

Componentes de dificultad de ítems para la evaluación de operaciones lógicas. Una aplicación del modelo LLTM¹

Item difficulty components for the assessment of logical operations. An application of the Linear Logistic Test Model

María Silvia Galibert², Jimena Cristina Picón Janeiro³, Gabriela Susana Lozzia⁴, María Ester Aguerri⁵ y Horacio Félix Attorresi⁶

Instituto de Investigaciones, Facultad de Psicología
Universidad de Buenos Aires, Argentina

(Recepción: Diciembre 2009 – Aceptación: Junio 2010)

Resumen

En este trabajo se analizan los componentes cognitivos implicados en la resolución de ítems de razonamiento deductivo mediante la aplicación del Modelo Logístico Lineal de Rasgo Latente (LLTM) de Fischer. Los ítems fueron aplicados a una muestra de 538 estudiantes de Psicología de la Universidad de Buenos Aires. Se describe el procedimiento de depuración del conjunto inicial de 47 ítems, del cual finalmente quedaron 27 a los que se ajustó el modelo. La correlación lineal de las estimaciones obtenidas bajo los modelos LLTM y Rasch fue $r = 0,8851$. El orden de dificultad de los componentes considerados, desde más difícil a más fácil, fue: Falacias de Negación del Antecedente y de Afirmación del Consecuente, Negación de la Conjunción y de la Disyunción, Silogismos, Cuantificadores, Doble Negación, Tipo de Contenido y Argumentos Modus Ponens y Tollens. Se discuten los alcances y limitaciones de la aplicación de este modelo a los ítems elaborados.

Palabras claves: Modelo LLTM de Fischer, Modelo de Rasch, Componentes de dificultad, Razonamiento deductivo, Teorías cognitivas.

Abstract

This paper analyzes the cognitive components involved in the resolution of deductive reasoning items through the application of Fischer's Linear Logistic Test Model (LLTM). Items were applied to a sample of 538 psychology students of the University of Buenos Aires. The paper describes the refinement procedure of the initial set of 47 items, which resulted in 27 items to which the model was adjusted. The linear correlation of the estimations obtained under the LLTM and Rasch Model was $r = 0.8851$. The order of difficulty of the components considered (from the most difficult to the easiest) was: Fallacies Denying the Antecedent and Affirming the Consequent, Fallacies Denying a Conjunct and Denying a Disjunct, Syllogisms, Quantificational Fallacies, Double Negation, Type of Content, and Modus Ponens and Modus

-
- 1 La investigación que se presenta en este artículo fue realizada con subsidios de la Universidad de Buenos Aires (UBACyT P043) y de la Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT PICT 20909).
 - 2 Correspondencia a: María Silvia Galibert. Tejedor 555, (1424). Ciudad de Buenos Aires. Teléfono: 4923-0605. E-mail: galibert@psi.uba.ar.
 - 3 Correspondencia a: Jimena Cristina Picón Janeiro. Castelli 35, 6° "N", (1031). Ciudad de Buenos Aires. Teléfono: 4951-1832. E-mail: jpicon@psi.uba.ar.
 - 4 Correspondencia a: Gabriela Susana Lozzia. Lope de Vega 1507. Dpto. 2, (1407). Ciudad de Buenos Aires. Teléfono: 4568-6172. E-mail: glozzia@psi.uba.ar.
 - 5 Correspondencia a: María Ester Aguerri. Agrelo 3555, (1224) Ciudad de Buenos Aires. Teléfono: 4931-5453. E-mail: maguerri@psi.uba.ar.
 - 6 Correspondencia a Horacio Félix Attorresi: Rivera Indarte 132 1° "A", (1406) Ciudad de Buenos Aires. Teléfono: 4637-0923. E-mail: horacioattorresi@fibertel.com.ar, hatorre@psi.uba.ar.

Tollens Arguments. The paper includes a discussion of the scope and limitations of the application of this Model to the items developed.

Key words: Fischer LLTM Model, Rasch Model, Difficulty components, Deductive reasoning, Cognitive theories.

Componentes de dificultad de ítems para la evaluación de operaciones lógicas. Una aplicación del modelo LLTM

Con el objeto de mejorar tanto los fundamentos teóricos como la eficiencia técnica de los tests en los últimos años se ha producido un acercamiento entre la Psicología Cognitiva y la Psicometría. La construcción de tests se ha visto influenciada por los planteamientos cognitivos, lo cual ha enriquecido el concepto de “validez”: esta no sólo implica la relación de la prueba con otras medidas sino que también incluye la determinación de los procesos, estrategias y estructuras de conocimiento necesarias a la hora de responder a los ítems (Embretson, 1983).

En este trabajo se analizarán los componentes cognitivos implicados en la resolución de ítems de razonamiento deductivo, tema ampliamente estudiado por la psicología cognitiva. Las investigaciones acerca de las dificultades de las operaciones lógicas se han llevado a cabo con tareas de tipo experimental y también en el contexto psicométrico a partir de la evaluación por parte de los sujetos de la validez de argumentos lógicos. Tal como se ha señalado en Attorresi, Picón Janeiro, Abal, Aguerrí y Galibert (2009) han sido documentadas las dificultades que exhiben los sujetos para ese tipo de razonamiento cuando éste implica distintos conectivos lógicos, tales como el condicional. Respecto de este, por ejemplo, se ha encontrado que los sujetos tienden a interpretar los argumentos condicionales, bajo ciertas circunstancias, como bicondicionales, interpretación que da lugar a las falacias llamadas de Negación del Antecedente (NA) y de Afirmación del Consecuente (AC) (Rips y Marcus, 1977; Marcus y Rips, 1979 y Dasí y Algarabel, 2003). Respecto de la combinación de la negación con la conjunción o con la disyunción (problemas que requieren para su correcta resolución de la aplicación de las leyes de De Morgan o de los silogismos conjuntivo / disyuntivo) se ha encontrado que los problemas que necesitan de las reglas de introducción de la negación o del silogismo conjuntivo son más difíciles de resolver que los que requieren de la conjunción o del Modus Ponens (Rips, 1994). Lin y Cheng (1997) afirman que las sentencias en las que aparecen negados una conjunción, una disyunción o un cuantificador resultan más difíciles tanto para estudiantes secundarios como universitarios (Citado por Lin y Yu, 1997). Por su parte Lin, Lee y Yu (2003) hallaron que el 40% de estudiantes negaba el “todos” como “ninguno”. En cuanto al efecto del contenido temático versus el contenido abstracto se han reportado resultados inestables: algunos investigadores hallaron un efecto facilitador del contenido temático sobre las tareas de selección (Bracewell y Hidi, 1974; Johnson-Laird, Legrenzi y Legrenzi, 1972; Van Duyne, 1974; Wason y Shapiro, 1971) y otros no encontraron diferencias significativas asociadas a material estimular concreto u abstracto en relación al condicional (Manktelow y Evans, 1979).

En el marco de la Teoría de Respuesta al Ítem (TRI) se han desarrollado modelos psicométricos que en su formulación incluyen variables de contenido cognitivo. En estos modelos se cuantifica el peso de dichas variables en las propiedades psicométricas del test, permitiendo así no sólo escalar a los sujetos y a los ítems sino también estudiar con mayor profundidad el constructo medido (López Jáuregui y Elosúa Oviden, 2002). El Modelo Logístico Lineal de Rasgo Latente LLTM (Fischer, 1973, 1997) permite determinar las fuentes de dificultad de los reactivos a partir del impacto de las características de esos ítems sobre el procesamiento y sobre las respuestas. El LLTM vincula las características del diseño de los ítems con sus características psicométricas y con los atributos cognitivos que ese ítem pretende medir, y hace posible analizar el tipo de procesos requeridos para resolver dicho ítem. La aplicación del LLTM requiere de una teoría cognitiva con la cual se puedan explicar los requerimientos de procesamiento de los ítems responsables de la variabilidad de la dificultad entre los mismos. Este modelo es una extensión del modelo de Rasch (1960) y permite

descomponer la dificultad de un ítem como suma de los efectos debidos a las fuentes de dificultad predichas por las teorías cognitivas, decidir si éstos son significativos y estimarlos. Su formulación es

$$\beta_i = \sum_{l=1}^p w_{il} \alpha_l + c \quad (1)$$

donde β_i son los parámetros de dificultad del modelo de Rasch, p expresa el número de componentes considerados, α_l la dificultad del componente l , w_{il} la frecuencia con la que ese componente está presente en el ítem i y c es una constante de escalamiento. Los α_l se denominan *parámetros básicos*.

De la ecuación (1) se sigue que la aplicación del LLTM tiene sentido sólo si el Modelo de Rasch se ajusta suficientemente bien a los datos. Por otra parte, si los componentes considerados explican de manera suficientemente exhaustiva las diferencias entre los ítems, una vez estimados los parámetros se deberían recuperar a través la ecuación (1) estimaciones β_i similares a las obtenidas directamente del modelo de Rasch, lo que implicaría una alta correlación entre los parámetros estimados bajo ambos modelos.

Revuelta y Ponsoda (1998) aplicaron el modelo LLTM en la primera fase de la transformación de un test tipo “papel y lápiz” a un test adaptativo informatizado (TAI) puesto que la delimitación de los componentes de dificultad y la estimación de sus efectos pueden ser útiles a la hora de generar ítems con parámetros de dificultad preestablecidos. Real, Olea, Ponsoda, Revuelta y Abad (1999) ajustaron el modelo LLTM para analizar la dificultad de un test de matemática.

Attorresi, Picón Janeiro, Abal, Aguerri y Galibert (2009) elaboraron ítems para la evaluación del razonamiento deductivo que modelizaron con el modelo LLTM considerando cinco componentes de dificultad. El presente trabajo es una continuación de este último, donde se amplía el conjunto de ítems por analizar y se añaden otras dos componentes de dificultad. Por eso tiene como objetivos, por un lado, volver a ajustar el modelo LLTM sobre el nuevo conjunto de ítems, comparar los resultados obtenidos con los de la etapa anterior y discutir los alcances y limitaciones de la aplicación de este modelo a los ítems elaborados. Por otro lado, al presentarse el contenido de los ítems ajustados con la información exhaustiva respecto de las estimaciones de los parámetros obtenidas tanto con el LLTM como con el modelo de Rasch, el propósito es aportar un material que puede ser de interés para aquellos investigadores interesados en el desarrollo de pruebas de razonamiento deductivo que pueden tomarlo como material inicial para continuar con su desarrollo y modelización.

Método

Participantes

Los ítems se administraron a 538 estudiantes de segundo año de la Facultad de Psicología (incluido el CBC) de la UBA distribuidos en tres muestras de 134, 197 y 207 con un 81,4 % de mujeres y 18,6 % de varones y una edad promedio de 22,2 años (DS = 5.91)

Material

Consiste de tres pruebas de 24 ítems con 12 en común entre las tres y otro en común entre dos de las pruebas; por lo que el total de ítems diferentes es 47. Cada ítem consiste de un grupo de premisas y una conclusión. El formato de respuesta es de verificación; es decir, el sujeto debe decidir si una conclusión se sigue válidamente o no de las premisas. Algunos de ellos son de contenido concreto neutro, para evitar el sesgo de creencias u opinión, y otros de contenido abstracto o simbólico. Los hay válidos e inválidos. Algunos ítems involucran operaciones simples como el Modus Ponens

o Modus Tollens mientras que otros combinan distintas operaciones. Algunos de esos ítems se presentan en el Apéndice habiéndose separado premisas y conclusión por un punto aunque en su administración éstas se hallaban dispuestas unas debajo de otras.

Procedimiento

Debido a que los datos corresponden a tres muestras con un conjunto de ítems de anclaje, se utilizó el programa BILOG-MG que permite procesar datos con este tipo de diseño. Su aplicación se hizo de manera iterativa con el fin de ir depurando el conjunto mediante la eliminación de los ítems de correlación negativa y de los que no ajustaban al modelo de Rasch. Para descartar un ítem por desajuste no se tenía en cuenta sólo el valor p del test χ^2 , lo cual es un criterio muy exigente, sino también sobre la observación del diagrama correspondiente y se mantenían los ítems que exhibían una gráfica aceptable. Un diagrama se consideró aceptable cuando, a pesar de las discrepancias entre los puntos que corresponden a lo observado y a lo esperado bajo el modelo, globalmente el gráfico de lo observado tiene una tendencia logística; por ejemplo que no decrezca. Sobre los ítems así seleccionados se aplicó el modelo LLTM mediante el programa LPCM-Win (Fischer & Ponocny-Seliger, 1997). Se examinó la correlación entre las estimaciones de los parámetros de dificultad obtenidas con el modelo de Rasch y las obtenidas con el modelo LLTM y se eliminaron aquellos ítems que claramente se alejaban de la linealidad. También, a fin de explorar la estabilidad de los resultados con relación al ajuste realizado en la etapa anterior, se correlacionaron los parámetros de dificultad obtenidos en aquélla bajo el modelo de Rasch con los predichos por el modelo LLTM en la presente etapa.

Delimitación de los componentes de dificultad. Se llevó a cabo teniendo en cuenta las investigaciones en el área de la psicología cognitiva en general arriba mencionadas y, en particular, los resultados preliminares de Attorresi et al. (2009) donde se consideraron los componentes *Contenido*, *Negación* \wedge / \vee , *Modus Ponens / Tollens*, *Falacias NA / AC* y *Cuantificadores*. En el presente trabajo se desglosó el componente *Cuantificadores* en dos: *Negación de cuantificadores* y *Silogismos* y también se añadió el componente de *Doble Negación*. Todos estos componentes se definen a continuación.

Contenido. Cada ítem puede estar formulado con contenido simbólico, abstracto o concreto neutro. Attorresi et al (2009) mostraron que los ítems de contenido abstracto o simbólico tendían a estar más próximos entre sí en cuanto a la dificultad y ser más difíciles que los de contenido concreto. Por esa razón se decidió dicotomizar el componente *Contenido* asignando los valores 0 si es concreto y 1 si es simbólico o abstracto.

Doble Negación. El acceso a la comprensión de una proposición formulada en términos de doble negación es obviamente más indirecto que si ésta se expresa de manera afirmativa, por lo que surge como hipótesis natural que la doble negación puede introducir dificultad en un ítem. Por esta razón se asignó 1 a los ítems donde esta dificultad estaba presente y 0 si no lo estaba.

Negación \wedge / \vee . Este componente se refiere a la negación en relación a la conjunción o a la disyunción. Se halla presente en aquellos que involucran a las Leyes de De Morgan o a las reglas del silogismo hipotético conjuntivo o disyuntivo. Se codificó con 1 si cualquiera de estas leyes o reglas está presente y 0 en el caso contrario.

Modus Ponens / Tollens. Se asignó el valor 1 a este componente cuando las premisas hacen posible el empleo de Modus Ponens o Modus Tollens y 0 en otro caso. Algunos de los ítems tienen la conclusión válida que corresponde a estos argumentos mientras que otros, que combinan esta dificultad con la negación del cuantificador universal, tienen una conclusión inválida al hacer esta operación de manera incorrecta.

Falacias NA / AC. Si un ítem contiene una falacia NA o AC el componente fue codificado con 1 y de lo contrario con 0.

Cuantificadores. Los ítems asociados a este componente evalúan la negación de los cuantificadores. De estar presente esta dificultad el componente se codificó con 1 y de estar ausente con 0.

Silogismos. Entre los ítems los hay con estructura de silogismos categóricos, válidos o no, o modales. Este componente se codificó con 1 si el ítem era un silogismo y 0 en caso contrario.

Resultados

Del primer proceso de depuración se retuvieron 30 ítems, de los cuales hubo 6 para los que se rechazó el ajuste al 5% aunque uno ajustó al 1% y 5 ajustaron al 1%. El valor del estadístico χ^2 de Andersen (1973) para evaluar el ajuste del modelo LLTM fue de 165,2638 ($gl = 19$), con lo que se rechaza el ajuste global al modelo al 5% siendo el valor crítico 30,1348. La correlación lineal de las estimaciones obtenidas bajo los modelos LLTM y Rasch fue $r = 0,8851$. Ambos resultados antedichos, de ajuste y de correlación, corresponden a haber excluido tres ítems que evidenciaban un mayor desajuste, es decir que se obtuvieron trabajando sobre 27 ítems. El resultado de correlacionar los parámetros de dificultad obtenidos en la primera etapa bajo el modelo de Rasch con los predichos por el modelo LLTM en la presente etapa fue $r = 0,8851$, correlación basada solamente sobre los 5 ítems que, después de la depuración y ajustes, quedaron en común, designados en el apéndice como L01, L02, L03, L06 y L09.

En la Tabla 1 se detalla, para cada uno de los 27 ítems, la estimación de los parámetros con el LLTM, con el modelo de Rasch, el error estándar de esta estimación y el valor z . Aquí se designan con las letras A, B y C los ítems que integraron cada una de las tres formas administradas. La letra L señala los ítems “Link”, es decir aquellos comunes a las tres formas, y que fueron seleccionados entre los que ajustaban de la administración anterior (Attorresi, Picón Janeiro, Abal, Aguerri y Galibert, 2009).

Tabla 1: Estimaciones de los parámetros de dificultad bajo los modelos LLTM y de Rasch, errores estándar.

Ítems	Items β por LLTM	β por Rasch	
		β	Errores estándar
L01	0.2182	0.2848	0.0909
L02	0.2951	0.2439	0.0905
L03	-0.8624	-0.9173	0.1037
L06	0.4647	0.8811	0.0940
L08	-0.1817	-0.1784	0.0724
L09	0.2789	-0.1784	0.0724
L12	-0.1817	-0.5500	0.0959
A09	0.2951	-0.1776	0.0906
A11	-0.2025	-0.4306	0.1115
A16	-0.4329	-0.3934	0.1068
A18	0.9462	0.5986	0.1092
A20	0.9103	1.3922	0.1925

A21	-0.2113	-0.2841	0.0968
A22	-0.1817	-0.0733	0.0869
B10	1.3177	1.6337	0.1354
B11	0.1093	0.7052	0.1223
B12	0.0573	-0.1591	0.0781
B16	-0.2633	-0.3024	0.0837
B21	-0.6928	-1.1701	0.1569
B22	0.2951	-0.2541	0.0815
B23	-0.1817	-0.1827	0.0788
C05	0.0573	-0.4146	0.1118
C10	1.5614	1.2218	0.1190
C16	-0.2113	0.0558	0.0740
C22	0.0573	0.7836	0.1161
C23	-1.7154	-1.0820	0.1317
C24	-1.5458	-1.7660	0.1686

En la Tabla 2 se exhiben las estimaciones de los componentes de dificultad, todos ellos significativos, con sus errores estándar y niveles de significación.

Tabla 2: Estimación de parámetros en el modelo LLTM.

Componente	Parámetros	Errores estándar	Valores z
Falacias NA / AC	1.5810	0.1194	13.2435**
Negación $\wedge\vee$	1.3432	0.0973	13.8071**
Silogismos	1.1042	0.1282	8.6122**
Cuantificadores	0.8530	0.0929	9.1855**
Doble Negación	0.4815	0.0759	6.3430**
Contenido	0.1696	0.0692	2.4508*
MP / MT	-0.4295	0.0745	5.7617**

Nota: *.01 < p < .05; ** p < .001.

El orden de dificultad resultante, desde más difícil a más fácil, fue: *Falacias NA / AC*, *Negación $\wedge\vee$* , *Silogismos*, *Cuantificadores*, *Doble Negación*, *Contenido* y *Modus Ponens / Tollens*. Este orden coincide, salvo una permutación del *Contenido con Cuantificadores* (que en la etapa actual se dividió en dos: *Silogismos* y *Cuantificadores*), respecto del obtenido en la etapa previa.

En la Tabla 3 se muestra el porcentaje de respuestas correctas para cada ítem y la Matriz de diseño.

Tabla 3: Porcentaje de respuestas correctas (p%) para cada ítem y Matriz de diseño.

Ítems	p%	Contenido	Doble Negación	Negación \wedge / \vee	Condicional	NA / AC	Cuantificadores	Silogismos
L01	51	1	1	0	0	0	1	0
L02	52	0	0	0	0	1	0	0
L03	76	0	0	0	1	0	1	0
L06	37	1	0	0	0	1	0	0
L08	53	0	0	0	0	0	0	1
L09	53	1	1	1	1	0	0	0
L12	69	0	0	0	0	0	0	1
A09	63	0	0	0	0	1	0	0
A11	69	1	0	1	1	0	0	0
A16	68	0	0	0	0	0	1	0
A18	46	1	1	0	0	1	0	0
A20	30	0	0	1	0	0	1	0
A21	66	1	1	0	1	0	1	0
A22	61	0	0	0	0	0	0	1
B10	24	1	0	0	0	1	1	0
B11	42	0	1	1	1	0	0	0
B12	61	0	0	1	0	0	0	0
B16	65	1	0	0	0	0	1	0
B21	80	1	0	0	1	0	1	0
B22	64	0	0	0	0	1	0	0
B23	62	0	0	0	0	0	0	1
C05	64	0	0	1	0	0	0	0
C10	28	1	1	1	0	0	1	0
C16	53	1	1	0	1	0	1	0
C22	37	0	0	1	0	0	0	0
C23	77	0	0	0	1	0	0	0
C24	87	1	0	0	1	0	0	0

Discusión

Comparación con resultados anteriores

Es satisfactoria la estabilidad en cuanto a la significancia y orden de los componentes de dificultad una vez ajustado el modelo LLTM como también las predicciones de los parámetros de dificultad para los ítems en común. En efecto, habiendo sido la permutación entre *Contenido* y *Cuantificadores* la única discrepancia en el orden de dificultad de los componentes en ambas etapas, puede afirmarse que los resultados fueron bastante estables; más aún teniendo en cuenta que se agregaron dos componentes, lo que podría haber alterado bastante más la estimación del resto de los parámetros básicos, y que el componente *Cuantificadores* fue redefinido en la segunda etapa restringiéndolo a

la dificultad de la negación. Asimismo, la alta correlación obtenida entre los parámetros estimados con el modelo de Rasch en la primera etapa y los predichos por el LLTM en la segunda también dan cuenta de cierta estabilidad; sin embargo, el conjunto de ítems que finalmente quedó en común es muy pequeño por lo que será de interés replicar este análisis en etapas posteriores con más ítems.

Dificultades y limitaciones

Se hallaron dificultades en la construcción de ítems creados con la intención de representar ciertas combinaciones de los componentes de dificultad; en su mayor parte éstos resultaron demasiado artificiosos o complicados y presentaron correlaciones negativas, por lo que no pudieron pasar el primer proceso de depuración.

Por lo que respecta al ajuste del modelo de Rasch, difícilmente puedan obtenerse ítems con igual discriminación por lo que se necesita apelar a la robustez. En cuanto al modelo LLTM, no se ha logrado el ajuste en términos inferenciales pero Fischer (1997) afirma que esto ocurre la mayoría de los casos, no obstante los cuales su aplicación puede ser útil porque conduce al investigador a explicitar sus hipótesis respecto del material, le permite ponerlas a prueba y aún si la explicación de la dificultad de los ítems en términos de los parámetros básicos no es perfecta al menos le permite predecirlos aproximadamente. Con respecto a la exhaustividad de la representación, debe decirse que el parámetro de dificultad del componente *Silogismos* no debe considerarse para predecir la dificultad de cualquier tipo de silogismos sino solamente de los que han sido incluidos en la prueba: EAO, OAO, IOO, EAE. Por este motivo, o hay que tener en cuenta lo limitado de la representación de este componente, o habría que agregar otro tipo de silogismos que lo representara mejor o reducir la prueba a argumentos que no tengan la forma de silogismos. Por todo ello, el alcance de los resultados es limitado; éstos son aproximados y provisionarios y tienen fundamentalmente el valor de orientar el curso de la investigación en la construcción de nuevos ítems teniendo en cuenta las características tanto de los ítems que quedaron ajustados como de los eliminados.

Referencias

- Andersen, E. (1973). A goodness of fit test for the Rasch model. *Psychometrika*, 38, 123-140.
- Attorresi, H., Picón, J., Abal, F., Aguerrí, M. y Galibert, M. (2009). Aplicación del modelo LLTM de Fischer al análisis de las fuentes de dificultad de ítems de razonamiento deductivo. *Interdisciplinaria*, 26 (1), 77-93.
- Bracewell, R. y Hidi, S. (1974). The solution of an inferential problema as a function of the stimulus materials. *Quarterly Journal of Experimental Psychology*, 26 (3), 480-88.
- Dasí, C. y Algarabel, S. (2003). Influencia del entrenamiento sobre el razonamiento deductivo: Importancia del contenido y transferencia entre dominios. *Psicothema*, 15, 3, 440-445.
- Embretson, S .E. (1983). Construct validity: construct representation versus nomothetic span. *Psychological Bulletin*, 93(1), 179-197.
- Fischer, G. (1973). The linear logistic test model as an instrument in educational research. *Acta Psychologica*, 37, 359 -374.
- Fischer, G. (1997). Unidimensional Linear Logistic Rasch Model. En W. J. Van Der Linden y R. K. Hambleton (1997). *Handbook of modern item response theory*. New York: Springer-Verlag.
- Fischer, G. y Ponocny-Seliger, E. (1997). *LPCM-WIN Program*. Groningen: IEE, ProGamma.
- Johnson-Laird, P., Legrenzi, P. y Legrenzi, M. (1972). Reasoning and sense of reality. *British Journal of Psychology*, 63,395-400.
- Lin, F. y Cheng, Y. (1997). An explorative teaching on the argument principle of proof by contradiction. *Chinese Journal of Science Education*, 5(4), 557-591.

- Lin, F. y Yu, J. (1997). False Proposition – As a means for making conjectures in mathematics classrooms. Asian Mathematical Conference, 20-23 July, Singapore. Recuperado el 19 de abril de 2007, de <http://www1.math.nus.edu.sg/AMC/papers/Lin-Fou-Lai.pdf>.
- Lin F., Lee, Y. y Yu, J. (2003). Students' understanding of proof. by contradiction. *La lettre de la Preuve*, 4, 443-450. International Newsletter on the Teaching and Learning of Mathematical Prof. Recuperado el 19 de abril de 2007, de http://www.lettredelapreuve.it/PME/PME27/RR_wu_yu.pdf
- López Jáuregui, A. y Elosúa Oliden, P. (2002) Formulación y validación de un modelo logístico lineal para la tarea de adición y sustracción de fracciones y números mixtos. *Psicothema*, 14, 4, 802-809.
- Manktelow, K. y Evans, J. (1979). Facilitation of reasoning by realism: effect or non effect?. *British Journal of Psychology*, 71, 227-31.
- Marcus, S. y Rips, L. (1979). Conditional reasoning. *Journal of Verbal Learning and Verbal Behavior*, 18, 199-223.
- Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests. Chicago, IL: University of Chicago Press.
- Real, E., Olea, J., Ponsoda, V., Revuelta, J. y Abad, F. (1999). Análisis de la dificultad de un test de matemática mediante un modelo componencial. *Psicológica* 20, 121-134.
- Revuelta, J. y Ponsoda, V. (1998). Un test adaptativo informatizado de análisis lógico basado en la generación automática de ítems. *Psicothema*, 10, 3, 709-716.
- Rips, L. (1994). *The psychology of proof: Deductive reasoning in human thinking*. Cambridge, Mass. M.I.T. Press, Bradford Books.
- Rips, L. y Marcus, S. (1977). Suppositions and the analysis of conditional sentences. M. A. Just y P. A. Carpenter (Comps.), *Cognitive Processes in Comprehension*, Nueva York: Willey.
- Van Duyne, P. (1974). Realism and linguistic complexity in reasoning. *British Journal of Psychology*, 65, 59-67.
- Wason, P. y Shapiro, D. (1971). Natural and contrived experience in a reasoning problem. *Quarterly Journal of Experimental Psychology*, 23, 63-71.

Apéndice

Ítems de Razonamiento Deductivo

- L01. No se cumple que ningún elemento difiere de A. Por tanto: Algún elemento no es A. (Válido)
- L02. Si se rompe el parabrisas suena la alarma. No se rompe el parabrisas. Por tanto: No suena la alarma. (Inválido)
- L03. Si todos los archivos están comprimidos caben juntos en un CD. Los archivos no caben todos juntos en un CD. Por tanto: Hay algún archivo que está descomprimido. (Válido)
- L06. Si hay un punto también hay una raya. Hay una raya. Por tanto: Hay un punto. (Inválido)
- L08. Algunos alfajores no son de chocolate. Todas las golosinas de chocolate se venden bien. Por tanto: Algunos alfajores no se venden bien. (Inválido)
- L09. Si hay un círculo y no hay una letra entonces hay una estrella. No hay una estrella. Por tanto: No hay un círculo o hay una letra. (Válido)
- L12. No le gusta ningún programa de chismes. Todos los programas de chismes tienen bastante rating. Por tanto: Algunos programas de bastante rating no le gustan. (Válido)
- A20. Es falso que todas las fichas sean redondas y coloradas. Por tanto: Algunas fichas no son ni redondas ni coloradas. (Inválido)
- A09. Si no se ingresa la clave de identificación personal el cajero no entrega dinero. El cajero no entrega dinero. Por tanto: No se ha ingresado la clave de identificación personal. (Inválido)
- A11. A o B están abajo. Si B está abajo C está arriba. A no está abajo. Por tanto: C está arriba. (Válido)
- A16. En esta caja hay clavos que no tienen la longitud adecuada. Por tanto: En esta caja no hay clavos con la longitud adecuada. (Inválido)
- A18. Si R es 1 entonces R no es negro. R no es 1. Por tanto: R es negro. (Inválido)
- A21. Si ningún A es distinto de B entonces C es D. C no es D. Por tanto: Hay algún A que no es B. (Válido)
- A22. Algunos empresarios son pilotos. Algunos empresarios no son coleccionistas. Por tanto: Algunos pilotos no son coleccionistas. (Inválido)
- B10. Si todo A es B entonces C es D. Si ningún A es B entonces C no es D. Es falso que todo A sea B. Por tanto: C no es D. (Inválido)
- B11. Si los aspirantes reúnen los requisitos para el cargo entonces no son menores sin experiencia. Raúl reúne las condiciones para el cargo aunque no tiene experiencia. Por tanto: Raúl no es menor. (Válido)
- B12. En el papel no hay un cuadrado y un círculo. Por tanto: En el papel no hay ni un cuadrado ni un círculo. (Inválido)
- B16. Es falso que todo A sea B. Por tanto: Ningún A es B. (Inválido)
- B21. Si todos los azules están arriba todos los verdes están abajo. Hay un verde que no está abajo. Por tanto: No hay azules arriba. (Inválido)
- B22. Si el estudiante no eleva la solicitud no obtendrá la beca. El estudiante no obtuvo la beca. Por tanto: El estudiante no elevó la solicitud. (Inválido)
- B23. Ninguna de las vasijas que ofrecemos a la venta es de barro. Todos los artículos de barro se vendieron esta mañana. Por tanto: Esta mañana no se vendió ninguna vasija. (Inválido)

- C05. Solicitaron para la conferencia una sala amplia y fresca, por lo que no podemos elegir ésta. Por tanto: La sala no es amplia o no es fresca. (Válido)
- C10. Es falso que ningún A es B ni es C. Por tanto: Algún A es B y es C. (Inválido)
- C16. Si ningún A es distinto de B, entonces C es D. C no es D. Por tanto: Todos los A son distintos de B. (Inválido)
- C22. Es falso que el discurso fue largo y aburrido. Por tanto: El discurso no fue ni largo ni aburrido. (Inválido)
- C23. La presión está elevada. Cuando la temperatura está alta la presión está baja. Por tanto: La temperatura no está alta. (Válido)
- C24. El rombo no es verde. Si el triángulo es amarillo el rombo es verde. Por tanto: El triángulo no es amarillo. (Válido)

