

Desarrollo del Test de Alfabetización Económica para Adultos

René Gempp^{1 2}

Universidad Alberto Hurtado, Santiago, Chile

Marianela Denegri

Nadia Caripán

Valentina Catalán

Solange Hermosilla

Universidad de La Frontera, Temuco, Chile

Cristina Caprile

Banco de Chile, Chillán, Chile

Compendio

Este trabajo presenta la elaboración del Test de Alfabetización Económica para Adultos, versión 25 Items (TAE-A-25), derivado de un instrumento homónimo de mayor longitud construido previamente por los autores. El TAE-A-25 es una prueba compuesta por ítems de elección múltiple, de 4 alternativas, y está inspirado en el Test of Economic Literacy (TEL) para adolescentes (Soper & Walstad, 1987). Sus ítems abarcan cuatro áreas de conocimientos básicos sobre economía y evalúan conceptos y habilidades procedimentales. Se presentan análisis realizados a partir de la Teoría Clásica de los Tests sobre las respuestas dadas al instrumento por una muestra de 840 adultos chilenos de ambos géneros, entre 18 y 50 años. Los análisis permitieron obtener un instrumento de 25 ítems, con adecuados niveles de dificultad ($.24 \leq p \leq .76$), discriminación ($.27 \leq p \leq .54$), y confiabilidad (GLB=.87), además de óptimo comportamiento de los distractores. Se discuten los resultados y se comentan las limitaciones del instrumento.

Palabras clave: Tests psicológicos; economía y psicología; adulto.

Development of the Test of Economic Literacy for Adults

Abstract

This paper presents the development of the Test of Economic Literacy for Adults, version of 25 items (TAE-A-25), derived from a test previously constructed by the authors. The TAE-A-25 is composed for multiple-choice items, with 4 options, and it's inspired in the Test of Economic Literacy (TEL) for adolescents (Soper & Walstad, 1987). Their items include four areas of basic knowledge on economy and evaluate conceptual understanding and procedural skills. The analyses were realized in the Classical Test Theory framework, based on a sample of 840 male and female Chileans, aged between 18 and 50 years. The analyses allowed to obtain a test of 25 items, with appropriate levels of difficulty ($24 \leq p \leq 76$), discrimination ($27 \leq p \leq 54$), and reliability (GLB=.87), in addition to optimal performance of the item options. The results are discussed and the limitations of the test commented.

Keywords: Psychological testing; psychology and economy; adults.

El rápido avance de la globalización requiere generar nuevos modelos de desarrollo económico que sean sustentables y que, a la vez, consideren como eje central el desarrollo humano. Superar la pobreza y la discriminación, reformar la educación para hacerla compatible con el desarrollo de las destrezas cognitivas, sociales y afectivas requeridas por nuevos escenarios económicos y políticos, e impulsar el desarrollo de la participación e integración social, son objetivos presentes en las agendas políticas de la mayoría de los países en desarrollo (Cumbre de las Américas, 1999). Sin embargo, la mayoría de las agendas

políticas contrastan fuertemente con las capacidades reales que tiene la población para poder asumir efectivamente los requerimientos crecientes de sociedades cada vez más complejas.

En el caso de Chile, durante las últimas décadas el país ha experimentado importantes cambios productivos, tecnológicos, institucionales y sociales, que han contribuido a generar un contexto sociopolítico y económico-cultural cambiante y complejo. Este escenario exige a los individuos nuevas competencias para participar eficazmente de los procesos de desarrollo y desenvolverse eficientemente en un mercado libre. En particular, resulta imprescindible que los ciudadanos cuenten con conocimientos básicos de economía, suficientes para comprender el funcionamiento del mercado y tomar así decisiones que maximicen sus beneficios. Desde este punto de vista, a la alfabetización tradicional, que posibilita a las personas participar en los sistemas de signos y símbolos culturales de una sociedad

¹ Dirección: Escuela de Psicología, Universidad Alberto Hurtado, Barroso 26, Santiago, Chile. *Email* rgempp@uahurtado.cl - rene.gempp@mineduc.cl, Web: <http://www.sigmas.cl/rgempp>

² Esta investigación ha sido financiada por el Fondo de Ciencia y Tecnología (FONDECYT) de Chile a través de los proyectos 1030271 y 1060303.

letrada, se agrega el requerimiento de una educación económica básica o *Alfabetización Económica* (AE) que entregue a los individuos herramientas para entender su mundo económico, interpretar los eventos que pueden afectarlos directa o indirectamente y optimizar sus decisiones personales y sociales sobre la multitud de problemas económicos que se encuentran en la vida cotidiana.

Este concepto, originado hace tres décadas (Stigler, 1970) delimita aquellos conocimientos económicos básicos que deberían incluirse en los currículum para formar ciudadanos informados y competentes. Más recientemente, Yamane (1997) ha definido la AE como un 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. En este sentido, una persona alfabetizada económicamente debería ser capaz, como mínimo, de: a) evaluar diferentes métodos de localización de bienes y servicios comparando sus costos y beneficios alternativos; b) identificar los incentivos económicos que afectan la conducta de la gente y explicar como estos incentivos pueden influir en su propia conducta; c) comprender la ley de la oferta y la demanda y como la interacción entre los intereses de compradores y vendedores, mediada por la competencia, puede afectar el valor de los bienes y servicios; d) describir el rol de las instituciones económicas públicas y privadas; e) comprender conceptos básicos tales IPC, tasas de interés, inflación, impuestos; y, f) identificar y evaluar beneficios económicos a mediano y largo plazo de distintas políticas públicas.

Un requisito básico para la investigación y diagnóstico de la AE es el desarrollo de medidas confiables y válidas del constructo. A la fecha, el instrumento que goza de mayor aceptación internacional es el TEL (*Test of Economic Literacy*; Soper & Walstad, 1987), desarrollado por el National Council of Economic Education, de Estados Unidos, como parte de su política de promoción de la AE en Norteamérica. Se trata de una prueba estandarizada para estudiantes secundarios, conformada por dos formas paralelas de 46 ítems de elección múltiple. Siguiendo los Standards del National Council of Economic Education, el TEL evalúa cuatro áreas de conocimientos económicos definidos como indispensables para un desenvolvimiento mínimamente funcional en una sociedad de libre mercado: a) *Conceptos Económicos Fundamentales*, que incluyen el conocimiento de los sistemas económicos y los conceptos de bien, servicio, recurso productivo, asignación de recursos y costo de oportunidad; b) *Conceptos Microeconómicos*, que comprenden temáticas como mercado, oferta y demanda, precio, costo y producción, bienes públicos, monopolio y subsidio, instituciones financieras y su funcionamiento; c) *Conceptos Macroeconómicos*, con materias tales como producto nacional, ingreso o renta nacional, consumo, inversión, gasto público y exportaciones netas, inflación, desempleo, ciclo económico, crecimiento económico, desarrollo económico, política

económica, oferta y emisión monetaria y usos del dinero; y, d) *Conceptos de Economía Internacional*, que abarcan los tópicos relativos a la importación y exportación, tipo de cambio y aranceles, entre otros.

Respecto a sus propiedades psicométricas, el TEL ha mostrado buenos indicadores de confiabilidad y validez. Respecto a la confiabilidad Soper y Walstad (1987) reportan coeficientes de consistencia interna alfa de Cronbach $\alpha=.87$ para la Forma A y $\alpha=.88$ para la Forma B. Al mismo tiempo, diversos tipos de evidencia avalan la validez del instrumento. Por ejemplo, en la muestra de estandarización hay una diferencia significativa en el rendimiento de alumnos con formación económica respecto de quienes no la tienen (Soper & Walstad, 1987).

En Chile, por otra parte, no existen instrumentos estandarizados para la evaluación de la AE, lo que ha obligado a los investigadores a utilizar estrategias cualitativas. Por ejemplo, Denegri y Palavecinos (2003) debieron recurrir a entrevistas semi estructuradas para explorar la comprensión del mundo económico de 60 hombres y mujeres, de clase media y media baja, con edades entre 25 a 45 años, que trabajaban en los sectores financiero, servicios y comercio. Sus resultados muestran que los participantes manifiestan una comprensión básica del carácter fiduciario del dinero y de la idea de ganancia, y un marcado apego a preceptos morales heterónomos centrados en la necesidad de un gobierno paternalista que regule y controle todo el sistema económico. La mayoría muestra dificultad para comprender las relaciones complejas entre emisión y circulación monetaria y los procesos productivos. Sólo un 11,7% del total de entrevistados, logra establecer una comprensión completa del ciclo de circulación del dinero y éstos son mayoritariamente los sujetos con estudios universitarios (Denegri & Palavecinos, 2003). Al mismo tiempo, los resultados evidencian que la escolaridad y con ello el nivel educacional, tienen un impacto directo sobre la calidad de las respuestas de los participantes y sobre el nivel de complejidad con que analizan el funcionamiento económico.

Este tipo de estudios, aunque teóricamente interesantes, están limitados por la falta de instrumentos estandarizados para medir AE. El uso de instrumentos psicométricos para evaluar AE ofrece evidentes ventajas tanto para la investigación como para la intervención en el área de la educación económica. Para los investigadores los instrumentos facilitan, entre otras cosas, el acceso a muestras más grandes, en condiciones rigurosamente objetivas y controladas, lo cual permite maximizar la validez interna y externa de los resultados, además de desarrollar y someter a prueba modelos explicativos complejos sobre la relación entre AE y socialización económica familiar, efectos de la formación escolar y actitudes y conductas de consumo, por ejemplo. Por su parte, en el terreno de la acción profesional, el uso de medidas estandarizadas de AE ofrece la posibilidad de efectuar mediciones diagnósticas o de evaluar el impacto de programas de intervención en educación económica, entre otras aplicaciones.

Considerando ambos tipos de necesidades – investigación e intervención profesional – nuestro equipo de trabajo se

propuso la meta de elaborar instrumentos psicométricos para medir AE en niños y en adultos chilenos, en el marco del Proyecto Fondecyt 1030271 (Denegri, Palavecinos, & Gempp, 2003, 2005) dirigido a estudiar las relaciones complejas entre el comportamiento económico de los padres, la socialización familiar y el comportamiento económico de los niños, en una ciudad multifinanciera. Como producto, se obtuvieron dos Tests de Alfabetización Económica, para Niños y para Adultos, denominados genéricamente TAE-N (Cortés, Quesada, & Sepúlveda, 2004) y TAE-A (Caripán, Catalán, & Hermosilla, 2004), cuyas versiones originales fueron ensambladas y calibradas a partir de la Teoría de Respuesta al Ítem (Modelo de Crédito Parcial y dicotómico de Rasch, respectivamente). Con posterioridad, y para facilitar su uso e interpretación con fines diagnósticos y de evaluación, se optó por desarrollar una versión mejorada del TAE-N (Gempp, Denegri, Caprile, Cortés, Quesada, & Sepúlveda, 2006) y una versión abreviada del TAE-A, compuesta por 25 ítems, y sustentada en la Teoría Clásica de los Tests (TCT). Como se recordará, el núcleo de esta aproximación teórica es el concepto de *fiabilidad* y el intento de diferenciar entre Puntuación Verdadera y Error presentes en una Puntuación Observada. Se decidió desarrollar una versión del TAE-A basada en la TCT, porque tradicionalmente ésta ha servido como fundamento técnico a la mayoría de los instrumentos psicométricos actualmente en uso en la comunidad psicológica, por lo que sus indicadores básicos (e.g. fiabilidad, dificultad del ítem, discriminación, error de medida, entre otros) son más familiares y fáciles de comprender para los usuarios de tests psicológicos. Una presentación sintética, en castellano, de los supuestos, fundamentos, limitaciones y modelos de la TCT puede encontrarse en la obra de Muñiz (2001).

El objetivo de este artículo es presentar los antecedentes psicométricos de la construcción de la versión de 25 ítems del Test de Alfabetización Económica para Adultos (TAE-A-25), elaborado en el período 2005-2006, en el marco de la TCT. Específicamente se presentan en este trabajo los análisis realizados para reducir el conjunto inicial de ítems, estimar la dificultad de las claves y el atractivo de los distractores, determinar la capacidad discriminativa de los ítems y sus alternativas, estimar la confiabilidad y errores estándar de medida de las puntuaciones totales, obtener evidencia preliminar sobre la validez de la prueba y construir normas provisionales en Puntuaciones T y estandares. Información más detallada sobre el desarrollo de la versión original del TAE-A puede consultarse en Caripán et al. (2004).

Método

Participantes

Se trabajó con una muestra no probabilística de 840 adultos, organizada en un diseño de cuotas independientes 3x4x2, controlando las variables nivel socioeconómico (alto, medio y bajo), grupo de edad (18-25, 26-30, 31-40 y 41-50) y género (femenino y masculino). La variable nivel socioeconómico (NSE) se midió individualmente mediante una adaptación de la Escala ESOMAR (Adimark, 2000). Los

grupos de edad fueron determinados de acuerdo a los criterios de Stacey (1987) y Webley (1999) mientras que el balanceo por género se basó en los hallazgos de Denegri y Palavecinos (2003), según los cuales las mujeres exhibían una comprensión más fragmentaria de los procesos económicos que los hombres, y por tanto una AE, más básica y menos elaborada, con numerosos vacíos conceptuales. Aplicando estos criterios, la muestra definitiva estuvo compuesta por un 50% de participantes de cada género, un 7.3% de individuos de nivel socioeconómico alto, un 9.4% de nivel bajo y 83.3% de clase media. Respecto a la edad, el 45.7% de los integrantes de la muestra pertenecía al rango entre 18 a 25 años, el 17.2% al segmento entre 26 a 30 años, el 23.1% al grupo entre 31 a 40 años y el 14% tenía entre 41 a 50 años. De acuerdo a los antecedentes revisados, se incluyó a participantes que trabajaban, que trabajaban y estudiaban o que sólo estudiaban.

Instrumento

Como ya se adelantó, el análisis que aquí se presenta tuvo el objetivo de obtener una versión abreviada del TAE-A original. Para ello se trabajó con el mismo conjunto 55 de ítems utilizado previamente para la elaboración de aquel instrumento. Estos habían sido seleccionados, a su vez, mediante sucesivos análisis cualitativos y cuantitativos llevados a cabo sobre un grupo inicial de 89 ítems construidos por especialistas en Psicología Económica y por Economistas.

El instrumento se compone de ítems de elección múltiple, con cuatro alternativas, de las cuales sólo una es correcta. Los 55 ítems incluidos en el análisis abarcan las mismas cuatro áreas que el TEL: conceptos de *Economía General*, *Microeconomía*, *Macroeconomía*, y *Economía Internacional*, en ese orden de relevancia. Una descripción breve de cada área fue presentada previamente. Un ejemplo de ítem es el siguiente: "Si el precio de la carne se duplica, y el precio del pollo se mantiene, las personas probablemente comprarán: a) Más pollo y menos carne. b) Más pollo. c) Más pollo y la misma cantidad de carne. d) Ni pollo, ni carne." En este ejemplo, la alternativa (a) corresponde a la respuesta correcta.

La prueba incluye ítems conceptuales e ítems de aplicación; los primeros dirigidos a evaluar el conocimiento de conceptos económicos, y los segundos, su comprensión en situaciones cotidianas. Por ejemplo, el siguiente es ítem de tipo conceptual: "¿Qué es un crédito? a) Una operación en que se piden préstamos a una casa comercial para su posterior pago; b) Utilización de los fondos de otra persona a cambio de la promesa de devolverlos más los intereses correspondientes; c) Concesión de un permiso dado por una persona a otra para obtener la posesión de algo; d) Un cambio en el que una de las partes entrega de inmediato un bien o servicio y recibe, más tarde, el pago correspondiente más los intereses devengados." En este ejemplo, la alternativa (d) corresponde a la respuesta correcta.

Por otra parte, el siguiente es un ejemplo de ítem de aplicación: "¿Cuál es el costo de oportunidad de comprar un Auto? a) Adquirir una Camioneta; b) Regalar el dinero; c)

Dejar el dinero bajo la almohada; d) El interés ganado de dejar el dinero en una cuenta de ahorro.”. En este ejemplo, la opción (d) corresponde a la alternativa correcta.

Procedimiento

Para seleccionar a los sujetos se tuvieron en cuenta los criterios señalados previamente al describir la muestra. Los participantes potenciales fueron contactados individualmente o en grupos pequeños, directamente en sus lugares de trabajo (empresas públicas y privadas) o de estudio (estudiantes universitarios). Allí se les informó de los objetivos de la investigación y se les invitó a participar. En general, se observó una alta tasa de colaboración. Cada sujeto firmó una carta de consentimiento informado, en que se garantizaba la confidencialidad de la información y el anonimato de las respuestas. La participación fue completamente voluntaria y los sujetos no recibieron ningún tipo de remuneración por su colaboración. La aplicación se desarrolló en el mismo sitio, previa autorización de los supervisores respectivos (Jefes o profesores según el tipo de participante), a cargo de alumnos de Tesis y ayudantes del Proyecto Fondecyt 1030271, quienes leyeron las instrucciones y respondieron las dudas que surgían. El tiempo promedio de respuesta fue alrededor de 30 o 40 minutos por persona. Además de la versión de 55 ítems del TAE-A, los participantes debieron responder una breve encuesta referida a sus hábitos de consumo y variables socioeconómicas.

Resultados

Reducción del conjunto inicial de ítems

Para reducir el conjunto inicial de ítems se utilizaron dos criterios. El criterio cuantitativo consistió en la aplicación de una serie de análisis factoriales exploratorios, mientras el cualitativo fue aportado por especialistas en educación económica y economía que tuvieron la misión de aprobar o rechazar la selección de ítems tomando en cuenta su contenido. Cabe advertir que, aunque resulta teóricamente correcto anticipar una estructura multidimensional para la AE, el objetivo de esta investigación era producir un instrumento que permitiera expresar en un índice simple el nivel de AE de un sujeto. Por ello los análisis estuvieron dirigidos a obtener un test unidimensional.

Los ítems fueron puntuados en forma binaria asignando 0 puntos a las alternativas incorrectas y 1 punto a la alternativa correcta. Se calculó el índice clásico de dificultad (proporción de respuestas correctas) obteniéndose valores en el rango $p=.80$ a $p=.13$. Considerando los criterios convencionales en este tipo de análisis, sólo 3 ítems resultaron demasiado difíciles y fueron, por tanto, eliminados. A continuación, se intentó reducir el conjunto de 52 ítems restantes utilizando análisis factorial exploratorio. En primer lugar, se construyó la matriz de correlaciones tetracóricas entre ítems y se calculó el índice de Kaiser-

Meyer-Olkin para evaluar la factorizabilidad de la matriz. Como el resultado fue aceptable ($KMO>.75$) se procedió factorizar sucesivamente por el método de mínimos cuadrados no ponderados (ULS), con rotación oblicua Promin. Luego de cada extracción se eliminaron progresivamente aquellos ítems que mostraron una comunalidad estimada muy baja y/o una saturación inferior a $|.30|$ en algún factor y/o que conformaban factores específicos (definidos como aquellos factores que agrupan a sólo 2 ítems, cuyas cargas en las restantes dimensiones son inferiores a $|.25|$). De esta manera fue posible aislar 25 ítems que fueron posteriormente aprobados por los especialistas en contenido.

Para garantizar la unidimensionalidad de la prueba resultante se realizaron varios análisis adicionales sobre la matriz de correlaciones tetracóricas entre estos 25 ítems, la que previamente fue examinada para evaluar su factorizabilidad. Los resultados ($Determinante<.001$; $KMO=.89$) indicaron que era viable aplicar los análisis.

En primer lugar, se utilizó la técnica MAP (Velicer, 1976) para decidir el número de dimensiones subyacentes. Este procedimiento, considerado uno de los más robustos para estimar el número de factores a retener (e.g. Gorsuch, 1983) indicó que un factor era suficiente para explicar la covarianza entre ítems (los promedios parciales para siete dimensiones fueron, respectivamente, de .0099, .0111, .0187, .0348, .0723, .2326 y .9999). En forma adicional, el examen del Gráfico de Cattell y la razón entre el primer y el segundo autovalor de la matriz ($7.24/1.76 = 4.1$) sugieren la misma conclusión. Finalmente, al extraer sólo un factor mediante el método de mínimos cuadrados no ponderados (ULS), se observa que la razón entre el primer y segundo autovalor de la matriz reducida es $6.53/0.99 = 6.59$, mientras los residuales muestran una media, varianza y Raíz Cuadrática Media de $M=0.05$, $s^2=0.001$ y $RMSR=0.06$, respectivamente. En conjunto, estos resultados pueden considerarse suficiente evidencia de unidimensionalidad para fines de un instrumento psicométrico. Las cargas factoriales de los 25 ítems en esta solución unidimensional exploratoria se presentan en la segunda columna de la Tabla 1.

Estimadores de Dificultad y Atractivo de los Distractores

Se estimó la proporción de elección para cada distractor y para las claves que indican, respectivamente, el atractivo de los distractores y la dificultad del ítem. La consideración conjunta de ambos tipo de indicadores entrega una visión más completa del funcionamiento de la prueba en la medida que permite observar el rango de habilidad en torno al cual se concentran los ítems y el valor de los distractores como alternativas plausibles a la respuesta correcta. Idealmente, en cada ítem los distintos distractores deberían mostrar proporciones de elección heterogéneas pero en torno a la proporción de elección de la clave. El resultado se presenta en la Tabla 2. Puede observarse que los niveles de dificultad se distribuyen en el rango entre $p=.24$ y $p=.76$.

Es importante advertir que un revisor anónimo de la primera versión de este artículo planteó su discrepancia

respecto al criterio empleado, comentando que el rango de valores de dificultad de los ítems implicada una “fuerte presencia de distractores”. Este comentario encierra un malentendido habitual: que deben seleccionarse únicamente ítems con valores medios de dificultad. En realidad, es preferible seleccionar ítems en todo el rango de dificultad de la escala y no sólo en los valores medios de dificultad (Wright & Stone, 1979), a no ser que se pretenda utilizar explícitamente un modelo psicométrico que exija lo contrario (e.g. el Modelo de Error Binomial). Por otro lado, la relación entre la proporción de elección de la clave y de los distractores no debe juzgarse sólo a partir de los porcentajes medios de elección, sino que considerando todo el rango de puntuación total en la prueba. Para contar con este último criterio, se graficaron *Curvas Empíricas de Respuesta* por alternativa, para cada ítem, como la presentada en la Figura 1 para el ítem 8.

En estas curvas se puede observar la proporción de sujetos que eligen la alternativa correcta o algún distractor (eje vertical), dependiendo de su puntuación total en el test (eje horizontal). Para construir las gráficas, la puntuación total se dividió en 10 segmentos iguales (deciles) y para cada uno de ellos se calculó

la proporción de elección de cada alternativa. Las curvas resultantes fueron suavizadas utilizando una técnica de regresión kernel, según el método propuesto por Ramsay (1997) como aproximación no paramétrica a la construcción de *Curvas Empíricas de Respuesta*. Tal como se esperaba teóricamente para un ítem con funcionamiento adecuado, en todos los ítems se observó que, a mayor puntuación total, aumenta monotónicamente la probabilidad de elegir la alternativa correcta y disminuye monotónicamente la probabilidad de elegir alguno de los distractores. Estos resultados, agregados a los presentados en la Tabla 2, se consideran evidencia sólida a favor del buen funcionamiento de las claves y distractores para cada ítem.

Capacidad Discriminativa de los Ítems y de los Distractores

La capacidad discriminativa se calculó a partir de la correlación ítem-total para las alternativas correctas y, además, para cada distractor y el total. Dado que ya habían sido examinadas las *Curvas Empíricas de Respuesta*, este análisis se consideró complementario al anterior. Dado que se trata de una prueba con menos de 40 ítems, se aplicó la

Tabla 1
Cargas Factoriales y Correlación Ítem-Total Corregida para Cada Alternativa

Ítem	Carga Factorial	Correlación ítem-total corregida para cada alternativa			
		A	B	C	D
1	.41	(.29)	-.18	-.18	-.12
2	.48	(.34)	.06	-.29	-.24
3	.67	-.08	-.33	-.24	(.48)
4	.50	-.25	-.16	-.15	(.37)
5	.53	-.27	-.08	-.21	(.38)
6	.40	(.28)	-.18	-.13	-.13
7	.47	(.33)	-.21	-.11	-.18
8	.50	-.30	.01	-.18	(.36)
9	.55	-.22	-.22	-.14	(.41)
10	.39	(.27)	-.20	-.02	-.17
11	.67	(.46)	-.32	-.18	-.20
12	.49	(.32)	-.18	-.20	-.19
13	.74	-.23	-.34	(.54)	-.15
14	.48	(.33)	-.19	-.26	-.14
15	.47	-.04	-.30	-.20	(.34)
16	.45	(.30)	-.18	-.13	-.17
17	.41	-.05	-.17	-.24	(.29)
18	.39	-.04	-.24	-.07	(.29)
19	.55	-.16	-.16	-.28	(.41)
20	.52	-.20	-.09	(.36)	-.17
21	.46	-.18	-.10	-.20	(.34)
22	.62	-.28	(.45)	-.22	-.10
23	.40	-.07	-.18	-.18	(.30)
24	.57	-.03	-.02	(.39)	-.33
25	.48	.10	-.18	-.30	(.31)

Notas. (1) Se indica entre paréntesis la alternativa correcta. (2) Para resguardar la confidencialidad de las claves de corrección, el orden de los ítems en esta Tabla no corresponde al orden de presentación en el instrumento.

Tabla 2
Proporción de Elección de la Alternativa Correcta (Índice de Dificultad) y de los Distractores

Item	Proporción de elección de cada alternativa			
	A	B	C	D
1	(.71)	.06	.05	.18
2	(.44)	.23	.24	.09
3	.05	.26	.09	(.59)
4	.30	.05	.05	(.60)
5	.16	.27	.07	(.50)
6	(.66)	.23	.09	.03
7	(.53)	.12	.26	.09
8	.38	.17	.03	(.42)
9	.30	.11	.06	(.54)
10	(.69)	.15	.06	.10
11	(.74)	.15	.03	.09
12	(.76)	.02	.04	.17
13	.16	.20	(.51)	.12
14	(.76)	.19	.04	.02
15	.18	.09	.14	(.59)
16	(.74)	.15	.06	.05
17	.49	.04	.12	(.35)
18	.15	.21	.21	(.43)
19	.22	.08	.12	(.57)
20	.12	.15	(.41)	.32
21	.21	.15	.10	(.54)
22	.25	(.55)	.16	.04
23	.13	.35	.05	(.47)
24	.21	.12	(.29)	.37
25	.41	.17	.18	(.24)

Notas. (1) Se indica entre paréntesis la alternativa correcta. (2) Para resguardar la confidencialidad de las claves de corrección, el orden de los ítems en esta Tabla no corresponde al orden de presentación en el instrumento.

corrección de Henrysson (1963) a las correlaciones entre cada alternativa y el total. Los resultados muestran que las correlaciones ítem-total se encuentran en el rango $r=.27$ a $r=.54$, con una mediana de $Md=.34$, indicando que los ítems del TAE-A-25 tienen una apropiada capacidad discriminativa y que ésta es siempre mayor para la alternativa correcta que para los distractores, lo cual puede interpretarse, nuevamente, a favor del buen comportamiento psicométrico de las claves y los distractores.

Confiabilidad y Errores Estándar de Medida

Se calculó la confiabilidad por consistencia interna, a través del coeficiente Alfa de Cronbach. Se observó un resultado $\alpha=.82$ ($SE=.01$), indicativo de buena consistencia interna. No obstante, considerando que el coeficiente Alfa es un estimador del límite inferior de la confiabilidad, y dado que la relativa heterogeneidad de las cargas factoriales en el primer factor y en las correlaciones ítem-total sugieren que los ítems no son tau-equivalentes, se calculó también el coeficiente de confiabilidad GLB (Jackson & Agunwamba,

1979), que resulta más apropiado, además, para escalas en las que no se justifica suponer una unidimensionalidad estricta. El resultado obtenido ($GLB=.87$) confirma que el test es capaz de producir puntuaciones altamente confiables.

Dado que el objetivo final de la estimación de la confiabilidad es cuantificar el error de medida (Gempp, 2006), se efectuaron dos cálculos adicionales. En primer lugar, se computó el *Error Estándar de Medida*, obteniéndose un valor de $EEM=2.19$. En segundo lugar, se calculó el *Error Estándar de Medida Condicional*, según el método de Lord (1955). Los resultados para los distintos rangos de puntuación total fueron: $EEM_{COND}=0.95$ para el rango entre 0 a 2 puntos, $EEM_{COND}=1.78$ para el rango entre 3 a 5 puntos, $EEM_{COND}=2.18$ para el intervalo de 6 a 8 puntos, $EEM_{COND}=2.38$ para el rango de 9 a 11 puntos, $EEM_{COND}=2.42$ para el intervalo entre 12 a 14 puntos, $EEM_{COND}=2.33$ para el rango de 15 a 17 puntos, $EEM_{COND}=2.07$ para el intervalo entre 18 y 20 puntos, y $EEM_{COND}=1.58$ para puntuaciones totales iguales o mayores a 21 puntos. Como puede observarse, los errores resultan de moderada magnitud.

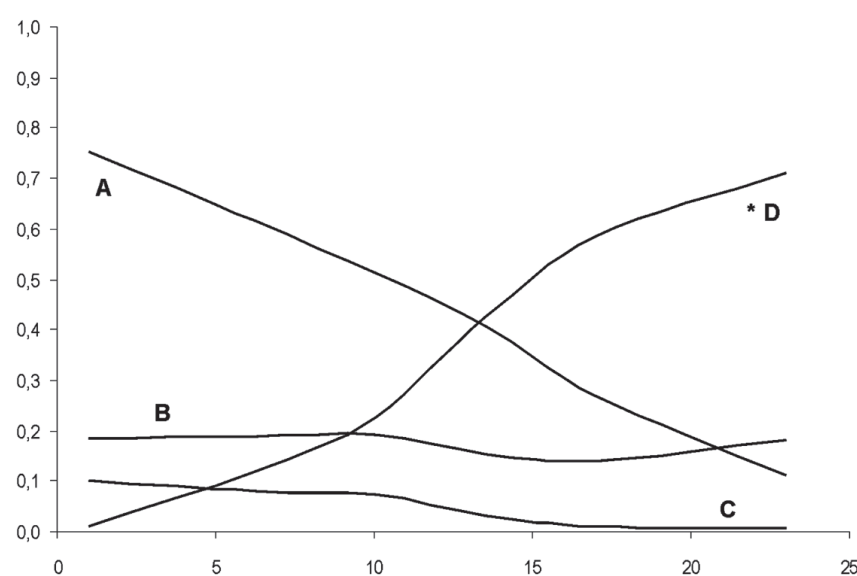


Figura 1. Curvas empíricas de respuesta al ítem 8, suavizadas con regresión kernel

Notas. (1) Se indica con un asterisco (*) la alternativa correcta. (2) La numeración del ítem corresponde a la utilizada en las Tablas 1 y 2 y no a la secuencia original en el cuadernillo de prueba.

Evidencia Empírica de validez

Los resultados reportados en este trabajo permiten avalar parcialmente la *validez de contenido* del TAE-A-25, a partir de *evidencia cualitativa*. Ésta se origina en el procedimiento de elaboración de ítems y en el apoyo de especialistas en Psicología Económica para la revisión de los mismos, cuyo propósito fue asegurar la representatividad de los contenidos evaluados respecto a las cuatro áreas de contenidos económicos mínimos definidos por el National Council of Economic Education, anteriormente discutidos. No obstante, estos procedimientos no representan evidencia cuantitativa sobre la validez de contenido de la prueba.

Sin perjuicio de lo anterior, varios resultados referidos a las relaciones observadas entre la puntuación total del TAE-A-25 y otras variables, aportan evidencia indirecta sobre la *validez convergente* de la prueba. Por ejemplo, se analizó la correlación entre la puntuación en el TAE-A-25 y el nivel educacional de los participantes en la investigación. De acuerdo a la teoría subyacente, es razonable anticipar que los sujetos expuestos a más años de educación formal deberían exhibir niveles de AE más altos que los individuos menos escolarizados. Corroborando esta expectativa, al definir seis grupos de escolaridad (primaria incompleta, primaria completa, secundaria incompleta, secundaria completa, superior incompleta, superior completa), y comparar sus respectivos promedios en el TAE-A-25, se observa una progresión claramente lineal ($F = 378.20$; $p < .001$) con medias iguales a $M = 10.28$ ($SD = 3.25$), $M = 10.67$ ($SD = 3.98$), $M = 11.29$, ($SD = 4.21$), $M = 13.85$ ($SD = 4.64$), $M = 18.38$ ($SD = 4.03$) y $M = 18.09$ ($SD = 4.31$). Un análisis minucioso indica que la diferencia entre educación superior completa e incompleta no es significativa ($p > .05$) pero que ambos tipos de educación marcan una diferencia sustantiva con los

restantes niveles educativos. Considerada en conjunto, la correlación entre escolaridad y la puntuación total en el TAE-A-25, alcanza a $r = .55$ ($p < .001$), lo que confirma el supuesto de *validez convergente* entre educación y nivel de AE.

En línea con lo anterior, se analizó también la relación entre los resultados del TAE-A-25 y el hecho de estar actualmente estudiando. Se asume que los sujetos que estudian tienen más probabilidad de exponerse a un entorno formal de AE y, por tanto, deberían obtener puntuaciones superiores en la prueba. Al comparar la media de puntuaciones totales en el TAE-A-25 de quienes actualmente estudian ($M = 14.39$; $SD = 5.26$) versus quienes no lo hacen ($M = 13.21$; $SD = 5.13$) se confirmó esta hipótesis ($t = 3.14$; $p = .002$) con un efecto estimado correspondiente a $d = .23$.

Además, se analizó la relación entre el resultado obtenido en el TAE-A-25 y el estatus ocupacional de los participantes, definido en un rango de cinco segmentos, que va desde oficios menores hasta cargos ejecutivos. Nuevamente de acuerdo a los antecedentes teóricos, se hipotetizó que los sujetos con un estatus ocupacional más elevado deberían obtener puntuaciones más altas en la prueba. Tal como se anticipó, los promedios observados para los cinco grupos ocupacionales exhibieron una tendencia lineal ($F = 128.13$; $p < .001$) equivalente a una correlación de Pearson de $r = .37$. Los promedios específicos fueron de $M = 10.72$ ($SD = 4.46$), $M = 11.52$ ($SD = 4.50$), $M = 13.09$ ($SD = 4.80$), $M = 14.97$ ($SD = 5.09$) y $M = 17.58$ ($SD = 4.92$).

Finalmente, se calculó la correlación entre la puntuación en el TAE-A-25 y las puntuaciones obtenidas en una escala de Hábitos de Compra Reflexivos (Denegri, Palavecinos, Ripio, & Yáñez, 1999) incluida en el cuestionario que respondieron los participantes. Teóricamente se espera que

los sujetos más alfabetizados económicamente desarrollen un patrón de gasto más reflexivo y eficiente que los sujetos con menor AE (Yamane, 1997). En consecuencia, una correlación entre las puntuaciones de ambos instrumentos puede interpretarse como evidencia de validez a favor del TAE-A-25. Corroborando lo anterior, se observó una correlación baja pero significativa entre ambas puntuaciones ($r = .10; p = .007$).

Considerados en conjunto, estos cuatro resultados representan evidencia sobre la *validez convergente* del TAE-A-25, en la medida que demuestran cómo las puntuaciones de la prueba son capaces de confirmar empíricamente correlaciones teóricamente previsibles entre la AE y variables afines.

Por otro lado, respecto a la *validez discriminante* (ausencia de correlaciones sustantivas entre el TAE-A-25 y constructos no relacionados con la AE) también se observaron un par de resultados interesantes que constituyen aval indirecto de este tipo de validez. En primer lugar, y tal como se esperaba teóricamente, al comparar las puntuaciones obtenidas en el TAE-A-25 por los sujetos de cada uno de los cuatro grupos de edad que se incluyeron en la muestra (18 a 25 años, 26 a 30 años, 31 a 40 años, 41 a 50 años) no se observan diferencias significativas entre los promedios ($F = 2.45; p = .06$) demostrándose así la ausencia de correlación entre ambas variables. En forma adicional, no se observó diferencia significativa entre géneros respecto a las puntuaciones totales en el TAE-A-25 ($t = 1.14; p = .25$), lo que indica que hombres y mujeres muestran un rendimiento similar en la prueba ($M = 13.83, SD = 5.06$ versus $M = 13.41, SD = 5.34$, respectivamente). Aunque este resultado contradice los hallazgos previos de Denegri y Palavecinos (2003), quienes sí encontraron diferencias de género en AE, hay que precisar que aquel estudio tuvo una naturaleza únicamente cualitativa y que hasta la fecha no existen antecedentes empíricos ni argumentos teóricos robustos que permitan presumir diferencias de género en los niveles de AE de sujetos adultos. En efecto, aunque las diferencias a favor de los varones en puntuaciones obtenidas en pruebas estandarizadas de economía son un hallazgo clásico en el campo de la educación económica (e.g. Siegfred, 1979) dichos resultados son válidos para adolescentes y estudiantes secundarios (e.g. Walstad & Robson, 1997) sin que existan pruebas indiscutibles de su validez en sujetos adultos. Al parecer, el género por sí mismo no tiene un valor explicativo en la AE si no se consideran el efecto de la escolaridad. Como prueba de lo anterior, se hizo un análisis factorial de la varianza con los datos obtenidos en el TAE-A-25, considerando como factores el género (hombres versus mujeres) y el hecho de estar o no estudiando actualmente. Aunque los efectos principales muestran los mismos resultados previamente obtenidos con las respectivas pruebas t (diferencia significativa entre quienes estudian y quienes no lo hacen, y ausencia de diferencia significativa por género), un resultado interesante fue la presencia de un efecto interactivo de primer orden entre el género y encontrarse estudiando

($F = 6.89; p = .009$). El examen cuidadoso de las medias marginales muestra que, en los varones, no existe una diferencia sustantiva de rendimiento en el TAE-A-25 entre quienes sólo estudian ($M = 13.96; SD = 5.06$) y quienes trabajan ($M = 13.74; SD = 5.07$), mientras en las mujeres, en cambio, se observa una diferencia de rendimiento apreciable entre aquellas que estudian ($M = 14.89; SD = 5.46$) y las que trabajan ($M = 12.71; SD = 5.14$). Este resultado sugiere que en el grupo de mujeres las diferencias de AE provocadas por el hecho de estar expuestas a la educación formal son más drásticas que las producidas en los varones. No obstante, estos datos también prueban que las diferencias de género en el rendimiento del TAE-A-25 son complejas y están mediadas por el nivel de estudio, lo cual aconseja explorar otros efectos antes de adelantar más conclusiones.

Finalmente, respecto a la *validez factorial* del instrumento, mientras no se realice un análisis factorial confirmatorio en el que se sometan a prueba hipótesis estructurales sobre el agrupamiento de los items, sería apresurado concluir evidencia de este tipo de validez. No obstante, el proceso utilizado para reducir el conjunto inicial de items y el ajuste exhibido por la solución factorial unidimensional exploratoria, descritos anteriormente, aportan evidencia preliminar en este sentido.

Normas provisionales en Puntuaciones T y Estaninos

Considerando que el propósito de desarrollar el TAE-A-25 fue contar con una medida breve de AE, fácil de aplicar y corregir, que permitiera su uso en investigación e intervención, y dadas las adecuadas propiedades psicométricas reportadas en los apartados anteriores, se procedió a calcular normas provisionales para la prueba, con la finalidad de facilitar la interpretación de sus puntuaciones.

La media y desviación típica de las puntuaciones totales para el TAE-A-25 correspondieron a $M = 13.62$ y $SD = 5.20$, observándose un valor mínimo de 0 puntos y un máximo de 25 puntos, con una moda igual a $Mo = 13$ puntos. La distribución de puntuaciones exhibe una curtosis levemente negativa ($Cu = -0.87$) y una simetría casi perfecta ($As = 0.01$), situándose el cuartil inferior, mediana y cuartil superior en $C_i = 10$, $Mdn = 13$ y $C_s = 18$ puntos, respectivamente. Como la distribución de puntuaciones observadas tiene una forma aproximadamente normal, se optó por calcular baremos estandarizados en lugar de percentiles, empleado puntuaciones típicas derivadas directamente, sin normalizar la distribución. Debido a que previamente no se encontraron diferencias significativas por género, las normas fueron obtenidas para el total de la muestra, sin distinguir entre hombres y mujeres.

Específicamente se obtuvieron baremos en Puntuaciones T ($M = 50; SD = 10$) y Estaninos ($M = 5; SD = 2$). Los resultados se presentan en la Tabla 3. Cabe advertir que estas normas pueden considerarse provisionales, hasta contar con una muestra aleatoria y de mayor tamaño, que

permita desagregar los resultados por nivel educacional u otras variables relevantes.

Discusión

El propósito de este trabajo fue elaborar una versión abreviada del Test de Alfabetización Económica para Adultos, en el marco de la Teoría Clásica de los Tests. En retrospectiva, los análisis realizados permiten concluir que este objetivo ha sido cumplido en forma satisfactoria.

En primer lugar, fue posible aislar 25 items que conforman una prueba unidimensional. En segundo lugar, estos 25 items exhiben un comportamiento de claves y distractores apropiado tanto en dificultad como en capacidad discriminativa. Lo anterior es válido además para todo el rango de la escala, como lo prueban las *Curvas Empíricas de Respuesta*. En tercer lugar, la prueba resultante es capaz de entregar resultados confiables y con niveles moderados de error de medida para todo el rango de puntuaciones observadas. En cuarto lugar, la evidencia existente hasta el momento es promisorio respecto a la validez de la prueba. Por último, fue posible elaborar baremos para asistir en forma provisional la interpretación de las puntuaciones producidas por el TAE-A-25.

No obstante, es necesario considerar las limitaciones de esta investigación y del instrumento derivado. La primera y más importante se relaciona con la validez de la prueba. Los análisis presentados aportan evidencia primitiva a favor de su *validez de contenido* y de su *validez de constructo*, esta última a través de los resultados que respaldan en forma rudimentaria la validez convergente, discriminante y factorial de la prueba. Sin embargo, debemos reconocer que hasta no contar con evidencia robusta (e.g. análisis de la red nomológica asociada al constructo, análisis factorial confirmatorio de los items, análisis multirasgo-multimétodo) no se puede garantizar la validez del TAE-A-25 aún cuando los resultados obtenidos constituyan un primer paso en esa dirección. Por lo tanto, una tarea pendiente para investigaciones posteriores es abordar en forma sistemática el estudio de la validez del instrumento. Mientras tanto, es prudente restringir su uso a tareas de investigación o análisis en el nivel grupal en lugar de utilizarlo para el diagnóstico individual.

En relación con lo anterior es necesario profundizar especialmente en la validez de constructo del instrumento y la pertinencia de sus items para predecir comportamientos económicos cotidianos. Esto es, esclarecer si efectivamente los contenidos incluidos en la prueba constituyen un buen indicador de comprensión del mercado y comportamiento económico

Tabla 3
Baremos Provisionales en Puntuaciones T y Estaninos

Puntuación Bruta	Puntuación T	Estaninos
0	20	1
1	20	1
2	23	1
3	28	1
4	31	1
5	34	2
6	37	2
7	39	3
8	41	3
9	43	4
10	45	4
11	47	4
12	48	5
13	50	5
14	51	5
15	53	6
16	55	6
17	56	6
18	58	7
19	60	7
20	62	7
21	64	8
22	68	9
23	74	9
24	80	9
25	97	9

eficiente, cuestión que representa el objetivo último de los estudios de AE. Aunque algunos resultados obtenidos con el TAE-A original parecen confirmar esta suposición (e.g. Denegri, Palavecinos, Gempp, & Caprile, 2005) es necesario investigar este punto con el TAE-A-25.

Otras limitaciones a tener en cuenta son el tipo y composición de la muestra. Aunque se extremaron los controles en la recolección de participantes, el hecho de no contar con una muestra aleatoria restringe la generalización de los resultados e impide calcular errores de estimación para los estadísticos calculados. En futuras investigaciones con el TAE-A-25 sería conveniente obtener muestras aleatorias y potenciar la inclusión de participantes de nivel socioeconómico bajo y alto, que en esta ocasión quedaron subrepresentados en comparación con los sujetos de clase media. Un aspecto pendiente sobre este mismo punto, es explorar la presencia de sesgos sistemáticos que favorezcan o perjudiquen a los miembros de algún grupo socioeconómico particular, a través de un análisis de *Funcionamiento Diferencial del Ítem* (DIF). El mismo alcance puede hacerse respecto al efecto del género sobre el rendimiento en el TAE-A, especialmente en virtud de los resultados que señalaron la existencia de interacción entre el género y la situación educacional.

En cualquier caso, los resultados obtenidos en este estudio resultan lo suficientemente prometedores como para recomendar el uso del TAE-A-25 y aconsejar que se continúen investigando sus características psicométricas.

Referencias

- Adimark (2000). *El nivel socioeconómico ESOMAR. Manual de aplicación*. Santiago: Adimark.
- Caripán, N., Catalán, V., & Hermosilla, S. (2004). *Construcción de un Test de Alfabetización Económica para Adultos (TAE-A)*. Tesis de Licenciatura no publicada, Universidad de La Frontera, Temuco, Chile.
- Cortés, L., Quesada, M. & Sepúlveda, J. (2004). *Construcción de un Test de Alfabetización Económica para Niños (TAE-N)*. Tesis de Licenciatura no publicada, Universidad de La Frontera, Temuco, Chile.
- Cumbre de las Américas (1999). *Organización de estados americanos*. Documento de trabajo no publicado.
- Denegri, M., & Palavecinos, M. (2003). Género y alfabetización económica: ¿oportunidades de desarrollo o nuevos caminos para la discriminación? *Revista Psicología desde el Caribe*, 12, 76-97.
- Denegri, M., Palavecinos M., & Gempp, R. (2003). *Socialización Económica: un estudio descriptivo de las estrategias y prácticas socializadoras y de alfabetización económica en las familias de una ciudad multifinanciera*. (Proyecto FONDECYT N° 1030271). Temuco: Universidad de la Frontera.
- Denegri, M., Palavecinos, M., Gempp, R., & Caprile, C. (2005). Socialización Económica en familias chilenas de clase media: educando ciudadanos o consumidores?. *Psicología e Sociedade*, 17(2), 88-98.
- Denegri, M.; Palavecinos, M.; Ripoll, M., & Yañez, V. (1999) Caracterización psicológica del consumidor de la IX Región. En Denegri, Fernández, Iturra, Palavecinos y Ripoll. *Consumir para vivir y no vivir para consumir*. Ediciones Universidad de La Frontera. Temuco.
- Gempp, F. (2006). El error estándar de medida y la puntuación verdadera de los tests psicológicos: Algunas recomendaciones prácticas. *Terapia Psicológica*, 24(2), 117-130.
- Gempp, R., Denegri, M., Caprile, C., Cortés, L., Quesada, M., & Sepúlveda, J. (2006). Medición de la Alfabetización Económica en niños: Oportunidades diagnósticas con el Modelo de Crédito Parcial. *Psyche*, 15(1), 13-27.
- Gorsuch, R.L. (1983). *Factor analysis* (2nd ed.). NJ: Lawrence Erlbaum.
- Henrysson, S. (1963). Correction of item-total correlations in item analysis. *Psychometrika*, 28, 211-218.
- Jackson, P.H., & Agunwamba, C.C. (1979). Lower bounds for the reliability of the total score on a test composed of non-homogeneous items: I. Algebraic lower bounds. *Psychometrika*, 42, 567-578.
- Lord, F.M. (1955). Estimating test reliability. *Educational and Psychological Measurement*, 15, 325-336.
- Muñiz, J. (2001). *Teoría Clásica de los Tests*. Madrid: Pirámide.
- Ramsay, J.O. (1997). A functional approach to modeling test data. In W. van der Linden, & R. K. Hambleton (Eds), *Handbook of modern item response theory*. New York, USA: Springer.
- Siegfried, J. J. (1979). Male-female differences in Economic Education. *Journal of Economic Education*, 10, 1-11.
- Soper, J.C., & Walstad, W.B. (1987) *Test of economic literacy. Examiner's manual*. (Second Edition) New York: Joint Council on Economics Education.
- Stacey, B. (1987). Economic socialization. *Annual Review of Political Science*, 2, 1-33.
- Stigler, G. J. (1970). The case, if any, for economic literacy. *Journal of Economic Education*, 1(2), 77-84.
- Velicer, W.F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, 41, 321-327.
- Walstad, W. B. & Robson, D. (1997). Differential Item Functioning and male-female differences on multiple-choice tests in economics. *Journal of Economic Education*, 28, 155-171.
- Webley, P. (1999). *The economic psychology of everyday life: Becoming an economic adult*. Documento de trabajo no publicado, Universidad de Exeter, Exeter, Inglaterra.
- Wright, B.D. & Stone, M.H. (1979). *Best test design*. Chicago: Mesa Press.
- Yamane, E. (1997). The meaning of "Economic Education" in Japanese elementary and secondary education: An historical perspective. *Proceedings of the Second Conference of the International Association for Children's Social and Economics Education, Sweden*, 101-104.

Received 24/03/2006

Accepted 23/01/2007

René Gempp. Psicólogo, Profesor Asociado Universidad Alberto Hurtado, Chile y Psicometrista, Sistema de Medición de la Calidad de la Educación (SIMCE), Ministerio de Educación, Chile.

Mariela Denegri. Psicóloga, Dra. en Psicología, Especialista en Psicología Económica, Profesora Asociada Universidad de la Frontera, Chile.

Cristina Caprile. Economista, Banco de Chile, Chillán, Chile.

Nadia Caripán, Valentina Catalán y Solange Hermosilla. Licenciadas en Psicología, Universidad de La Frontera, Chile.