

Medición de la Alfabetización Económica en Niños: Oportunidades Diagnósticas con el Modelo de Crédito Parcial

Measurement of Economic Literacy in Children: Diagnostic Opportunities With Partial Credit Model

René Gempp
Universidad Alberto Hurtado

Marianela Denegri
Universidad de la Frontera

Cristina Caprile
Banco de Chile

Lorna Cortés, Mariela Quesada y Jocelyne Sepúlveda
Universidad de la Frontera

Este trabajo describe el uso del Modelo de Crédito Parcial (MCP) para medir Alfabetización Económica en niños, a través de ítems de Elección Múltiple Ordenada (EMO). Elaboramos ítems EMO basados en un modelo piagetano que propone tres niveles progresivos en el desarrollo de conceptos y razonamiento sobre la economía. En el formato EMO, cada una de las posibles alternativas de respuesta estuvo ligada a una etapa cognitiva de desarrollo del pensamiento económico propuesta por la teoría. Los ítems fueron administrados a 1035 niños, entre 10 a 15 años de edad, y calibrados utilizando el MCP. Los resultados muestran tanto la utilidad de los ítems EMO como la eficiencia del MCP como herramientas para puntuar, analizar e interpretar evaluaciones diagnósticas construidas en un marco de referencia piagetano.

Palabras Claves: modelo de crédito parcial, modelo de Rasch, alfabetización económica, ítems de elección múltiple ordenada, niños.

This paper describes the use of the Partial Credit Model (PCM) to measure Economic Literacy in children, with Ordered Multiple-Choice Items (OMC). We developed OMC items based on a Piagetian model that proposes three progressive levels in the development of concepts and reasoning about the economy. In OMC format, each of the possible answer choices was linked to a cognitive stage of economical thinking proposed by the theory. The items were administered to 1035 children, ages 10 to 15 years, and calibrated using the PCM. The results show both the utility of OMC items and the efficiency of PCM as tools for the score, analysis, and interpretation of diagnostic assessments constructed in a Piagetian framework.

Keywords: partial credit model, Rasch model, economic literacy, ordered multiple-choice items, children.

René Gempp, Escuela de Psicología, Universidad Alberto Hurtado. Marianela Denegri, Departamento de Psicología, Universidad de la Frontera. Cristina Caprile, Economista, Banco de Chile. Lorna Cortés, Departamento de Psicología, Universidad de la Frontera. Mariela Quesada, Departamento de Psicología, Universidad de la Frontera. Jocelyne Sepúlveda, Departamento de Psicología, Universidad de la Frontera.

La correspondencia relativa a este artículo deberá ser dirigida a René Gempp, Escuela de Psicología, Universidad Alberto Hurtado, Barroso 26, Santiago, Chile, o a la dirección electrónica rgempp@uahurtado.cl

Esta investigación ha sido financiada por el Fondo de Ciencia y Tecnología (FONDECYT) de Chile mediante el Proyecto N° 1030271. Adicionalmente, el primer autor contó con aportes parciales del Fondo de Investigación Interno de la Universidad Alberto Hurtado, para financiar la adquisición de softwares especializados.

Los autores desean expresar su agradecimiento a dos evaluadores anónimos por sus valiosos comentarios y sugerencias a la primera versión de este trabajo.

Sucesivas investigaciones sobre educación económica y competencias ciudadanas, desarrolladas en las últimas décadas, han constatado las serias dificultades que presentan niños, adolescentes y adultos tanto para comprender la economía cotidiana como para actuar eficientemente en ella (e.g. Davis, Howie, Mangan & Shqiponja, 2002). En este marco, se ha propuesto la noción de “Alfabetización Económica” [AE] (Yamane, 1997) para definir aquel conjunto de conceptos, habilidades, destrezas y actitudes que permiten al individuo la comprensión del entorno económico cercano y global y la toma de decisiones eficientes de acuerdo a sus recursos financieros. Un individuo económicamente alfabetizado es aquel que cuenta con el mínimo suficiente de herramientas conceptuales y procedimentales para entender su mundo económico, in-

interpretar los eventos que pueden afectarlo directa o indirectamente y tomar decisiones personales y sociales sobre la multitud de problemas económicos que se encuentran en la vida cotidiana (Buckles & Melican, 2002).

Varios autores han destacado la importancia de estudiar la AE infantil en la medida que esta variable aglutina, bajo un mismo constructo, los problemas en la comprensión del mundo económico exhibidos por niños y por adultos (e.g. Crick, 1999; Davis, Gregory & Riley, 1999; Walstad & Rebeck, 2001). Desde este punto de vista, la medición de la AE infantil resulta, además, un recurso atractivo para los investigadores en el campo de la socialización económica (Günter & Furnham, 1998). Obviamente, una condición previa para el estudio de la AE y sus variables asociadas, es el desarrollo de medidas confiables y válidas del constructo.

No obstante, la medición de la AE en la infancia tropieza con un importante problema. Diversos estudios muestran que los niños y adolescentes no están económicamente inertes, sino que construyen activamente explicaciones sobre el mundo económico (e.g. Abramovitch, Freedman & Pliner, 1991; Berti & Grivet, 1990; Delval, Enesco & Navarro, 1994; Denegri, Delval, Ripoll, Palavecinos & Keller, 1998; Leiser, Sevón & Levi, 1990; Sonuga-Barke & Webley, 1993). Estas explicaciones, generadas espontáneamente, están ancladas en los recursos cognitivos disponibles según el nivel de desarrollo del niño. De esta forma, expuestos a la misma información, un niño pequeño y otro mayor construirán representaciones y explicaciones del mundo económico muy distintas.

Lo anterior tiene dos consecuencias. La primera es que, a diferencia de la AE adulta, el núcleo de la AE infantil no es el contenido o información, sino los esquemas cognitivos que el niño utiliza para representarse y pensar sobre el mundo económico. De este modo, la AE infantil es más un problema de razonamiento o pensamiento económico, es decir de estructuras cognitivas de procesamiento de la información económica, que de presencia o ausencia de conocimiento como tal. La segunda consecuencia es que, tal como se ha visto en estudios previos (Denegri, 1998) sujetos de distintas edades formularán explicaciones con niveles muy distintos de complejidad para un mismo problema, por lo que la variable evolutiva no puede soslayarse al evaluar la AE en niños.

Todo esto ocasiona, además, dificultades de orden metodológico cuando se intenta medir la AE en

niños. Primero, si en la AE infantil el contenido es menos importante que la estructura del pensamiento, la evaluación debe formularse como problemas a resolver que pongan a prueba estas estructuras, más que como una “prueba de conocimientos” tradicional. Segundo, si niños de distintas edades difieren sustancialmente en su manera de responder, la práctica tradicional aconsejaría varias pruebas diferentes, con distinto nivel de complejidad, para sujetos de diferentes segmentos de edad. Obviamente, esto acarrea costos y problemas de comparación intrasujeto en caso que deseen hacerse evaluaciones longitudinales.

El propósito de este artículo es presentar una aproximación empírica a la evaluación de la AE infantil que permite una solución simple pero potente a estos problemas. Básicamente, la estrategia propuesta consiste en el evaluar AE en niños mediante una prueba de elección múltiple, en que los ítems constituyen problemas a resolver vinculados a formas de procesamiento cognitivo de la información económica y, por lo mismo, las alternativas representan distintos grados de aproximación a la respuesta correcta. En la práctica esto significa valorar las respuestas parcialmente correctas, a través de *ítems de elección múltiple ordenada (Ordered Multiple-Choice ítems)*; Briggs, Alonzo, Schwab & Wilson, 2006). Este tipo de ítems entrega más información diagnóstica que los ítems de elección múltiple habituales pero conserva la eficiencia de aquellos en relación a otros formatos (respuesta abierta, por ejemplo). En esencia, se trata de ítems de elección múltiple en que cada alternativa está ligada a un nivel de desarrollo del constructo evaluado, cuya secuencia está avalada por investigaciones previas. Una característica única de este formato es que requiere simultáneamente de un modelo cognitivo que explique el desarrollo del constructo medido y de un modelo psicométrico capaz de analizar las respuestas considerando su naturaleza graduada. Esta última demanda excede con creces las posibilidades de la Teoría Clásica de los Test, lo que obliga a buscar un modelo psicométrico en el marco de la Teoría de Respuesta al Ítem.

En el caso de la medición de AE infantil las exigencias anteriores suponen: a) contar con un marco conceptual que describa el desarrollo del pensamiento económico infantil, a partir del cual establecer criterios para construir alternativas de respuesta y b) utilizar un modelo psicométrico apropiado para calificar el rendimiento parcial. En este trabajo se ha escogido como marco conceptual el Modelo de

Psicogénesis del Pensamiento Económico propuesto por Denegri (1995a, 1995b, 1998) y como modelo psicométrico la versión ordinal del Modelo de Rasch (1960), habitualmente conocida como Modelo de Crédito Parcial (Master, 1982).

El Modelo de Psicogénesis del Pensamiento Económico fue propuesto originalmente por Denegri (1995a) para describir la aparición, en niños y adolescentes, de una secuencia evolutiva de razonamiento sobre los eventos económicos. Ésta se caracteriza por la presencia de un patrón de cambio conceptual característico de una construcción progresiva, en la que aparecen nuevas explicaciones ante un mismo problema y desaparecen aquellas que no se relacionan con la nueva forma de conceptualizarlos. Este patrón culmina con la adquisición de un razonamiento sistémico para comprender la variabilidad de relaciones entre los eventos que constituyen un “hecho económico”.

En esta secuencia evolutiva se han podido identificar (Denegri, 1995b) tres niveles diferenciados de pensamiento económico que van desde una representación fragmentaria, con una concepción más bien preeconómica, hasta la elaboración de una compleja red de conceptualizaciones que le permiten al individuo representarse el mundo de la economía desde una perspectiva de sistemas en interacción. Sucesivos estudios cualitativos, realizados con muestras españolas, chilenas y colombianas, muestran invariablemente la presencia de los tres niveles de desarrollo e indican una tendencia secuencial asociada a la maduración cognitiva. Así, en el Nivel I se ubican mayoritariamente pre escolares y escolares hasta los 9-10 años, en el Nivel II los preadolescentes y en el Nivel III los adolescentes y adultos. Sin embargo, la distribución de sujetos al interior de cada nivel es heterogénea dependiendo de la escolaridad, nivel socioeconómico y oportuni-

dades de interacción con el sistema económico (Denegri & Delval, 2002). Por ello, el énfasis del modelo no está puesto en las edades de adquisición sino más bien en la secuencia psicogenética de aparición de las distintas conceptualizaciones. Un punto aún en discusión es la subdivisión del Nivel I en dos subniveles, debido a que se han observado dos tipos de conceptualizaciones con muy distinto grado de complejidad, que aparecen consecutivamente al interior de esa etapa (Denegri, 1998).

En el Cuadro 1 se describen, brevemente, las características definitorias de cada nivel. Una presentación más detallada puede encontrarse en Denegri (1995a, 1995b, 1998).

Por otro lado, el Modelo de Crédito Parcial [MCP] (Master, 1982) es un modelo psicométrico de la familia de modelos de Rasch (1960) diseñado explícitamente para analizar ítems que admiten respuestas graduadas en varios niveles de calidad y/o respuestas parcialmente correctas. Como en otros modelos desarrollados en el marco de la Teoría de Respuesta al Ítem, en el MCP se estima la *probabilidad* de emitir una respuesta de cada nivel de calidad de acuerdo a la ecuación:

$$(1) \quad P(x=c | \theta, \xi) = \frac{\exp \sum_{j=0}^c (\theta - \delta_j)}{\sum_{k=0}^m \exp \sum_{j=0}^k (\theta - \delta_j)}$$

Explicada simplemente, esta ecuación establece que la probabilidad de dar una respuesta de una calidad determinada ($X=c$) para un sujeto, dado su nivel de habilidad (θ) y el conjunto de dificultades asociadas a cada paso del ítem ($x=d_{11}, d_{12}, d_{13}$) corresponde a la diferencia entre la habilidad del sujeto y la dificultad del paso asociado a cada categoría, considerando todos los pasos del ítem en conjunto. Un ítem de c categorías tendrá un número de pasos $m=c-1$.

Tabla 1
Composición de la muestra definitiva

Nivel Socioeconómico						
Bajo		Medio		Alto		
Edad	Hombre	Mujer	Hombre	Mujer	Hombre	Mujer
10	35	30	21	23	31	23
11	35	42	61	59	32	35
12	38	54	47	39	35	26
13	33	44	39	46	35	35
14	10	13	17	21	28	34
15	2	8	4	0	0	0

Una descripción completa del MCP puede consultarse en Master (1982).

El MCP y los modelos de Rasch, en general, son ampliamente utilizados en psicometría debido a que poseen un buen número de propiedades atractivas para la construcción y análisis de pruebas. Una de ellas es que cuentan con estadísticos suficientes, lo que en términos prácticos se traduce en que, si los datos ajustan al modelo, la puntuación bruta total (la suma de respuestas correctas) contiene toda la información necesaria para estimar la habilidad de los sujetos. Esto no sucede en los modelos ajenos a la tradición de Rasch, en los cuales es necesario recurrir a métodos complejos de estimación para calcular la habilidad de los sujetos, lo que resulta problemático para tests convencionales. Otras características atractivas de los modelos Rasch son la posibilidad de producir medidas aditivas (con propiedades de intervalo y no sólo ordinales como las puntuaciones habituales de los tests) y específicamente objetivas (los parámetros de los ítems son independientes de la habilidad de los sujetos con los cuales se los estima). Éstas y otras propiedades del conjunto de modelos psicométricos de la familia Rasch han sido materia de extensas discusiones que exceden con creces el propósito de este trabajo. Una revisión clara y sintética, en castellano, puede consultarse en Prieto y Delgado (2003).

En el presente caso, la elección del MCP se enmarca en el creciente interés por utilizar modelos de Rasch para analizar tareas de evaluación construidas en un marco piagetano (ver, por ejemplo, Andrich & Styles, 1994; Bond, 1994; Bond & Bunting, 1995; Dawson, 2000; Dawson, Commons, Wilson & Fischer, 2005; Müller, Sokol & Overton, 1999; Sijtsma & Verweij, 1999; Wilson, 1989). De hecho, Briggs et al. (2006) utilizan una versión generalizada del MCP (Wilson, 1992) para el análisis de ítems de elección múltiple ordenados, como los empleados en la presente investigación. Aunque la TRI ofrece otros modelos para evaluar respuestas graduadas de manera confirmatoria (e.g. Modelo de Respuesta Graduada de Samejima, Modelo de Crédito Parcial Generalizado) o exploratoria (e.g. Modelo de Respuesta Nominal), el MCP ofrece la ventaja, inherente a todo Modelo de Rasch, de contar con estadísticos suficientes, que garantizan mejores índices de ajuste aún con muestras pequeñas y la posibilidad de utilizar la puntuación total del test como estimador de la puntuación verdadera.

En suma, la presente investigación se propuso explorar empíricamente la factibilidad de evaluar AE en niños mediante ítems de elección múltiple ordenada (Briggs et al., 2006) construidos a partir de un modelo teórico sobre el desarrollo del pensamiento económico (Denegri, 1995b). Para ello se utilizó un modelo psicométrico para ítems graduados (Modelo de Crédito Parcial; Master, 1982) con los objetivos específicos de: a) identificar un conjunto de ítems que funcionaran adecuadamente de acuerdo al modelo; b) examinar el funcionamiento empírico de las alternativas graduadas según el modelo teórico; c) proponer, si fuera necesario, un sistema de puntuación de alternativas más discriminativo empíricamente; y d) evaluar la calidad psicométrica de la prueba resultante.

Método

Participantes

Para la calibración de los ítems se trabajó con una muestra no probabilística de 1035 escolares, planificada originalmente en un diseño de cuotas independientes 3x2x5, controlando las variables nivel socioeconómico (alto, medio y bajo), género (hombre y mujer) y grupo de edad (10, 11, 12, 13 y 14). El nivel socioeconómico no se midió individualmente, sino por establecimiento, utilizando conjuntamente el tipo de dependencia administrativa del centro educativo (municipal, particular subvencionado o particular pagado) y el método ESOMAR. La edad fue medida en meses, calculándola a partir de la diferencia entre la fecha de aplicación y la fecha de nacimiento de cada evaluado. La muestra definitiva estuvo compuesta por 503 varones (48.6%) y 532 niñas (51.4%), cuyas edades fluctuaron entre los 120 y los 189 meses (10 y 15.75 años, respectivamente), con una media $M = 148.48$ meses ($ds = 15.24$), equivalente a $M = 12.37$ años ($ds = 1.27$). La composición definitiva de la muestra se presenta en la Tabla 1. Allí puede observarse que finalmente se incluyeron 14 casos fuera de rango respecto a edad (mayores de 180 meses, es decir 15 años), principalmente de nivel socioeconómico bajo. El análisis posterior de los resultados demostró que estos casos (un 1.3% de la muestra) tienen un patrón de respuestas consistente con los sujetos de 14 años, por lo cual no fueron excluidos de los análisis.

Respecto a la justificación del rango de edad definido para la muestra, es legítimo preguntarse si la utilización de sujetos de hasta 14 años pudiese limitar el análisis de los datos y la interpretación de los resultados, por la escasez de casos correspondientes al Nivel III del Modelo de Psicogénesis del Pensamiento Económico. Es importante insistir en que el énfasis del modelo está puesto en la secuencia de aparición de los distintos conceptos y no prescribe rangos de edad asociados a cada nivel (Denegri & Delval, 2002). Además, los resultados obtenidos en el estudio, que se expondrán más adelante, indican que se presentaron suficientes casos en el Nivel III, sin observarse anomalías en los resultados.

Instrumentos

Se comenzó elaborando un conjunto inicial de 100 ítems, cuyo contenido se basó en las transcripciones de entrevistas clínicas piagetanas realizadas a sujetos de 6 a 18 años de edad, en el marco del Proyecto Fondecyt 1970364, “*El desarrollo de conceptos económicos en la infancia. Estudio evolutivo con niños y adolescentes chilenos*” (Denegri, Delval, Palavecinos, Keller & Gempp, 2000).

A partir de las entrevistas se confeccionaron ítems de elección múltiple con cuatro alternativas ordenadas según su aproximación progresiva a la respuesta correcta. En la práctica, cada alternativa representa un nivel distinto de desarrollo del pensamiento económico, considerando una alternativa para los niveles Ia, Ib, II y III. Respecto al contenido, se balancearon los temas referidos a microeconomía y a macroeconomía.

Por ejemplo, el ítem 6, correspondiente al área de microeconomía, fue redactado como sigue:

En una tienda ubicada en el centro de la ciudad una camisa cuesta 15 mil pesos y en una tienda ubicada lejos del centro, la misma camisa cuesta 10 mil pesos. ¿Por qué ocurre esta diferencia de precio si es la misma camisa?

- a) Porque el dueño de la tienda del centro trabaja más, porque va más gente, entonces debe cobrar más.
- b) Porque la tienda del centro está mejor ubicada, por lo tanto se paga más por la ubicación y es esto lo que incrementa el precio de la camisa.
- c) La camisa del centro es mejor, por eso es más cara.
- d) Porque el dueño de la tienda del centro quiere cobrar más caro.

En este ejemplo, la alternativa c) corresponde al subnivel Ia (explicación simple, con atribución de una característica concreta, sin reparar en que se trata de la misma camisa), la alternativa d) corresponde al nivel Ib (aparece la voluntad de los agentes como un elemento determinante), la alternativa a) corresponde al nivel II (el valor está asociado al trabajo; hay un elemento de “justicia” en la respuesta) y, por último, la alternativa b) corresponde al nivel III (hay una análisis más complejo del mercado). En todos los ítems las alternativas fueron elaboradas a partir de respuestas literales consignadas en transcripciones de las entrevistas realizadas a 500 niños en el estudio Fondecyt 1970364, antes referido (Denegri et al., 2000).

Luego de varios análisis cuantitativos y cualitativos se redujo el conjunto inicial de 100 ítems y se seleccionaron 57 ítems para aplicarse en la muestra de calibración. Los detalles pueden consultarse en Cortés, Quesada y Sepúlveda (2004).

Procedimiento

La aplicación del instrumento se llevó a cabo en los establecimientos educacionales, en las salas de clases respectivas. La aplicación fue colectiva, anónima y voluntaria, registrándose el género, edad y establecimiento. Después de concluida la aplicación se obsequiaron caramelos a los sujetos, sin que esto implicase un incentivo para que aceptaran participar. El equipo que administró el test se encargó de resolver dudas de comprensión lectora. Otros pormenores del procedimiento pueden consultarse en Cortés et al. (2004).

Resultados

Para alcanzar el primer objetivo específico el estudio, se procedió en dos fases; una exploratoria y otra de calibración.

La exploración de los ítems comenzó revisando la proporción de respuestas omitidas y calculando el índice clásico de dificultad para cada ítem. A partir de los resultados se eliminaron seis ítems y los restantes fueron sometidos a sucesivos análisis factoriales exploratorios con el propósito de aislar un conjunto unidimensional. Los análisis se llevaron a cabo con el método de mínimos cuadrados no ponderados y se eliminaron progresivamente aquellos ítems que mostraron una carga factorial inferior a .30 en el primer factor sin rotar y/o una comunalidad muy baja. Una vez seleccionado un conjunto inicial de ítems, se procedió a la etapa de calibración.

En la etapa de calibración se aplicó el MCP a los ítems puntuados en cuatro categorías: 0 puntos para la alternativa de nivel Ia, 1 punto para la alternativa nivel Ib, 2 puntos para la alternativa nivel II y 3 puntos para la alternativa nivel III. Se estimaron los parámetros de los ítems por el método de Máxima Verosimilitud (ML) utilizando integración de Monte Carlo, con 1000 nodos de cuadratura y logits mínimos y máximos de -6 y 6, respectivamente. Se revisó el ajuste del modelo considerando sucesivamente los siguientes criterios:

1. Número de sujetos que respondieron a cada alternativa superior a 15 casos.
2. Índices de ajuste *infit/outfit* para cada ítem entre 0.8 y 1.2.
3. Índice de ajuste Q para cada ítem cercano a 0 e inferior a 0.5.
4. Índices de ajuste *infit/outfit* para cada alternativa entre 0.6 y 1.4.

A partir de los resultados, se eliminaron en forma iterativa aquellos ítems que no ajustaron al modelo, recalibrando los ítems restantes. En cada proceso de estimación, se introdujo el requerimiento adicional de observar un decremento sustantivo en los índices de información (BIC, AIC y CAIC) asociados al ajuste global del modelo.

De esta manera se logró aislar un conjunto de 22 ítems que ajustaron satisfactoriamente al modelo. De éstos, el 50% evalúa contenidos de microeconomía y los restantes de macroeconomía.

A continuación, y de acuerdo al segundo objetivo específico de la investigación, se utilizaron los resultados obtenidos en la calibración de estos 22 ítems para estudiar el comportamiento de las alter-

nativas y decidir si estaban funcionando apropiadamente de acuerdo al modelo teórico subyacente que, como se recordará, propone una secuencia ordinal en las alternativas. Dado que los ítems ajustan al modelo psicométrico (por tanto, son capaces de medir en forma confiable la variable latente subyacente), esta revisión cumplió con la finalidad de corroborar si los ítems ajustaban, además, al modelo teórico con el cual se generaron los ítems.

Se analizaron dos aspectos del comportamiento de las alternativas. En primer lugar, se revisó si cada alternativa consecutivamente graduada representaba *más* nivel de la variable latente. En la práctica, este análisis significa comparar, ítem a ítem, el nivel de habilidad estimado promedio para los sujetos que eligieron cada alternativa. Si la habilidad promedio se incrementa con las alternativas asociadas a los niveles más desarrollados de pensamiento económico, se puede concluir que las opciones de los ítems están correctamente ordenadas. En segundo lugar, se analizó si cada alternativa estaba asociada una zona de habilidad en la cual fuera más probable, lo que equivale a examinar si los pasos de cada ítem se encuentran empíricamente ordenados.

La revisión de ambos tipos de indicadores mostró un resultado similar en todos los ítems: las categorías de respuesta se encontraban ordenadas, pero los pasos de los ítems estaban desordenados, indicando que había niveles de habilidad en los cuales era improbable que algunas alternativas fueran utilizadas. En todos los ítems este problema afectó a las alternativas correspondientes a los niveles Ib y II propuestos por el modelo de desarrollo del pensamiento económico (puntuaciones 1 y 2). A como de ejemplo, la Tabla 2 exhibe los resultados obtenidos por el ítem 6, cuyas alternativas fueron descritas previamente en el apartado *Instrumentos*.

Los resultados del ítem 6 son prototípicos del comportamiento de las alternativas de los 22 ítems analizados. Puede observarse que la media de habilidad de los sujetos que escogen cada alternativa se incrementa en la medida que se elige una respuesta

correspondiente a un nivel más desarrollado de pensamiento económico; sin embargo, los pasos de los ítems se encuentran desordenados. En este ítem, un sujeto que supere el paso 1 (i.e. tenga más de 0.08 logits de habilidad) tiene una alta probabilidad de elegir cualquiera de las alternativas superiores y no necesariamente la alternativa Ib.

Lo anterior puede comprenderse más claramente al examinar las Curvas Características de las alternativas del ítem 6, presentadas en la Figura 1. Allí puede observarse que el desorden de los pasos del ítem implica que no hay una zona de habilidad para la cual sean más probables las respuestas correspondientes a los niveles Ib y II (puntuaciones 1 y 2, respectivamente). La conclusión es evidente: aunque las alternativas estén empíricamente ordenadas el ítem funciona, en la práctica, en forma dicotómica, diferenciando sólo entre las alternativas de nivel I y de nivel III. Las alternativas de nivel Ib y II son, por lo tanto, insuficientemente discriminativas. Un patrón de resultados similar se observó en todos los ítems.

Atendiendo a que el tercer objetivo específico del estudio era proponer, si fuera necesario, un sistema de puntuación de alternativas que optimizara la capacidad discriminativa de los ítems, se sometieron a prueba dos métodos alternativos para colapsar categorías. El primero, teóricamente guiado, consistió en fusionar las alternativas correspondientes a los subniveles del nivel I. En otras palabras, se asignó 0 puntos a las alternativas de nivel Ia y Ib, 1 punto a la alternativa de nivel II y 2 puntos a la alternativa de nivel III. Esta opción resultaba razonable en la medida que los niveles Ia y Ib corresponden, según el modelo teórico subyacente, a subetapas de un mismo estadio. Simultáneamente se contrastó un método guiado empíricamente, es decir, un método cuyo ajuste podía anticiparse como satisfactorio a partir de los resultados ya obtenidos. Específicamente, este método colapsó la alternativa del nivel Ib con la alternativa del nivel II, asignando puntuaciones de la siguiente manera: 0 puntos para la alternativa de nivel Ia, 1 punto para las alternativas de nivel Ib y nivel II, y 2 puntos para la alternativa del nivel

Tabla 2

Comportamiento de las alternativas para el ítem 6 puntuado con cuatro niveles de calidad

Nivel	<i>f</i>	%	Media	<i>SD</i>	Paso	S.E Paso
Ia	156	15.07	0.00	0.20	—	—
Ib	166	16.04	0.20	0.21	0.08	0.06
II	227	21.93	0.39	0.22	-0.05	0.10
III	486	46.96	0.52	0.24	-0.37	0.12

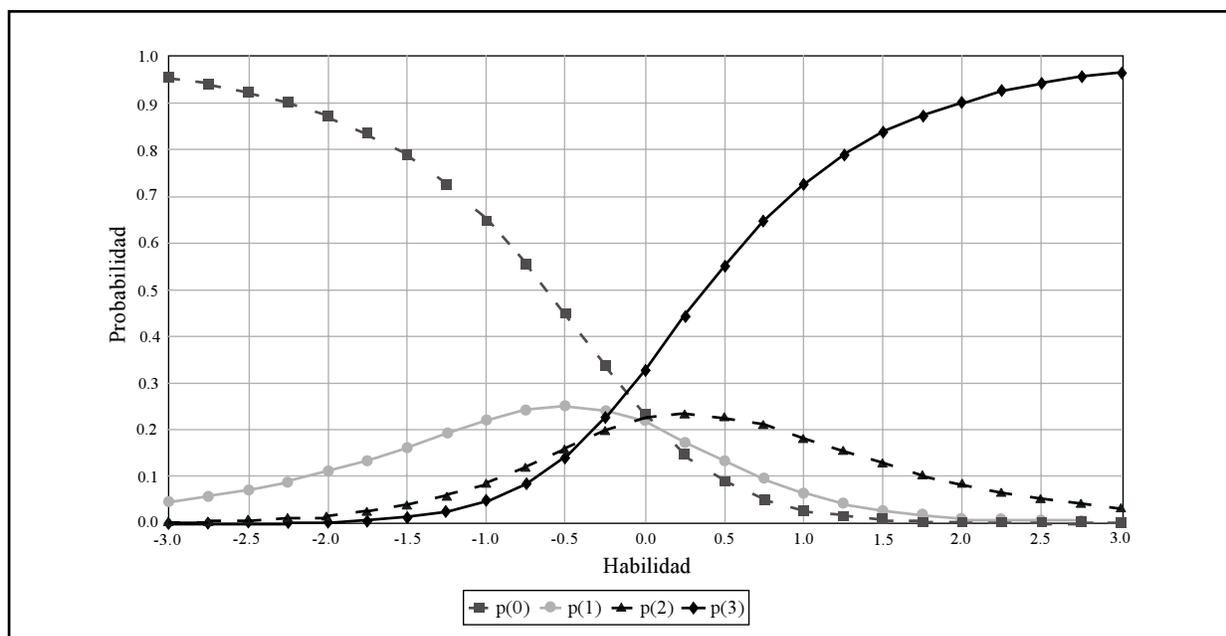


Figura 1. Curvas características de las alternativas del Ítem 6, considerando cuatro niveles de calidad.

III. Como puede verse, en ambos métodos de puntuación la alternativa III recibe 2 puntos y sólo cambia la distribución de puntaje para las alternativas inferiores.

La comparación de métodos significó recalibrar los 22 ítems con ambas modalidades de puntuación y comparar los resultados. El análisis se realizó con las mismas especificaciones técnicas utilizadas en la calibración original y se emplearon idénticos criterios para evaluar el ajuste de las soluciones. Además, se introdujeron tres consideraciones adicionales para comparar los resultados obtenidos con ambos métodos de fusión de alternativas. En primer lugar, se revisó cada solución para comprobar que los promedios de dificultad de cada alternativa resultaran ordenados consecutivamente. En segundo lugar, se examinó si los pasos de cada ítem aparecían graduados secuencialmente. Por último se compararon los índices de información (BIC, AIC y CAIC) correspondientes a los ajustes globales de ambos modelos, los cuales son presentados en la Tabla 3.

La consideración conjunta de estos tres criterios indicó que la mejor solución correspondía al método de puntuación guiado empíricamente, es decir, otorgar 0 puntos a la alternativa Ia, 1 punto a las alternativas Ib y II, y 2 puntos a la alternativa III. Tal como ocurrió con el método de cuatro puntos, el método de puntuación guiado teóricamente no logró satisfacer los criterios de ajuste recién comentados.

Con fines ilustrativos, la Figura 2 muestra las Curvas Características de las alternativas del ítem 6, una vez colapsadas las alternativas. La comparación de estas funciones de respuesta con las originales, exhibidas en la Figura 1, muestra cómo al fusionar las alternativas de nivel Ib y II el ítem presenta un comportamiento psicométrico más parsimonioso. Tal como antes, los resultados del ítem 6 son representativos del comportamiento observado en las categorías de los restantes 21 ítems. A modo de resumen, la Tablas 4 y 5 presentan los índices de ajuste y los estadísticos de cada ítem.

Tabla 3
Índices de ajuste global para el modelo original (4 alternativas), el modelo guiado teóricamente (fusión de alternativas Ia-Ib) y el modelo orientado empíricamente (fusión de alternativas Ib-II)

Modelo	Log L	Parámetros	BIC	AIC	CAIC
Original	-27969.04	67	56403.21	56072.08	56470.21
Teórico	-22858.22	45	46028.84	45806.45	46073.84
Empírico	-22034.33	45	44381.06	44158.67	44426.06

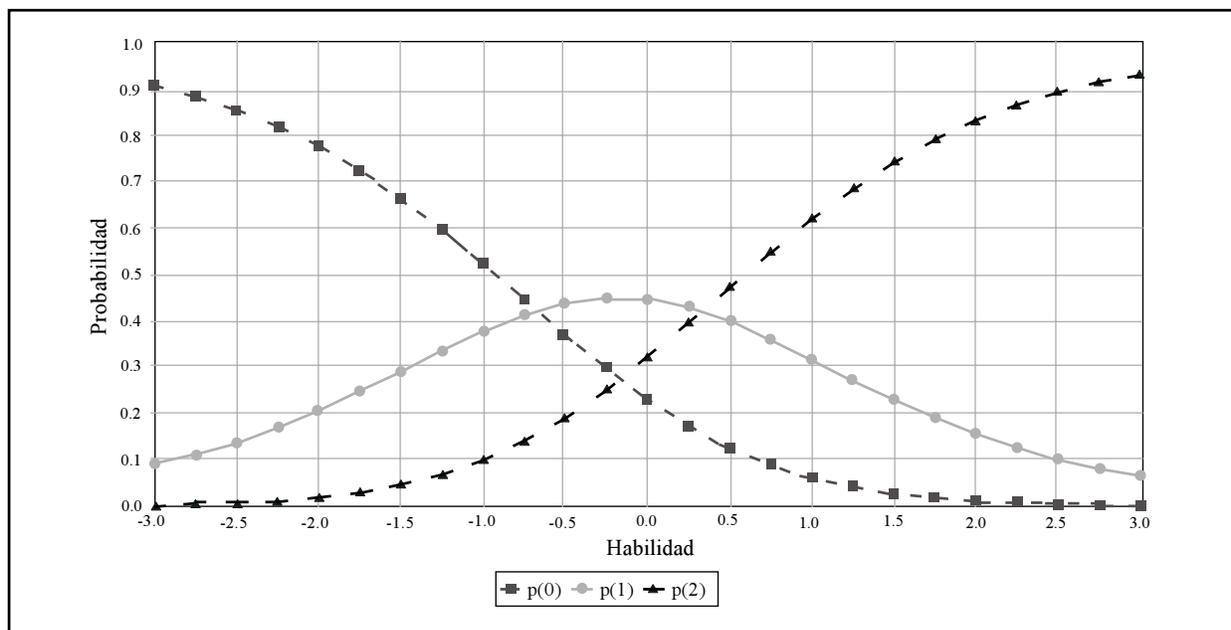


Figura 2. Curvas características de las alternativas del Ítem 6, considerando tres niveles de calidad (fusión de alternativas Ib-II).

Tabla 4

Índices de ajuste para los 22 ítems seleccionados y sus alternativas

Ítem	Ajuste global			Nivel Ia		Nivel Ib +II		Nivel III	
	<i>Infit</i>	<i>Outfit</i>	<i>Q</i>	<i>Infit</i>	<i>Outfit</i>	<i>Infit</i>	<i>Outfit</i>	<i>Infit</i>	<i>Outfit</i>
1	0.98	.97	0.25	0.96	0.96	1.07	1.10	0.99	0.99
2	1.02	1.02	0.38	1.03	1.03	1.29	1.35	1.18	1.18
3	1.03	1.05	0.22	1.01	1.08	0.86	0.79	0.93	0.95
4	1.03	1.03	0.31	0.94	0.94	1.21	1.26	1.13	1.12
5	1.04	1.04	0.25	1.03	1.04	0.94	0.88	0.99	0.99
6	1.03	1.03	0.26	0.92	0.88	1.22	1.43	1.03	1.03
7	1.02	1.02	0.24	0.97	0.94	1.03	1.02	0.96	0.97
8	1.01	1.00	0.30	1.08	1.08	0.96	0.92	1.06	1.06
9	1.02	1.02	0.24	0.97	0.96	1.14	1.18	0.98	0.97
10	0.99	0.99	0.16	0.93	0.89	0.76	0.60	0.85	0.89
11	1.09	1.08	0.24	1.12	1.14	0.81	0.72	0.95	0.96
12	0.99	0.99	0.25	1.03	1.01	1.00	1.01	1.00	1.00
13	0.94	0.94	0.24	1.05	1.05	0.80	0.72	0.92	0.93
14	0.94	0.94	0.22	0.95	0.94	0.97	0.94	0.95	0.95
15	1.04	1.04	0.24	1.11	1.13	0.76	0.70	0.94	0.97
16	1.05	1.04	0.28	1.14	1.18	0.98	0.97	1.04	1.03
17	0.98	0.96	0.19	1.00	0.94	0.74	0.60	0.89	0.92
18	0.95	0.94	0.29	1.07	1.07	1.03	1.01	1.08	1.07
19	1.02	1.01	0.22	0.97	0.94	0.88	0.79	0.94	0.95
20	0.92	0.92	0.30	0.94	0.94	1.22	1.36	1.12	1.11
21	0.96	0.96	0.24	0.97	0.94	0.95	0.90	0.97	0.98
22	1.00	1.00	0.27	1.06	1.09	0.88	0.81	1.00	1.00

Tabla 5

Porcentaje de respuestas por alternativa, promedio de habilidad por alternativa y parámetros de dificultad del ítem y sus pasos, para los 22 ítems seleccionados

Ítem	% de Respuestas			Media de Habilidad			Parámetro de Dificultad		
	Ia	Ib+II	III	Ia	Ib+II	III	Ítem	Paso1	Paso2
1	15.46	42.80	41.74	0.02	0.39	0.83	-0.84	-0.74	0.58
2	11.11	63.86	25.02	0.14	0.51	0.70	0.04	-1.45	1.53
3	17.39	31.98	50.63	0.06	0.28	0.82	-0.14	-0.35	0.07
4	15.46	58.45	26.09	0.03	0.53	0.79	0.20	-1.02	1.41
5	14.80	37.91	47.29	0.08	0.32	0.81	-0.19	-0.69	0.31
6	15.07	37.97	46.96	-0.04	0.45	0.75	-0.17	-0.67	0.32
7	11.50	41.93	46.57	-0.01	0.34	0.81	-0.31	-1.05	0.42
8	24.64	53.04	22.32	0.21	0.47	0.97	0.54	-0.42	1.50
9	23.48	35.46	41.06	0.07	0.42	0.86	0.15	-0.12	0.42
10	9.66	31.98	58.36	-0.09	0.15	0.82	-0.55	-0.99	-0.12
11	13.91	38.45	47.63	0.19	0.23	0.84	-0.23	-0.77	0.31
12	13.62	36.43	49.95	0.08	0.32	0.78	-0.27	-0.74	0.20
13	15.85	46.57	37.58	0.13	0.31	0.94	-0.01	-0.79	0.78
14	38.36	36.62	25.02	0.16	0.51	1.08	0.74	0.42	1.05
15	21.26	38.84	39.90	0.19	0.28	0.92	0.12	-0.31	0.54
16	21.26	34.69	44.06	0.20	0.33	0.81	0.06	-0.21	0.32
17	16.33	31.50	52.17	0.04	0.21	0.85	-0.20	-0.41	0.02
18	25.70	44.83	29.47	0.18	0.46	0.89	0.41	-0.22	1.04
19	16.52	38.55	44.93	0.03	0.31	0.87	-0.09	-0.58	0.39
20	19.61	53.72	26.67	0.05	0.54	0.81	0.32	-0.68	1.31
21	13.04	45.41	41.55	0.02	0.35	0.86	-0.17	-0.98	0.64
22	11.79	51.30	36.91	0.13	0.35	0.87	-0.15	-1.19	0.89

Considerando las buenas propiedades psicométricas de los 22 ítems seleccionados, es razonable analizar su comportamiento conjunto para utilizarlos como prueba de evaluación de AE en niños, tal como lo solicita el último objetivo específico del estudio.

Para ello se comenzó analizando la Función de Información de la prueba (homólogo de la fiabilidad) y el error estándar de medida asociado. Ambas funciones, graficadas en la Figura 3, indican que los ítems seleccionados arrojan medidas altamente fiables, sobre todo para aquellos sujetos con niveles medios de habilidad. Con fines comparativos, la fiabilidad calculada desde la perspectiva tradicional, con el método Alfa de Cronbach, es igual a 0.76 mientras que el error clásico de medida es de 3.12 puntos para todo el rango de habilidad evaluado.

La conclusión anterior es consistente, además, con la distribución de las dificultades de los ítems, presentada en la octava columna de la Tabla 5. Pue-

de observarse que el nivel de dificultad de los ítems se encuentra entre -0.84 y 0.74 logits, con una media de -0.03 logits. La misma conclusión puede obtenerse examinando el *Mapa de Wright* presentado en la Figura 4.

Por último, para analizar la relación entre las estimaciones de habilidad producidas por la prueba y las variables extrañas controladas en el diseño, se examinaron las diferencias en las medias de habilidad, en métrica logits, según género, edad y nivel socioeconómico.

En el caso del género, no se presentan diferencias significativas ($t = 0.52$; $p = 0.60$) entre la media de habilidad de las mujeres ($M = 0.51$; $sd = 0.69$) y la de los varones ($M = 0.49$; $sd = 0.69$), lo cual es coherente con los hallazgos anteriores sobre pensamiento económico en niños (e.g. Denegri, 1998).

Por otro lado, y también de acuerdo a lo anticipado por estudios previos (Denegri 1995a, 1995b), existen diferencias significativas en el nivel de AE esti-

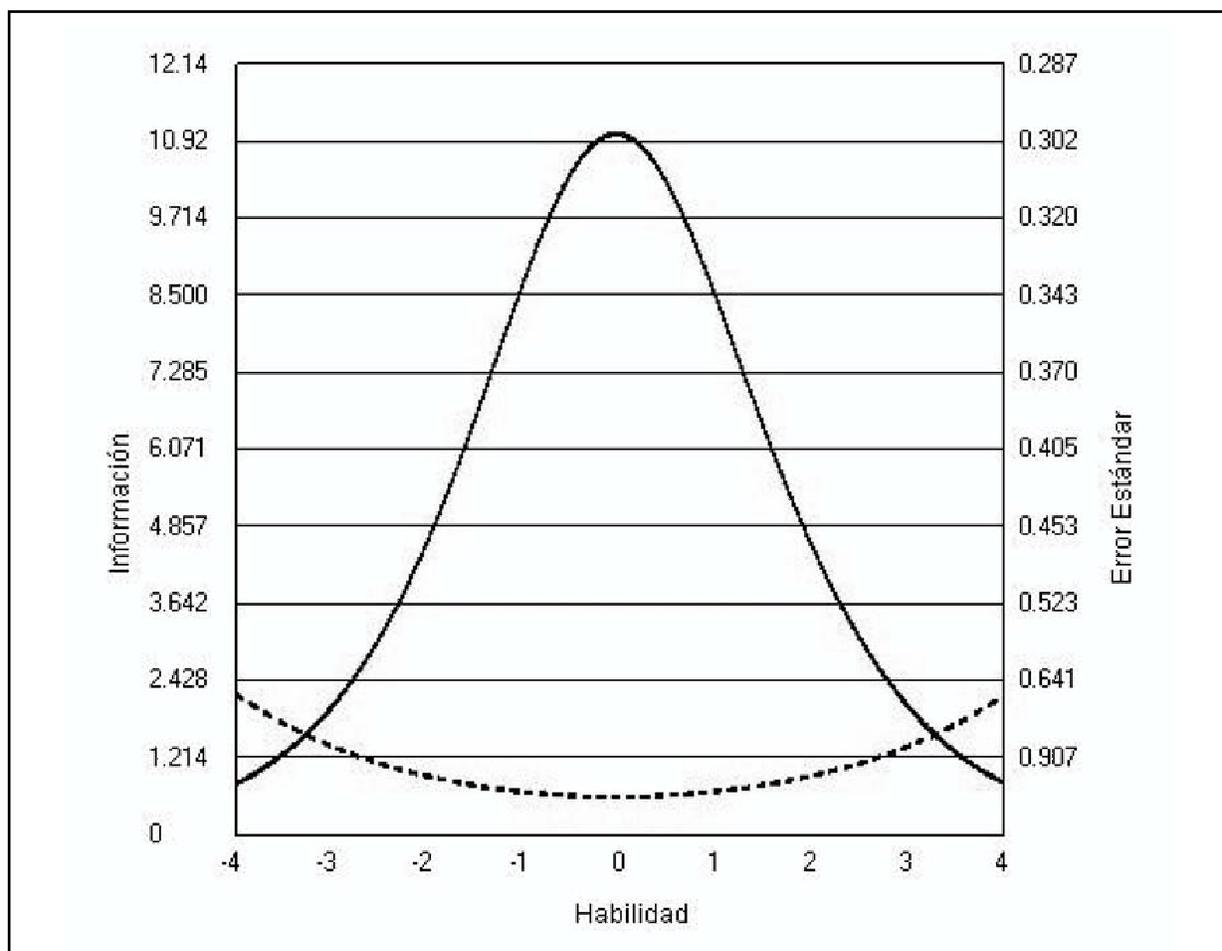


Figura 3. Función de información (línea continua) y error estándar de medida (línea segmentada) para distintos niveles de habilidad.

mado para niños de diferente grupo socioeconómico ($F = 32.67$; $p < 0.001$). Específicamente se observa que las medias de los niños de niveles socioeconómicos bajo, medio y alto son, respectivamente, de $M = 0.28$ ($sd = 0.62$), $M = 0.50$ ($sd = 0.66$) y $M = 0.72$ ($sd = 0.73$). Las diferencias a posteriori entre cada par de grupos también resultan significativas ($p < 0.001$) y la progresión es claramente lineal ($F = 65.32$; $p < 0.001$) indicando que, a mayor nivel socioeconómico, mayor nivel de AE ($r = 0.24$).

En tercer lugar, se comparó la media de habilidad obtenida por niños de los seis grupos de edad que finalmente integraron la muestra: 10, 11, 12, 13, 14 y 15 años. Los promedios observados corresponden, en orden sucesivo, a $M = 0.11$ ($sd = 0.51$), $M = 0.41$ ($sd = 0.58$), $M = 0.50$ ($sd = 0.68$), $M = 0.75$ ($sd = 0.74$), $M = 0.71$ ($sd = 0.76$) y $M = 0.62$ ($sd = 0.70$). La diferen-

cia global entre las medias es significativa ($F = 21.49$; $p < 0.001$) y muestra una asociación no lineal significativa ($h = 0.31$ versus $r = 0.28$; $F = 4.01$, $p < .003$) entre la edad y la habilidad estimada. Es interesante acotar que, pese al reducido número de grupos de edad utilizados en la investigación, la relación observada muestra una tendencia no lineal equivalente a la encontrada en otros estudios en que se analiza la asociación entre niveles de desarrollo y edad, en un marco de referencia piagetano (e.g. Dawson, 2002). Esta tendencia, graficada en la Figura 5, indica que el promedio de habilidad estimada (con intervalos de confianza de un 95%) se incrementa progresiva y significativamente desde los 10 hasta los 13 años ($p < 0.05$), momento en el que se produce un efecto techo a partir del cual las diferencias dejan de ser significativas ($p > 0.05$).

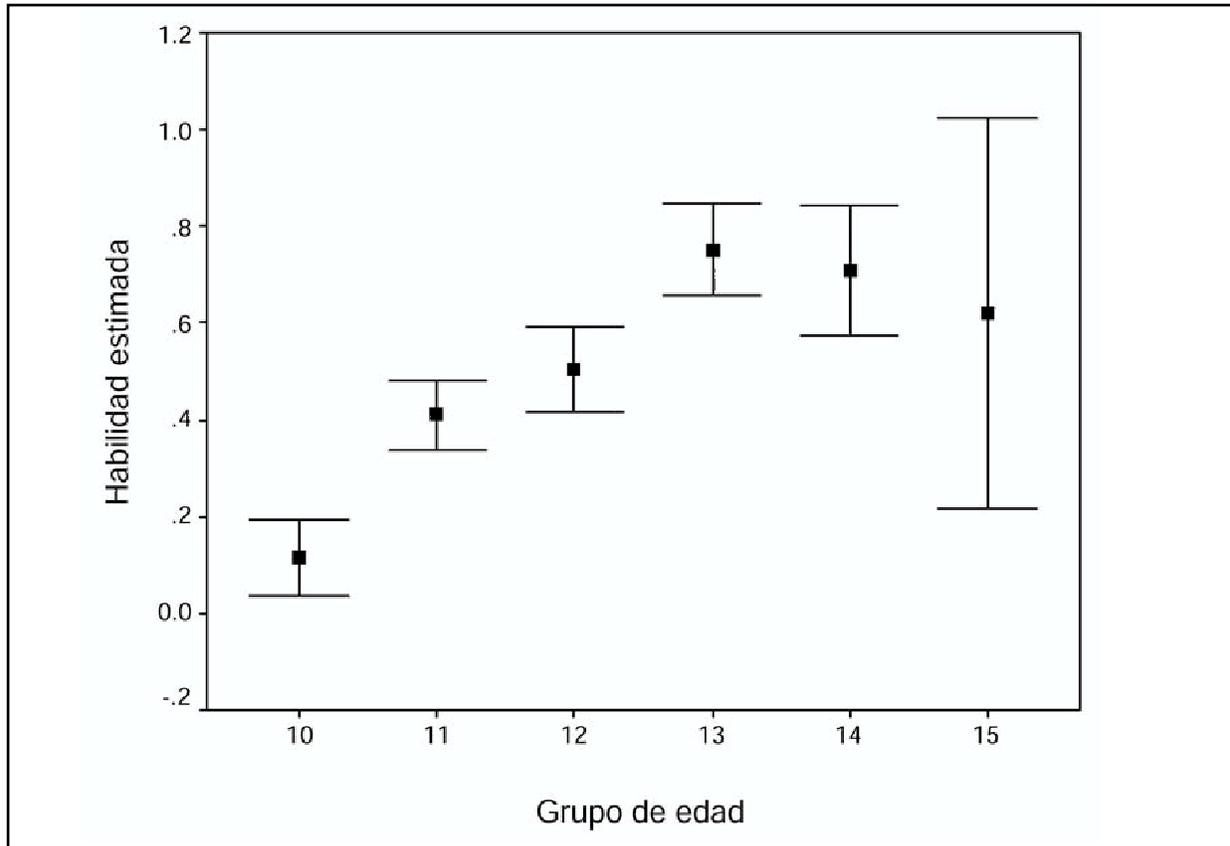


Figura 5. Promedios de habilidad estimada para niños de diferentes grupos de edad, considerando un intervalo de confianza del 95% para la estimación.

Discusión

Este trabajo se propuso como objetivo general evaluar empíricamente la viabilidad de medir AE en niños mediante ítems de elección múltiple ordenada, calibrados con el MCP. En retrospectiva, puede concluirse que el objetivo se cumplió ampliamente y que los resultados sugieren una respuesta positiva a la pregunta de investigación.

Los resultados ofrecen, cuando menos, tres líneas de discusión. En primer lugar, las ventajas asociadas al uso de ítems de elección múltiple ordenada. En segundo lugar, las oportunidades diagnósticas del MCP aplicado a ítems con respuestas graduadas según calidad. En tercer lugar, las consecuencias de los resultados para la teoría subyacente sobre el desarrollo del pensamiento económico. Por limitaciones de espacio, nos centraremos sólo en este último punto.

Al respecto, es interesante observar que los resultados aportan una corroboración, aunque indirecta y parcial, de la premisa básica de modelo: Es

posible observar una progresión empírica en la habilidad para comprender situaciones y resolver problemas asociados a la representación del mundo económico. Al menos tres tipos de resultados permiten formular esta interpretación. La primera evidencia es que se logró identificar un grupo de ítems en que las alternativas resultaron graduadas en la misma secuencia propuesta por el modelo (ver, por ejemplo, cuarta columna de la Tabla 2). Aunque por motivos de espacio se ha omitido la presentación detallada de esos resultados, análisis previos han demostrado que la secuencia se cumplía para todos los ítems calibrados con cuatro alternativas (Gempp, et al., 2005). El segundo argumento empírico es la correlación significativa que existe entre la habilidad estimada y la edad de los sujetos, lo cual es consistente con la idea de progresión evolutiva que propone el modelo. Además, y pese a la restricción de rango en la edad de los niños que participaron en el estudio, el resultado sugiere una relación funcional equivalente a la observada en otros estudios que utilizan modelos de Rasch para evaluar rendi-

miento en un marco piagetano (e.g. Dawson, 2002). El tercer resultado que justifica nuestra interpretación puede encontrarse en la distribución conjunta de sujetos e ítems, cuya comprensión resulta más simple revisando el *Mapa de Wright*. Una inspección cuidadosa de la gráfica muestra que el logro de los pasos de los ítems se encuentra claramente ordenado según el nivel de habilidad de los sujetos, corroborando nuevamente la idea de una progresión en la habilidad subyacente.

No obstante, también se observan otros resultados que parecen contradecir algunos componentes específicos de la teoría, elaborados a partir de datos obtenidos con investigaciones cualitativas (Denegri, 1995a, 1995b, 1998). Estos hallazgos representan, desde nuestro punto de vista, oportunidades de investigación futura con el modelo.

El primero de ellos es que la distribución de sujetos e ítems (ver *Mapa de Wright*) muestra una progresión intervalar de los pasos de los ítems a través de la habilidad latente. Cuando la variable latente está definida por categorías ordenadas, es esperable que los ítems se agrupen en clusters relativamente homogéneos a lo largo del continuo de habilidad, entre los cuales se observen “saltos” o zonas vacías, sin ítems (Wilson, 1989). Esto sugiere que la AE, medida por estos ítems, puede caracterizarse mejor como una variable continua que como una variable discreta definida por etapas. Sin embargo, más que evidencia contra del modelo este resultado sólo sugiere que los ítems seleccionados discriminan en un continuo de AE más que en una secuencia ordenada de etapas, lo cual simplemente indica una característica de la prueba elaborada y no necesariamente del modelo subyacente. Por otro lado, y desde un punto de vista estadístico, la alta variabilidad de las estimaciones de habilidad para cada categoría (ver, por ejemplo, la cuarta y quinta columna de la Tabla 2) indican que la distribución es heterogénea en las zonas de habilidad definidas por cada alternativa. Una sugerencia razonable, entonces, es explorar la presencia de subgrupos definidos por su perfil de respuesta en la variable latente utilizando, por ejemplo, modelos ordinales de Rasch que incorporan clases latentes (Rost, 1990). Otra posibilidad es utilizar modelos desarrollados específicamente para analizar discontinuidad en el desarrollo desde la perspectiva de la Teoría de Respuesta al Ítem (e.g. Wilson, 1989, 1992). Si el objetivo es validar el Modelo de Psicogénesis del Pensamiento Económico propiamente tal, ambas estrate-

gias podrían aplicarse, además, a los datos obtenidos con los estudios cualitativos previos.

Otro resultado relevante para el Modelo de Psicogénesis de Pensamiento Económico es la alta frecuencia de sujetos que logran alcanzar el nivel III. Los estudios cualitativos anteriores sugerían que las respuestas del nivel III son propias de adolescentes y adultos. Por ello resulta contradictorio que en esta muestra, que incluyó sólo 137 sujetos en el rango de 14 a 15 años, se observen entre un 22.32% a un 58.36% de respuestas de nivel III, para los ítems 8 y 10, respectivamente (ver columna 4 de la Tabla 5). Este resultado amerita tres consideraciones. Una, es que los estudios cualitativos previos utilizaron el Método Clínico Piagetano, que solicita a los niños argumentos para sus respuestas y permite al investigador contra preguntar. Por ello, ese método de evaluación es típicamente más exigente que las pruebas de papel y lápiz en cuanto a demandas cognitivas para el respondiente y más estricto para calificar el nivel de las respuestas. Es posible que los ítems escritos, con presentación visual de las alternativas de respuesta, resultaran más fáciles que su contraparte cualitativa, con respuesta construida, y ello permitiera que niños de menor edad alcanzaran más respuestas de nivel III. Otra explicación es que los ítems fueron contruidos para medir AE a partir del modelo, pero no son una prueba del modelo como tal. Éste propone una serie de relaciones complejas entre las explicaciones que los niños utilizan para construir estructuras de razonamiento económico. En cambio, los ítems fueron diseñados para evaluar conocimientos y habilidades de resolución de problemas económicos utilizando el modelo como referente, así que no abordan toda la serie de estructuras cognitivas que propone la teoría. Por último, no debe olvidarse que, tal como se comentó en la introducción de este trabajo, el modelo no prescribe edades asociadas a cada etapa sino que enfatiza la secuencia psicogenética de aparición de los distintos tipos de conceptualización. Por estas razones es conveniente hacer otros estudios, con sujetos en un rango de edad más extenso, antes de adelantar conclusiones. Una sugerencia alternativa es reexaminar los datos de los estudios originales (e.g. Denegri, 1995a, 1995b) utilizando modelos de TRI, para contar con una triangulación de método respecto a los análisis cualitativos.

El último resultado llamativo fue la necesidad de colapsar las alternativas correspondientes a los niveles Ib y II. Este tipo de análisis es una de las prin-

cipales ventajas diagnósticas del MCP para analizar ítems graduados. En la práctica este resultado *no significa* que las alternativas no estén secuencialmente ordenadas sino que, aún cuando esas alternativas aparezcan progresivamente según habilidad, son respuestas poco probables en las respuestas empíricas. En otras palabras, se trata de alternativas que no discriminan de manera fiable entre sí. Para entender las implicaciones de este resultado para el modelo de desarrollo del pensamiento económico es preciso entender que, tal como puede inferirse a partir de la Ecuación (1), en el MCP la probabilidad de elección de cada alternativa depende de la dificultad de todos los pasos del ítem. Ello tiene una importante consecuencia para comprender el proceso de respuesta que postula el MCP: enfrentado a todas las alternativas el sujeto elige aquella más cercana a su nivel de habilidad. El MCP intenta entonces cuantificar la probabilidad de elección de cada alternativa y deducir así su atractivo para el sujeto que responde. En nuestros resultados los sujetos en el rango de habilidad correspondiente a los niveles Ib y II tenían más probabilidad de elegir inmediatamente la alternativa de nivel III. La conclusión es que las alternativas nivel Ib y nivel II son poco atractivas para los sujetos que se encuentran en esos niveles de habilidad, lo que aconseja colapsarlas. La consecuencia para el modelo es que el nivel Ib funciona mejor asociado al nivel II que asociado al nivel Ia, como postulaba hasta ahora el modelo. Esta conclusión es, hasta cierto punto, congruente con observaciones teóricas previas a propósito de que en el nivel Ib ya comienzan a aparecer en el niño indicios de pensamiento económico que podrían ser un precursor de éste, lo que define al nivel Ib como un periodo inestable y transicional hacia la consolidación del pensamiento económico que se logra en el nivel II (Denegri, 1998). No obstante, es preciso seguir investigando con miras a aclarar este punto, cuestión que representa un desafío para futuras investigaciones con la teoría.

Finalmente, un producto adicional de este trabajo es la obtención de la primera versión de un instrumento para evaluar AE en niños, compuesto por 22 ítems, al cual hemos denominado *Test de Alfabetización Económica para Niños* [TAE-N]. Los resultados psicométricos reportados en este artículo sugieren que el TAE-N muestra adecuados niveles de fiabilidad y validez (ver la discusión de Bond, 2003, a propósito de porqué el ajuste a modelos de Rasch puede considerarse un antecedente de validez de

constructo). Nuestra recomendación final, por tanto, es continuar investigando sus propiedades psicométricas.

Referencias

- Abramovitch, R., Freedman, J. & Pliner, P. (1991). Children and money: Getting an allowance, credit versus cash, and knowledge of pricing. *Journal of Economic Psychology*, 12, 27-46.
- Andrich, D. & Styles, I. (1994). Psychometric evidence of intellectual growth spurts in early adolescence. *Journal of Early Adolescence*, 14, 328-344.
- Berti, A. E. & Grivet, A. (1990). The development of economic reasoning in children from 8 to 13 years old: Price mechanism. *Contributi di Psicologia*, 3, 37-47.
- Bond, T. (1994). Piaget and measurement II: Empirical validation of the Piagetian model. *Archives de Psychologie*, 63, 155-185.
- Bond, T. (2003). Validity and assessment: A Rasch measurement perspective. *Metodología de las Ciencias del Comportamiento*, 5(2), 179-194.
- Bond, T. & Bunting, E. (1995). Piaget and measurement III: Reassessing the méthode clinique. *Archives de Psychologie*, 63, 231-255.
- Briggs, D. C., Alonzo, A. C., Schwab, C. & Wilson, M. (2006). Diagnostic assessment with ordered multiple-choice items. *Educational Assessment*, 11(1), 33-63.
- Buckles, S. & Melican, C. (2002). Assessment framework: 2006 National Assessment of Educational Progress in Economics. Washington: National Assessment Governing Board.
- Cortés, L., Quesada, M. & Sepúlveda, J. (2004). *Construcción de un Test de Alfabetización Económica para Niños (TAE-N)*. Tesis para optar al grado de Licenciatura en Psicología no publicada, Universidad de La Frontera, Temuco, Chile.
- Crick, B. (1999). The presuppositions of citizenship education. *Journal of Philosophy of Education*, 33(3), 337-352.
- Davies, I., Gregory, I. & Riley, S. C. (1999). *Good citizenship and educational provision*. Londres: Falmer Press.
- Davies, P., Howie, H., Mangan, J. & Shqiponja, T. (2002). Economic aspects of citizenship education: An investigation of students' understanding. *The Curriculum Journal*, 13(2), 201-223.
- Dawson, T. L. (2000). Moral reasoning and evaluative reasoning about the good life. *Journal of Applied Measurement*, 1, 346-371.
- Dawson, T. L. (2002). New tools, new insights: Kohlberg's moral judgement stages revisited. *International Journal of Behavioral Development*, 26(2), 154-166.
- Dawson, T. L., Commons, M., Wilson, M. & Fischer, K. W. (2005). The shape of development. *European Journal of Developmental Psychology*, 2(2), 163-195.
- Delval, J., Enesco, I. & Navarro, A. (1994) La construcción del conocimiento económico. En M. Rodrigo, *Contexto y desarrollo social* (pp.345-383). Madrid: Síntesis.
- Denegri, M. (1995a). *El desarrollo de las ideas acerca del origen y circulación del dinero: Un estudio evolutivo con niños y adolescentes*. Madrid: Ediciones Universidad Autónoma de Madrid.

- Denegri, M. (1995b). El desarrollo de las ideas acerca de la emisión monetaria en niños y adolescentes: Estudio exploratorio. *Revista del Instituto de Ciencias de la Educación, Enero-Abril*, 47-62.
- Denegri, M. (1998). La construcción de nociones económicas en la infancia y adolescencia. En J. Ferro & J. Amar (Eds.). *Desarrollo humano, perspectiva para el Siglo XXI* (pp.203-219). Colombia: Ediciones UNINORTE.
- Denegri, M. & Delval, J. (2002). Concepciones evolutivas acerca de la fabricación del dinero: Los tipos de respuesta. *Investigación en la Escuela*, 48, 55-70.
- Denegri, M., Delval, J., Palavecinos, M., Keller, A. & Gempp, R. (2000). *El desarrollo de conceptos económicos en la infancia: Estudio evolutivo con niños y adolescentes chilenos* (Proyecto Fondecyt 1970364). Santiago: Fondo Nacional de Ciencia y Tecnología.
- Denegri, M., Delval, J., Ripoll, M., Palavecinos, M. & Keller, A. (1998). Desarrollo del pensamiento económico en la infancia y adolescencia. *Boletín de Investigación Educativa*, 13, 291-308.
- Gempp, R., Denegri, M., Caprile, C., Cortés, L., Quezada, M. & Sepúlveda, J. (2005, julio). *Medición de la alfabetización económica en niños: Una aplicación del Modelo de Crédito Parcial*. Ponencia presentada en el Simposio "La comprensión del mundo económico como necesidad de adaptación", XXX Congreso Interamericano de Psicología, Buenos Aires, Argentina.
- Günter, F. & Furnham, A. (1998). *Children as consumer. A psychological analysis of the young people's market*. Londres: Routledge.
- Leiser, D., Sevon, G. & Levi, D. (1990). Children's economic socialization: Summarizing the cross-cultural comparison of ten countries. Special issue: Economic socialization. *Journal of Economic Psychology*, 11(4) 591-631.
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149-174.
- Müller, U., Sokol, B. & Overton, W. F. (1999). Developmental sequences in class reasoning and proportional reasoning. *Journal of Experimental Child Psychology*, 74, 69-106.
- Prieto, G. & Delgado, A. (2003). Análisis de un test mediante el modelo de Rasch. *Psicothema*, 15(1), 94-100.
- Rasch, G. (1960). *Probabilistic models for some intelligence and achievement tests*. Copenhagen: Danish Institute for Educational Research (Expanded Edition, 1980. Chicago: University of Chicago Press).
- Rost, J. (1990). Rasch models in latent classes: An integration of two approaches to item analysis. *Applied Psychological Measurement*, 14, 271-282.
- Sijtsma, K. & Verweij, A. (1999). Knowledge of solution strategies and IRT modeling of items for transitive reasoning. *Applied Psychological Measurement*, 23(1), 55-68.
- Sonuga-Barke, E. & Webley, P. (1993). *A child's saving: A study in the development of economic behavior*. NJ: Lawrence Erlbaum.
- Walstad, W. & Rebeck, K. (2001). Teacher and student economic understanding in transition economies. *Journal of Economic Education*, 58-67.
- Wilson, M. (1989). Saltus: A psychometric model of discontinuity in cognitive development. *Psychological Bulletin*, 105, 276-289.
- Wilson, M. (1992). The ordered partition model: An extension of the Partial Credit Model. *Applied Psychological Measurement*, 16, 309-325.
- Yamane, E. (1997). The meaning of Economic Education in Japanese Elementary and Secondary Education: An historical perspective. En *Proceeding of the Second Conference of the International Association for Children's Social and Economic Education* (pp. 101-104). Malmo: Edge Hill University College/IACSEE.