



Laboratorio
Interdisciplinario de
Estadística Social



Discussion Paper N° 2021 | 01

Rasch y Skinner
¿Psicometría sin Psicología?
Ernesto San Martín

Facultad de Matemáticas
Pontificia Universidad Católica de Chile
Av. Vicuña Mackenna 4860, Macul
lies.mat.uc.cl/

Rasch y Skinner
¿Psicometría sin Psicología?

Ernesto San Martín¹

Laboratorio Interdisciplinario de Estadística Social LIES
Facultad de Matemáticas, Pontificia Universidad Católica de Chile

1. Introducción

La Teoría de Respuesta al Item (TRI) es ampliamente usada no solo en la investigación social y educacional, sino también en los protocolos de análisis de pruebas estandarizadas. En Latinoamérica tenemos dos casos importantes, Chile y Colombia, que utilizan modelos TRI para reportar los resultados de las pruebas SIMCE y SABER 11, respectivamente; para detalles, ver ACE (2015) e ICSES (2018). Tal es su relevancia, que se han establecido una serie de estándares para reportar los resultados que se obtienen con este tipo de modelos; ver por ejemplo, AERA (2014). Sin embargo, tanto sus usos concretos, como los estándares no permiten responder la siguiente pregunta: *¿Cuál es el estatuto teórico de los modelos TRI?* Con la expresión *estatuto teórico* hacemos referencia a la significación de los modelos TRI, en contraste con los aspectos relacionados con la estimación estadística.

1.1. Prueba SIMCE de Matemáticas 2011

A fin de aclarar el tenor teórico de esta pregunta, vamos a introducir un ejemplo concreto. En el año 2011, se aplicó en Chile la prueba SIMCE de Matemáticas a estudiantes que cursaban 8° año de la enseñanza básica. La prueba consistió de 39 preguntas con 4 alternativas, una de las cuales era la correcta. Una de las formas de dicha prueba fue aplicada a 57.725 estudiantes. Comencemos

¹ Email: esanmart@mat.uc.cl

haciendo un análisis descriptivo y exploratorio de estos datos. Para ello, definimos $Y_{ij} = 1$ si el estudiante i responde correctamente el ítem j ; e $Y_{ij} = 0$ en cualquier otro caso. Esta variable aleatoria permite construir una matriz de 0s y 1s, de dimensión $N = 57.725$ filas (que corresponden a los estudiantes) y $J = 39$ columnas (que corresponden a los ítems). Si definimos Y_{i+} como el total de respuestas correctas obtenidas por el estudiante i , y por Y_{+j} como la proporción de estudiantes que responden correctamente el ítem j , entonces podemos describir tanto la variabilidad inducida por Y_{i+} por medio de un histograma (Figura 1), como aquella inducida por Y_{+j} por medio de una representación gráfica como la de la Figura 2.

INSERTAR AQUÍ FIGURAS 1 y 2

De la Figura 1 se puede apreciar que hay una mayor proporción de estudiantes con un total de respuestas correctas “bajo” que “altos”. De la Figura 2, se puede apreciar que hay varios ítems tales que entre un 40% y un 70% de los estudiantes los respondieron correctamente. También se aprecia que hay varios ítems con un porcentaje bajo de estudiantes que lo respondieron correctamente. Es importante hacer notar que la variable aleatoria Y_{i+} clasifica a los estudiantes: todos los estudiantes tales que $Y_{i+} = t$, con $t = 0,1,2, \dots, 39$, son indistinguibles; de hecho, las “barras” del histograma corresponden a los conjuntos $I_t = \{i: Y_{i+} = t\}$ con $t = 0,1,2, \dots, 39$.

El análisis descriptivo anterior puede complementarse combinando la información provista por los estudiantes con la información provista por los ítems de la siguiente manera: para cada ítem j , se cuenta el total de estudiantes que, teniendo un total de respuestas correctas igual a t , respondieron correctamente el ítem j . Denotemos esta cantidad por n_{tj} . Entonces podemos definir una función que a cada $t \in \{0,1, \dots, 39\}$ le asocia n_{tj}/n_t . Esta función suele llamarse *Curva Característica Empírica del Ítem*. Las Figuras 3, 4 y 5 muestran esta función para un ítem que el 84% de los estudiantes lo respondió correctamente, que el 28% lo respondió correctamente, y que el 51% lo

respondió correctamente. Se puede apreciar que cuanto mayor es el total de respuestas correctas, mayor es la proporción de estudiantes que responde correctamente el ítem en cuestión. Por otro lado, estas figuras sugieren que menor es la proporción de estudiantes que responde correctamente el ítem bajo análisis, más a la derecha está el punto de inflexión de la curva característica empírica de dicho ítem.

1.2. ¿Cómo se interpreta este análisis descriptivo/exploratorio?

Hasta aquí la descripción y exploración de los datos. El lector debe respetar cuidadosamente el significado de las variables aleatorias Y_{i+} e Y_{+j} , y de la Curva Característica Empírica del Ítem: en estos objetos estadísticos no hay *ni puntaje total, ni proxy de habilidad matemática, ni dificultad o facilidad de un ítem, ni una función de distribución de un ítem con respecto a un proxy de una habilidad*. La interpretación sustantiva, que es de orden psicológico-educacional, *no está en las observaciones* (como, digámoslo al pasar, se suele decir en esta “era de la Ciencia de datos”): esta solo puede posiblemente encontrarse en un modelo teórico que eventualmente puede interpretarse en sentido psicológico-educacional. Por tanto, cuando preguntamos por el estatuto teórico de los modelos TRI, estamos inquiriendo en primer lugar por la adecuación teórica de usar un determinado modelo psicométrico para darle un posible sentido sustantivo a las observaciones. Enfatizamos que esto *nada tiene que ver con la adecuación estadística de un modelo psicométrico a los datos; esto tiene que ver con una opción de modelamiento estructural* (Koopmans & Reirersol, 1950; San Martín, 2016; 2018), es decir, con una posible teoría sustantiva formalizada en términos probabilísticos. En este artículo queremos discutir el estatuto teórico de uno de los modelos TRI, el modelo Rasch. Vamos a comenzar discutiendo la motivación que Rasch tuvo para desarrollar este tipo de modelos, para lo cual recurriremos a Skinner. Luego vamos a mostrar que en dicho modelo *no hay psicología*. En la discusión final mencionaremos qué *tipo* de modelo Rasch podría tener sentido psicológico,

pero para ello será necesario resolver una pregunta previa, que creemos podemos enunciar correctamente, de modo que “si una pregunta puede siquiera formularse, también *puede responderse*” (Wittgenstein, 2010) .

2. La ciencia de lo individual

Como es sabido, una de las fuentes de los modelos TRI la encontramos en George Rasch, un matemático danés que, por falta de oportunidades académicas como matemático, se dedicó a la psicometría; para detalles, ver Olsen (2003). En dos de sus trabajos, Rasch (1961) y Rasch (1980), Rasch parte mencionando que los “métodos estadísticos han sido severamente criticados por algunos sectores de la psicología” (Rasch, 1961, p. 321), haciendo referencia a Skinner (1956). Resulta relevante describir la crítica de Skinner a fin de entrever la recepción de Rasch y su subsecuente desarrollo del modelo Rasch.

Skinner (1954) discute acerca de la posibilidad de formar a un estudiante universitario en un científico. Para ello, Skinner se pregunta si hay suficiente experiencia para poder describir el *comportamiento* de lo que llamamos un *hombre de ciencia*. Responde negativamente, pero a su vez afirma que es posible entrever algo de dicho comportamiento describiendo lo que un científico hace e hizo: toma como caso de estudio su propia experiencia. Esta opción la justifica distinguiendo entre la *práctica científica misma*, que es el comportamiento manifestado, y la *formalización de prácticas científicas* usando la estadística y el método científico. Para Skinner, enseñar estadística y método científico no asegura “dar forma al comportamiento científico”. Es importante recordar que Skinner, como él mismo lo menciona, escribe en un momento en que el Diseño de Experimentos tal y como fue desarrollado por R. A. Fisher es la moda metodológica del momento. De hecho, Skinner se dedica a describir una serie de experimentos

con ratas y otros animales que realizó cuando desarrollaba su tesis de doctorado, y en otros momentos posteriores.

Revisando su propio comportamiento como científico, Skinner propone tres principios “no formalmente reconocidos por los metodólogos científicos”: (1) “Cuando te encuentras con algo interesante, deja todo lo demás y estúdialo”; (2) “Algunas formas de hacer investigación son más fáciles que otras”; (3) “Algunas personas tienen suerte”; (4) “Los aparatos a veces se rompen”. Cada uno de estos principios son motivados por experiencias concretas de Skinner en su trabajo de laboratorio, lo que corresponde en algún grado a las reacciones de un organismo (el científico) frente a determinadas experiencias, en que ciertas condiciones cambian (lo que le ocurría en los laboratorios).

Skinner enfatiza que dichas experiencias de laboratorio eran con un número pequeño de ratas, incluso una rata. Y constata que desde cierta perspectiva eso podía ser considerado como una práctica científica mediocre dado que la estadística sugería “usar más datos”. Describe entonces su colaboración con Heron, quien estaba al día con los avances del diseño experimental desarrollado por Fisher. Llevaron a cabo un experimento con 95 ratas, y la conclusión fue que esto no facilitó la detección de determinados comportamientos, al contrario, a dificultó. La razón se debe a que los experimentos desarrollados por Skinner consistía en variar determinadas características del medio ambiente, de forma de observar reacciones comportamentales ante tal cambio. En esto Skinner usa el aforismo de Pavlov, el cual recuerda en este trabajo de 1954: “Controla tus condiciones y verás orden”. Controlar condiciones con 95 ratas fue algo imposible, por lo que Skinner abogó por otro camino: controlar las diferencias individuales *previo a medir reacciones comportamentales*; ese control lo puede hacer el experimentador porque sabe lo que está haciendo.

La conclusión de todo esto la expresa Skinner de forma clara: “Estamos al alcance de una ciencia del individuo. Esto se logrará, no recurriendo a alguna teoría especial del conocimiento en la que la intuición o la comprensión sustituya a la observación y el análisis, sino mediante una comprensión cada vez mayor de las condiciones pertinentes para producir un orden en el caso individual” (Skinner 1954, p. 231). De ahí entonces que “tanto el tratamiento estadístico de las medias de grupo como el promedio de las curvas alientan la creencia de que de alguna manera estamos yendo detrás del caso individual a un otro proceso inaccesible, pero más fundamental”. Pero no es el caso: en los grupos, no se encuentra el orden de comportamiento. No deja de ser interpelante, aún hoy en día, las palabras finales de Skinner: “No tenemos más razones para decir que todos los psicólogos deben comportarse como yo me he comportado, que como R. A. Fisher. El científico, como cualquier organismo, es el producto de una historia única. Las prácticas que encuentre más apropiadas dependerán en parte de esta historia. Afortunadamente, las idiosincrasias personales suelen dejar una marca insignificante en la ciencia como propiedad pública. Sólo son importantes cuando se trata de alentar a los científicos y proseguir la investigación. Cuando tengamos por fin una adecuada cuenta empírica del comportamiento del Hombre Pensante, comprenderemos todo esto. Hasta entonces, puede ser mejor no intentar encajar a todos los científicos en un solo molde” (p.233).

Rasch considera su contribución (lo que hoy llamamos modelo Rasch) como una forma de responder a este desafío de la ciencia individual; esta contribución la resume en los siguientes términos: “El modelo especifica una función de distribución de las potenciales respuestas de una persona determinada a un determinado estímulo de un cierto conjunto de estímulos, y *esta función de distribución depende de un parámetro que caracteriza a la persona y de un parámetro que caracteriza al estímulo*. [El modelo tiene] una notable propiedad en común que hace

posible, en el análisis de los datos, separar los parámetros personales de los parámetros del estímulo, y viceversa” (Rasch, 1961, p. 321; los subrayados son nuestros).

3. Modelar significa focalizarse en parámetros

¿Qué significar modelar estadísticamente? Siguiendo a Fisher (1922) y Koopmand & Reiersol (1950), especificar un modelo estadístico significa explicitar la distribución de probabilidades que genera los datos bajo estudio. Esta distribución describe la variabilidad que se observa en los datos. Ciertamente esto significa aceptar que todos los fenómenos exhiben variabilidad y que, por tanto, son susceptibles de ser representados por una distribución de probabilidad. Pero hay algo más en el proceso de modelización: la distribución de probabilidad está caracterizada por parámetros, donde estos parámetros describen determinadas propiedades de las observaciones. Puede ocurrir (y lo veremos en el caso del modelo Rasch) que dichos parámetros no coincidan con los parámetros de interés, es decir, con propiedades de las observaciones que interesen al investigador. Como se puede apreciar, y es necesario en esta “época de la Ciencia de Datos”, enfatizar que dichos parámetros *no están en los datos*, sino en las ideas que el investigador tiene acerca de cómo se generaron dichos datos; para detalles, ver San Martín (2016, 2018).

Estas ideas son las que precisamente expresa Rasch en la cita antes mencionada; de hecho, es importante recordar que el modelamiento estructural se desarrolló en lo que se llamó la Cowless Commission, que albergó un grupo de econometras como Koopmans, Haavelmo, Hurwickz, entre otros, y psicómetras como Reiersol y Rasch.

Entonces para responder a la pregunta inicial, a saber, cuál es el estatuto teórico del modelo Rasch, es necesario exhibir el *significado que tiene los parámetros de interés del modelo*, que corresponden a un parámetro que caracteriza la persona y otro que caracteriza el estímulo.

4. El significado de los parámetros del modelo Rasch

Siguiendo (y mejorando) a San Martín, González & Tuerlinckx (2009, 2015), examinemos cómo Rasch especifica su modelo: en primer lugar, asume que las respuestas Y_{ij} que cada persona da a cada estímulo son mutuamente independientes. Hagamos notar una primera confusión en la presentación estándar de los modelos TRI, en particular del modelo Rasch (y que subyace a los protocolos de análisis de pruebas estandarizadas, tanto en Colombia como en Chile, y también en la Prueba PISA): se habla de algo que se llama *Axioma de Independencia Local*, que técnicamente corresponde a una independencia *condicional*; ver, por ejemplo, Embretson & Reise, 2000; Fan & Sun, 2013). Sin embargo, en la presentación de Rasch, no existe tal axioma: solo existe una independencia *marginal* de las respuestas de las personas a los estímulos. La pregunta relevante es qué aspecto sustantivo y a qué tipo de modelos conduce dicho axioma. Para más detalles, ver Del Pino, San Martín, González & De Boeck (2008).

Dada la independencia mutua, basta especificar la distribución de probabilidad de cada Y_{ij} . Rasch quiere especificar un modelo probabilístico de forma que le permita representar lo que significa *medir*, que se reduce a *comparar*. De ahí entonces que propone la siguiente especificación:

$$P(Y_{ij} = 1) = \frac{\lambda_{ij}}{1 + \lambda_{ij}} \quad (1)$$

En efecto, el sentido de la comparación es precisamente el significado del parámetro λ_{ij} pues

$$\frac{P(Y_{ij}=1)}{P(Y_{ij}=0)} = \lambda_{ij} \quad (2)$$

Así, $\lambda_{ij} > 1$ significa que la persona i tiene mayor probabilidad de reaccionar positivamente (responder correctamente) al estímulo j , que de reaccionar negativamente (responder

incorrectamente). De manera similar, si $\lambda_{ij} < 1$, la persona i tiene mayor probabilidad de reaccionar negativamente (responder incorrectamente) al estímulo j , que de reaccionar positivamente (responder correctamente). Si $\lambda_{ij} = 1$, la persona i tiene la misma probabilidad de reaccionar positivamente (responder correctamente) al estímulo j , que de reaccionar negativamente (responder incorrectamente). Notemos que las afirmaciones que se realizan con respecto a la especificación (1) son *relativas*.

Por otro lado, sabemos que el objetivo de Rasch es satisfacer el desafío de *la ciencia de lo individual* propuesto por Skinner. De ahí entonces que Rasch introduce la siguiente especificación:

$$\lambda_{ij} = \frac{\varepsilon_i}{\eta_j} \quad (3)$$

El parámetro ε_i es lo que Rasch llama “parámetro que caracteriza a la persona i ”, mientras que η_j es el “parámetro que caracteriza el estímulo j ”. Una presentación más habitual de este modelo es la siguiente:

$$P(Y_{ij} = 1) = \frac{\exp(\theta_i - \beta_j)}{1 + \exp(\theta_i - \beta_j)} \quad (4)$$

La especificación (1) y (3) es equivalente a (4) pues $\varepsilon_i = \exp(\theta_i)$ y $\eta_j = \exp(\beta_j)$. Como es sabido, en esta presentación se dice que θ_i (o ε_i) corresponde a la *habilidad de la persona i* , mientras que β_j (o η_j) corresponde a la *dificultad del estímulo o item j* . Sin embargo, estas denominaciones son solo eso, denominaciones pues no hay nada aún que permita atribuir a los datos generados por (4) una propiedad llamada “habilidad” o “dificultad”. Dicho más directamente, dichas designaciones son meras etiquetas *vacías de todo significado y que crean la ilusión de asumir que el modelo Rasch tiene algo que ver con alguna idea psicológica de habilidad o dificultad*. De hecho, estas consideraciones son del todo correctas pues la

especificación (3) incluye una ambigüedad intrínseca, esto es, un *problema de identificación*. En efecto, sabemos lo que significa λ_{ij} , pero los parámetros que nos interesan son ε_i y η_j : a un mismo valor de λ_{ij} corresponde una infinidad de distintos pares de valores ε_i y η_j . Por ejemplo, si $\lambda_{ij} = 2$ (es decir, la persona i tiene el doble de probabilidad de responder correctamente el ítem j que de responderlo incorrectamente), entonces

$$2 = \frac{2}{1} = \frac{0.5}{0.25} = \frac{200}{100} = \dots$$

Es decir, las personas que tienen el doble de probabilidad de responder correctamente el ítem j que de responderlo incorrectamente pueden tener una característica individual igual a 2 y una característica del ítem igual a 1, o igual a 0.5 y 1 0.25, o a 200 y a 100, etc. Es decir, a un único valor de λ_{ij} no corresponden únicos valores de ε_i y η_j , lo que hace que la interpretación de dichos parámetros sean completamente arbitraria. Esto es lo que se conoce como problema de identificación.

¿Cómo se resuelve el problema? Una posibilidad, conocida en la literatura, es fijar el parámetro que caracteriza el ítem de etiqueta 1, por ejemplo $\eta_1 = 1$. Usando esta restricción de identificación, entonces se sigue que

$$\varepsilon_i = \lambda_{i1} = \frac{P(Y_{i1} = 1)}{P(Y_{i1} = 0)}$$

Esto permite interpretar el parámetro ε_i en términos de la distribución de probabilidad que genera los datos: ε_i corresponde al cociente de la probabilidad de responder correctamente el ítem 1 y de responderlo incorrectamente. Enfatizamos que ε_i tiene sentido con respecto al ítem 1. Que a esto se le llame “habilidad de la persona i ” es algo completamente arbitrario y no tiene asidero en el modelo Rasch. La característica de la persona i simplemente se reduce a lo que en estadística se llama *betting odd* con respecto al ítem 1. Nada más ni nada menos.

Este parámetro permite *comparar persona en relación a esta característica*. Así, se puede probar que

$$\varepsilon_i > \varepsilon_l \Leftrightarrow P(Y_{i1} = 1) > P(Y_{l1} = 1) \quad (5)$$

Esto es, la persona i tiene una característica individual mayor que la persona l si la persona i tiene una mayor probabilidad de responder correctamente el ítem 1 que la persona l . Nuevamente, aquí no hay nada de “una persona más hábil que otra”; lo que sí hay es que dos personas se pueden comparar entre sí con respecto a una referencia, a saber, el ítem 1: esto es lo que significa medir. Y dicha comparación es en términos probabilísticos.

Usando un argumento en la línea del anterior, se deduce que

$$\eta_j = \frac{\lambda_{i1}}{\lambda_{ij}} = \frac{P(Y_{i1} = 1) P(Y_{ij} = 0)}{P(Y_{i1} = 0) P(Y_{ij} = 1)}$$

Para los lectores con experiencia en tablas de contingencia, esta expresión les puede sugerir un *odd ratio* entre el ítem 1 y el ítem j ; de hecho, es lo que afirman San Martín, González & Tuerlinckx (2009). Pero no hay que olvidar que un *odd ratio* es una medida de asociación (dependencia), mientras que en el modelo Rasch Y_{ij} e Y_{i1} son mutuamente independientes, por lo que no hay asociación alguna. Pero por otro lado esta igualdad tiene una característica interesante, a saber, es válida *para toda persona i* . Dado que para Rasch se trata de medición, una manera de interpretar η_j es la siguiente:

$$\eta_j > \eta_k \Leftrightarrow P(Y_{ik} = 1) > P(Y_{ij} = 1) \text{ para toda persona } i \quad (6).$$

Es decir, el ítem j tiene una característica mayor que el ítem k si, para toda persona, la probabilidad de responder correctamente el ítem k es mayor que la de responder correctamente el ítem j . Es posible argüir que esto sugiere la idea intuitiva (o lo que sea) de

dificultad. Pero es mejor decir que esa comparación de probabilidades de responder correctamente es lo que *podemos llamar dificultad*. Nuevamente, aquí no hay psicología alguna.

Finalmente, es posible comparar una característica de una persona con una característica de un ítem:

$$\varepsilon_i > \eta_j \Leftrightarrow P(Y_{ij} = 1) > P(Y_{ij} = 0) \quad (7)$$

Es decir, la característica de la persona i es mayor que la característica del ítem j si la probabilidad que dicha persona tiene de responder correctamente el ítem j es mayor que la probabilidad de responderlo incorrectamente. Es decir, $\lambda_{ij} > 1$, algo que sabíamos sin introducir la restricción de identificabilidad. Pero no hay contradicción, ni argumento circular: dejamos al lector argumentar por qué.

5. Discusión

Ya es una tradición que los métodos cuantitativos forman parte de la formación de todo cientista social, en particular de psicólogos. Sin embargo, lo que es deplorable es la enorme confusión que se transmite: por un lado, gracias a la eficiencia de los métodos computacionales y a la disponibilidad de software psicométricos, se enseña a los estudiantes a usarlos, enfatizando la potencia en estimar parámetros: con esto solo se forman estadísticos mediocres. Y por otro lado, no se discute el posible significado sustantivo de los modelos psicométricos, lo que significa la necesidad imperiosa de interpretar los parámetros de las distribuciones de probabilidad que generan las observaciones. Digamos de forma enfática: *no hay otra manera de interpretar sustantivamente un modelo psicométrico*.

Pues bien, al hacer este ejercicio con el modelo Rasch, caemos en la cuenta que no hay interpretación alguna psicológica de los parámetros que caracterizan a las personas y a los ítems o

estímulos. Ciertamente sería posible atribuirles un significado psicológico a las desigualdades (5) y (6), pero eso requiere una interpretación psicológica del concepto de probabilidad. Sin este paso epistemológico de enorme envergadura, constatamos que el modelo Rasch no tiene relación alguna con términos (quizás conceptos) como *habilidad de una persona*, *dificultad de un ítem*. Sin embargo, a pesar de esta profunda disociación entre psicometría y psicología, este modelo Rasch es el *único que permite comparar características de personas con características de ítems* (desigualdad (7)), que es lo que en medición educacional se conoce como Standard Setting, con una técnica asociada: Bookmark. Pero nuevamente cuando en este contexto se les pide a jueces expertos comparar esas características, ¿qué está entendiendo cada uno? Solo la desigualdad (7) tiene sentido probabilístico y cualquier protocolo de Standard Setting debería verbalizar dicha desigualdad, cosa que, hasta donde sabemos, no ocurre.

Aparte de la disociación cognitiva que subyace al *uso del modelo Rasch*, hay otra dificultad que se ha confundido en la literatura psicométrica: en esta discusión que hemos desarrollado del modelo Rasch, no hemos hablado de variables latentes. Pero existe una presentación del modelo Rasch en base a variables latentes, y que tiene una implicancia no reconocida en la literatura: que el modelo Rasch, y otros modelos TRI, son *casos particulares del modelo de Teoría Clásica de Tests*; para detalles, ver Lord (1952). En este contexto, y solo en este, tiene sentido el Axioma de Independencia Local.

Una forma de ilustrar la diferencia entre la presentación del modelo Rasch por Rasch mismo, y la debida a Lord, es la siguiente: el modelo Rasch es similar a tener una caja de fichas negras y blancas; se conoce la configuración de la caja. Se extraen fichas con reposición (así se asegura la independencia mutua del modelo Rasch). La pregunta que se intenta responder es la siguiente: ¿cuál es la probabilidad que una determinada secuencia de fichas ocurra? Pero cuando introducimos las variables latentes, entonces la situación cambia: tenemos por ejemplo dos cajas, con fichas blancas

y negras cada una. Conocemos la composición de ellas; conocemos en principio la probabilidad de escoger una u otra caja. Dada una caja, se extraen fichas con reposición (esto asegura la independencia condicional o Axioma de Independencia Local). Ahora bien: antes de escoger una caja, le vendamos los ojos; extraemos con reposición 5 fichas; le sacamos la venda de sus ojos, le mostramos las 5 fichas. La pregunta es: ¿cuál es la caja más probable de dónde fueron extraídas estas 5 fichas?

La pregunta que dejamos al cerrar estas reflexiones es la siguiente: ¿por qué en psicometría interesó resolver un problema de clasificación como el expuesto con la metáfora de las cajas?

Agradecimientos: Parte de este trabajo ha sido parcialmente financiado por el Proyecto FONDEF ID17110220 Espacios de Conocimiento en Matemáticas: Cognición, Medición y Evaluación de Aprendizajes al Servicio de la Diversidad en el Aula.

Referencias

ACE (2015). *Informe Técnico SIMCE 2015*. Agencia de Calidad de la Educación, Santiago, Chile.

AERA (2014). *Standards for Educational and Psychological Testing*. American Educational Research Association, Washington DC, U.S.A.

Del Pino, G. San Martín, E., González, J. & De Boeck, P. (2008). On the Relationships Between Sum Score Based Estimation and Joint Maximum Likelihood Estimation. *Psychometrika* 73, 145-151.

ICFES (2018). *Saber a Detalle. ¿Cómo se generan los puntajes en las pruebas SABER del ICFES?*. Instituto Colombiano para la Evaluación de la Educación, ICFES, Bogotá, Colombia.

Fan, X. & Sun, S (2013). Item Response Theory. In: T. Teo (editor), *Handbook of Quantitative Methods for Educational Research*, Chapter 3. Sense Publishers.

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 222, 309-368.

Koopmans, T. C., & Reiersol, O. (1950). The identification of structural characteristics. *The Annals of Mathematical Statistics*, 21(2), 165-181.

Lord, F. (1952). A Theory of Test Scores. *Psychometric Monographs*.

Olsen, L. W. (2003). *Essays on Georg Rasch and his contributions to statistics*. Københavns Universitet, Økonomisk Institut.

Rasch, G. (1961). On general laws and the meaning of measurement in psychology. In Proceedings of the fourth Berkeley symposium on mathematical statistics and probability (Vol. 4, pp. 321-333).

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. The University Chicago Press, Chicago, U.S.A.

San Martín, E. (2016). Identification of Item Response Theory Models. In: W. van der Linden (Ed.), *Handbook of Item Response Theory, Volumen 2, Statistical Tools*, Chapter 8. CRC Press Taylor & Francis Group, New York.

San Martín, E. (2018). Identifiability of Structural Characteristics: How relevant is in the Bayesian Approach? *Brazilian Journal of Probability and Statistics* 32, 346-373.

San Martín, E., González, J. & Tuerlinckx, F. (2009). Identified Parameters, Parameters of Interest and Their Relationships. *Measurement: Interdisciplinary Research and Perspective* 7, 95-103.

San Martín, E., González, J. & Tuerlinckx, F. (2015). On the Unidentifiability of the Fixed-Effects 3PL Model. *Psychometrika* 80, 450-467.

Skinner, B. F. (1956). A Case History in Scientific Method. *American Psychologist*, 11(5), 221-233.

Wittgenstein, L. (2010). *Tractatus lógico-philosophicus*. Alianza Editorial, Madrid, España.

Figura 1

Matemáticas SIMCE 2001

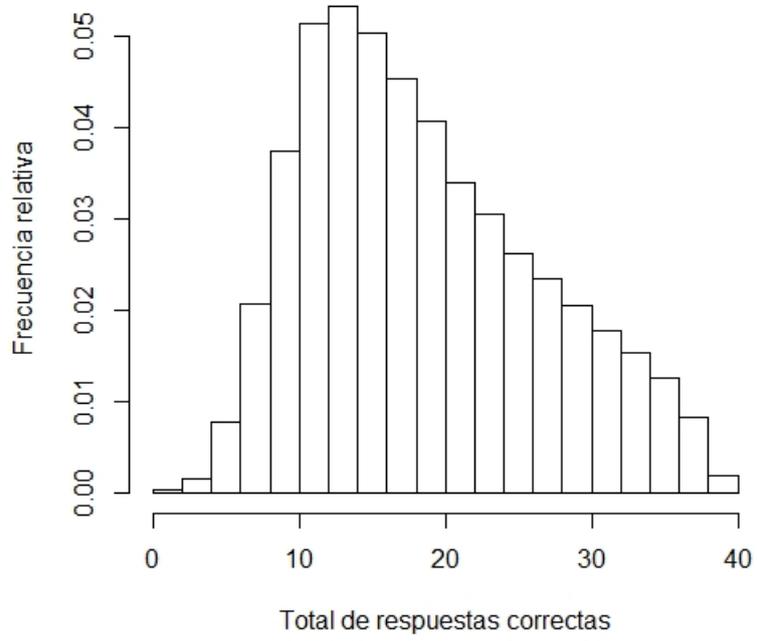


Figura 2

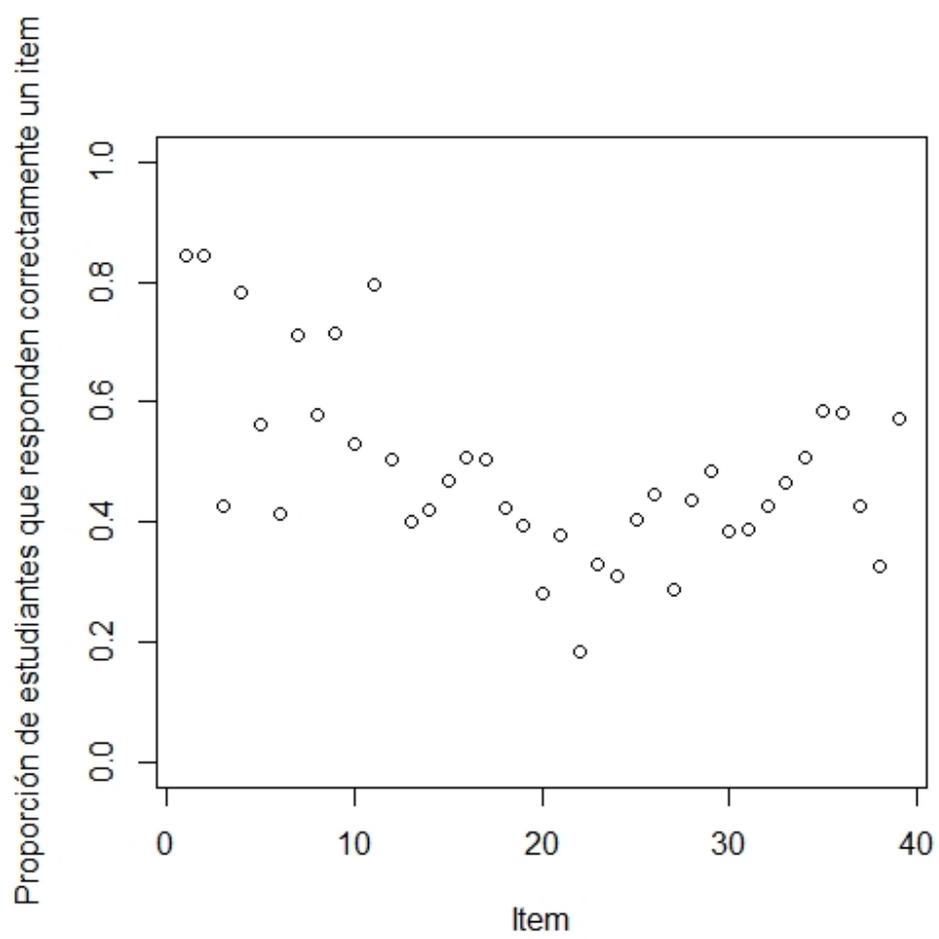


Figura 3

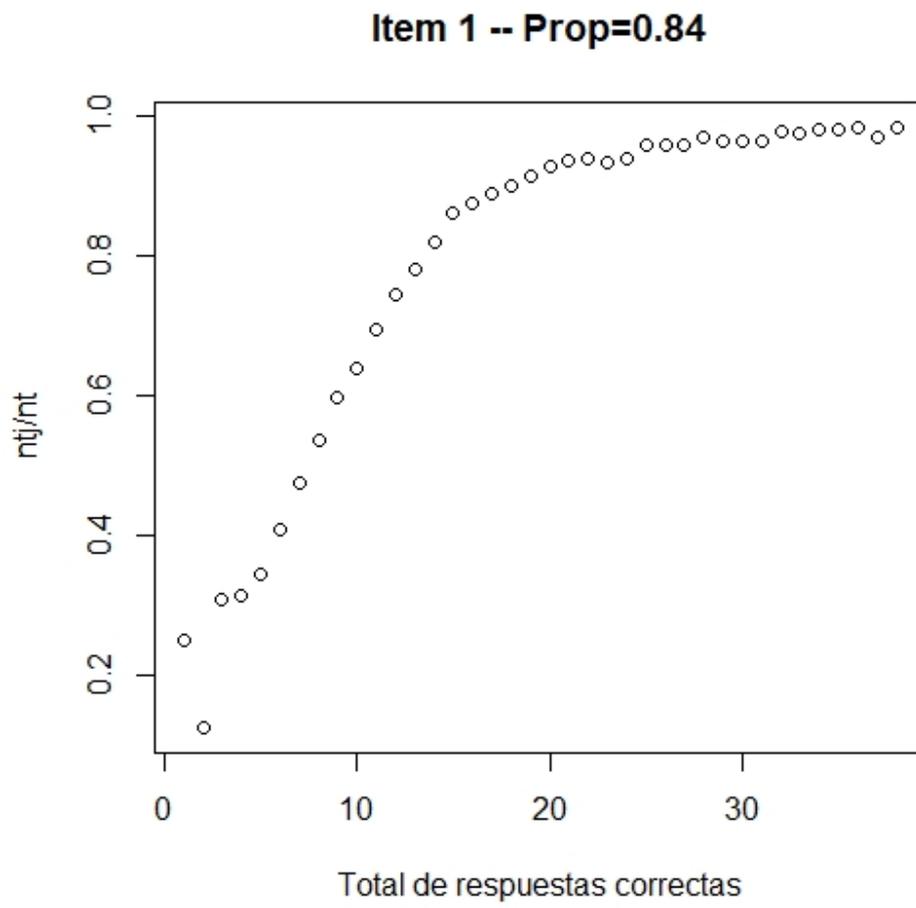


Figura 4

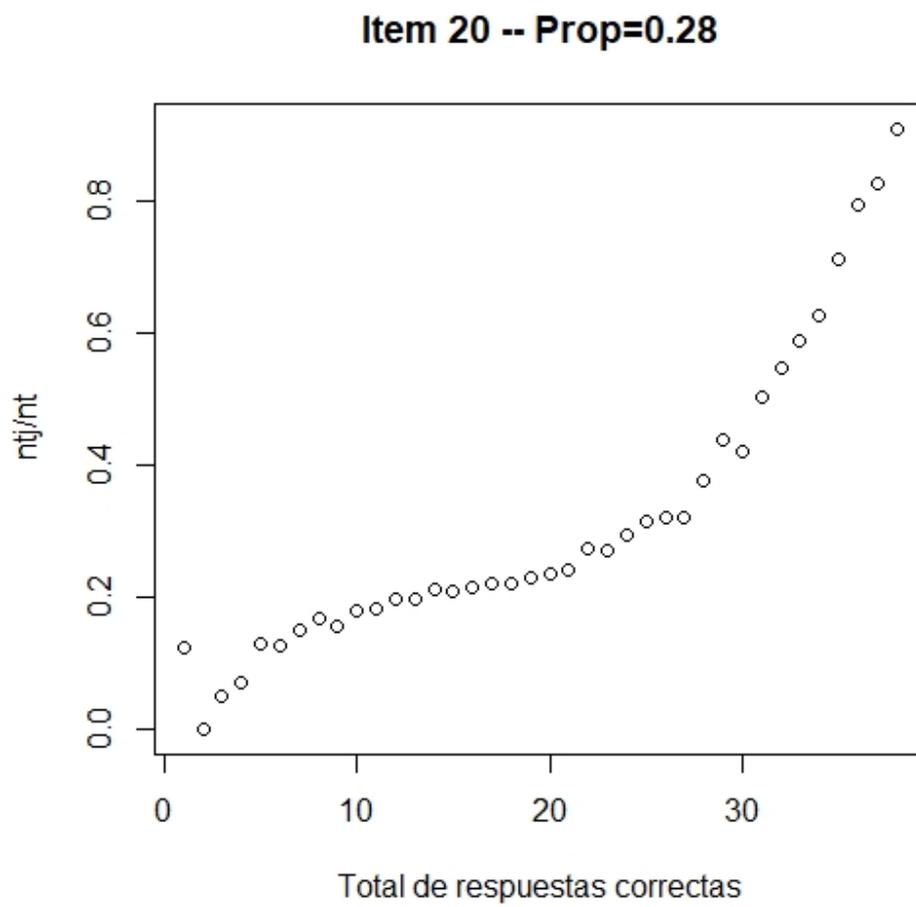


Figura 5

