

UNIVERSIDAD DE COSTA RICA
SISTEMA DE ESTUDIOS DE POSGRADO

ESTUDIO DE LOS JUICIOS METACOGNITIVOS EN LA SOLUCIÓN DE ÍTEMS
MEDIANTE SIMULACIÓN.

Tesis sometida a la consideración de la Comisión del Programa de Posgrado en Ciencias
Cognoscitivas para optar al grado y título de Maestría Académica Ciencias Cognoscitivas

DIANA LUCIA CHACON CAMACHO

Ciudad Universitaria Rodrigo Facio, Costa Rica

2024

DEDICATORIA

A mi hija Aileen que con sus risas y juegos inocentes me dieron la fuerza que me hacía falta para cerrar este ciclo.

A Helen por siempre creer en mí.

A David porque su amor y dedicación lo hacen el compañero perfecto para transitar por esta vida.

A mi mamá y mi suegra por ser mi red de apoyo, sin ustedes esto no sería posible.

A mi papá porque el trabajo de sus manos me sustentó para poder estudiar.

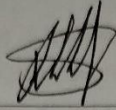
A toda mi familia porque saben dibujar una sonrisa en mi rostro a pesar de las dificultades.

AGRADECIMIENTO

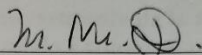
Al profesor Mauricio Molina Delgado por compartir su conocimiento conmigo, por crear para mí una ruta de aprendizaje que da como fruto esta investigación, por la paciencia y el tiempo dedicado. Eternamente agradecida.

A mis estimadas lectoras Alejandra Alfaro y Vanessa Smith-Castro, gracias por aceptar ser parte de mi trabajo, por sus aportes, pero sobre todo por el respeto y la paciencia.

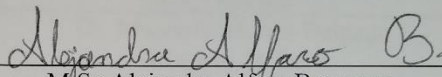
Esta tesis fue aceptada por la Comisión del Programa de Posgrado en Ciencias Cognoscitivas de la Universidad de Costa Rica, como requisito parcial para optar al grado y título de Maestría Académica en Ciencias Cognoscitivas.



Dr. Rodolfo Fallas Soto
**Representante de la Decana
Sistema de Estudios de Posgrado**



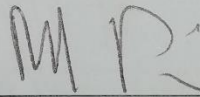
Dr. Mauricio Molina Delgado
Director de Tesis



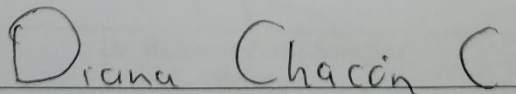
M/Sc Alejandra Alfaro Barquero
Asesora



Dra. Vanessa Smith-Castro
Asesora



Dr. Rolando Pérez Sánchez
Director
Programa de Posgrado en Ciencias Cognoscitivas



Diana Lucía Chacón Camacho
Candidata

TABLA DE CONTENIDO

DEDICATORIA	ii
AGRADECIMIENTO	iii
TABLA DE CONTENIDO	v
RESUMEN	vii
LISTA DE CUADROS.....	ix
LISTA DE FIGURAS.....	x
LISTA DE ABREVIATURAS.....	xi
CAPÍTULO I. INTRODUCCIÓN.....	1
1.1. Antecedentes.....	1
1.2. Marco conceptual.....	2
1.3. Justificación, relevancia y descripción del estudio.....	5
1.4. Problema de Investigación.....	5
CAPÍTULO II. OBJETIVOS.....	7
CAPÍTULO III. MÉTODO.....	8
3.1. Procedimiento de la simulación.....	8
3.1.1 Simulación 1	8
3.1.2 Simulación 2	10
3.2. Procedimiento de análisis	11
CAPÍTULO IV. HALLAZGOS Y CONCLUSIONES	13
4.1. Resultados de la simulación 1.....	13
4.2. Resultados de la simulación 2.....	14
4.3. Discusión y conclusiones.....	15
REFERENCIAS.....	18

ANEXOS	21
Anexo 1. Artículo.	21
Anexo 2. Código del programa generado.	53

RESUMEN

Uno de los procesos metacognitivos que las personas realizan es crear estimaciones sobre su rendimiento en la solución de ítems, estas estimaciones se conocen como juicios metacognitivos y pueden presentar sesgos respecto al desempeño real por diversos factores como la ansiedad, el género entre otros. El modelo de Rasch permite estimar la probabilidad de acertar un ítem en una prueba estandarizada basándose en la habilidad del sujeto resolutor en el constructo y la dificultad del ítem. El presente estudio supone que los humanos realizamos un proceso análogo al descrito por la ecuación de Rasch para establecer los juicios metacognitivos, por lo que propone 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. Dado el modelo computacional basado en la ecuación de Rasch se realizaron dos simulaciones computacionales en la cuales se agregaron errores aditivos a los parámetros de habilidad del sujeto y dificultad del ítem y se determina la estimación metacognitiva del desempeño en los ítems. Finalmente se compararon juicios metacognitivos emitidos por estudiantes en pruebas anteriores con una simulación encontrándose resultados similares a la de los datos reales.

ABSTRACT

One of the metacognitive processes that people perform is creating estimations about their performance in solving items; these estimations are known as metacognitive judgments and may present biases regarding actual performance due to various factors such as anxiety, gender, among others. The Rasch model allows estimating the probability of correctly answering an item in a standardized test based on the subject's ability in the construct and the item's difficulty. The present study assumes that humans undergo a process analogous to the one described by the Rasch equation to establish metacognitive judgments, therefore it proposes a computation model of the relationship between item performance on standardized math tests and metacognitive judgments of performance prediction. Given the computational model based on the Rasch equation, two computational simulations were conducted, adding additive errors to the subject's ability parameters and item difficulty, for determining the metacognitive estimation of item performance. Finally, metacognitive judgments issued by students in previous tests were compared with a simulation, finding similar results to those of real data.

LISTA DE CUADROS

Cuadro 1	39
Cuadro 2	39
Cuadro 3	40
Cuadro 4	41
Cuadro 5	44

LISTA DE FIGURAS

Figura 1	36
Figura 2	44

LISTA DE ABREVIATURAS

Abreviatura	Significado
JOL	Juicios que las personas emiten sobre su aprendizaje.
LOESS	Modelo de regresión local (locally estimated scatterplot smoothing).
MK	Conocimiento metacognitivo.
ME	Experiencias metacognitivas.
MS	Habilidades metacognitivas.
PISA	Programa para la Evaluación Internacional de Estudiantes, por sus siglas en inglés se denomina PISA.
POP	Estimaciones que realizan las personas sobre el desempeño que tendrán en el futuro al realizar una tarea.

CAPÍTULO I. INTRODUCCIÓN

1.1. Antecedentes

Realizar estimaciones sobre la posibilidad de resolver con éxito un ítem es un proceso metacognitivo y estas estimaciones se denominan juicios metacognitivos (Alfaro, 2014; Macbeth, et.al., 2009; Valenzuela, 2018), sin embargo, los juicios emitidos suelen presentar sesgos (errores sistemáticos) con respecto al desempeño real de las personas.

Kepecs y Mainen (2012) proponen que las estimaciones de confianza en la toma de decisiones pueden ser atribuidos a animales no humanos, y sugieren el uso de modelos computacionales para comprender estos procesos. Estos autores definen la confianza como la estimación de la probabilidad de que una decisión tomada sea correcta. Proponen que la confianza puede ser modelada utilizando técnicas bayesianas.

El principal aporte de la investigación de Kepecs y Mainen (2012) a este estudio es la posibilidad de abordar el fenómeno de toma de decisiones mediante el cálculo de probabilidades, ver más detalle en el anexo 1 (secciones 1.3 y 1.4).

En este mismo sentido, el modelo de Rasch ofrece una fórmula para determinar la probabilidad de acertar un ítem en una prueba estandarizada, basado en la relación entre la habilidad del individuo en el constructo y la dificultad del ítem (ver anexo 1, sección 1.4). Este modelo es fundamental para el presente estudio pues, suponiendo que los seres humanos realizamos un proceso similar para evaluar las probabilidades de éxito en la resolución de

ítems, basándose en procedimientos análogos al modelo de Rasch, se propone el modelo implementado computacionalmente para estudiar los juicios metacognitivos del desempeño.

Alfaro (2014) realiza un estudio empírico en el cual recolecta juicios metacognitivos de estimación del desempeño en tareas de matemática por parte de estudiantes del Instituto Tecnológico De Costa Rica. Ella 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 (ver anexo 1, sección 1.2).

El estudio de Alfaro (2014) brinda a esta investigación datos empíricos sobre estimaciones metacognitivas de predicción de éxito en la solución de ítems de matemática. Estos datos son de utilidad para compararlos con las estimaciones metacognitivas generadas por el modelo propuesto.

1.2. Marco conceptual

Flavell (1976) plantea que la metacognición se compone del conocimiento de los procesos y productos cognitivos de uno mismo y el proceso de examinar, regular y organizar ese conocimiento de manera consciente.

Durante la etapa del monitoreo del procesamiento cognitivo, se emiten juicios o estimaciones metacognitivas que pueden referir a la calidad del aprendizaje, de la información, o del desempeño en una tarea (Efklides, 2006; Efklides y Tsiora, 2002). Allueva (2002) destaca que estos juicios están relacionados con la capacidad de predecir si se tienen los

conocimientos necesarios para la solución de un problema. Además, estos juicios influyen en el procesamiento cognitivo de una tarea y su respuesta (Efklides y Petkaki, 2005),

Existen diversos factores que influyen en los juicios metacognitivos provocando sesgos, por ejemplo: la forma lógica del enunciado, el área de estudio de los estudiantes, su estado de ánimo, incluso el género. En el anexo 1, sección 1.2 se describe la influencia de esos factores sobre los juicios metacognitivos.

En las investigaciones sobre solución de tareas se reporta un efecto denominado efecto difícil-fácil, en el cual, en una tarea difícil se presenta un sesgo de sobre-confianza y en una tarea fácil el sesgo de sub-confianza (Macbeth, et.al., 2009; Alfaro, 2014).

En una investigación sobre inversiones, Camerer y Lovallo (1999) no encontraron evidencia de que se disminuya la sobre-confianza después de obtener repetidamente resultados negativos. Macbeth y Razumiejczyk (2008) encontraron una disminución en los sesgos cuando se dan intervenciones para mejorar las estrategias sobre el monitoreo metacognitivo.

Macbeth, et al. (2009) consideran un proceso psicológico denominado calibración. La variable calibración es la diferencia entre el éxito subjetivo y el éxito objetivo. Esta variable será positiva cuando hay sesgo de sobre-confianza y será negativa en el caso de que exista sub-confianza. Finalmente, la calibración será nula en ausencia de sesgos.

Respecto a los juicios metacognitivos, Kepecs y Mainen (2012) proponen que el uso de modelos computacionales constituye una herramienta para entender los juicios de confianza en diversas especies, bajo el supuesto de que estos juicios corresponden a un proceso de bajo nivel que podrían ser atribuidos a animales no humanos.

La confianza desde un punto de vista teórica se puede entender como “una estimación por parte del responsable de la toma de decisiones sobre la probabilidad de que una decisión tomada sea correcta” (Kepecs y Mainen, 2012, p.1328). Kepecs y Mainen (2012) generan una definición matemática de confianza en la toma de decisiones (ver ecuación 1, anexo 1, sección 1.4), la cual facilita la posibilidad de desarrollar simulaciones y programas computacionales sobre este constructo, así como su influencia en la toma de decisiones.

De manera similar al modelo de Kepecs y Mainen (2012) el modelo de Rasch puede ser utilizado para estudiar la confianza de acertar un ítem mediante el cálculo de probabilidad. En el modelo de Rasch esta probabilidad está en función al nivel de dificultad del ítem y de la habilidad de la persona que responde el ítem en el constructo correspondiente (Montero y Rojas, 2012).

La ecuación (1) representa el modelo de Rasch, descrito por Prieto y Delgado (2003). De esta ecuación se tiene como variable dependiente la probabilidad de acertar una respuesta y como variables independientes la habilidad de la persona en el constructo (θ_i) y la dificultad del ítem (β_j).

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

El supuesto de esta investigación es que los humanos realizamos un procesamiento análogo al descrito por el modelo de Rasch para establecer una estimación de nuestra probabilidad subjetiva de resolverlo correctamente.

1.3. Justificación, relevancia y descripción del estudio

Las pruebas estandarizadas son un referente de la calidad de la educación (Programa del Estado de la Educación, 2017); por lo tanto, es necesario comprender los procesos cognitivos involucrados en la solución de estas pruebas, entre ellos, los juicios metacognitivos.

Se ha elegido utilizar simulaciones computacionales pues estas brindan una metodología especialmente útil ya que permiten realizar una gran cantidad de iteraciones del proceso y es posible contrastar los resultados del modelo con los datos empíricos. Estos modelos computacionales han sido utilizados para estudiar fenómenos cognitivos (Curşeu, Krehel, Evers, y Muntean, 2014; Kaber y Kim, 2011; Thilakarathne y Treur, 2015; Vancouver y Purl, 2017, sin embargo, se ha utilizado menos en el campo de la metacognición (Kepecs y Mainen, 2012).

Esta investigación busca contribuir al entendimiento de los procesos cognitivos involucrados en la evaluación del rendimiento, así como entender como los sesgos metacognitivos pueden afectar dicha evaluación. Para esto se propone un modelo computacional para estudiar la generación de juicio metacognitivos de predicción del rendimiento, y posteriormente se realizan modificaciones a los parámetros del modelo para estudiar las variaciones en las predicciones metacognitivas generadas mediante la simulación, finalmente se realiza una comparación entre los resultados de las simulaciones con una serie de datos empíricos.

1.4. Problema de Investigación

Comprender los procesos cognitivos involucrados en la resolución de los ítems de una prueba es crucial dado el importante papel que desempeñan las pruebas estandarizadas en los

sistemas educativos y los modelos de evaluación. Los juicios metacognitivos, particularmente aquellos relacionados con la predicción del desempeño, son vitales en este contexto, ya que influyen en cómo los individuos resuelven la prueba y modifican a largo plazo el concepto que tienen de sí mismos (Efklides y Petkaki, 2005).

Para avanzar en conocimiento de cómo las personas realizamos nuestras estimaciones sobre el rendimiento en el desempeño en pruebas estandarizadas es necesario abordar las siguientes dos preguntas: ¿cómo se pueden modelar la generación de juicios metacognitivos y los sesgos asociados en un contexto de solución de ítems en pruebas estandarizadas?, ¿en qué medida coinciden las predicciones generadas por ese modelo con los datos reales de desempeño en pruebas estandarizadas de matemáticas?

Para responder a estas preguntas y tener una mejor comprensión de los juicios metacognitivo esta investigación plantea los objetivos que se presentan en el siguiente capítulo.

CAPÍTULO II. OBJETIVOS

A continuación, se describen el objetivo general y los objetivos específicos de esta investigación.

Objetivo general:

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.

Objetivos específicos:

1. Generar un programa que simule juicios metacognitivos de predicción de éxito a partir de la ecuación del modelo de Rasch.
2. Realizar modificaciones en los parámetros del modelo para analizar cómo afectan los resultados de la predicción.
3. Comparar los resultados de las simulaciones con un conjunto de datos reales.

CAPÍTULO III. MÉTODO

Se creó un programa computacional que utiliza la ecuación del modelo de Rasch para calcular la probabilidad de éxito en la solución de ítems en función de la dificultad de ítems y niveles de habilidad. Además, el programa considera la asignación aleatoria de sesgos tanto para la estimación de la habilidad como la dificultad y así obtener una estimación del juicio metacognitivo.

Se aplicaron dos simulaciones distintas. La Simulación 1 se realizó empleando 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).

3.1. Procedimiento de la simulación

3.1.1 Simulación 1

En esta investigación, en la tarea simulada, para cada ítem simulado se consideró un único juicio, además se empleó un valor real entre 0 y 1 para cuantificar la probabilidad de éxito y los juicios de confianza.

Posteriormente se dicotomizaron las respuestas para producir predicciones de éxito de los ítems. Esto permitía comparar predicciones de la simulación contra los juicios de confianza en estudios empíricos.

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

Con el modelo de Rasch se determina la probabilidad de que un sujeto i resuelva correctamente un ítem j en función (P_{ij}) (ver ecuación 3, anexo 1, sección 2.1). Asumimos en esta investigación que los juicios de probabilidad de éxito son estimados de manera similar al modelo descrito anteriormente, tenemos 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} (ver ecuación 4, anexo 1, sección 2.1).

En esta simulación se crearon 100000 sujetos y 30 ítems. Para la creación de los sujetos se establecieron los valores β_i que representan la habilidad real del sujeto en el constructo, para la creación de los ítems se establecieron los valores δ_j que representan la dificultad real del ítem. Tanto β_i como δ_j son valores aleatorios de una distribución normal con media 0 y desviación estándar 1.

Posteriormente se calculó la probabilidad real de acertar cada ítem (P_{ij}) para cada sujeto i y cada ítem j aplicando el modelo de Rasch a los valores β_i y δ_j mediante la ecuación 3 (ver anexo 1, sección 2.1). Luego se dicotomizaron estas probabilidades utilizando una distribución de Bernoulli (ver anexo 1, sección 2.1) para obtener dos posibles valores, el 0 (respuesta incorrecta) o 1 (respuesta correcta). Finalmente, para medir el desempeño real en la prueba se calculó para una nota cada sujeto.

Para la estimación de los aspectos metacognitivos se agregaron sesgos aditivos tanto a la habilidad del sujeto como a la dificultad del ítem. La estimación de la habilidad del sujeto se determinó en función del valor real de su habilidad y del sesgo ($\widehat{\beta}_i = \beta_i + a_i$). De igual forma, la dificultad estimada para cada ítem fue definida por su dificultad real y un sesgo k_i ($\widehat{\delta}_{ij} = \delta_j + k_i$), aclarando que es el mismo sesgo k_i para los n ítems, fijado el sujeto 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.

Posteriormente, el juicio metacognitivo o la probabilidad estimada por un sujeto i respecto de acertar un ítem j (\widehat{P}_{ij}) fue generada aplicando la ecuación 4 (ver anexo 1, sección 2.1) a los valores $\widehat{\beta}_i$ y $\widehat{\delta}_{ij}$.

Finalmente, al igual que como se realizó para P_{ij} , para \widehat{P}_{ij} se dicotomizaron las respuestas y se generó una nota en la prueba.

Todas las simulaciones y los respectivos análisis fueron realizados mediante R versión 4.3.0 y la biblioteca *ggplot2*.

3.1.2 Simulación 2

En la investigación de Alfaro (2014) los participantes resolvieron 9 problemas de matemática realizando juicios metacognitivos (prospectivos y retrospectivos) para cada uno de ellos. La simulación 2 fue realizada para comparar los datos simulados con los recolectados por Alfaro (2014).

Para realizar esta comparación fue necesario realizar algunos ajustes a la base de datos que se tenía de la investigación de Alfaro (2014) y como a los parámetros del modelo. (ver anexo 1, sección 3.1).

La muestra simulada de la simulación dos consistió en ejecutar el programa computacional para 200 sujetos y 9 ítems

3.2. Procedimiento de análisis

Para la simulación 1, se establecieron tres intervalos para el sesgo al estimar la dificultad del ítem y otros tres para el sesgo al estimar su propia habilidad. El intervalo 1 representaba sujetos simulados con sesgos en el extremo inferior (2 desviaciones estándar debajo del promedio), el intervalo 2 con sesgo cercanos a cero (1 desviación estándar alrededor del promedio), y en el intervalo 3 con sesgo en el extremo superior (2 desviaciones estándar por encima del promedio) (ver Figura 1, anexo 1, sección 2.3). A partir de una población total de 100,000 casos generados, la delimitación de tres intervalos por cada sesgo permitió identificar 9 subpoblaciones según la magnitud de estos sesgos (ver Cuadro 1, anexo 1, sección 2.3).

En cada una de las subpoblaciones se analizó la calibración entre los promedios de la nota y la nota estimada con el fin de observar el comportamiento de las estimaciones metacognitivas según el tipo y tamaño de los sesgos.

Para la simulación 2, de los datos de Alfaro (2014) una submuestra de 100 casos se utilizó para analizar la calibración promedio, otra submuestra de 200 casos se utilizó para comparar con los datos de la simulación 2 mediante el modelo de regresión local o LOESS (locally estimated scatterplot smoothing).

Todas las simulaciones y los respectivos análisis fueron realizados mediante R versión 4.3.0 y la biblioteca *ggplot2*.

CAPÍTULO IV. HALLAZGOS Y CONCLUSIONES

4.1. Resultados de la simulación 1

Los cuadros 2, 3, y 4 (anexo 1, sección 2.3) muestra las estadísticas descriptivas de la simulación con respecto a la nota real y la nota estimada para: la población total y cada una de las 9 subpoblaciones, para las subpoblaciones agregadas por el criterio de los sesgos de estimación de dificultad de los ítems y para las subpoblaciones agregadas por el criterio de sesgos de estimación de la habilidad respectivamente.

Se destaca que, en general, las puntuaciones simuladas tienen un promedio similar en todos los grupos, alrededor de 50 puntos en una escala de 0 a 100. Sin embargo, se observan diferencias significativas en las estimaciones según el sesgo. Los optimistas respecto a la dificultad de los ítems o la habilidad muestran las estimaciones más altas, mientras que los pesimistas en alguno de estos sesgos muestran las estimaciones más bajas. Estas diferencias se traducen en los valores extremos en la calibración, con una diferencia de alrededor de 12 puntos para los optimistas y -12 para los pesimistas.

Respecto a las otras subpoblaciones, se destaca que el grupo de personas realistas respecto a ambos sesgos (subpoblación 5) presenta estimaciones cercanas a los valores reales, con una diferencia mínima. Por otro lado, el grupo de optimistas en ambos sesgos (subpoblación 3) tiende a sobreestimar el rendimiento, mientras que el grupo de pesimistas en ambos sesgos tiende a subestimar (subpoblación 7).

Para las subpoblaciones 1 y 9, las cuales son optimistas en uno de los sesgos, pero pesimistas en el otro tienen una calibración cercana al cero. Una posible explicación es que las direcciones de ambos sesgos tienden a compensarse en estas subpoblaciones específicas.

4.2. Resultados de la simulación 2

El análisis de los datos de la submuestra de 100 sujetos de Alfaro (2014) evidencia una fuerte tendencia al optimismo con una sobrestimación del 32% respecto a la nota real, estos resultados son similares a la subpoblación 2 de la simulación.

Respecto a la comparación de la submuestra de 200 sujetos de Alfaro (2014) con los sujetos simulados en la simulación 2 se encuentra una tendencia moderada al optimismo con una sobrestimación cercana al 10% (ver cuadro 5 en el anexo 1, sección 3.2).

Se puede observar una notoria diferencia entre los datos simulados y reales (ver Cuadro 5, anexo 1, sección 3.2). Una posible explicación a esta diferencia puede ser que la prueba simulada podría ser más difícil, aclarando que no existe una prueba en realidad para los datos simulados.

Para poder obtener una idea de la calidad de la simulación, se utilizó una regresión local con datos estandarizados de nota y su estimación (tanto para datos reales como simulados). Los patrones resultantes fueron similares al eliminar las diferencias de escala, y los límites de confianza al 95% se superponen, indicando que las diferencias no son significativas entre -2 a 1.5 desviaciones estándar (ver Figura 2, anexo 1, sección 3.2).

4.3. Discusión y conclusiones

Para la simulación 1 se determinaron 9 subpoblaciones. Las subpoblaciones con sesgos pesimistas tuvieron estimaciones de nota promedio más bajas que las notas reales, mientras que aquellas con sesgos optimistas mostraron estimaciones más altas. La población considerada realista presentó estimaciones cercanas a las notas reales. Este comportamiento en los resultados permite observar que existe una coherencia interna en el modelo.

Los datos generados en la simulación 2 fueron comparados con los datos reales de juicios metacognitivos de una prueba de matemáticas provenientes de la investigación de Alfaro (2014), una vez que a esos últimos se les realizó una serie de transformaciones.

Los datos reales mostraron más optimismo en comparación con los datos de la simulación 2, con resultados similares a la subpoblación realista en términos de habilidad, pero optimistas en cuanto a la dificultad de los ítems. Sin embargo, es cuestionable la validez de esta comparación debido a que la simulación no ajustó la dificultad de los ítems a los datos empíricos. La mejor opción fue comparar los datos estandarizados, revelando un traslape en los límites de confianza para todos los puntos entre 2 y 1.5 desviaciones estándar. A pesar de las diferencias en la asimetría de los datos, las curvas mostraron un comportamiento muy similar, sugiriendo similitudes en los mecanismos hipotetizados por la simulación y los utilizados por los participantes reales en el estudio de Alfaro (2014).

A pesar de que los resultados sugieren similitudes entre la simulación y los datos empíricos, es necesario contar con datos empíricos medidos en la misma métrica que la simulación. Además, es necesario realizar comparaciones con datos empíricos variados, abarcando diferentes poblaciones y pruebas, para establecer la validez externa del modelo, sin embargo,

los datos obtenidos sugieren una coincidencia razonable entre los mecanismos hipotetizados en la simulación y los seguidos por individuos reales al realizar estimaciones metacognitivas.

Respecto al alcance de esta coincidencia, podría interpretarse que los resultados representen una generalización de los hallazgos de Kepecs y Mainen (2012). Estos autores postulan que los juicios de confianza se deben a la implementación de algoritmos de razonamiento probabilístico en el sistema nervioso, lo que podría llevar a considerar a los seres humanos como una suerte de máquina bayesiana.

Sin embargo, en esta investigación pensamos que esta coincidencia puede ser explicada como la existencia de un isomorfismo débil entre el cálculo de probabilidades y la ley de Bayes, en relación con la forma en que los seres humanos evalúan su rendimiento en un examen. Esta diferencia de enfoque se asemeja a la distinción planteada por Searle (1980) entre la inteligencia artificial fuerte y débil.

En ese sentido, sostenemos que los seres humanos tienen la capacidad de aproximar los cálculos proporcionados por la ley de Bayes o el modelo de Rasch mediante la aplicación de múltiples heurísticas, imperfectas pero eficaces de manera razonable y rápida. Esta interpretación se alinea con la perspectiva de Gigerenzer (2021) sobre el enfoque de la racionalidad limitada, para el cual, en situaciones de incertidumbre, los humanos empleamos heurísticas que no requieren la aplicación precisa de algoritmos matemáticos, sino que representan estrategias de otra naturaleza que permiten una aproximación eficaz a la solución del problema.

Nosotros argumentamos que las estimaciones realizadas por sujetos humanos en pruebas de resolución de problemas son procesos de alto nivel que requieren conciencia y reflexión consciente, además del uso de marcadores emocionales para evaluar el desempeño.

Existen mejoras que se pueden realizar al modelo de esta simulación, por ejemplo, probar la influencia de la independencia entre la predicción de un ítem y la siguiente, así como explorar la correlación de sesgos entre habilidad y dificultad de los ítems en futuros estudios.

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, San José, Costa Rica.
- Allueva P. (2002). Conceptos básicos sobre metacognición. En P. Allueva, *Desarrollo de habilidades metacognitivas: programa de intervención* (pp.59-85). Zaragoza: Consejería de Educación y Ciencia. Diputación General de Aragón.
- Camerer, C., & Lovallo, D. (1999). Overconfidence and Excess Entry: An Experimental Approach. *The American Economic Review*, 89(1), 306-318.
- Curşeu, P. L., Krehel, O., Evers, J. H., & 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., & Petkaki, C. (2005). Effects of mood on students' metacognitive experiences. *Learning and Instruction*, 15(5), 415-431.
- Efklides, A., & 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: Resnick, L.B., Ed., *The Nature of Intelligence*, Lawrence Erlbaum Associates, Hillsdale, 231-235.

- Gigerenzer, G. (2021). Axiomatic rationality and ecological rationality. *Synthese*, 198, 3547–3564. <https://doi.org/10.1007/s11229-019-02296-5>
- Kaber, D. B., & 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., & 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., & Razumiejczyk, E. (2008). Disolución del sesgo de subconfianza en tareas verbales. *Anales de Psicología/Annals of Psychology*, 24(1), 143-149.
- Macbeth, G., López Alonso, A. O., Razumiejczyk, E., Sosa, R. A., Pereyra, C. I., & 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. *El modelo de Rasch: una herramienta esencial para la evaluación educativa de gran escala*.
- Prieto, G. y Delgado A.R. (2003). Análisis de un test mediante el modelo de Rasch. *Psicothema*, 15(1), 94-100.

- Searle, J. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3(3), 417-424. doi:10.1017/S0140525X00005756
- Thilakarathne, D. J., & 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., & 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. November, Hong Kong.

ANEXOS

Anexo 1. Artículo.

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

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 la 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.

Palabra clave: Juicios metacognitivos, ecuación de Rash, modelación computacional.

1. 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 juegan las pruebas estandarizadas en los sistemas de educación y los modelos de evaluación. Las pruebas estandarizadas como PISA, bachillerato internacional y exámenes de admisión a las universidades son consideradas un referente de la calidad de la educación (Programa del Estado de la Educación, 2017); por lo tanto, es necesario comprender los procesos cognitivos involucrados en la solución de estas pruebas.

En la solución de ítems, uno de los procesos metacognitivos que las personas realizan son 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, esto 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, Krehel, Evers, y Muntean, 2014; Kaber y Kim, 2011; Thilakarathne y Treur, 2015; Vancouver y Purl, 2017). Estos modelos son especialmente útiles porque permiten contrastar los resultados del modelo con los datos empíricos, permiten realizar una gran cantidad de iteraciones de los procesos, y según Summerfield y Yeung, (2014) permite 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, basándonos en 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) en el sentido de que 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.

1.1. 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 proceso de examinar, regular y organizar 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:

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

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

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 (Efklides, 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 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 la misma. Además, los autores afirman que las experiencias metacognitivas pueden tener lugar antes, durante o después de finalizado el procesamiento de una tarea.

Allueva (2002) considera también la habilidad de la predicción dentro de los procesos metacognitivos. Esta habilidad consiste en prever si se es capaz de resolver un problema, lo cual incluye poder 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 los juicios sobre su aprendizaje (JOL) y los segundos son las predicciones sobre su rendimiento (POP). 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, parte importante de la metacognición es la estimación que los sujetos hacen en su desempeño, por tanto, es importante 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.

1.2.Sesgos Metacognitivos

Como se evidenció anteriormente, cuando se resuelven tareas existen una serie de factores que influyen en los juicios metacognitivos provocando sesgos. Un sesgo es la diferencia entre la esperanza del estimador de un valor con el valor real representando 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 hacer 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 sobre-confianza, 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 sub-confianza. 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 sobre-confianza, lo cual presumiblemente se debe al citado efecto de simplicidad.

Se ha encontrado que la resolución de problemas matemáticos genera un sesgo de sobre-confianza 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 sobre-confianza y sub-confianza (Macbeth y Razumiejczyk, 2008).

Se conoce también un efecto de las tareas denominado efecto difícil-fácil, en el cual, en una tarea difícil se presenta un sesgo de sobre-confianza y en una tarea fácil el sesgo de sub-confianza (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 sobre-confianza, 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 sobre-confianza 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 el estado de ánimos 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.

1.3. 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 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 se pueden observar, con variables externas que por el contrario sí pueden ser observadas de manera cuantitativa. Estos enfoques permiten crear modelos que pueden ser “probados, refutados y mejorados iterativamente” (p.1322) y que permiten avanzar el debate científico.

1.4. 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

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

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 en cuanto puede ser interpretada desde una perspectiva bayesiana. Mientras que estos autores se centran en un problema de naturaleza discreto donde se debe decidir entre 2 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 mediante el cual podrían 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).

El modelo de Rasch (Prieto y Delgado, 2003) define la probabilidad de acertar una respuesta mediante la ecuación (2) donde θ_i representa la habilidad de la persona en el constructo mientras β_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 humanos realizamos un procesamiento análogo al descrito por el modelo de Rasch para establecer una especie de cálculo subjetivo de nuestra 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 Rasch agregando una serie de posibles sesgos.

Se aplicaron dos simulaciones distintas. La Simulación 1 se realizó empleando 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).

2. Simulación 1

2.1. 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 tiene 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 resolver los problemas (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.) debería 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 como predicciones en el sentido de que la respuesta estaba incorrecta) y aquellos superiores al mismo (considerados como 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 aplicado a ninguna serie de datos, si no que se utilizó 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 β_i y la dificultad del ítem δ_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 asumimos que los juicios de probabilidad de éxito son estimados de manera similar al modelo descrito anteriormente, tenemos 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:

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

Entendemos aquí β_i como la habilidad real del sujeto i en el constructo correspondiente y $\widehat{\beta}_i$ como la estimación que realiza el sujeto de su propia habilidad. De igual manera δ_j es la dificultad real del ítem y $\widehat{\delta}_{ij}$ es la dificultad estimada por el sujeto i respecto del ítem j . Nótese que, a diferencia del valor real, $\widehat{\delta}_{ij}$ 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*.

2.2. Procedimiento de la simulación

Inicialmente se realizó la simulación de 100000 sujetos y 30 ítems. La simulación de los sujetos consistió en asignar una habilidad real para cada uno, es decir, los valores β_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 δ_j . Los valores δ_j se generaron igualmente mediante una distribución normal con media 0 y desviación estándar 1.

A partir de los parámetros β_i y δ_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 $\widehat{\beta}_i = \beta_i + a_i$.

Similarmente, 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 $\widehat{\beta}_i$ y $\widehat{\delta}_{ij}$.

Finalmente, asumiendo que las predicciones de los sujetos sobre si habían acertado o no un ítem también seguían una distribución de Bernoulli con $p = \widehat{P}_{ij}$, 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.

2.3.Resultados

A partir de la población total de los 100000 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. Finalmente, 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 (ver Figura 1). Nótese que los sesgos fueron generados de una normal (0, 1/3) por lo que 1 desviación estándar corresponde a 1/3.

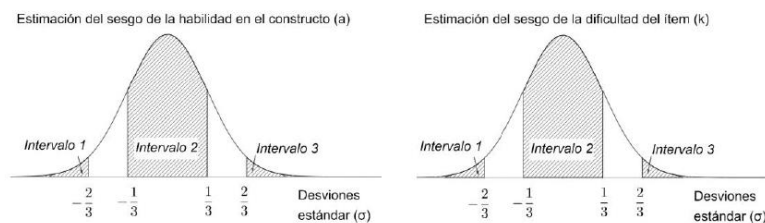


Figura 1

De la combinación de los 3 intervalos de cada uno de los sesgos aparecen las 9 subpoblaciones que se presentan en el Cuadro 1. El Cuadro 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 el Cuadro 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 (intervalo3) respecto a la dificultad de los ítems. Por su parte, el Cuadro 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 1b, 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 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.

Cuadro 1

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

	Sesgo de la habilidad			
	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
Sesgo dificultad del ítem				
Menor a -2/3 Optimistas	Subpoblación 1 (N=49)	Subpoblación 2 (N=1566)	Subpoblación 3 (N=54)	N=2311
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

Cuadro 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

Cuadro 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

Cuadro 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

3. Simulación 2

3.1. Procedimiento la simulación

Una segunda simulación fue realizada 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 realizar esta comparación fue necesario realizar 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, convirtiéndose ambas 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 9 ítems.

Se trabajó con una submuestra (submuestra 1) de 100 casos de los datos de Alfaro (2014) para analizar la calibración promedio. Posteriormente, 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 o LOESS (locally estimated scatterplot smoothing). Este modelo fue asumido dado que la naturaleza de los datos hace que presumiblemente la relación entre los mismos podría ser no lineal, en particular por el hecho de que las estimaciones tenían un piso y un techo. Esto hace que, excepto en el caso donde la recta de regresión tuviera una pendiente de 1 y un intercepto de 0, cualquier otra función lineal llevara a valores de y fuera del rango esperado.

3.2. 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, siendo la media de la calibración 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 solamente 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, representando un 24% de la nota (ver Cuadro 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 (ver Cuadro 5). Mientras que no es fácil interpretar esta diferencia en el caso presente donde comparamos una muestra real con otra simulada, podríamos interpretar que la prueba correspondiente a la simulación es de mayor dificultad que la otra, haciendo la salvedad de que en realidad no existe tal prueba. Para poder 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 ($\text{span}=.75$). 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 las mismas no resulta significativa.

Cuadro 5

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

	Mínimo	1 ^{er} cuartil	Mediana	Media	3 ^{er} cuartil	Máximo
Nota	0.00	22.22	44.44	40.06	55.56	88.89
Nota estimada	0.00	22.22	44.44	44.00	66.67	100.00

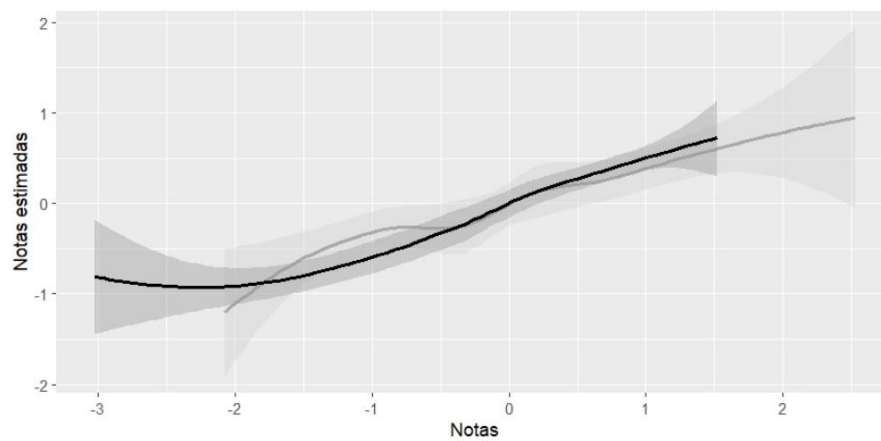


Figura 2. La curva loess de los datos simulados (línea gris) muestra un traslape con la curva correspondiente a los datos empíricos (línea negra) en todos los puntos del intervalo entre -2 y 1.5 desviaciones estándar.

4. Discusión

A partir de los resultados obtenidos 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. De este modo, la subpoblación simulada con parámetros correspondientes a sesgos pesimistas en relación tanto con su habilidad como con la dificultad de los ítems presentó una nota estimada promedio inferior al promedio real de la nota. De igual manera, 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. Estos resultados muestran que 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. En primer lugar, 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 la suma fue

considerada como una estimación global del puntaje en escala de 0 a 9 que fue finalmente convertida 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.

Mientras que la muestra empírica tratada mediante el procedimiento anterior mostró resultados más optimistas que la muestra simulada, sugiriendo que la muestra empírica se comportaba aproximadamente como el grupo de realistas respecto a la habilidad- optimistas respecto a la dificultad de los ítems, 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) pero 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).

Aunque los resultados sugieren coincidencias el comportamiento de la simulación y los datos empírico, 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 tests, para establecer la validez externa del modelo. Aunque resulta necesario realizar nuevas comparaciones con datos reales que consideren

este aspecto, consideramos que 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, podríamos pensar que nuestros resultados resultan ser una generalización de los de Kepecs y Mainen (2012). Estos autores afirman que 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 podríamos pensar que los seres humanos conforman 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 al de las dos formas en que Searle (1980) concebía la inteligencia artificial hace más de 30 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, estos heurísticos 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 consiente, 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, aún 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, estudios posteriores podrían realizarse modificando el supuesto de que los sesgos ocurren de manera aditiva o permitiendo que los sesgos con respecto a la habilidad y la dificultad de los ítems estuvieran 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, San José, Costa Rica.
- Allueva P. (2002). Conceptos básicos sobre metacognición. En P. Allueva, *Desarrollo de habilidades metacognitivas: programa de intervención* (pp.59-85). Zaragoza: Consejería de Educación y Ciencia. Diputación General de Aragón.
- Budiño Sánchez, M., & 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., & Lovallo, 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.

- Corrales, E. (2015). *Ocurrencia del juicio metacognitivo en la resolución de problemas por Insight* (tesis de maestría). Universidad de Costa Rica, San José, Costa Rica.
- Curşeu, P. L., Krehel, O., Evers, J. H., & 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., & Petkaki, C. (2005). Effects of mood on students' metacognitive experiences. *Learning and Instruction*, *15*(5), 415-431.
- Efklides, A., & Tsiora, A. (2002). Metacognitive experiences, self-concept, and self-regulation. *Psychologia*, *45*(4), 222-236.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American psychologist*, *34*(10), 906.
- Flavell, J.H. (1976) Metacognitive Aspects of Problem Solving. En: Resnick, L.B., Ed., *The Nature of Intelligence*, Lawrence Erlbaum Associates, Hillsdale, 231-235.
- Gigerenzer, G. (2021). Axiomatic rationality and ecological rationality. *Synthese*, *198*, 3547–3564. <https://doi.org/10.1007/s11229-019-02296-5>
- Gómez, V. M. Á. (2005). *Inferencia estadística*. Revisado de <https://ebookcentral-proquest-com.ezproxy.sibdi.ucr.ac.cr>

- Insabato, A., Pannunzi, M., Rolls, E. T., & Deco, G. (2010). Confidence-related decision making. *Journal of neurophysiology*, *104*(1), 539-547.
- Kaber, D. B., & 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., & 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., & Razumiejczyk, E. (2008). Disolución del sesgo de subconfianza en tareas verbales. *Anales de Psicología/Annals of Psychology*, *24*(1), 143-149.
- Macbeth, G., López Alonso, A. O., Razumiejczyk, E., Sosa, R. A., Pereyra, C. I., & 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. *El modelo de Rasch: una herramienta esencial para la evaluación educativa de gran escala*.
- Prieto, G. y Delgado A.R. (2003). Análisis de un test mediante el modelo de Rasch. *Psicothema*, *15*(1), 94-100.

- Programa Estado de la Nación. (2017). *Cuarto Informe Estado de la Educación*. San José, Costa Rica: Programa Estado de la Nación. Recuperado de <https://www.estadonacion.or.cr/educacion2017/assets/ee6-informe-completo.pdf>
- Rolls, E. T., Grabenhorst, F., & 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. doi:10.1017/S0140525X00005756
- Summerfield, C., y Yeung, N (2014). Shared Mechanisms for Confidence Judgements and Error Detection in Human Decision Making. En Fleming, S. M., & Frith, C. D. (Eds.), *The cognitive neuroscience of metacognition*. (pp. 147-167). London, UK: Springer.
- Thilakarathne, D. J., & 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., & 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. November, Hong Kong.

Anexo 2. Código del programa generado.

```

---
title: "Simulacion Modelo Rasch"
author: "Diana Chacón & Mauricio Molina"
date: "21/3/2022"
output: html_document
---

```{r paquetes}
#install.packages("purrr")
#install.packages("ggplot2")
#install.packages('readr')
```

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```

```{r librerias}
library("purrr")
library("ggplot2")
library('readr')
```

```{r semilla}
set.seed(1)
```

## Simulación de ítemes mediante valores dificultad

Esta función simula un conjunto de m ítemes. Para ello se generan m valores de dificultad al azar de una  $N(0,1)$ . Estas dificultades se asumen como las habilidades reales dentro de esta simulación
```{r ítemes}
ítemes<-function(m)
{
 ítemes.D<- rep(0,m)

 for (j in 1:m)
 {
 D<-rnorm(1,0,1)
 ítemes.D[j]<- D
 }
 ítemes.D
}
```

## Simulación de sujetos mediante valores habilidad

Esta función simula un conjunto de n sujetos. Para ello se generan n valores de habilidad al azar de una  $N(0,1)$ . Estas habilidades se asumen como las habilidades reales dentro de esta simulación

```{r habilidades}
sujetos<-function(n)
{

```

1/11/23, 9:37

Posit Cloud

```

 sujetos.H<-rep(0,n)

 for (i in 1:n)
 {

 H<-rnorm(1,0,1)
 sujetos.H[i]<-H
 }
 sujetos.H
}

...

##Entrada de datos
```{r num_sujetos_iteles}
as.numeric(readline("Indique el número de sujetos "))->n
as.numeric(readline("Indique el número de ítemes "))->m
#n<-100000
#m<-30
```

Matriz de probabilidades de éxito de un sujeto frente a un ítem

La siguiente sección genera una matriz con las probabilidades de que un sujeto j (filas)
responda correctamente
un ítem i (columnas)

```{r generacion_probabilidades}

sujetos(n)->suj
iteles(m)->it
p.exito<-matrix(0,n,m)
for (i in 1:n)
{
  for (j in 1:m)
  {
    p.exito[i,j]= exp(suj[i]-it[j])/(1+exp(suj[i]-it[j])) # Probabilidad de exito real
  }
}

p.exito
```

#Generación de la matriz de respuestas

```{r generación_respuestas}

exito<-matrix(0,n,m)

for (i in 1:n)
{
  for (j in 1:m)
  {
    exito[i,j]= rbernoulli(1,p.exito[i,j])
  }
}

exito
```

#Calificación de cada estudiante
```{r nota, echo=FALSE}
apply(exito,1,sum)->notas
100*notas/m->notas
table(notas)
```

Simulación de ítemes con sesgo en la dificultad del ítem

```

<https://posit.cloud/content/6943540>

2/4

1/11/23, 9:37

Posit Cloud

Esta sección toma el conjunto de  $m$  ítems y genera una estimación metacognitiva de la dificultad. Estas estimaciones se centran en la dificultad real más un cierto sesgo que es generado al azar.

```

```{r itemes.Est}
nombres<-"Itmen 1"
for (i in 2:m) {
  nombres<-c(nombres, paste("Itmen ", as.character(i)))
}
print(nombres)

itemes.est.met<-matrix(0,n,m)
sesgo.D<-rep(0,n)

for (i in 1:n)
{
  sesgo.D[i]<-rnorm(1,0,1/3) #se genera un sesgo al azar de una normal con d.e.=.3333 lo que
  garantiza que la mayoría están entre -1 y 1 d.e. El sesgo es por sujeto, aunque se refiere a la
  dificultad de los ítems

  for (j in 1:m)
    itemes.est.met[i,j]<- rnorm(1,+it[j]+sesgo.D[i],1)
}

colnames(itemes.est.met)<-nombres
itemes.est.met

```

```

...

```

```

## Simulación de sujetos con sesgo en la habilidad del sujeto

```

Esta sección toma el conjunto de m sujetos y genera una estimación metacognitiva de la habilidad del sujeto. Estas estimaciones se centran en la habilidad real más un cierto sesgo de la habilidad generado al azar.

```

```{r habilidades.Est}

sesgo.H<-rep(0,n)
sujeto.est.met<-suj

for (j in 1:n)
{
 sesgo.H[j]<-rnorm(1,0,1/3)
 #se genera un sesgo al azar de una normal con d.e.=.3333 lo que garantiza que la mayoría
 están entre -1 y 1 d.e.
 sujeto.est.met[j]<- rnorm(1,+suj[j]+sesgo.H[j],1)
}
sujeto.est.met

```

```

...

```

```

Matriz de la estimación metacognitiva de la probabilidad de éxito de un sujeto frente a un
ítem

```

La siguiente sección genera una matriz con las probabilidades de que un sujeto  $j$  (filas) estime responder correctamente un ítem  $i$  (columnas)

```

```{r generacion_probabilidades_estimadas}

p.exito.est.met<-matrix(0,n,m)
for (i in 1:n)
{
  for (j in 1:m)
  {

```

```

1/11/23, 9:37 Posit Cloud
  p.exito.est.met[i,j]= exp(sujeto.est.met[i]-itemes.est.met[i,j])/(1+exp(sujeto.est.met[i]-
itemes.est.met[i,j]))
  # Probabilidad de exito estimada
}
}

p.exito.est.met
...

#Generación de la matriz de respuestas con sesgos
```{r generación_respuestas_estimadas}

exito.est.met<-matrix(0,n,m)

for (i in 1:n)
{
 for (j in 1:m)
 {
 exito.est.met[i,j]= rbernoulli(1,p.exito.est.met[i,j])
 }
}

exito.est.met
...

#Estimación de la calificación de cada estudiante
```{r nota.est, echo=FALSE}
apply(exito.est.met,1,sum)->notas.est.met
100*notas.est.met/m->notas.est.met
table(notas.est.met)
...

```{r exportar_datos}
id<-c(1:n)
df<-data.frame(id,notas, notas.est.met,suj,sujeto.est.met,sesgo.D,sesgo.H)
exito_df<-data.frame(id, exito)
exito_est_met_df<-data.frame(id,exito.est.met)
write_csv(df,)
write_csv(exito_df,)
write_csv(exito_est_met_df,)
...

```