Revista Electrónica de Metodología Aplicada
2012, Vol. 17 nº 1, pp. 18-34

Validación y propiedades psicométricas de la prueba de pensamiento crítico PENCRISAL

Silvia F. Rivas y Carlos Saiz

Universidad de Salamanca

RESUMEN

El propósito de nuestro estudio ha sido validar la prueba de pensamiento crítico PENCRISAL en población española. Esta prueba es una herramienta adecuada para evaluar las competencias de razonamiento (de diferente índole, como argumentación, razonamiento causal, analógico…), de solución de problemas y de toma de decisiones. El estudio psicométrico se realizó con una muestra de 715 adultos españoles, de nivel cultural universitario, con edades comprendidas entre los 18 y 53 años, y de ambos sexos. La fiabilidad como consistencia interna alcanza un nivel aceptable dada la complejidad del modelo teórico que subyace bajo el constructo Pensamiento Crítico (alfa de Cronbach: ,632). Por su parte, la fiabilidad como estabilidad temporal, según el método test-retest, ha resultado ser elevada (r = ,786). En cuanto a la fiabilidad interjueces ha demostrado un elevado índice de concordancia entre los correctores (valores de Kappa entre ,600- ,900). El análisis factorial ha revelado un conjunto de factores y subfactores que se ajustan al modelo teórico planteado y los resultados obtenidos de las correlaciones con otras pruebas apoyan la validez divergente, pero no la convergente. El PENCRISAL se presenta como un instrumento novedoso, validado en población española cuyos resultados muestran una elevada precisión y eficacia como instrumento de medida de los factores que componen el constructo de Pensamiento Crítico.

Palabras clave: pensamiento crítico, evaluación, fiabilidad, validez, análisis factorial.

ABSTRACT

The purpose of our study was to validate the critical thinking test PENCRISAL in the Spanish population. This test is an appropriate tool to assess reasoning skills (of various kinds, such as argumentation, causal reasoning, analog...), problem solving and decision making. The psychometric study was conducted with a sample of 715 Spanish adults, with college cultural level, aged between 18 and 53, and of both sexes. Reliability in terms of internal consistency achieved an acceptable level, if we consider the complexity of the theoretical model of the construct is under critical thinking (Cronbach alpha: ,632). In turn, the reliability in terms of temporal stability, according to the test-retest method, this has proven to be high (r = ,786). And finally, the reliability between judges has reached a high level of agreement between the correctors (Kappa values between ,600 - ,900). Factor analysis has shown a number of factors and subfactors that fit the theoretical model proposed and the results we obtained from the correlations with other tests support the divergent validity, but not convergent. The PENCRISAL is presented as a novel instrument, validated for a Spanish population, whose results show high accuracy and effectiveness as an instrument for measuring the factors that make up the construct of critical thinking.

Keywords: critical thinking, assessment, reliability, validity, factorial analysis.

Contacto:
Silvia Fernández Rivas y Carlos Saiz Sánchez
Dpto. Psicología Básica, Psicobiología y Metodología de las CC.
Universidad de Salamanca
Avda. de la Merced, 109-131
37005 Salamanca (España)
Email: silviaferivas@usal.es; csaiz@usal.es
Tlf.: 923 29 45 00. Ext. 3278 Fax: 923 29 46 08

1.- Introducción

Son muchas las concepciones que hay sobre pensamiento crítico, por lo que es necesario precisar cuál es la que nosotros defendemos. Nuestra tesis es que razonamos y tomamos decisiones para resolver problemas o lograr metas. Dentro de este planteamiento concebimos el pensamiento crítico como una teoría de la acción. Pensar críticamente no es solo profundizar en el terreno del buen juicio y de la buena argumentación. Es imprescindible que esa buena reflexión demuestre que sirve para resolver problemas o alcanzar metas, considerando así a la argumentación como un medio, no un fin. Concebimos el pensamiento crítico como una acción que nos obliga a poner en práctica nuestros planes.

Desde esta perspectiva, el pensamiento crítico descansa en tres habilidades fundamentales: razonamiento, solución de problemas y toma de decisiones. El pensamiento tiene que cambiar la realidad, no solo nuestras ideas, debe servir para algo más que producir conocimiento, debe resolver problemas. La vertiente aplicada del pensamiento crítico, termina en la acción, en resolver los problemas con eficacia y en tomar decisiones sólidas. Y para esto, es imprescindible una buena reflexión. Por lo tanto, razonar, decidir y resolver deben plantearse como mecanismos de pensamiento inseparables y dependientes unos de otros. Con una buena reflexión se diseña un buen plan de acción, que se ejecuta con buenas estrategias de decisión y de solución de problemas.

La necesidad o la importancia de la evaluación del pensamiento crítico en la vida diaria provienen de si social o personalmente se desea que estas competencias se mejoren. Saber si dicha mejora existe precisa de la cuantificación de la misma. Por ello la razón para desarrollar la prueba PENCRISAL nace de la necesidad de evaluar nuestro programa de intervención ARDESOS que llevamos a cabo en este ámbito (Saiz y Rivas, 2011) y de la carencia de instrumentos adecuados para ello. Las principales dificultadas en la evaluación del pensamiento crítico son tanto conceptuales como metodológicas. Las primeras, provienen de la diversidad en la conceptualización del pensamiento crítico. Y las metodológicas tienen su origen en que la mayoría de las pruebas que evalúan pensamiento crítico (Ennis, 2003) son instrumentos de formato de respuesta cerrada, que impiden la exploración de los mecanismos fundamentales del pensamiento implicados en la tarea de responder a un test. El test HCTAES (Halpern Critical Thinking Assessment Using Everyday Situations; Halpern, 2006) permite solventar esta dificultad. Este instrumento se centra en los procesos de pensamiento y los ítems que se proponen en la prueba son situaciones que describen problemas cotidianos que se deben resolver mediante respuestas abiertas y cerradas. Nuestra prueba PENCRISAL tiene su origen en el planteamiento de esta autora. Hemos mantenido parte de sus principios, pero hemos modificado algunos que no resultan muy apropiados (Saiz y Rivas, 2008). Los principios que fundamentan nuestra prueba son: 1) la utilización de ítems que sean situaciones cotidianas, 2) el uso de diferentes dominios, con la intención de valorar el grado de generalización de las habilidades, 3) un formato de respuesta abierta, que posibilita la exploración de los procesos de pensamiento, y 4) el empleo de situaciones-problema de respuesta única que permite evaluar el mecanismo de pensamiento correspondiente y facilita la cuantificación de los ítems (ver también, Rivas y Saiz, 2010).

El objetivo del presente estudio es la validación en población española de la prueba de Pensamiento Crítico PENCRISAL.

2.- Metodología

2.1.- Elaboración del instrumento y procedimiento

En una 1ª fase se confeccionó un amplio banco de ítems para poder hacer una buena selección. Esta 1ª 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 reelaborando aquellos que podían aún alcanzar las propiedades que se precisan.

En función de estos resultados, se elaboró una 2ª versión, que se aplicó en una muestra de 313 estudiantes universitarios. Los resultados del estudio psicométrico mediante análisis factorial mostraron ya un conjunto de factores y subfactores que explicaban el 59,35% de la variabilidad total. La mayoría de los ítems (un 80%) demostraron correctamente su pertenencia a los factores teóricos esperados. En general, se puede considerar que la escala ya estaba demostrando unos buenos resultados. Este análisis nos permitió realizar las modificaciones necesarias para ajustar las propiedades de los ítems, fundamentalmente reducir el índice de dificultad de algunos ítems que era muy elevado, y mejorar el ajuste al modelo factorial teórico (Saiz y Rivas, 2011).

El presente estudio se enmarca dentro de la 3ª fase de la investigación, donde se presenta la validación de la 3ª versión de la prueba.

2.2. - Características del instrumento

El PENCRISAL es una prueba que consta de 35 situaciones-problema de producción de respuesta abierta. Los enunciados se han diseñado de tal manera que no requieren que la respuesta se elabore y se exprese en términos técnicos, más bien al contrario, se pueden redactar sin dificultad en lenguaje coloquial. Estos 35 ítems se configuran en torno a 5 factores: razonamiento deductivo, inductivo y práctico, toma de decisiones, y solución de problemas, a razón de 7 ítems por factor (véase Anexo I). 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. 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 de administración del PENCRISAL podría realizarse en formato lápiz y papel de forma colectiva, sin embargo hemos optado por la aplicación de la prueba informatizada, vía internet y de manera individual, por ser la que más ventajas ofrece. Estas se dan tanto para los correctores, ya que facilita la tediosa tarea del volcado de datos, así como para la persona que realiza la prueba, puesto que el sistema permite que ésta pueda realizarse en varias sesiones, reduciendo así los posibles efectos del cansancio que pueda darse, especialmente en el rendimiento de los últimos ítems. El sistema también permite controlar todos los aspectos relevantes de la prueba, tales como evitar que dejen ítems en blanco y que puedan corregir respuestas o volver a realizar la prueba una vez terminada. La versión de internet permite a los participantes poder realizar la prueba en cualquier lugar donde se disponga de una conexión a la red de redes. Otras ventajas de la recogida de datos on line son sobradamente conocidas y no vamos a detenernos en ellas. Por lo tanto, la aplicación de la prueba a través de internet parece el medio más conveniente.

El formato de los ítems es abierto, de manera que la persona debe responder a una pregunta concreta añadiendo a ésta una justificación del porqué 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 el porqué (en donde se hace uso de procesos más complejos que implican verdaderos mecanismos de producción).

De esta manera se está utilizando 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, nuestra 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 (2008).

Las dimensiones del test deben considerarse de manera multidimensional, en los siguientes términos. El pensamiento crítico tal como lo concebimos es lo que 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 con una de estas actividades fundamentales solo. Se necesita de la cooperación de todas o una parte, según situaciones. Por esta razón, las dimensiones de nuestra 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 sean algo difíciles de entender los resultados multidimensionales de nuestra 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.

2.3.- Instrumentos utilizados

Cornell Critical Thinking Test (Level Z) (Ennis et al., 1985). Consta de 52 ítems con tres alternativas de respuesta. Evalúa las siguientes habilidades: inducción, deducción, observación, credibilidad, suposiciones, y el significado. Se realizó la traducción del test y se

aplicó a través de internet, manteniendo todas las exigencias de la prueba original (rxx entre ,500 y ,770).

PMA, Test de Aptitudes Mentales Primarias (Thurstone ,1976). Consta de 5 factores básicos de inteligencia: Verbal (rxx=,910), Espacial (rxx=,730), Numérico (rxx=,990), Razonamiento (rxx=,920) y Fluidez Verbal (rxx=,730).

3.- Participantes

Para la validación final de la versión española de PENCRISAL, se diseñó una muestra con un tamaño mínimo de 784 casos para una confianza del 95%, una potencia del 80%, p=q=0,50 y un error máximo del 3,5%. Se decidió emplear un método de muestreo intencionado y por conveniencia, ante la imposibilidad logística de encontrar sujetos por m.a.s. (muestreo aleatorio simple). La muestra finalmente conseguida fue semejante en tamaño, 753 casos, aunque en el análisis exploratorio previo hubo que eliminar algunos casos, debido a cuestionarios incompletos, a respuestas malintencionadas que denotaban falta de participación y a valores extremos. El número total de casos analizados, finalmente, fue de 715 (el 91,1% del diseñado) que representa perfectamente a la población española adulta, de nivel cultural universitario.

De estos 715 participantes, un 30,8% (220) son hombres y un 69,2% (495) mujeres. La edad media de todos ellos es de 24,35 años (IC 95%: 23,88–24,81) con desviación típica de 6,28 años. Esta variable no se distribuye normalmente con p<,050 (Test K-S: Z=5,89; p=,000) debido a una marcada asimetría positiva (As=1,502) y a una altura mayor a la normal (K=2,33). Con una mediana de 21 años, el 50% central se encuentra comprendido entre los 20 y los 28 años, siendo el rango completo: 18–53 años. La edad media de los varones es 24,90 (IC al 95%: 24,03–25,76) y la de las mujeres es 24,10 (IC 95%: 23,56–24,64); esta diferencia no es significativa con p>,050 (T de Student: t=1,56; 713 gl; p=,118). Por rangos, un 57,5% (411 casos) están aún en edad universitaria de grado (hasta 22 años), un 42,5% (304) entre universitarios de post-grado y profesionales en ejercicio.

Esta muestra de 715 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 la estabilidad temporal, la fiabilidad interjueces y la validez convergente-discriminante, se han empleado diferentes submuestras, extraídas aleatoriamente de entre los 753 participantes iniciales, antes de comenzar con los análisis estadísticos, intentando con ello evitar posibles sesgos.

4. – Análisis estadístico

El análisis de datos se realizó con el paquete estadístico IBM-SPSS Statistics-19. Se emplearon pruebas de bondad de ajuste de Kolmogorov-Smirnov (K-S) para comprobar que las diferentes variables numéricas seguían el modelo de la campana normal de Gauss. 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, estimado mediante Pearson. Para el análisis de la fiabilidad, se empleó: coeficiente alfa de Cronbach y Pearson para la estabilidad temporal. 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 similaridad, se decidió finalmente por optar por las que se encontraron a traés del método Varimax. 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. Las correlaciones para la validez convergente y divergente se realizaron mediante coeficientes de Pearson.

5.- Resultados

5.1.- Puntuaciones del Pencrisal y baremación

Las puntuaciones totales del Pencrisal, en la muestra de 715 participantes analizada, se distribuyen con media 27,48 (IC 95%: 27,00–27,95) y desviación típica 6,49 para un rango de puntuaciones: 12–44. La distribución de estos valores presenta una muy ligera desviación del modelo normal de la campana de Gauss con p<,050 pero tolerable (p=,039>,001 en el test K-S). Se ha construido un baremo en percentiles para la población general, dado que no existen diferencias significativas ni por sexos ni por edad, y cada uno de los factores (ver tabla 1).

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
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

Tabla 1. PENCRISAL: Baremos para población general

5.2.- Análisis de items

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 podemos demostrar el efecto de la intervención, sin necesidad de diseñar otro instrumento paralelo para este propósito. Advertido lo anterior, la dificultad de los ítems, varía entre 0,80–0,06 con

media 0,39 (IC 95%: 0,34–0,45) y desviación estándar de 0,16. De ellos, 18 ítems (el 51,4%) presentan un rango de dificultad media, 3 (8,6%) son fáciles (ID>0,65) y los 14 restantes (40%) pueden ser considerados como de dificultad alta (ID<0,35).

El índice de homogeneidad corregido de cada uno de ellos con respecto a la escala total, es altamente significativo en todos ellos con p<,001. El rango de estos índices es: ,172–,383.

5.3.- Consistencia interna y fiabilidad

El estudio de la fiabilidad se ha realizado desde las perspectivas de consistencia interna, estabilidad temporal y concordancia entre jueces, cuestión ésta 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 ,632 altamente significativo con p<,001 (n=715; Anova: F=174,73; 34 y 24276 gl; p=,000), lo que indica que el grado de homogeneidad entre los ítems es bastante aceptable.

La fiabilidad como estabilidad temporal se estimó mediante el método del retest. Se seleccionó una submuestra aleatoria de 130 casos, a quienes se les aplicó de nuevo la prueba entre 4 y 5 semanas después de la primera aplicación. Los resultados demuestran una buena estabilidad con coeficiente de Pearson elevados y significativos tanto en la puntuación total (r=,786; p<,001) como para cada una de las subescalas. Ver tabla 2.

Variables	1ª aplicación		Retest		Correlación	test-retest
	Med.	d.s.	Med.	d.s.	r	p
P.TOTAL	26,44	5,49	26,61	5,31	,786	,000
R.D.	3,93	1,93	3,83	1,95	,599	,000
R.I.	5,12	1,34	5,18	1,63	,467	,000
R.P.	5,52	2,07	5,79	1,92	,465	,000
T.D	5,94	1,87	5,78	1,97	,548	,000
S.P.	5,93	2,08	6,04	2,11	,556	,000

Tabla 2. Fiabilidad según método del retest

Para la fiabilidad interjueces, dada la complejidad que requiere la corrección de los ítems del test, se seleccionó otra submuestra aleatoria de 100 participantes. Dichos cuestionarios fueron corregidos de forma independiente por 3 jueces debidamente formados en esta tarea. Durante este proceso se observaron algunos cuestionarios incompletos, por lo que el número de casos analizado para esta parte del estudio, varía entre 91 y 96. Posteriormente se cruzaron los resultados de los 3 jueces entre sí y se estimaron todos los coeficientes Kappa de Cohen. Los resultados obtenidos se pueden ver en tabla 3 e indican que en todos los casos se han encontrado coeficientes mayores a 0,500 y la mayoría de ellos tiene valores por encima de 0,600 por lo que pueden ser calificados de concordancia buena según el criterio de Landis y Koch (1977). La media de la concordancia entre los jueces 1 y 2 es de 0,738 (IC 95%: 0,70–0,78) con un rango de ,515 a ,970. La media de la coincidencia de los correctores 1 y 3 es ,677 (IC 95%: 0,64–0,72) con un rango de ,510 a ,979. Y por último la media de la fiabilidad entre los jueces 2 y 3 es 0,627 (IC 95%: 0,59–0,66) con un rango de ,503 a ,939. Todos estos índices han resultado ser altamente significativos con p<,001.

Deducción	C1-C2	C1-C3	C2-C3
1	,727	,586	,594
3	,970	,587	,564
5	,716	,657	,546
8	,637	,606	,503
16	,827	,821	,664
23	,834	,539	,535
28	,862	,597	,666
Inducción	C1-C2	C1-C3	C2-C3
2	,553	,662	,519
4	,919	,838	,738
6	,630	,769	,572
9	,659	,622	,556
10	,608	,628	,657
24	,565	,510	,552
29	,658	,590	,580
R.P.	C1-C2	C1-C3	C2-C3
7	,667	,716	,547
11	,752	,637	,606
21	,674	,647	,646
25	,630	,758	,677
30	,760	,828	,711
31	,718	,598	,569
34	,908	,536	,519
T.D.	C1-C2	C1-C3	C2-C3
14	,785	,692	,581
17	,721	,827	,605
18	,844	,663	,752
19	,515	,670	,540
20	,672	,558	,601
27	,742	,643	,609
32	,699	,665	,661
S.P.	C1-C2	C1-C3	C2-C3
12	,835	,949	,879
13	,747	,590	,661
15	,959	,979	,939
22	,733	,632	,522
26	,729	,516	,615
33	,858	,903	,901
35	,717	,665	,544
Media de índices K	,738	,677	,626
P	<,000	<,000	<,000

Tabla 3. Valores de coeficientes Kappa inter-jueces

5.4.- Validez de constructo

Los diferentes niveles de nuestra actividad mental deben relacionarse, integrarse para así ser eficaces en la acción. Por ello, y dada la demostrada multidimensionalidad del constructo de pensamiento crítico se decidió comenzar el estudio de la validez de constructo aplicando el Análisis Factorial de forma independiente a cada uno de ellos. En todos estos análisis se han cumplido satisfactoriamente las condiciones previas de adecuación muestral (KMO>,500) y esfericidad (test de Bartlett con p<,001), con determinantes de las matrices de correlación próximos a 0. Los valores concretos de cada caso se encuentran en las tablas respectivas, donde se puede comprobar que en todos los casos se cumplen las condiciones necesarias para la utilización de esta técnica estadística.

A continuación se exponen los resultados para cada una de las 5 dimensiones, que demuestran el adecuado ajuste al modelo teórico de partida y que ya había aparecido en los estudios realizados con las versiones anteriores de la prueba.

a) *Deducción*. Ver tabla 4. Se demuestra que 4 ítems se agrupan en torno al factor deducción *Proposicional*, con una saturación que se encuentra en el rango ,495–,720. Los otros 3 ítems se han agrupado en torno al subfactor deducción *Categórica* con cargas factoriales entre ,597–,706. La variabilidad total interna explicada por los ítems de esta dimensión es del 44,83%. La dimensión deducción explica muy cerca de un 10% de la variabilidad total del Pencrisal.

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	25,43	5.48
Ded. Categórico	3	5; 23; 28	-	1	19,41	4.18
Total Dimensión	7	-	,371	2	44,83	9.66

Tabla 4. Estructura Factorial y Fiabilidad de la dimensión: DEDUCCIÓN (7 ítems). *Condiciones*: KMO=,634; Bartlett p<,001

b) *Inducción*. Ver tabla 5. Tres ítems con pesos factoriales comprendidos en el rango ,562–,674 se configuran como razonamiento *Analógico*. Otros dos definen el factor inductivo *Causal*, con saturaciones de ,649 y ,816. Y los dos últimos, con cargas de ,680 y ,765 constituyen los procedimientos de *Verificación* (comprobación de hipótesis y generalizaciones inductivas). La variabilidad interna explicada por todos ellos alcanza el 50,59%; mientras que el factor inducción explica casi un 11% de la variabilidad total de la prueba.

Componente	Nº ítems	Ítems	α Cronbach	Factores	% de Varianza explicada en su Dimensión	% de Varianza total explicada
Induct. Razon. Analóg.	3	6; 9; 24	-	1	19,23	4,14
Induct. Causal	2	2; 10	-	1	16,02	3,46
Induct. Proc. Verificación (CH y GI)	2	4; 29	-	1	15,32	3,30
Total Dimensión	7	-	,250	3	50,59	10,90

Tabla 5. Estructura Factorial y Fiabilidad de la dimensión: INDUCCIÓN (7 ítems). *Condiciones*: KMO=,575; Bartlett p<,001

c) *Razonamiento práctico*. Ver tabla 6. Cuatro ítems se agrupan en la dimensión *Argumentación*, con cargas factoriales incluidas en el rango ,525–,753; mientras que los otros tres configuran el componente *Falacias* con saturaciones entre ,483–,634. La variabilidad total explicada internamente alcanza el 40,38%. La dimensión razonamiento práctico explica aproximadamente un 9% de la variabilidad total.

Componente	Nº ítems	Ítems	α Cronbach	Factores	% de Varianza explicada en su Dimensión	% de Varianza total explicada
Argumentación	4	7; 21; 25; 30	-	1	24,05	5,18
Raz. Práct.: Falacias	3	11; 31; 34	-	1	16,32	3,52
Total Dimensión	7	-	,425	2	40,38	8,70

Tabla 6. Estructura Factorial y Fiabilidad de la dimensión: RAZONAMIENTO PRÁCTICO (7 ítems). *Condiciones*: KMO=,624; Bartlett p<,001

d) *Toma de decisiones.* Ver tabla 7. En este componente se identifican 4 subfactores, todos ellos constituidos por 2 ítems, ya que uno de ellos, satura en dos de los subfactores identificados. El factor de *TD General* con saturaciones superiores a ,806 explica un 19,70 de la variabilidad interna del factor. La *TD Probabilidad*, con pesos de ,512 y ,859, explica un 15,02%. *TD Heurísticos generales* (representatividad y disponibilidad) con saturaciones de ,523 y ,698 explica un 17,21%. Y por último, la *TD Heurísticos específicos* (disponibilidad y coste de inversión), con saturaciones de ,527 y ,905 explica un 15,94%. Como se ve, el ítem de disponibilidad satura en estos dos últimos subfactores. En el de heurísticos generales se explica ya que, aunque los dos ítems conceptualmente sean diferentes, lo que ponen en funcionamiento es el mismo tipo de estrategias generales para estimar probabilidades de ocurrencia de acontecimientos. Sin embargo, el coste de inversión, del factor TD heurísticos específicos, es una estrategia que depende en una parte de la disponibilidad y no de la representatividad. Por esta razón, se agrupa como factor distinto del general. La variabilidad total interna explicada llega hasta el 67,87%. El componente toma de decisiones es el que mayor peso tiene dentro de la prueba completa ya que explica un 14,61% de la variabilidad total.

Componente	Nº ítems	Ítems	α Cronbach	Factores	% de Varianza explicada en su Dimensión	% de Varianza total explicada
TD General	2	14; 27	-	1	19,70	4,24
TD: Heurísticos generales (REPy DIS)	2	19; 20	-	1	17,21	3,71
TD: Heurísticos específicos (DIS y CI)	2	18; 20		1	15,94	3,43
TD Probabilidad	2	17; 32	-	1	15,02	3,23
Total Dimensión	7	-	,213	4	67,87	14,61

Tabla 7. Estructura Factorial y Fiabilidad de la dimensión: TOMA DECISIONES (7 ítems). *Condiciones*: KMO=,575; Bartlett p<,001

e) *Solución de problemas (S.P.).* Ver tabla 8. En el subfactor *S.P. general* se han encuadrado 4 ítems con cargas de valores en el rango ,511–,710; mientras que los otros 3 ítems constituyen el componente *S.P. específico* (búsqueda de regularidades y análisis medio-fin) con cargas factoriales entre ,548–,705. Estos ítems explican un 38,96% de la variabilidad total específica y el factor S.P. un 8,4% de la variabilidad total del Pencrisal.

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	19,56	4,21
S.P. Especifico (RGLy MF)	3	12; 15; 33	-	1	19,40	4,18
Total Dimensión	7	-	,373	2	38,96	8,39

Tabla 8. Estructura Factorial y Fiabilidad de la dimensión: SOLUCIÓN PROBLEMAS (7 ítems). *Condiciones*: KMO=,624; Bartlett p<,001

Se calcularon las correlaciones entre los cinco factores anteriormente descritos y con la puntuación total (ver tabla 9). Se obtienen coeficientes de correlación estadísticamente significativos dado el tamaño de la muestra, pero de intensidades entre factores (desde ,103 hasta ,291). Esto apoya la multidimensionalidad del constructo y la independencia entre factores.

		RD	RI	RP	SP	TD	Total
RD	Correlación de Pearson	____					
	Sig.						
	N						
RI	Correlación de Pearson	,204	____				
	Sig.	,000					
	N	715					
RP	Correlación de Pearson	,254	,289	____			
	Sig.	,000	,000				
	N	715	715				
SP	Correlación de Pearson	,103	,235	,291	____		
	Sig.	,003	,000	,000			
	N	715	715	715			
TD	Correlación de Pearson	,115	,149	,176	,206	____	
	Sig.	,001	,000	,000	,000		
	N	715	715	715	715		
Total	Correlación de Pearson	,558	,569	,713	,638	,516	
	Sig.	,000	,000	,000	,000	,000	____
	N	715	715	715	715	715	

Tabla 9. Matriz de intercorrelaciones de los factores y con el total del PENCRISAL

En cuanto al análisis factorial del conjunto completo de los 35 ítems, (KMO=,683; test de Bartlett: Chi2=1988,39: 595 gl; p=,000) revela la existencia entre factores y subfactores de 13 componentes que coinciden con el desglose anterior: 2 en deducción, 3 en inducción, 2 en razonamiento práctico, 2 en solución de problemas y los 4 restantes en toma de decisiones. Las saturaciones de los ítems se encuentran en el rango ,400–,762. La variabilidad total de la prueba explicada por este conjunto de factores y subfactores se acerca al 53% como se observa en tabla 10.

	Suma de las saturaciones al cuadrado de la rotación		
Componente	Total	% de la varianza	% acumulado
1	1,914	5,467	5,467
2	1,692	4,835	10,303
3	1,664	4,755	15,057
4	1,469	4,198	19,255
5	1,464	4,184	23,439
6	1,396	3,988	27,427
7	1,369	3,911	31,338
8	1,299	3,713	35,051
9	1,281	3,661	38,712
10	1,277	3,647	42,359
11	1,275	3,642	46,001
12	1,201	3,433	49,434
13	1,153	3,293	52,727

Tabla 10. Variabilidad explicada en el A.F. de C.P. con rotación Varimax de la prueba completa (35 ítems)

5.5.- Validez convergente y divergente

Para el análisis de ambos tipos de validación se tomó una nueva submuestra aleatoria de 130 participantes. En el estudio exploratorio previo se decidió eliminar algún caso debido a valores extremos, pero la pérdida es mínima.

Para esta parte del estudio, se aplicó el test de Cornell de pensamiento crítico, por ser uno de los más utilizados.

Tras comprobar la linealidad de la relación, se procedió a correlacionar las puntuaciones del Pencrisal con las puntuaciones del Cornell (ver tabla 11). Los coeficientes obtenidos no son en su mayoría estadísticamente significativos (p>,050). Estos resultados no apoyan la validez convergente

		RD PENCRISAL	RI PENCRISAL	RP PENCRISAL	SP PENCRISAL	TD PENCRISAL	TOTAL PENCRISAL
RI CORNELL	Correlación de Pearson	,080	,101	,128	,192	,130	,211
	Sig.	,372	,258	,150	,031	,146	,017
	N	127	127	127	127	127	127
RD CORNELL	Correlación de Pearson	,099	,125	,000	-,083	,092	,059
	Sig.	,269	,161	,998	,355	,301	,513
	N	127	127	127	127	127	127
TOTAL CORNELL	Correlación de Pearson	,066	,220	,197	,152	,046	,224
	Sig.	,461	,013	,026	,088	,611	,001
	N	127	127	127	127	127	127

Tabla 11. Correlaciones entre Cornell y Pencrisal

Para la validez divergente se les administró el test de inteligencia PMA (Aptitudes Mentales Primarias). Las correlaciones encontradas son mayoritariamente no significativas (p>,050) y las que resultaron serlo presentan intensidades muy bajas (r<,200) lo que demuestra claramente la ausencia de asociación teórica entre las pruebas, y defiende la divergencia (tabla 12).

		RD PENCRISAL	RI PENCRISAL	RP PENCRISAL	SP PENCRISAL	TD PENCRISAL	TOTAL PENCRISAL
PMA.V	Correlación de Pearson	-,067	,165	,114	,198	,109	,169
	Sig.	,454	,063	,202	,025	,221	,057
	N	127	127	127	127	127	127
PMA.E	Correlación de Pearson	,072	-,010	,140	,141	,157	,174
	Sig.	,418	,910	,117	,114	,077	,050
	N	127	127	127	127	127	127
PMA.R	Correlación de Pearson	,025	,109	,001	,199	,204	,169
	Sig.	,778	,221	,992	,025	,021	,057
	N	127	127	127	127	127	127
PMA.N	Correlación de Pearson	-,093	,031	-,002	-,028	-,020	-,040
	Sig.	,298	,733	,987	,754	,824	,652
	N	127	127	127	127	127	127
PMA.F	Correlación de Pearson	,157	-,058	,137	-,120	-,033	,037
	Sig.	,078	,519	,124	,180	,716	,682
	N	127	127	127	127	127	127
PMA TOTAL	Correlación de Pearson	,057	,049	,159	,126	,143	,181
	Sig.	,525	,583	,074	,157	,110	,041
	N	127	127	127	127	127	127

Tabla 12. Correlaciones entre PMA y Pencrisal

6.- Conclusiones

El PENCRISAL se presenta como un instrumento útil y novedoso para la evaluación de las habilidades de pensamiento crítico, que demuestra su validez en población española con nivel educacional universitario.

El PENCRISAL aporta una serie de ventajas en la evaluación: 1) esta medida, muy innovadora, junto con el HCTAES, son las únicas pruebas de pensamiento crítico enfocadas hacia los procesos de pensamiento crítico, 2) contribuye a mejorar la evaluación de las habilidades de pensamiento crítico de manera integrada, ya que en la actualidad no existen instrumentos de esta naturaleza en español, y 3) el utilizar como ítems situaciones cotidianas que se puedan resolver de una única forma y el formato de respuesta abierta, hacen del PENCRISAL una herramienta precisa de medida en pensamiento crítico.

En cuanto al estudio de las propiedades psicométricas de la prueba se ha conseguido demostrar estadísticamente el adecuado ajuste de la estructura factorial del test al modelo teórico planteado en población española adulta de nivel cultural universitario. Así mismo, en lo relativo a la validez convergente y divergente, la prueba PENCRISAL ha demostrado una elevada potencia divergente respecto a constructos teóricos de capacidades intelectuales. Por su parte, la ausencia de otras pruebas específicas que midan los mismos rasgos y con el mismo tipo de formato de respuesta abierta dificulta alcanzar una sólida validez convergente. La ausencia de validez convergente con respecto al Cornell se debe a la naturaleza de los instrumentos. El Cornell es una prueba cerrada y de comprensión, el pencrisal, por el contrario, es abierta y de producción. Esto hace que la forma de responder sea lo suficientemente distinta como para arrojar resultados diferentes. Lo que se pide en cada una marca esta diferencia. Nuestra prueba exige desarrollar una explicación para cada respuesta, en el Cornell no, solo marcar. Por lo que el rendimiento y su naturaleza dificulten la obtención de esa validez. Sin embargo, esto pone de manifiesto el aspecto diferencial e innovador de esta prueba con respecto a las existentes actualmente en el campo de la evaluación de las habilidades de pensamiento crítico.

En cuanto al estudio de la fiabilidad, ha quedado probada una elevada estabilidad temporal con el procedimiento retest del instrumento de medida.

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. Estas correlaciones tanto en la puntuación total del test como en los 5 factores del mismo, son altamente significativas y con valores de correlación altos.

Entre las limitaciones de la prueba podemos destacar, en primer lugar, que el constructo que evalúa, las habilidades de pensamiento crítico es un constructo muy complejo que puede ser definido desde marcos teóricos muy diversos, dando como resultado instrumentos de diferente naturaleza. En segundo lugar, el PENCRISAL presenta las limitaciones propias de las pruebas de respuesta abierta. El sistema de corrección requiere de evaluadores expertos, y el tiempo de corrección de los protocolos de respuesta es elevado. Por último, somos conscientes de que el procedimiento de análisis factorial por dimensiones no es muy común y se ha realizado de esta manera por la peculiaridad y complejidad de la prueba. Como se ha podido comprobar, la composición del análisis factorial conjunto con la totalidad de los 35 ítems se corresponde exactamente con cada uno de los factores y subfactotres descritos en los análisis realizados por dimensiones. Resultaría mucho más complejo para el lector interpretarlo y comprenderlo desde la matriz de los 13 componentes, que desde cada uno de los factores por separado. Y por eso se presenta de esta manera. Es evidente, que visto en su globalidad, 35 ítems agrupados en 13 factores y subfactores implican 2 ó 3 ítems por factor, que no es lo ideal. Pero dado el tiempo que exige la realización de la prueba no es aconsejable añadir más ítems, ya que sería un instrumento inaplicable en cuanto al tiempo que requeriría puesto que, recordemos, quien contesta, debe justificar lo que responde, esto es, debe producir una respuesta extensa. Desde estas reflexiones nos estamos planteando para el fututo reconvertir la prueba en una batería que esté compuesta por 5 subescalas que se correspondan con los 5 constructos teóricos estudiados y que pudiera tener un mayor número de ítems para cada una de ellas.

Dadas las características de la prueba PENCRISAL consideramos 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. Sin embargo, para el futuro, debemos mejorar aún algunos aspectos del test que vienen determinados por las limitaciones señaladas anteriormente. Es importante trabajar el instrumento con el fin de conseguir una mayor precisión dimensional, fusionando algunos de los subfactores propuestos. También, dada la complejidad y naturaleza de la prueba, y siendo conscientes de que los índices psicométricos son mejorables, sería conveniente hacer un esfuerzo mayor en esta dirección. Y finalmente, sería necesario desarrollar una automatización de la corrección de la prueba, por procedimientos de categorización semántica. Todas estas mejoras están ya en marcha en diferentes proyectos que estamos desarrollando.

7.- Referencias

Ennis, R. H. (2003). Critical thinking assessment. En D. Fasko (Ed.), *Critical thinking and reasoning. Current research, theory, and practice.* (pp. 293-313). Cresskill, NJ: Hampton Press.

Ennis, R.H., Millman, J., & Tomko, T.N. (1985). *Cornell Critical Thinking Test, Level X & Level Z-Manual* (3rd ed.). Pacific Grove, CA: Midwest.

Halpern, D.F. (2006). *Halpern Critical Thinking Assessment Using Everyday Situations: Background and scoring standards*. Unpublished report.

Rivas, S.F. y Saiz, C. (2010). ¿Es posible evaluar la capacidad de pensar críticamente en la vida cotidiana? En Jales, H.R. y Neves, J. (Eds.), *O Lugar da Lógica e da Argumentação no Ensino da Filosofia* (53-74). Coimbra: Unidade I&D, Linguagem, Interpretação e Filosofia

Saiz, C. y Rivas, S.F. (2008). Evaluación en pensamiento crítico: una propuesta para diferenciar formas de pensar *Ergo, Nueva Época, 22-23*, 25-66.

Saiz, C. y Rivas, S.F. (2011). Evaluation of the ARDESOS program: an initiative to improve critical thinking skills. *Journal of the Scholarship of Teaching and Learning, Vol. 11, No. 2*, 34-51.

Thurstone, L.L., &Thurstone. T.G. (1976). *PMA: Aptitudes Mentales Primarias.* Madrid: TEA

ANEXO I

DISTRIBUCIÓN DE ITEMS Y FACTORES DEL PENCRISAL

	FACTORES				
Item	**DEDUCCIÓN**	**INDUCCIÓN**	**RZ. PRÁCTICO**	**TOMA DECISIONES**	**SOLUCIÓN DE PROBLEMAS**
1	R. Proposicional				
2		R. Causal			
3	R. Proposicional				
4		Comprobación de Hipótesis			
5	R. Categórico				
6		R. Causal			
7			Argumentación		
8	R. Proposicional				
9		R. Analógico			
10		R. Causal			
11			Falacia		
12					Regularidades
13					General
14				General	
15					Regularidades
16	R. Proposicional				
17				Probabilidad	
18				Coste de Inversión	
19				Representatividad	
20				Disponibilidad	
21			Argumentación		
22					General
23	R. Categórico				
24		R. Analógico			
25			Argumentación		
26					General
27				General	
28	R. Categórico				
29		Generalización Inductiva			
30			Argumentación		
31			Falacia		
32				Probabilidad	
33					Medio Fin
34			Falacia		
35					General

Deducción 7 ítems	**Inducción 7 ítems**	**Raz. Práctico 7 ítems**	**Toma Decisiones 7 ítems**	**Solución de Problemas 7 ítems**
RPR = 4 **RCT = 3**	**RC = 3** **CH = 1** **RA = 2** **GI = 1**	**ARG = 4** **FAL = 3**	**GRAL = 2** **PRB = 2** **CI = 1** **REP = 1** **DIS = 1**	**GRAL = 4** **RGL = 2** **MF = 1**