. Posteriormente se dicotomizaron las respuestas para producir predicciones de éxito de los ítems. Esto permitía comparar predicciones de la simulación (modeladas como un resultado de un proceso de Bernoulli con  $p$  igual a las

probabilidades estimadas), contra los juicios de confianza en estudios empíricos al dividir estos entre aquellos menores o iguales a un valor central (considerados predicciones, en el sentido de que la respuesta estaba incorrecta) y aquellos superiores (considerados predicciones correctas).

Cabe recalcar que esta investigación no se realizó mediante la aplicación de una prueba estandarizada, de modo que el modelo de Rasch no fue usado sobre ninguna serie de datos, sino más bien para definir las probabilidades de resolver correctamente una serie de ítems (simulando valores de dificultad) por un conjunto de sujetos (simulando valores de habilidad).

El modelo de Rasch describe la probabilidad de que un sujeto  $i$  resuelva correctamente un ítem  $j$  en función del nivel de habilidad de la persona en el constructo  $\beta_j$  y la dificultad del ítem  $\delta_j$ . Esta probabilidad se calcula de la siguiente manera:

$$P_{ij} = \frac{e^{\beta_i - \delta_j}}{1 + e^{\beta_i - \delta_j}} \quad (3)$$

Si se asume que los juicios de probabilidad de éxito son estimados de manera similar al modelo descrito anteriormente, se tiene que, dado un sujeto  $i$  y un ítem  $j$ , la probabilidad estimada por este sujeto de acertar la respuesta correcta en el ítem  $j$  sería:

$$\widehat{P}_{ij} = \frac{e^{\widehat{\beta}_i - \widehat{\delta}_{ij}}}{1 + e^{\widehat{\beta}_i - \widehat{\delta}_{ij}}} \quad (4)$$

Aquí se entiende  $\beta_i$  como la habilidad real del sujeto  $i$  en el constructo correspondiente y  $(\beta_i)^\wedge$  como la estimación que realiza el sujeto de su propia habilidad. De igual manera,  $\delta_j$  es la dificultad real del ítem y  $(\delta_{ij})^\wedge$  es la dificultad estimada por el sujeto  $i$  respecto del ítem  $j$ . Nótese que, a diferencia del valor real,  $(\delta_{ij})^\wedge$  depende no solo del ítem sino también del sujeto, ya que se refiere a una evaluación metacognitiva.

Considerando las ecuaciones (3) y (4), se realiza el proceso de simulación que se describe a continuación. Las simulaciones y los respectivos análisis fueron realizados mediante R, versión 4.3.0 y la biblioteca ggplot2.

### Procedimiento de la simulación

Inicialmente se realizó la simulación de 100 000 sujetos y 30 ítems. La simulación de los sujetos consistió en asignar una habilidad real para cada uno, es decir, los valores  $\beta_i$ . Estos valores fueron generados a partir de una distribución normal con media 0 y desviación estándar 1. Luego se crearon 30 ítems. La simulación de cada ítem equivale a la asignación de una dificultad real  $\delta_j$ . Los

valores  $\delta_j$  se generaron igualmente mediante una distribución normal con media 0 y desviación estándar 1.

A partir de los parámetros  $\beta(i)$  y  $\delta_j$  se calculó para cada sujeto y en cada ítem, la probabilidad real de acertar cada ítem  $P_{ij}$  mediante la ecuación (3).

Asumiendo que el evento consistente en que el sujeto  $i$  acertara el ítem  $j$  seguía una distribución de Bernoulli con parámetro  $p = P_{ij}$ , se simularon los resultados de cada ítem para cada sujeto, los cuales podían asumir valores 0 (respuesta incorrecta) o 1 (respuesta correcta).

Finalmente, se calculó para cada sujeto, una nota que representaba el desempeño real en la prueba (es decir el total de respuestas correctas en los 30 ítems).

Posteriormente se realizó la estimación de los aspectos metacognitivos. Se asumió que las estimaciones que un sujeto  $i$  tenía de su propia habilidad dependían tanto del valor real de su habilidad como de un sesgo aditivo, es decir que  $(\beta_i)^\wedge = \beta_i + a_i$ .

De manera similar, se asumió que la dificultad estimada por un sujeto  $i$  a un ítem  $j$  dependía tanto del valor real de esta dificultad, así como de un sesgo  $k_i$ :

$$\widehat{\delta}_{ij} = \delta_j + k_i$$

Los valores de  $a_i$  y  $k_i$  fueron generados a partir de una distribución normal con media cero y desviación estándar 1/3. Dichos valores presumiblemente podrían estar asociados en el mundo real a distintos factores que influirían en que un sujeto tenga algún nivel de subconfianza o sobreconfianza, por ejemplo, el género, el grupo social, el tipo de ítems empleados, etc.

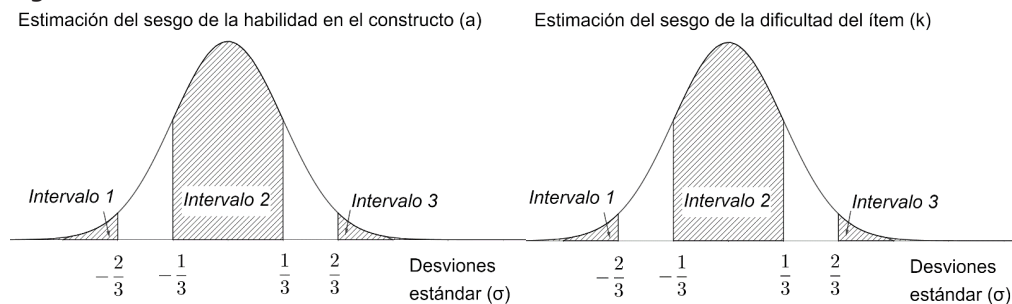
Posteriormente, el juicio metacognitivo o la probabilidad estimada por un sujeto  $i$  respecto de acertar un ítem  $j$  ( $P_{ij}$ ) fue generada aplicando la ecuación (4) a los valores  $(\beta_i)^\wedge$  y  $(\widehat{\delta}_{ij})^\wedge$ .

Finalmente, asumiendo que las predicciones de los sujetos sobre si habían acertado o no un ítem también seguían una distribución de Bernoulli con  $p = (P_{ij})^\wedge$ , se generaron las respectivas predicciones, donde 1 significaba que el sujeto creía haber resuelto correctamente el ítem y 0 no haberlo conseguido. La suma de estas predicciones por sujeto producía el puntaje esperado en la prueba.

## Resultados

A partir de la población total de los 100 000 casos generados, se establecieron 9 subpoblaciones caracterizadas por su magnitud en los diferentes tipos de sesgo. Para ello, se tomaron 3 intervalos para el sesgo de los sujetos al estimar la dificultad del ítem y 3 intervalos para el sesgo del sujeto al estimar su propia habilidad en el constructo. En ambos casos, el primer intervalo correspondía a aquellos sujetos simulados con sesgos en el extremo inferior de la distribución debajo de 2 desviaciones estándar, es decir, aquellos que tenían una alta tendencia a subvalorar el parámetro correspondiente (la habilidad o la dificultad, según el caso). El segundo intervalo correspondía al grupo que tenía una mejor calibración, con valores de sesgos cercanos a 0, entre -1 y 1 desviación estándar. El tercer intervalo se caracteriza por la tendencia a sobrevalorar los parámetros correspondientes, con valores de sesgo en el extremo superior, arriba de 2 desviaciones estándar (figura 1). Nótese que los sesgos fueron generados de una normal (0, 1/3) por lo que 1 desviación estándar corresponde a 1/3.

**Figura 1**



De la combinación de los 3 intervalos de cada uno de los sesgos, aparecen las 9 subpoblaciones que se presentan en la tabla 1. La tabla 2 muestra las estadísticas descriptivas para cada una de las subpoblaciones de la simulación con respecto a la nota real y la nota estimada. La misma información se muestra en la tabla 3 para las subpoblaciones agregadas por el criterio de los sesgos de estimación de dificultad de los ítems según los 3 intervalos mostrados en la figura 1, es decir optimistas (intervalo 1), realistas (intervalo 2) y pesimistas (intervalo 3) respecto a la dificultad de los ítems. Por su parte, la tabla 4 muestra dicha información para las subpoblaciones agregadas por el criterio de sesgos de estimación de la habilidad de los sujetos

según los intervalos de la figura 1, es decir, optimistas (intervalo 3), realistas (intervalo 2) y pesimistas (intervalo 1) respecto a su propia habilidad.

Como se esperaba, mientras que el promedio de las puntuaciones simuladas fue similar en todos los grupos (aproximadamente 50 puntos en escala de 0 a 100), los optimistas con respecto a la dificultad de los ítems, así como los optimistas respecto a la habilidad propia mostraron las estimaciones más altas de dicha puntuación (63,65 y 63,76 respectivamente), mientras los pesimistas en ambos grupos obtuvieron las estimaciones más bajas (39,17 y 38,53 respectivamente). Estas diferencias, en términos de la calibración, corresponden respectivamente a una diferencia de 12,55 (para dificultad) y 12,45 (para habilidad) para los optimistas, y de -12,51 (para dificultad) y -12,29 (para habilidad) para los pesimistas. También es previsible el hecho de que los realistas mostraron diferencias muy pequeñas en el caso de los sesgos de estimación de la dificultad (-0,2) y de la habilidad (-0,19).

En el caso de las 9 subpoblaciones generadas a partir de la combinación de esos intervalos, también se encontró que la subpoblación 5 (realistas respecto a ambos sesgos) obtuvo estimaciones en promedio muy similares a los valores reales (con una diferencia de 0,2). En cambio, el grupo 3 (optimistas en ambos sesgos) mostró una clara tendencia a sobreestimar el rendimiento obteniendo la diferencia positiva más amplia (24,94); mientras que la subpoblación 7 (pesimistas en ambos sesgos), hacia subestimar el rendimiento con la diferencia más alta (-20,36).

El caso de las subpoblaciones 1 y 9 resulta interesante, pues, a pesar de contar con sesgos importantes respecto a la habilidad y la dificultad, muestran una calibración bastante alta, lo cual puede explicarse por el hecho de que las direcciones de ambos sesgos tienden a compensarse, dado que la subpoblación 1 se caracteriza por ser pesimista con respecto a la habilidad, pero optimista respecto a los ítems, mientras la subpoblación 9 corresponde al patrón opuesto.

**Tabla 1**

*Subpoblaciones definidas por los tres intervalos tanto para el sesgo de habilidad como para el sesgo de dificultad*

	<b>Sesgo de la habilidad</b>			
	Menor a -2/3 Interpretación: pesimistas respecto a su habilidad	Entre -1/3 y 1/3 Interpretación: realistas respecto a su habilidad	Mayor a 2/3 Interpretación: optimistas respecto a su habilidad	Cualquier valor
<b>Sesgo dificultad del ítem</b>	Subpoblación 1 (N = 49)	Subpoblación 2 (N = 1566)	Subpoblación 3 (N = 54)	N = 2311

Menor a -2/3 Optimistas				
Entre -1/3 y 1/3. Realistas	Subpoblación 4 (N = 1520)	Subpoblación 5 (N = 46357)	Subpoblación 6 (N = 1609)	N = 68136
Mayor a 2/3. Pesimistas	Subpoblación 7 (N = 47)	Subpoblación 8 (N = 1510)	Subpoblación 9 (N = 43)	N = 2208
Cualquier valor	N = 2253	N = 68109	N = 2310	N = 10000

**Tabla 2**

*Estadísticas descriptivas de la nota y la nota estimada para las nueve subpoblaciones*

	Nota		Nota estimada	
	M	SD	M	SD
Total	51,26	20,88	51,07	24,50
Sub. 1	49,59	19,32	48,50	20,19
Sub. 2	51,19	20,72	63,29	22,77
Sub. 3	47,96	21,69	72,90	21,49
Sub. 4	50,76	20,78	38,46	23,28
Sub. 5	51,29	20,82	51,09	23,94
Sub. 6	51,36	20,82	63,79	22,72
Sub. 7	50,99	20,36	30,63	22,49
Sub. 8	51,64	20,54	38,91	22,74
Sub. 9	53,80	21,93	55,89	25,00

**Tabla 3**

*Estadísticas descriptivas de la nota y la nota estimada, según el sesgo de la dificultad del ítem en cada uno de los tres intervalos*

Subpoblación	Nota		Nota estimada	
	M	SD	M	SD
Optimistas	51,10	20,69	63,65	22,65
Realistas	51,32	20,86	51,12	24,24
Pesimistas	51,68	20,71	39,17	23,25

**Tabla 4**

*Estadísticas descriptivas de la nota y la nota estimada,*

*según el sesgo de la habilidad en el constructo en cada uno de los tres intervalos*

Subpoblación	Nota		Nota estimada	
	M	SD	M	SD
Pesimistas	50,82	20,79	38,53	23,53
Realistas	51,23	20,84	51,04	24,21
Optimistas	51,31	20,89	63,76	23,01

## Simulación 2

### Procedimiento la simulación

Se adelantó una segunda simulación con el fin de comparar los resultados con los datos recolectados por Alfaro (2014). En esta investigación un grupo de participantes resolvieron una serie de 9 problemas de matemática realizando juicios metacognitivos (prospectivos y retrospectivos) para cada uno de ellos. Para esta comparación fue necesario adelantar algunos ajustes a la base de datos que se tenía de la investigación de Alfaro (2014), así como a los parámetros de la simulación. Los cambios se detallan a continuación:

1. Alfaro (2014) tenía para cada ítem una nota del 0 a 100 y estimaciones metacognitivas sobre el rendimiento en el ítem también de 0 a 100. Se dicotomizaron tanto la nota obtenida en el ítem como la estimación de dicha nota; ambas se convirtieron a una escala de 0 o 1, dependiendo de si la nota o su estimación era mayor o menor a 70.
2. Se trabajó con una estimación de la puntuación total de acuerdo con el número de ítems que cada participante consideró con un rendimiento superior a 70, de modo que se obtuvo una puntuación en un rango de 0 a

- 9, la cual fue transformada a una escala de 0 a 100. Igualmente se calculó una puntuación del rendimiento de cada participante del estudio de Alfaro, sumando los ítems con puntuaciones superiores a 70.
3. Se trabajó únicamente con las estimaciones retrospectivas del estudio de Alfaro (2014).
  4. Como la prueba de Alfaro (2014) tenía únicamente 9 ítems, se generó una nueva simulación también de 9 ítems.

Se trabajó con una submuestra (submuestra 1) de 100 casos de los datos de Alfaro (2014) para analizar la calibración promedio. Luego, mediante una nueva submuestra de 200 casos se analizaron en detalle los patrones de los juicios y se compararon con una muestra simulada de 200 sujetos y 9 ítems. Ambos conjuntos de datos fueron analizados mediante un modelo de regresión local (*locally estimated scatterplot smoothing*, LOESS). Se adoptó este modelo, debido a que la naturaleza de los datos hace que presumiblemente la relación entre ellos podría ser no lineal, en particular por el hecho de que las estimaciones tenían un piso y un techo. Esto hace que cualquier otra función lineal llevara a valores dentro y fuera del rango esperado —excepto en el caso donde la recta de regresión tuviera una pendiente de 1 y un intercepto de 0—.

## Resultados

La submuestra 1 de los datos de Alfaro (2014) arrojó una nota promedio de 62,5, mientras que el promedio de las estimaciones metacognitivas de 82,28. La media de la calibración fue de 19,78. Mientras estos datos muestran una alta tendencia al optimismo, la simulación obtuvo un optimismo más moderado, con una media de 40,06 para la nota y de 44 para la estimación metacognitiva; la media de la calibración es de 3,94. En términos relativos, la sobrestimación en la muestra de personas reales fue de un 32 % de la nota, mientras en la muestra simulada fue solo de un 10 %. De hecho, este patrón es más bien cercano al grupo 2 de la simulación 1 (realistas respecto a la habilidad, optimistas respecto a la dificultad de los ítems), donde la calibración resultó de 12,10, lo que representó un 24 % de la nota (véase tabla 2).

Los datos también muestran que el rango en el que se mueven tanto las notas como las estimaciones es muy diferente en los datos simulados en relación con los datos reales (véase tabla 5). Mientras que no es fácil interpretar esta diferencia en el caso presente, donde se comparó una muestra real con otra simulada, podría interpretarse que la prueba correspondiente a

la simulación es de mayor dificultad que la otra, salvo que en realidad no existe tal prueba.

Para obtener una idea de la plausibilidad de la simulación, quitando el efecto de las diferencias en las notas, se decidió trabajar con los datos estandarizados de la nota (variable independiente) y la estimación (variable dependiente) en una regresión local (span = 0,75). La figura 2 muestra que, al quitar el efecto de las diferencias de escala, los patrones son similares y de hecho los límites de confianza al 95% se traslapan en todos los puntos del intervalo entre -2 y 1,5 desviaciones estándar, de modo que la diferencia entre ellas no resulta significativa.

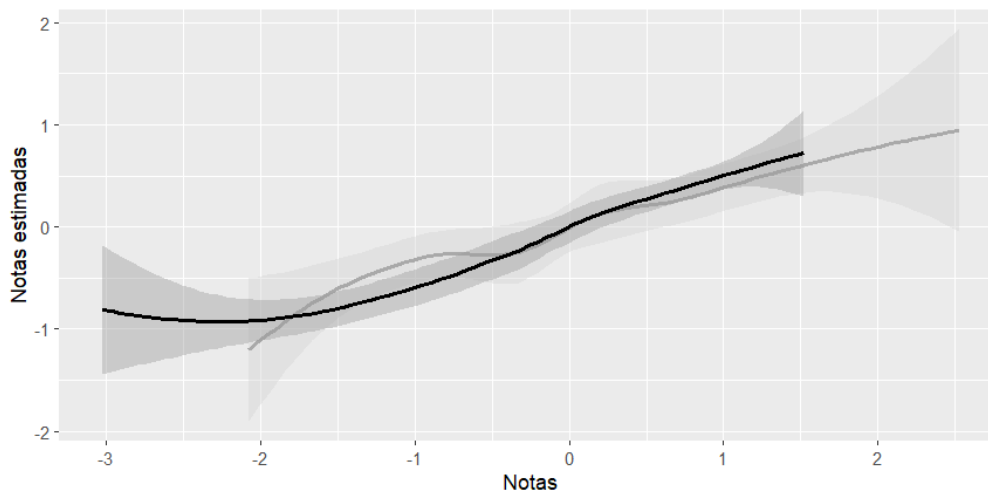
**Tabla 5**

*Estadísticas descriptivas de la nota y la nota estimada para los sujetos de la simulación 2*

	Mínimo	Primer cuartil	Mediana	Media	Tercer cuartil	Máximo
<b>Nota</b>	0,00	22,22	44,44	40,06	55,56	88,89
<b>Nota estimada</b>	0,00	22,22	44,44	44,00	66,67	100,00

**Figura 2**

*Curvas de regresión local para los datos empíricos y simulados*



*Nota: la curva LOESS de los datos simulados (línea gris) muestra un traslape con la curva correspondiente a los datos empíricos (línea negra) en todos los puntos del intervalo entre -2 y 1,5 desviaciones estándar.*

## Discusión

A partir de los resultados de la simulación 1 se reconocieron 9 subpoblaciones correspondientes a diferentes niveles de sesgo en la evaluación metacognitiva, tanto de la dificultad de los ítems como de la habilidad de los sujetos simulados. Los resultados son coherentes con respecto a la dirección de las discrepancias entre los valores de la nota y su estimación metacognitiva. En este sentido, la subpoblación simulada con parámetros correspondientes a sesgos pesimistas, en relación tanto con su habilidad como con la dificultad de los ítems, registró una nota estimada promedio inferior al promedio real de la nota. También, la población cuyos parámetros correspondían a sesgos optimistas en ambos casos mostró una nota estimada promedio superior al promedio de la nota real. Igualmente, la población denominada realista, simulada mediante sesgos cercanos a 0 en cuanto a la dificultad y la habilidad, mostró notas estimadas cercanas a la nota real. Según estos resultados, la simulación mostró buenos niveles de consistencia interna.

Con respecto a la validez externa del modelo, una simulación independiente fue realizada con el fin de comparar los resultados con un conjunto de datos empíricos reportados por Alfaro (2014), a partir de juicios metacognitivos prospectivos y retrospectivos de una prueba de matemática aplicada a estudiantes universitarios. Para valorar esta comparación es importante considerar que se realizaron una serie de transformaciones en los datos empíricos para poder contrastarlos con la simulación, lo que significa que a este procedimiento le subyacen una serie de supuestos que deben hacerse explícitos. Así, se asumió que una puntuación de confianza superior a 70 sobre 100 en el estudio de Alfaro (2014) equivalía a que la persona consideraba que su solución era correcta. Este supuesto se siguió del hecho de que las calificaciones en la universidad tienen como valor mínimo aceptable la puntuación de 70. Después de dicotomizar las respuestas mediante el procedimiento anterior, se sumaron estas respuestas de modo que el resultado fue considerado una estimación global del puntaje en escala de 0 a 9. Luego, este fue convertido a puntajes de 0 a 100. De conformidad con el supuesto anterior, se utilizaron solo los juicios retrospectivos, los cuales previsiblemente serían más cercanos a la estimación que los sujetos realizarían después de finalizar la prueba.

La muestra empírica generada mediante el procedimiento anterior produjo estimaciones más optimistas que la muestra simulada, de modo que la primera se comportó aproximadamente como el grupo realista respecto a la habilidad-optimista respecto a la dificultad de los ítems. Sin embargo, esta comparación podría no ser correcta dado que la simulación no consideró ajustar la dificultad de los ítems a los datos empíricos. En este sentido, la mejor opción era comparar los datos estandarizados. Al comparar las curvas LOESS de las dos muestras se encontró que para todos los puntos entre 2 y 1,5 desviaciones estándar existía un traslape para los límites de confianza estimados. Las diferencias en el rango de los valores reflejan diferencias en cuanto a la asimetría de los datos (positiva para los datos simulados y negativa para los datos reales). Aparte de este aspecto, las curvas tuvieron un comportamiento muy similar, lo que sugiere coincidencias en los mecanismos hipotetizados mediante la simulación y aquellos empleados por los participantes reales del estudio de Alfaro (2014).

A pesar de que los resultados sugieren coincidencias entre el comportamiento de la simulación y los datos empíricos, lo ideal sería haber contado con datos empíricos medidos en la misma métrica de los datos simulados, es decir, con estimaciones metacognitivas globales donde los participantes expresaran su nota. Además, sería importante realizar comparaciones contra datos empíricos de diversa naturaleza, incluyendo diferentes poblaciones y pruebas, para establecer la validez externa del modelo. Aunque resulta necesario realizar nuevas comparaciones con datos reales que consideren este aspecto, los resultados del presente estudio sugieren una coincidencia razonable entre los mecanismos hipotetizados en la simulación y aquellos seguidos por sujetos reales al realizar estimaciones metacognitivas globales en la solución de pruebas. Pero ¿cómo puede interpretarse el alcance de esta posible coincidencia? Bajo una interpretación fuerte, podría pensarse que nuestros resultados constituyen una generalización de los de Kepecs y Mainen (2012), para quienes los juicios de confianza, específicamente la confianza en la toma de decisiones, son el resultado de la implementación de algoritmos de razonamiento probabilístico en el sistema nervioso. En ese sentido, los seres humanos conformarían una suerte de máquina bayesiana.

Mientras que estos autores piensan que los procesos por los cuales los animales no humanos estiman la confianza en la toma de decisiones constituyen mecanismos automáticos de cálculo probabilístico mediante un correlato neuronal, nosotros pensamos que simplemente existe un isomorfismo débil entre el cálculo de probabilidades y la ley de Bayes y el modo en que los seres humanos estiman su rendimiento en un examen. La

diferencia entre ambos enfoques es análoga a la de las dos formas en que Searle (1980) concebía la inteligencia artificial hace más de treinta años: la IA fuerte y la IA débil. Nosotros pensamos que los seres humanos somos capaces de aproximar los cálculos que puede proporcionar la ley de Bayes o el modelo de Rasch, aplicando múltiples heurísticos, imperfectos pero que funcionan de forma razonable y rápida.

Nuestra interpretación es coherente con Gigerenzer (2021), el cual considera el enfoque de la racionalidad limitada para explicar cómo los humanos realizamos algunos procesos metacognitivos. Este principio propone que los humanos realizamos heurísticos cuando nos enfrentamos a situaciones de incertidumbre, los cuales no requieren la aplicación de los algoritmos matemáticos exactos para la solución del problema, sino que representan estrategias de otra naturaleza que les permite aproximar la solución.

Mientras Kepecs y Mainen (2012) piensan que la plausibilidad de este cálculo requiere de demostrar la posibilidad de implementar su modelo en una red neuronal, nuestra posición es que las estimaciones metacognitivas de rendimiento, por ejemplo, aquellas realizadas por sujetos humanos al resolver una prueba de matemática, son procesos de alto nivel que requieren de manera fundamental de la conciencia. A diferencia de los paradigmas empleados en el estudio de la metacognición en animales no humanos, el tipo de estudios aquí considerados, en donde sujetos humanos realizan juicios sobre rendimiento en solución de problemas, requiere de un alto nivel de reflexión consciente, pero también del empleo de marcadores corporales emocionales para evaluar el desempeño.

Lo anterior no significa que tengamos necesariamente que considerar dos conceptos diferentes de metacognición, si pensamos en las tareas aplicadas a animales no humanos y otra al caso de resolución de problemas en pruebas con sujetos humanos. Al respecto, considerando que la conciencia no es privativa de los seres humanos, no debería descartarse la posibilidad de que otros animales emplearan también estrategias metacognitivas complejas en las tareas empleadas por Kepecs y Mainen (2012), y que el resultado de esas estrategias fuera también isomórfico con respecto a la ley de Bayes.

Finalmente, aun si aceptamos que nuestro modelo basado en Rasch es análogo al modo en que los seres humanos realizan estimaciones metacognitivas en el contexto aquí estudiado, eso no excluye la posibilidad de que modificaciones en los supuestos empleados en nuestra simulación nos lleven a mejores aproximaciones de los procedimientos empleados por sujetos reales. Por ejemplo, nuestro modelo asume que la predicción de un ítem no afecta la predicción del siguiente, pero podría probarse también si

eliminando este supuesto de independencia se obtienen simulaciones que se ajusten mejor con los datos empíricos. Igualmente, podrían realizarse estudios posteriores, que modifiquen el supuesto de que los sesgos ocurren de manera aditiva o permitan que los sesgos con respecto a la habilidad y la dificultad de los ítems estén correlacionados.

## Referencias

- Alfaro, A. (2014). *Experiencias metacognitivas y su relación con la autoeficacia, la habilidad matemática, la dificultad de la tarea y el rendimiento académico* [Tesis de maestría, Universidad de Costa Rica].
- Allueva, P. (2002). Conceptos básicos sobre metacognición. En *Desarrollo de habilidades metacognitivas: programa de intervención* (pp. 59-85). Consejería de Educación y Ciencia. Diputación General de Aragón.
- Budiño Sánchez, M. y Sánchez Fuentes, D. (2016). La paradoja del tabaco. *Revista Clínica Española: publicación oficial de la Sociedad Española de Medicina Interna*, 216(6), 308-310.
- Camerer, C. y Lovo, D. (1999). Overconfidence and excess entry: An experimental approach. *The American Economic Review*, 89(1), 306-318.
- Castro, E. M. (2019). Bioestadística aplicada en investigación clínica: conceptos básicos. *Revista Médica Clínica Las Condes*, 30(1), 50-65.
- Consejo Nacional de Rectores (Conare). (2017). *Cuarto Informe Estado de la Educación*. <https://repositorio.conare.ac.cr/handle/20.500.12337/672>
- Corrales, E. (2015). *Ocurrencia del juicio metacognitivo en la resolución de problemas por insight* [Tesis de maestría, Universidad de Costa Rica].
- Curşeu, P. L., Krehel, O., Evers, J. H. y Muntean, A. (2014). Cognitive distance, absorptive capacity and group rationality: A simulation study. *PloS one*, 9(10), e109359.
- Efklides, A. (2006). Metacognition and affect: What can metacognitive experiences tell us about the learning process. *Educational Research Review*, 1(1), 3-14.
- Efklides, A. y Petkaki, C. (2005). Effects of mood on students' metacognitive experiences. *Learning and Instruction*, 15(5), 415-431.
- Efklides, A. y Tsiora, A. (2002). Metacognitive experiences, self-concept, and self-regulation. *Psychologia*, 45(4), 222-236.
- Flavell, J. H. (1976) Metacognitive aspects of problem solving. En L. B. Resnick (ed.), *The nature of intelligence* (pp. 231-235). Lawrence Erlbaum Associates.

- Gigerenzer, G. (2021). Axiomatic rationality and ecological rationality. *Synthese*, 198, 3547-3564. <https://doi.org/10.1007/s11229-019-02296-5>
- Insabato, A., Pannunzi, M., Rolls, E. T. y Deco, G. (2010). Confidence-related decision making. *Journal of Neurophysiology*, 104(1), 539-547.
- Kaber, D. B. y Kim, S. H. (2011). Understanding cognitive strategy with adaptive automation in dual-task performance using computational cognitive models. *Journal of Cognitive Engineering and Decision Making*, 5(3), 309-331.
- Kepecs, A. y Mainen, Z. (2012). A computational framework for the study of confidence in humans and animals. *Philosophical Transactions of the Royal Society*, 367, 1322-1337.
- Macbeth, G. (2009). Distorsiones del éxito subjetivo en la resolución de problemas matemáticos. *Boletín de Psicología*, 95(1), 59-72.
- Macbeth, G. y Razumiejczyk, E. (2008). Disolución del sesgo de subconfianza en tareas verbales. *Anales de Psicología*, 24(1), 143-149.
- Macbeth, G., López Alonso, A. O., Razumiejczyk, E., Sosa, R. A., Pereyra, C. I. y Fernández, H. (2009). Sesgos de la calibración en tareas de razonamiento lógico. *SUMMA Psicológica UST*, 6(2), 19-30.
- Montero, E. y Rojas, S. (2012). *Cuarto Informe Estado de la Educación. Aporte especial: El modelo de Rasch: una herramienta esencial para la evaluación educativa de gran escala*. Consejo Nacional de Rectores. <https://repositorio.conare.ac.cr:8443/rest/bitstreams/40c056af-acac-4f10-8f9a-6071a69b842b/retrieve>
- Oskamp, S. (1965): Overconfidence in case-study judgments. *The Journal of Consulting Psychology*, 29, 261-265.
- Prieto, G. y Delgado, A. R. (2003). Análisis de un test mediante el modelo de Rasch. *Psicothema*, 15(1), 94-100.
- Rolls, E. T., Grabenhorst, F. y Deco, G. (2010). Choice, difficulty, and confidence in the brain. *Neuroimage*, 53(2), 694-706.
- Searle, J. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3(3), 417-424. <https://doi.org/10.1017/S0140525X00005756>
- Summerfield, C. y Yeung, N. (2014). Shared mechanisms for confidence judgements and error detection in human decision making. En S. M. Fleming y C. D. Frith (eds.), *The cognitive neuroscience of metacognition* (pp. 147-167). Springer.
- Thilakarathne, D. J. y Treur, J. (2015). Computational cognitive modelling of action awareness: Prior and retrospective. *Brain Informatics*, 2(2), 77.

- 
- Valenzuela, Á. (2018). La metacognición en los procesos de lectura y escritura académica: ¿qué nos dice la literatura? *Lenguaje*, 46(1), 69-93.
- Vancouver, J. B. y Purl, J. D. (2017). A computational model of self-efficacy's various effects on performance: Moving the debate forward. *Journal of Applied Psychology*, 102(4), 599.
- Wilson, J. (1997) *Beyond the basics: Assessing student's metacognition* [Paper presented at The Hong Kong Educational Research Conference]. Hong Kong.