

Propiedades psicométricas de la adaptación peruana de la prueba de pensamiento crítico PENCRISAL¹

Silvia F. Rivas²

Universidad de Salamanca, Salamanca, España

Patricia Morales Bueno

Pontificia Universidad Católica del Perú, Lima, Perú

Carlos Saiz

Universidad de Salamanca, Salamanca, España

RESUMEN

El objetivo principal de este estudio fue estudiar, en población peruana, las propiedades psicométricas de la prueba de pensamiento crítico, PENCRISAL, desarrollada y validada originalmente en población española. Participaron respondiendo voluntariamente la prueba 422 estudiantes de Ciencias e Ingeniería de una universidad peruana. Se analizó el nivel de fiabilidad con el alfa de Cronbach; la fiabilidad interjueces con coeficientes de concordancia Kappa de Cohen para cada uno de los ítems; la validez de constructo con Análisis Factorial de Componentes Principales, método Varimax y la validez divergente mediante el estudio de correlaciones con otras pruebas a través del coeficiente de Pearson. Los resultados obtenidos permiten concluir que el test PENCRISAL reúne los requisitos psicométricos necesarios para la evaluación de habilidades de pensamiento crítico en población peruana.

Palabras clave: Pensamiento crítico, instrumentos de medición, evaluación, educación superior.

ABSTRACT – Psychometric properties of Peruvian adaptation of the critical thinking test PENCRISAL

The main objective of this study was to study, in Peruvian population, the psychometric properties of the critical thinking test, PENCRISAL, originally developed and validated in Spanish population. 422 students of Science and Engineering of a Peruvian university participated voluntarily answering the test. Reliability was analyzed with Cronbach's alpha; interrater reliability was studied by mean of Cohen's Kappa coefficients for each of the items; construct validity was analyzed with Factor Analysis of Principal Components, Varimax method. Finally, divergent validity was studied by mean of the correlations with other tests through Pearson coefficient. It is concluded that PENCRISAL test fits with the psychometric requirements for the assessment of critical thinking skills in Peruvian population.

Keywords: Critical thinking, measurement instrument, validity, evaluation, higher education.

RESUMO – Propriedades psicométricas da adaptação peruana de pensamento crítico teste PENCRISAL

O objetivo principal deste trabalho foi estudar, em população peruana, as propriedades psicométricas do teste de pensamento crítico, PENCRISAL, originalmente desenvolvida e validada na população espanhola. Respondendo o teste participaram voluntariamente 422 estudantes de Ciência e Engenharia de uma universidade peruana. O nível de confiabilidade foi analisada com alfa de Cronbach; a confiabilidade inter-juizes foi analisada com coeficientes Kappa de Cohen para cada um dos itens; a validade do construto foi analisada por meio da análise fatorial de componentes principais, método Varimax. Finalmente, a validade divergente foi analisada por meio do estudo de correlações com outros testes utilizando o coeficiente de Pearson. Conclui-se que o teste PENCRISAL tem os requisitos psicométricos necessários para a avaliação de habilidades de pensamento crítico na população peruana.

Palavras-chave: pensamento crítico, instrumentos de medição, avaliação, ensino superior.

Reflexionar de manera crítica o ser capaz de tomar decisiones sólidas son algunas de las habilidades de pensamiento más deseadas en la sociedad del siglo XXI. Los cambios tan enormes que está experimentando nuestro mundo exige del buen juicio para alcanzar un mínimo

bienestar personal y una razonable competencia profesional, en cualquier ámbito. No es casual que haya una preocupación importante por mejorar las competencias intelectuales, como las citadas. Existe un deseo cada vez mayor, en la enseñanza superior, de mejorar

¹ Este proyecto multi-institucional fue financiado por el Programa de Ayudas para Programas de Cooperación Inter-universitaria e Investigación Científica (AECID, proyecto N° A/023725/09). La primera fase de la validación en el contexto peruano se realizó en la PUCP con apoyo del Vicerrectorado de Investigación (Proyecto DGI 2010.0001).

² Endereço para correspondência: Departamento de Psicologia Básica, Psicobiologia y Metodología, Universidad de Salamanca, Av. De la Merced, 109-131, 37005, Salamanca, España. Tel.: +(34) 92329-4500, ext. 3278. E-mail: silviaferivas@usal.es

las denominadas competencias transversales u horizontales, esto es, la buena reflexión o el pensamiento crítico. Se incorporan estas competencias como objetivos explícitos en los planes de estudios universitarios de muchos países. Sin embargo, hay un desinterés generalizado por diagnosticar el nivel de estas competencias, con el fin de saber si son razonablemente buenas o no. Hay estudios preocupantes que muestran que no son tan buenas como nos gustaría, incluso después de tres años de estudio (Arum & Roksa, 2011). Una de las razones de esta carencia preocupante es la falta de sistemas de evaluación eficaces de habilidades tan complejas como las de pensamiento. Nuestra investigación se dirige a llenar esta grave laguna, aún mayor en idioma español.

El instrumento de medida que hemos desarrollado se apoya en el enfoque del pensamiento crítico, que tiene como referentes más representativos a autores de diferentes disciplinas (Ennis, 1996; Facione, 2011; Halpern, 1998, 2003; Walton, 2006). En concreto, nuestro enfoque sobre el pensamiento crítico se ha desarrollado en trabajos precedentes, que son los que fundamentan el trabajo actual (Saiz & Rivas, 2008a, 2008b, 2011, 2012). Pensar críticamente implica reflexión y acción, todo ello encaminado a lograr nuestros fines. Alcanzar nuestras metas está promovido por alguna necesidad, buscar algo que no tenemos. Dicho de otro modo, es resolver un problema, eliminar esa carencia. De un modo sencillo, podemos decir que pensar críticamente es razonar y decidir para resolver problemas del modo más eficaz posible. Esta es la idea esencial que guía nuestra investigación. La evaluación, lo que aquí nos importa, consiste en medir el grado de eficacia en lograr nuestras metas. Y este grado no siempre es el adecuado. Por ello, ese interés cada vez mayor, en la enseñanza, por desarrollar esas competencias transversales. El deseo de mejorar, cambiar o desarrollar el pensamiento crítico es inherente a su concepción, ya que no puede separarse del de eficacia y, desafortunadamente, esta, las más de las veces, no es la deseable. Solo hay que mirar a nuestro alrededor para ver el mundo que estamos creando.

En educación, hay esfuerzos interesantes por mejorar estas competencias. No es objeto de nuestro trabajo revisarlas, pero sí lo es el considerar su eficacia, el grado en el que logran algún cambio en esas competencias o en la transferencia de esas habilidades. Como decíamos, esta valoración de los logros por mejorar nuestras habilidades de pensamiento es más bien escasa, al no disponer de buenas herramientas de evaluación. Existen algunos trabajos de investigación que aportan evidencias de que la enseñanza explícita de habilidades de pensamiento, por ejemplo, incluso a través de los contenidos de asignaturas propias del plan de estudios de diferentes disciplinas, como Metodología de Investigación en Psicología y Física 1 en una malla de formación tecnológica, logra un desarrollo mayor en

algunas habilidades de pensamiento crítico, en relación a un grupo control (Bensley, Crowe, Bernhardt, Buckner, & Allman, 2010; Poveda, 2010). Sin embargo, como decíamos, uno de los mayores inconvenientes para realizar este tipo de estudios continúa siendo la falta de instrumentos confiables y válidos que permitan evaluar con certeza los logros en las distintas propuestas de intervención. Algunos esfuerzos en esta dirección se están llevando a cabo, con resultados muy positivos, gracias, como decíamos, al empleo de medidas adecuadas de evaluación de las competencias de pensamiento. En un estudio realizado con el fin de probar un programa de instrucción en pensamiento crítico, ARDESOS (Saiz & Rivas, 2011), se ha podido demostrar su eficacia al incorporar medidas psicométricamente adecuadas, como de la que nos ocupamos en este estudio.

No es nuestro objetivo aquí discutir sobre la naturaleza de nuestra prueba, pues la revisión de las pruebas de pensamiento crítico, sus deficiencias, y la fundamentación de nuestro instrumento ya ha sido publicada en otros trabajos. A este respecto, pueden consultarse (Rivas & Saiz, 2012; Saiz, 2002; Saiz & Rivas, 2008b). Por todo ello, el objetivo de nuestro trabajo es paliar parte del problema de evaluación, con la adaptación y validación, en población universitaria peruana, de la prueba de pensamiento crítico PENCRISAL, validada ya en población española.

Método

Elaboración del instrumento y procedimiento

Este instrumento ha sido elaborado y aplicado con fines de validación a través de varias fases. En la primera, se confeccionó un banco de ítems amplio para poder hacer una buena selección. Esta primera versión del test, se aplicó en una prueba piloto a una muestra de 469 universitarios de diferente procedencia, con el objetivo de realizar el análisis psicométrico de los ítems. A partir de estos análisis se descartaron aquellos ítems que no cumplían satisfactoriamente las propiedades psicométricas necesarias para permanecer en la escala final, sustituyéndolos por nuevos ítems y se reelaboraron aquellos que podían aún alcanzar las propiedades que se buscaban. En función de estos resultados, se elaboró una segunda versión, la cual se aplicó a una muestra de 938 universitarios: 313 de España, 300 de México, 145 de Chile y 180 de Perú. Las versiones aplicadas en estos países fueron adaptadas lingüísticamente a las peculiaridades del idioma en cada lugar. Los resultados de esta fase permitieron realizar la versión final de la prueba.

La aplicación de la tercera versión del PENCRISAL, involucró instituciones de cuatro países: Universidad Pontificia de Salamanca (España), Universidad Diego Portales (Chile), Universidad Veracruzana (México) y Pontificia Universidad Católica del Perú (PUCP, Perú).

Características del instrumento

El test PENCRISAL (Rivas & Saiz, 2012), es una batería que consta de 35 ítems que plantean problemas de situaciones cotidianas, tienen formato de respuesta abierta, proponen problemas de diferentes temáticas de conocimiento y tienen respuestas únicas.

Los ítems están configurados en 5 factores, con 7 ítems cada uno de ellos. En la distribución de las situaciones-problema, en cada factor, se ha tenido en cuenta la selección de las estructuras más características de cada uno de ellos. Razonamiento deductivo: evalúa el razonamiento proposicional y el razonamiento categórico; Razonamiento inductivo: evalúa el razonamiento analógico, el hipotético y las generalizaciones inductivas; Razonamiento práctico: evalúa las habilidades de argumentación y la identificación de falacias; Toma de decisiones: evalúa el uso de procedimientos generales de decisión, lo que implica la elaboración de juicios precisos de probabilidad y el uso de heurísticos adecuados para adoptar decisiones sólidas; Solución de problemas: evalúa la puesta en marcha de estrategias específicas de solución a las situaciones planteadas.

Estos factores representan las habilidades fundamentales de pensamiento y, dentro de cada uno de ellos, encontraremos las formas de reflexión y resolución más relevantes en nuestro funcionamiento cotidiano. El orden de presentación de los ítems ha sido aleatorio.

La forma más habitual de administración del PENCRISAL es la informatizada, vía internet y de manera individual, por ser la que más ventajas ofrece. Pero podría realizarse en formato lápiz y papel de forma colectiva. Esta fue la modalidad de aplicación con la muestra peruana debido a las dificultades que presentaban por la falta de recursos informáticos. La prueba fue aplicada en las clases en cada uno de los grupos de los participantes.

El formato de los ítems es abierto, de manera que la persona debe responder a una pregunta concreta añadiendo a ésta una justificación del por qué de su respuesta. Por esta razón, se han establecido unos criterios de corrección estandarizados que asignan valores entre 0 y 2 puntos, en función de la calidad de la respuesta: 0 puntos: cuando la respuesta dada como solución del problema es incorrecta; 1 punto: cuando solamente la solución es correcta, pero no se argumenta adecuadamente (identifica y demuestra la comprensión de los conceptos fundamentales); 2 puntos: cuando además de dar la respuesta correcta, justifica o explica por qué (en donde se hace uso de procesos más complejos que implican verdaderos mecanismos de producción).

De esta manera se utiliza un sistema de escalamiento cuantitativo, cuyo rango de valores se sitúa entre 0 y 70 puntos como límite máximo, para la puntuación global de la prueba y entre 0-14 para cada una de las cinco escalas.

A continuación mostramos un ejemplo del tipo de ítems utilizados en la prueba:

Juan necesita utilizar el transporte público todos los días para ir a trabajar y tarda aproximadamente unas dos horas. Estos últimos días, con la huelga de autobuses, ha habido problemas de tráfico, por lo que siempre ha llegado tarde. Hoy tiene una reunión muy importante y su jefe está intranquilo por si llegará a tiempo. Le pregunta a un compañero por Juan y éste le dice que no se preocupe que hoy no hay huelga, luego no tendrá problemas de tráfico, así que llegará a tiempo para la reunión.

¿Es correcta la conclusión del compañero de Juan? Justifica tu respuesta

En cuanto al tiempo de administración, esta prueba se define como un test psicométrico de potencia, es decir, sin limitación de tiempo. La duración promedio estimada para la realización completa es de 60 a 90 minutos. Para una información más detallada sobre los fundamentos de la prueba véase Saiz y Rivas (2008b).

Las dimensiones del test deben considerarse de manera multidimensional, en los términos que se explican a continuación. El pensamiento crítico tal como lo concebimos tiene que ver con lo que es razonar y decidir para resolver. Estas habilidades deben entenderse como interrelacionadas. El alcanzar una meta o resolver un problema implica reflexión, elección y utilización de buenas estrategias de solución. El fin deseado no se alcanza solamente con una de estas actividades fundamentales. Se necesita de la cooperación de todas o una parte de ellas, según la situación dada. Por esta razón, las dimensiones de esta prueba deben entenderse en los mismos términos. Deducción e inducción, con sus diferentes modos, no son otra cosa que formas particulares de razonamiento. El razonar o explicar siempre consiste en establecer una conclusión a partir de unas razones. La diferencia descansa en el modo de lograrlo. Usar analogías o relaciones de contingencia exige mecanismos de pensamiento lo suficientemente distintos, como para dar sentido a conceptos tales como razonamiento analógico o causal. Pero el propósito general es el mismo en ambos. Esta interdependencia entre los diferentes mecanismos de pensamiento hace que sea algo difícil el entendimiento de los resultados multidimensionales de la validación. Según este planteamiento, lo esperable serían relaciones entre las dimensiones, mayores entre deducción e inducción, y entre toma de decisiones y solución de problemas. Y siempre con relación entre todas.

Los resultados del estudio psicométrico de la prueba en el contexto español revelaron un conjunto de factores y subfactores que explican el 53% de la varianza empírica. La fiabilidad como consistencia interna alcanza un nivel aceptable, dada la complejidad del modelo teórico subyacente bajo el constructo de *Pensamiento Crítico* (alfa de Cronbach 0,632). En cuanto a la fiabilidad como estabilidad temporal, según el método test-retest es elevada ($r=0,786$). La fiabilidad interjueces manifiesta un elevado índice de concordancia entre los correctores, con

coeficientes de Kappa de Cohen, en la mayoría de los ítems, por encima de 0,600, llegando en algunos casos a superar el 0,900. Estos datos proporcionan evidencias sólidas del ajuste de la estructura factorial de la prueba al modelo teórico subyacente, así como también de sus adecuados índices de fiabilidad. Con ello se garantiza su validez de constructo y su estabilidad, sustentada sobre la base del alto nivel de concordancia logrado entre los evaluadores españoles.

Otros instrumentos utilizados

En el estudio de la validez divergente se utilizó el Test de Aptitudes Mentales Primarias, PMA (Thurstone & Thurstone, 1976). Esta batería consta de 5 pruebas que detectan aisladamente 5 factores básicos de inteligencia: Verbal ($r_{xx} = ,910$), Espacial ($r_{xx} = ,730$), Numérico ($r_{xx} = ,990$), Razonamiento ($r_{xx} = ,920$) y Fluidez Verbal ($r_{xx} = ,730$). Se realizó colectivamente en formato de lápiz y papel. Las medidas de inteligencia han demostrado ser diferentes a las de pensamiento. Por ello, es una buena medida divergente (véase, por ejemplo, Detterman & Sternberg, 1993; Saiz, 1994).

Muestra

Para la validación peruana del test PENCRISAL se utilizó una muestra de 422 sujetos que respondieron el test voluntariamente. El método de muestreo empleado es (como en la española) un muestreo por conveniencia.

Los participantes de la muestra fueron estudiantes del área de Ciencias e Ingeniería de la PUCP, de los cuales un 28,7% (121) fueron mujeres y un 71,1 % (300) varones. La media de edad de los participantes era de 19,41; IC 95% [19.18-19.64] con una desviación típica de 2,42 años, situándose el rango de edad entre los 16-29 años. Esta variable no se distribuye normalmente con $p < ,050$ (Test K-S: $Z = 0,243$; $p = 0,000$) debido a una clara asimetría positiva ($As = 1,202$) y a una curva claramente leptocúrtica ($K = 1,242$).

Esta muestra de 422 casos se ha empleado para el análisis de ítems, la consistencia interna, la validación factorial y el estudio descriptivo junto a la construcción del baremo. Para los estudios de fiabilidad interjueces y la validez divergente, se han empleado diferentes submuestras, extraídas aleatoriamente de entre los 422 participantes iniciales, antes de comenzar con los análisis estadísticos, intentando con ello evitar posibles sesgos.

Análisis de Datos

El análisis de datos se realizó mediante el paquete estadístico IBM-SPSS Statistics-19. Se han realizado las pruebas de bondad de ajuste de Kolmogorov-Smirnov (K-S) para comprobar la normalidad de las diferentes variables numéricas. El análisis de los ítems se realizó mediante el índice de dificultad y el índice de homogeneidad corregido entre el ítem y la puntuación total en la escala. Para el análisis de la fiabilidad, se empleó: coeficiente alfa

de Cronbach. La fiabilidad interjueces se comprobó con coeficientes de concordancia Kappa de Cohen para cada uno de los ítems. La validez de constructo se analizó con Análisis Factorial de Componentes Principales, probando con diferentes métodos de rotación, tanto ortogonal como oblicua; comparando soluciones y viendo su similitud, se decidió finalmente optar por las que se encontraron a través del método Varimax, al igual que ocurría en la versión española. Previamente se habían comprobado las condiciones de factorización con las pruebas de Bartlett y Kaiser-Meier-Olkin, junto al determinante de la matriz de correlaciones. Por último, las correlaciones para la validez divergente se realizaron mediante coeficientes de Pearson.

Resultados

Análisis de ítems

El PENCRISAL se configura como una prueba difícil en cuanto a su nivel de ejecución. Esto es algo necesario en este tipo de pruebas ya que sólo de esta manera se puede demostrar el efecto de la intervención, sin necesidad de diseñar otro instrumento paralelo para este propósito. Advertido lo anterior, la dificultad de los ítems, varió entre 0,06 y 0,72 con media 0,36; IC 95% [0,30-0,41] y desviación estándar de 0,16. De ellos, 18 ítems (el 45,7%) presentaron un rango de dificultad media, 1 de los ítems (2,8%) fue fácil ($ID > 0,65$) y los 16 restantes (51,4%) mostraron dificultad alta ($ID < 0,35$).

El índice de homogeneidad corregido de cada uno de los ítems con respecto a la escala total, fue altamente significativo en todos ellos con $p < 0,001$. El rango de estos índices fue [0.131-0.394].

Consistencia interna y fiabilidad

El estudio de la fiabilidad se realizó desde las perspectivas de consistencia interna, estabilidad temporal y concordancia entre jueces, siendo esta última fundamental, dada las peculiaridades de la forma de corrección de la prueba. La consistencia interna de los 35 ítems se ha estimado mediante el método alfa de Cronbach. El coeficiente de fiabilidad obtenido es de 0,734 altamente significativo con $p < 0,001$ ($n = 422$; Anova: $F = 102,999$; 34 y 14314 gl; $p < 0,001$), lo que indica que el grado de homogeneidad entre los ítems es bastante aceptable.

La fiabilidad interjueces se realizó con una submuestra aleatoria de 100 participantes de la muestra total. Los jueces que participaron en el estudio fueron 3 expertos en el ámbito del pensamiento crítico. Dos de ellos españoles y el tercero peruano. Así mismo, todos ellos recibieron una instrucción para la corrección de la prueba. Los 3 jueces corrigieron los cuestionarios de forma independiente. Durante este proceso algunos cuestionarios estaban incompletos, por lo que el número de casos analizado para esta parte del estudio varía entre 91 y 96.

Se cruzaron los datos de los 3 jueces entre sí y se calcularon todos los coeficientes de Kappa de Cohen.

Todos estos índices han resultado ser altamente significativos con $p < 0,001$. Los resultados que se muestran en la Tabla 1 indican que todos los ítems presentaron coeficientes mayores a 0,500 y la mayoría de ellos tuvo valores por encima de 0,600, por lo que se puede indicar

que existe una buena concordancia. La media de concordancia entre los jueces 1 y 2 fue de 0,738 con un rango de 0,515 a 0,970. La media de los correctores 1 y 3 fue de 0,624 con un rango de 0,511 a 0,879. Y por último la media de fiabilidad entre los jueces 2 y 3 fue de 0,602 con un rango de 0,514 a 0,861. Todos estos índices resultaron ser altamente significativos con $p < 0,001$.

Tabla 1
Valores de coeficientes Kappa inter-jueces

Deducción	C1-C2	C1-C3	C2-C3
1	,727	,569	,619
3	,970	,603	,581
5	,716	,656	,540
8	,637	,632	,695
16	,827	,584	,565
23	,834	,552	,550
28	,862	,525	,526
Inducción	C1-C2	C1-C3	C2-C3
2	,553	,549	,584
4	,919	,859	,849
6	,630	,737	,526
9	,659	,607	,686
10	,608	,550	,541
24	,565	,543	,568
29	,658	,687	,572
R.P.	C1-C2	C1-C3	C2-C3
7	,667	,567	,514
11	,752	,640	,652
21	,674	,599	,567
25	,630	,818	,648
30	,760	,718	,654
31	,718	,636	,669
34	,908	,530	,580
T.D.	C1-C2	C1-C3	C2-C3
14	,785	,583	,573
17	,721	,549	,525
18	,844	,511	,520
19	,515	,643	,533
20	,672	,634	,573
27	,742	,583	,572
32	,699	,543	,592
S.P.	C1-C2	C1-C3	C2-C3
12	,835	,879	,861
13	,747	,576	,517
15	,959	,805	,766
22	,733	,511	,534
26	,729	,543	,570
33	,858	,681	,718
35	,717	,653	,543
Media de índices K	,738	,624	,602
p	<,000	<,000	<,000

Validez de constructo

El estudio de validez de constructo se realizó aplicando el Análisis Factorial de forma independiente en cada una de las dimensiones ya que está demostrada la multidimensionalidad del constructo del pensamiento crítico.

En todos los análisis se cumplieron satisfactoriamente las condiciones previas de adecuación muestral ($KMO > 0,500$) y esfericidad (test de Bartlett con $p < 0,001$), con determinantes de las matrices de correlación próximos a 0. Los resultados del análisis factorial demostraron el adecuado ajuste al modelo teórico de partida y que es consistente con el estudio realizado con la versión original de la prueba. A continuación se exponen los resultados para cada una de las 5 dimensiones:

a) *Deducción*. Ver Tabla 2. El factor deducción *Proposicional* agrupó 4 ítems con saturaciones comprendidas entre 0,552-0,696. Los otros 3 ítems se agruparon en torno al subfactor deducción *Categorico* con cargas factoriales entre 0,615-0,758. La variabilidad total interna

explicada por los ítems de esta dimensión fue del 43,4%. Esta dimensión explicó un 9,77% de la variabilidad total del PENCRISAL.

b) *Inducción*. Ver Tabla 3. Los ítems se configuraron en torno al factor inductivo *Causal* con una saturación que se encuentra en el rango 0,616-0,704. Otros 2 ítems definieron el factor inductivo *Analógico* con cargas de 0,732 y 0,815. Los dos últimos, con cargas de 0,769 y 0,759 constituyen los procedimientos de *Verificación* (generalizaciones inductivas y comprobación de hipótesis). La variabilidad interna explicada por todos ellos alcanzó el 55,80%. Este componente explicó un 12,69% de la variabilidad total de la prueba.

c) *Razonamiento práctico*. Ver Tabla 4. Se demostró que 4 ítems, con pesos factoriales comprendidos en el rango 0,543-0,764 se agruparon en la dimensión *Argumentación*; mientras que los otros 3 configuraron el componente *Falacias* con saturaciones entre 0,537-0,750. La variabilidad total explicada alcanzó el 44,26% mientras que el factor razonamiento práctico explicó un 10,43% de la variabilidad total de la prueba.

Tabla 2
Estructura Factorial y Fiabilidad de la dimensión: DEDUCCIÓN (7 ítems)

Componente	Nº ítems	Ítems	α Cronbach	Factores	% de Varianza explicada en su Dimensión	% de Varianza total explicada
Ded. Proposicional	4	1; 3; 8; 16	-	1	21,69	5,19
Ded. Categorico	3	5; 23; 28	-	1	21,71	4,58
Total Dimensión	7	-	,521	2	43,40	9,77

Nota. Condiciones: $KMO = ,667$; Bartlett $p < ,001$.

Tabla 3
Estructura Factorial y Fiabilidad de la dimensión: INDUCCIÓN (7 ítems)

Componente	Nº ítems	Ítems	α Cronbach	Factores	% de Varianza explicada en su Dimensión	% de Varianza total explicada
Induct. Razon. Analóg.	2	9; 24	-	1	17,92	4,13
Induct. Causal	3	2; 6; 10	-	1	20,22	4,64
Induct. Proc. Verificación (CH y GI)	2	4; 29	-	1	17,65	3,92
Total Dimensión	7	-	,382	3	55,80	12,69

Nota. Condiciones: $KMO = ,595$; Bartlett $p < ,001$.

Tabla 4
Estructura Factorial y Fiabilidad de la dimensión: RAZONAMIENTO PRÁCTICO (7 ítems)

Componente	Nº ítems	Ítems	α Cronbach	Factores	% de Varianza explicada en su Dimensión	% de Varianza total explicada
Argumentación	4	7; 21; 25; 30	-	1	23,27	5,39
Raz. Práct.: Falacias	3	11; 31; 34	-	1	20,98	5,04
Total Dimensión	7	-	,549	2	44,26	10,43

Nota. Condiciones: $KMO = ,665$; Bartlett $p < ,001$.

d) *Toma de decisiones*. Ver Tabla 5. En esta dimensión se identificó 3 subfactores, los dos primeros constituidos

por 2 ítems cada uno de ellos, y el último por 3 ítems. El factor de TD *General* explicó un 19,15% de la

variabilidad interna del factor con pesos de 0,717 y 0,808. La TD *Heurísticos específicos* presentó pesos factoriales de 0,691 y 0,765 y explicó un 17,26%. Por último, la TD *Probabilidad*, con saturaciones comprendidas en el rango 0,539-0,731 explicó un 19,67%. La variabilidad total interna explicada llegó hasta el 56,08%. El componente toma de decisiones es el que mayor peso tuvo dentro de la prueba completa ya que explicó un 12,32% de la variabilidad total.

e) *Solución de problemas (S.P.)*. Ver Tabla 6. En el subfactor *S.P. general* se agruparon 4 ítems con cargas de valores en el rango 0,516-0,656; mientras que los otros 3 ítems constituyeron el componente *S.P. específico* (búsqueda de regularidades y análisis medio-fin) con cargas factoriales entre 0,455-0,741. Estos ítems explicaron un 41,70% de la variabilidad total específica y el factor S.P. un 8,77% de la variabilidad total del PENCRISAL.

Tabla 5

Estructura Factorial y Fiabilidad de la dimensión: TOMA DECISIONES (7 ítems)

Componente	Nº ítems	Ítems	α Cronbach	Factores	% de Varianza explicada en su Dimensión	% de Varianza total explicada
TD General	2	14; 27	-	1	19,15	4,74
TD: Heurísticos generales (REP y DIS)	2	19; 20	-	1	17,26	3,65
TD Probabilidad	3	17;18; 32	-	1	19,67	3,93
Total Dimensión	7	-	,309	3	56,08	12,32

Nota. Condiciones: KMO=,599; Bartlett $p<,001$.

Tabla 6

Estructura Factorial y Fiabilidad de la dimensión: SOLUCIÓN PROBLEMAS (7 ítems)

Componente	Nº ítems	Ítems	α Cronbach	Factores	% de Varianza explicada en su Dimensión	% de Varianza total explicada
S.P. General	4	13; 22; 26; 35	-	1	20,01	3,73
S.P. Especifico (RGL y MF)	3	12; 15; 33	-	1	21,69	5,04
Total Dimensión	7	-	,467	2	41,70	8,77

Nota. Condiciones: KMO=,624; Bartlett $p<,001$.

Se calcularon las correlaciones entre los cinco factores anteriormente descritos y con la puntuación total (ver Tabla 7). Se obtuvieron coeficientes de correlación estadísticamente significativos dado el tamaño de la muestra, pero de intensidades entre factores desde 0,211 hasta 0,368. Esto apoya la multidimensionalidad del constructo y la independencia entre factores, pues los valores Pearson entre factores son de escasa intensidad.

En cuanto al análisis factorial del conjunto completo de los 35 ítems, (KMO=0,791; test de Bartlett:

$\chi^2=1731,138$: 595 *gl*; $p=0,000$), este reveló la existencia entre factores y subfactores de 12 componentes que coincidieron con el desglose anterior: 2 en deducción, 3 en inducción, 2 en razonamiento práctico, 3 en toma de decisiones y los 2 restantes en solución de problemas. Los pesos factoriales de los ítems se encontraron en el rango 0,403-0,831. La variabilidad total de la prueba explicada por este conjunto de factores y subfactores se acercó al 54% como se observa en Tabla 8.

Tabla 7

Matriz de intercorrelaciones de los factores y con el total del PENCRISAL

		RD	RI	RP	SP	TD	Total
RD	Correlación de Pearson	—					
	<i>p</i>						
	N						
RI	Correlación de Pearson	0,211	—				
	<i>p</i>	0,000					
	N	422					

Tabla 7 (continuação)
Matriz de intercorrelaciones de los factores y con el total del PENCRISAL

		RD	RI	RP	SP	TD	Total
RP	Correlación de Pearson	0,277	0,276	—			
	<i>p</i>	0,000	0,000				
	N	422	422				
SP	Correlación de Pearson	0,266	0,230	0,361	—		
	<i>p</i>	0,003	0,000	0,000			
	N	422	422	422			
TD	Correlación de Pearson	0,286	0,242	0,368	0,252	—	
	<i>p</i>	0,000	0,000	0,000	0,000		
	N	422	422	422	422		
Total	Correlación de Pearson	0,623	0,540	0,740	0,671	0,660	
	<i>p</i>	0,000	0,000	0,000	0,000	0,000	—
	N	422	422	422	422	422	

Tabla 8
Variabilidad explicada en el A.F. de C.P. con rotación Varimax de la prueba completa (35 ítems)

Componente	Suma de las saturaciones al cuadrado de la rotación		
	Total	% de la varianza	% acumulado
1	1,886	5,389	5,389
2	1,817	5,192	10,581
3	1,766	5,047	15,629
4	1,763	5,037	20,665
5	1,646	4,704	25,369
6	1,625	4,642	30,011
7	1,604	4,583	34,594
8	1,446	4,132	38,726
9	1,374	3,926	42,652
10	1,373	3,922	46,574
11	1,305	3,729	50,302
12	1,277	3,649	53,951

Validez divergente y convergente

La validez convergente no se estudió porque los resultados de la versión española mostraron ausencia de validez convergente que está justificada por la naturaleza de los instrumentos, ya que estos son muy diferentes tanto en estructura, como en el constructo a evaluar.

Para el análisis de la validación divergente se tomó una nueva submuestra aleatoria de 142 participantes. En esta parte del estudio, se aplicó el test de inteligencia PMA (Aptitudes Mentales Primarias). Las correlaciones encontradas no son significativas ($p > 0,050$) lo que demuestra claramente la ausencia de asociación teórica entre las pruebas, y defiende la divergencia (Tabla 9).

Puntuaciones del pencrisal y baremación

Las puntuaciones totales del PENCRISAL en la muestra peruana analizada se distribuyeron con una media de 24,80; IC 95% [24.10-25.51] y desviación típica de 7,35 para un rango de puntuaciones de 4-43. La distribución de estos valores se ajustó al modelo normal de la campana de Gauss con $p > 0,050$ (Test K-S: $Z = 0,914$; $p = 0,373$). Se obtuvieron las puntuaciones normalizadas y percentiles correspondientes para cada uno de los factores y la puntuación total, para la población general, dado que los análisis previos indicaron que no existían diferencias significativas ni por sexo ni por edad (ver Tabla 10).

En la Tabla 11 se presenta los baremos obtenidos en la validación con muestra española.

Como se puede observar al comparar ambas poblaciones, no existen grandes diferencias entre las puntuaciones requeridas por una y otra población para alcanzar

determinados centiles. En términos generales, la población peruana mostró un rendimiento ligeramente menor que la española, no obstante, no sería necesario la realización de un baremo específico para la población peruana pudiéndose utilizar el baremo español.

Tabla 9
Correlaciones entre PMA y PENCRISAL

		RD	RI	RP	SP	TD	TOTAL
PMA. Verbal	Correlación de Pearson	-0,100	0,049	0,059	0,097	0,016	0,040
	<i>p</i>	0,239	0,559	0,489	0,252	0,849	0,633
	N	142	142	142	142	142	142
PMA. Espacial	Correlación de Pearson	-0,108	-0,095	-0,067	-0,030	-0,086	-0,118
	<i>p</i>	0,202	0,260	0,429	0,720	0,308	0,161
	N	142	142	142	142	142	142
PMA. Razonamiento	Correlación de Pearson	-0,008	0,064	0,011	0,086	0,073	0,068
	<i>p</i>	0,929	0,447	0,897	0,308	0,389	0,419
	N	142	142	142	142	142	142
PMA. Numérico	Correlación de Pearson	0,029	-0,149	-0,039	-0,024	-0,120	-0,088
	<i>p</i>	0,298	0,077	0,648	0,776	0,155	0,299
	N	142	142	142	142	142	142
PMA. Fluidez verbal	Correlación de Pearson	-0,162	-0,090	-0,058	0,075	0,106	0,033
	<i>p</i>	0,733	0,285	0,490	0,372	0,210	0,695
	N	142	142	142	142	142	142
PMA TOTAL	Correlación de Pearson	-0,129	-0,066	-0,029	0,070	-0,006	0,181
	<i>p</i>	0,126	0,437	0,729	0,411	0,941	0,041
	N	142	142	142	142	142	142

Tabla 10
PENCRISAL: Baremos para población universitaria (Muestra peruana)

Centiles	Puntuaciones Directas					
	RD	RI	RP	SP	TD	PT
99	10	9	12	12	10	41
95	7	7	10	9	9	37
90	7	7	9	9	8	35
85	6	6	8	8	8	33
80	6	6	8	8	7	31
75	5	6	7	7	7	30
70	5	6	7	7	7	29
65	4	5	6	7	6	28
60	4	5	6	6	6	27
55	4	5	6	6	6	26
50	4	5	5	6	6	25

Tabla 10 (continuação)
PENCRISAL: Baremos para población universitaria (Muestra peruana)

Centiles	Puntuaciones Directas					
	RD	RI	RP	SP	TD	PT
45	3	5	5	5	5	24
40	3	4	4	5	5	23
35	3	4	4	5	5	22
30	2	4	4	4	4	21
25	2	4	3	4	4	20
20	2	3	3	4	3	19
15	1	3	2	3	3	17
10	1	3	2	2	2	15
5	0	2	1	1	2	13
1	0	0	0	0	0	7
N	422	422	422	422	422	422
M	3,66	4,75	5,32	5,66	5,41	24,80
DS	2,21	1,67	2,66	2,21	2,3	7,35

Nota. RD: Razonamiento deductivo; RI: Razonamiento inductivo; RP: Razonamiento práctico; SP: Solución de problemas; TD: Toma de decisiones; PT: Puntuación total.

Tabla 11
PENCRISAL: Baremos para población general (Muestra española)

Centiles	Puntuaciones Directas					
	RD	RI	RP	SP	TD	PT
99	10	9	12	11	10	41
95	8	8	10	10	9	38
90	7	7	9	9	9	36
85	7	7	9	9	8	35
80	6	6	8	8	8	33
75	6	6	8	8	8	32
70	5	6	7	7	7	31
65	5	6	7	7	7	30
60	5	5	6	7	7	29
55	4	5	6	7	7	29
50	4	5	5	6	6	28
45	4	5	5	6	6	27
40	4	5	5	5	6	26
35	3	4	4	5	5	25
30	3	4	4	5	5	24
25	3	4	4	4	5	23
20	3	4	3	4	5	22
15	2	3	3	3	4	20

Tabla 11 (continuação)
PENCRISAL: Baremos para población general (Muestra española)

Centiles	Puntuaciones Directas					
	RD	RI	RP	SP	TD	PT
10	2	3	3	3	4	18
5	1	2	2	2	3	16
1	0	1	1	1	2	13
N	715	715	715	715	715	715
Media	4,42	5,03	5,78	6,04	6,21	27,48
DS	2,16	1,63	2,58	2,39	1,91	6,49

Nota. RD: Razonamiento deductivo; RI: Razonamiento inductivo; RP: Razonamiento práctico; SP: Solución de problemas; TD: Toma de decisiones; PT: Puntuación total.

Los resultados de esta investigación indican que la versión peruana del PENCRISAL posee unas propiedades psicométricas aceptables que corroboran los resultados obtenidos en la versión original, mejorando incluso algunos aspectos de esta. Por tanto, este instrumento nos proporciona una medida del constructo de pensamiento crítico como un conjunto de habilidades de razonamiento, toma de decisiones y solución de problemas. Asimismo, el PENCRISAL mostró un buen comportamiento con una muestra Peruana de universitarios del área de ciencias e ingeniería, lo cual amplía sus posibilidades de utilización con muestras de otros países y de áreas de conocimiento específicas.

Al comparar la estructura factorial obtenida en la muestra peruana con el test original (Rivas & Saiz, 2012), se comprobó la correspondencia de los factores en ambas escalas. Tan solo en el componente toma de decisiones, hay un factor menos con respecto a la versión española, en donde este factor estaba constituido por un ítem que saturaba en dos. Por ello la solución actual es más sencilla que la que se tenía en la versión española, ya que se reduce un factor. Los valores y porcentajes de los análisis son muy parecidos en ambas poblaciones.

Además el instrumento ha demostrado una muy buena consistencia interna, comparable a la de la versión original española, siendo incluso ligeramente superior en la muestra peruana. Finalmente, uno de los aspectos más

importantes del instrumento es el estudio de la fiabilidad interjueces, puesto que, dadas las especiales características del tipo de prueba, el sistema de corrección requiere imprescindiblemente de un elevado grado de acuerdo entre los correctores. Se ha conseguido demostrar un elevado índice de concordancia con cada uno de los 3 evaluadores. Estos valores son consistentes con los obtenidos en la prueba original.

No obstante, nuestro instrumento presenta una serie de limitaciones que ya han sido discutidas y justificadas en el artículo de la validación española de la prueba. Por todo ello podemos decir que las propiedades psicométricas se mantienen estables transculturalmente presentando una mejora en la estructura factorial y en el porcentaje de varianza total explicada que es ligeramente superior en la versión peruana.

A la vista de los resultados satisfactorios obtenidos y dadas las características de la prueba PENCRISAL creemos que su aplicabilidad es amplia, abarcando ámbitos educativos, sociales, personales y de investigación, siendo además un instrumento apropiado para evaluar la eficacia de programas de instrucción y mejora de las habilidades de pensamiento crítico. En este sentido, consideramos de interés que se amplíe la muestra de estudio a otros países para comprobar que las propiedades psicométricas del instrumento se mantienen estables.

Referencias

- Arum, R., & Roksa, J. (2011). *Academically Adrift: Limited Learning on College Campuses*. Chicago: The University of Chicago Press.
- Bensley D. A. , Crowe D. S. , Bernhardt P. , Buckner C., & Allman A. L.(2010). Teaching and Assessing Critical Thinking Skills for Argument Analysis in Psychology. *Teaching of Psychology*, 37(2), 91-96.
- Detterman, D. K., & Sternberg, R. J. (1982). *Transfer on trial: Intelligence, cognition, and instruction*. Norwood, NJ: Ablex.
- Ennis, R. H. (1996). *Critical thinking*. Upper Saddle River, NJ: Prentice-Hall.
- Facione, P. A. (2011). *Think critically*. New York: Prentice Hall.

- Halpern, D. (1998). Teaching critical thinking for transfer across domains. Dispositions, skills, structure training, and metacognitive monitoring. *American Psychologist*, 53(4), 449-455.
- Halpern, D. (2003). *Thought and knowledge. An introduction to critical thinking*, (4a. ed.). New Jersey: Erlbaum.
- Poveda I. L. (2010). Formación de pensamiento crítico en estudiantes de primeros semestres de educación superior. *Revista Iberoamericana de Educación*, 53(3). Recuperado de <http://www.rioei.org/3263.htm>
- Rivas, S.F., & Saiz, C. (2012). Validación y propiedades psicométricas de la prueba de pensamiento crítico PENCRISAL. *Revista Electrónica de Metodología Aplicada*, 17(1), 18-34.
- Saiz, C. (1994). Pensamiento e instrucción. En M.A. Verdugo (Ed.) Evaluación curricular. *Una guía para la intervención psicopedagógica* (p. 613-689). Madrid: Siglo XXI.
- Saiz, C. (2002). Enseñar o aprender a pensar. *Escritos de Psicología*, 6, 53-72.
- Saiz, C., & Rivas, S.F. (2008a). Intervenir para transferir en pensamiento crítico. *Praxis*, 10(13), 129-149.
- Saiz C., & Rivas, S. (2008b). Evaluación en pensamiento crítico: una propuesta para diferenciar formas de pensar. *Ergo, Nueva Época*, 22-23, 25-66.
- Saiz, C., & Rivas, S.F. (2011). Evaluation of the ARDESOS program: an initiative to improve critical thinking skills. *Journal of the Scholarship of Teaching and Learning*, 11(2), 34-51.
- Saiz, C., & Rivas, S.F. (2012). Pensamiento crítico y aprendizaje basado en problemas. *Revista de Docencia Universitaria*, 10(3), 325-346.
- Thurstone, L.L., & Thurstone. T.G. (1976). *PMA: Aptitudes Mentales Primarias*. Madrid: TEA
- Walton, D. (2006). *Fundamentals of critical argumentation*. Cambridge, UK: Cambridge University Press.

Recebido em agosto de 2013
Reformulado em novembro de 2013
Aprovado em fevereiro de 2014

Sobre os autores

Silvia F. Rivas. Doctora en Psicología. Profesora de Pensamiento Crítico de la Facultad de Psicología. Universidad de Salamanca (España).

Patricia Morales Bueno. Doctora en Ciencias de la Educación. Profesora del Departamento de Ciencias, Sección Química. Pontificia Universidad Católica del Perú (PUCP).

Carlos Saiz. Doctor en Psicología. Profesor de Pensamiento Crítico de la Facultad de Psicología. Universidad de Salamanca (España).