



Evaluación de destrezas de pensamiento crítico: validación de instrumentos libres de cultura

- **Assessment of Critical Thinking Skills: Validation of Free-Culture Tools**
- **Avaliação de habilidades de pensamento crítico: validação de instrumentos livres de cultura**

Resumen

Es un lugar común que el pensamiento crítico es una capacidad importante en la educación STEM (ciencia, tecnología, ingeniería y matemáticas). Sin embargo, la investigación ha dedicado más atención a los programas de enseñanza del pensamiento crítico que a evaluar los efectos de esos programas, de modo que los estudios empíricos constituyen la excepción en este tema. Este estudio presenta la validación piloto de dos pruebas de evaluación de destrezas de pensamiento crítico con dos grupos naturales de estudiantes de primaria y secundaria, que participan en un programa de aprendizaje basado en pensamiento. Las pruebas se diseñaron adaptadas a las destrezas enseñadas y a la edad de los estudiantes, combinando cuestiones de instrumentos ya publicados y cuestiones figurativas de nuevo diseño, independientes de la cultura y de los conocimientos escolares. Los resultados muestran una relación positiva entre las puntuaciones de las pruebas y los criterios de validación independientes y externos (calificaciones escolares y notas del programa de pensamiento de los estudiantes). Se informan una correlación excepcionalmente alta en una de las pruebas y valores aceptables de la fiabilidad de las escalas; además, las nuevas cuestiones figurativas presentan mejor fiabilidad y validez que las verbales. Finalmente, se discuten las aportaciones de este estudio piloto, la toma de decisiones futuras para mejorar los instrumentos presentados y el desarrollo del diseño de instrumentos libres de cultura a medida, iniciados en este estudio.

Palabras clave

pensamiento; evaluación; enseñanza; método de evaluación

María-Antonia Manassero-Mas*
Ángel Vázquez-Alonso**

* Doctora en Psicología, Catedrática de Psicología Social, Departamento de Psicología, Universidad de las Islas Baleares, Palma, España.
ma.manassero@uib.es
ORCID: <http://orcid.org/0000-0002-7804-7779>

** Doctor en Ciencias de la educación, profesor e investigador, Centro de Estudios de Posgrado, Universidad de las Islas Baleares, Palma, España.
angel.vazquez@uib.es
ORCID: <http://orcid.org/0000-0001-5830-7062>



Abstract

It is a commonplace that critical thinking is an important ability for STEM (science, technology, engineering and mathematics) education. Nevertheless, research has devoted more attention to critical thinking programs than to evaluating the effects of these programs, so that the empirical studies are the exception in this area. This study presents the pilot validation of two assessment tools on critical thinking skills using two natural groups of elementary and secondary students, who participate in a thinking-based learning program. The tools are designed to fit the skills developed in the program and the students' age, and combine some items drawn from already published critical thinking tests and some newly designed figurative items, which are independent of school's culture and knowledge. The results show a positive relationship between the assessment test scores and the external and independent validation criteria (students' school grades and thinking program marks). There is an extremely relevant correlation for one of the tools, acceptable reliability indices for the scales and better reliability and validity indices of the newly designed figurative items than the classical verbal items are the main contributions of the study. Finally, the future decision making to improve the tools, the development of research on free-culture customized tests, as initiated with this study, and the contributions of this pilot study are discussed.

Keywords

thinking; evaluation; teaching; assessment methods

Resumo

É um lugar comum que o pensamento crítico é uma capacidade importante da educação STEM (ciência, tecnologia, engenharia e matemática). No entanto, a pesquisa tem dedicado mais atenção aos programas de ensino do pensamento crítico que à tarefa de avaliar os efeitos desses programas, de modo que os estudos de avaliação empírica são a exceção neste tópico. Este estudo apresenta a validação piloto de dois testes de avaliação de habilidades de pensamento crítico com dois grupos naturais de alunos do ensino fundamental e médio, que participam de um programa de aprendizagem baseado em pensamento. Os testes são projetados de acordo com as habilidades ensinadas e a idade dos alunos, combinando questões de instrumentos já publicados e questões figurativas recém-projetadas, independente da cultura e do conhecimento da escola. Os resultados mostram uma relação positiva entre os resultados dos testes e os critérios de validação independentes e externos (notas escolares e anotações do programa de pensamento do aluno). Uma correlação excepcionalmente alta é fornecida em um dos testes e valores aceitáveis da confiabilidade das escalas; além disso, as novas questões figurativas apresentam melhor confiabilidade e validade que as verbais. No final, são discutidas as contribuições deste estudo piloto, futuras tomadas de decisão para melhorar os instrumentos apresentados e o desenvolvimento do desenho de instrumentos livres de cultura para medir, iniciados neste estudo.

Palavras-chave

pensamento; avaliação; ensino; métodos de avaliação

Introducción

El pensamiento es una capacidad y un rasgo distintivo de la naturaleza humana. Sin embargo, la mayor parte del pensamiento espontáneo suele ser de baja calidad y defectuoso (desinformado, parcial, tendencioso, injusto, cargado de prejuicios o intereses, erróneo, inválido, etc.). Por ello, una característica consensual y básica del pensamiento es su normatividad, es decir, debería ser sinónimo de pensamiento de calidad (correcto, bueno) y esta normatividad se considera incluida ya en la denominación pensamiento crítico (Bailin, Case, Coombs y Daniels, 1999).

Pensar correctamente es central en las comunidades científico-técnicas y en el aprendizaje de materias STEM (ciencia, tecnología, ingeniería y matemáticas), pues pensamiento científico y pensamiento crítico son entidades equivalentes (Vázquez-Alonso y Manassero-Mas, 2018). Pero pensar también debería ser un valor sistemática y socialmente cultivado y promovido; por ello, desde los años 1980 se han generado programas para la enseñanza del pensamiento crítico con diversas orientaciones y prácticas. En todos los casos, parten de la hipótesis básica que el pensamiento puede ser mejorado mediante programas educativos adecuados. Sin embargo, la variedad en los programas e incluso la propia conceptualización de pensamiento crítico es tan amplia entre los especialistas, que resulta difícil advertir posibilidades de conseguir una síntesis virtuosa en la gran variedad de los programas de enseñanza (Follmann, Mattos, Güllich, 2018; Saiz, 2017; Swartz et ál., 2013).

Antecedentes

En general, las definiciones y la literatura sobre pensamiento crítico son muy extensas y complejas, pues abarcan desde los planteamientos generales, que consideran la dimensión de pensamiento, la crítica y los procesos, hasta los reduccionistas, que se centran en alguna parte de esa complejidad, donde quizá los más

comunes son los que identifican pensamiento con razonamiento. Algo análogo sucede en los planteamientos del pensamiento crítico como un conjunto articulado de destrezas, con planteamientos globales y extensos, que incluyen habilidades cognitivas y disposiciones (por ejemplo, <http://www.criticalthinking.org/> y <http://criticalthinking.net>), y otros planteamientos más concretos centrados en algunas destrezas (Fisher, 2009).

La mejora del pensamiento tiene tres puntos clave: la mejora de las destrezas de pensamiento, la permanencia a lo largo del tiempo de las mejoras alcanzadas y la extensión o generalización de las mejoras a contextos y situaciones diferentes de donde se aprendieron. La apreciación de los niveles, cambios y mejoras del pensamiento requieren instrumentos apropiados de evaluación que permitan una medida fiable de esas magnitudes. La eficacia de los programas educativos de pensamiento, para cambiar y mejorar, para hacer permanente la mejora y para extenderla a otras situaciones, requiere la creación de instrumentos para la evaluación del pensamiento crítico que produzcan medidas fiables y válidas de esos tres criterios básicos para el cambio y mejora del pensamiento. Sin embargo, en la investigación sobre educación del pensamiento crítico, las actividades de diseño de programas de intervención para la enseñanza de este han sido dominantes sobre las evaluaciones de los efectos de esos programas, hasta el punto que los estudios empíricos de evaluación que acrediten los efectos de los programas son la excepción más que la regla. Según Saiz (2017), el programa de filosofía para niños desarrollado por Lipman es uno de los programas más evaluados empíricamente (Colom, García, Magro y Morilla, 2014), pero otros como el programa canadiense razonamiento, argumentación y retórica (Walton y Macagno, 2015) carece de evaluaciones de sus efectos, y otros muy extendidos, como el programa de aprendizaje basado en pensamiento (TBL) (Swartz et ál., 2013) solo han sido evaluados ocasionalmente.

Marco teórico

El diseño y construcción de instrumentos de evaluación es una actividad de investigación que requiere tomar numerosas decisiones y está plagada de dificultades. El primer grupo de estas se asocia a cuestiones psicométricas básicas, acerca de la validez y fiabilidad del instrumento de medida; en el caso del pensamiento crítico, las controversias en su definición sugieren ya las previsible grandes dificultades a la hora de superar las cuestiones de validez de los instrumentos. El segundo grupo de dificultades surge de la naturaleza estadística de la medida de la eficacia del tratamiento educativo y está relacionada con cuestiones de atribución de los efectos estadísticos observados a las variables propias del tratamiento; la significación y la potencia de las medidas obtenidas (tamaño del efecto) deben ser suficientes para valorar la eficacia de los tratamientos, desde los tres criterios de cambio, mantenimiento y generalización. El tercer grupo surge de las necesidades de clarificación conceptual y teórica en la investigación del pensamiento crítico, como evaluar la eficacia de los programas dependientes frente a los independientes (del contexto o dominio, específico frente a general), de las metodologías directas o indirectas de enseñanza del pensamiento crítico, de las distintas destrezas cognitivas enseñadas y, en general, la comparación de distintos programas de enseñanza.

Además, por las exigencias de viabilidad y utilidad de su propia finalidad, los instrumentos de evaluación del pensamiento crítico adoptan restricciones que se traducen en una selección estricta de las destrezas que evalúan, de modo que elaboran modelos más concretos, sintéticos y prácticos, los cuales proyectan una imagen más convergente que las divergencias expuestas en las contribuciones mencionadas.

Entre las propuestas de evaluación más amplias, Ennis (2019) sugiere las siguientes dimensiones (compuestas a su vez de varias destrezas): clarificación básica de ideas, bases de decisión, inferencias, clarificación avanzada y metacognición. En el plan nacional para la evaluación del pensamiento crítico, Paul y Nosich (2019) proponen una extensa lista de destrezas de pensamiento crítico (88), agrupadas en las dimensiones siguientes: propósitos, metas o fines, pregunta en cuestión o problema a resolver, punto de vista o marco de referencia, la dimensión empírica del razonamiento (información), la dimensión conceptual del razonamiento, presuposiciones y supuestos, implicaciones y consecuencias, interpretación e inferencias, dimensiones transversales y dimensiones afectivas.

La evaluación propuesta por James Madison (2004) incluye las destrezas siguientes: reconocer y clarificar problemas, reclamaciones, argumentos y explicaciones, distinguir condiciones necesarias y suficientes, identificar y evitar errores y falacias, evaluar si un argumento es fuerte o débil, distinguir conclusiones, premisas (razones), argumentos, explicaciones, supuestos, problemas, conclusiones, conclusiones tácitas, premisas no declarados e implicaciones, reivindicaciones, argumentos, explicaciones, descripciones, representaciones, etc., describir la estructura de argumentos y explicaciones, evaluar argumentos analógicos y generalización inductiva, evaluar un argumento deductivo, interpretar y aplicar textos instrucciones, ilustraciones, etc., reconocer la ambigüedad, evaluar la rele-

vancia de las declaraciones para otras, evaluar argumentos en términos de criterios y discernir si pares de afirmaciones son consistentes, contrarias, contradictorias o paradójicas.

Entre las propuestas de evaluación más sencillas, el cuestionario de situaciones cotidianas de Halpern (2010) evalúa las destrezas resolución de problemas, razonamiento verbal, utilizar probabilidad e incertidumbre, pensamiento como prueba de hipótesis, análisis de argumentos y toma de decisiones. El cuestionario de pensamiento crítico de Rivas y Saiz (2012) de la universidad de Salamanca (Pencrisal) evalúa resolución de problemas, razonamiento inductivo, razonamiento deductivo, razonamiento práctico o informal, y toma de decisiones. El instrumento de Watson y Glaser (2002) propone evaluar deducción, reconocimiento de supuestos, interpretación, inferencia y evaluación de argumentos.

En este marco teórico de gran diversidad, el enfoque del pensamiento crítico desde los procesos y destrezas de pensamiento constituyentes ofrece un planteamiento concreto de formas de razonar o argumentar, de tomar decisiones eficaces, de resolver problemas sólidamente y, en general, de evaluar y juzgar ideas, conclusiones y procesos de razonamiento y argumentación. Elaborando una síntesis empírica de todos esos modelos, estudios e instrumentos de evaluación, hemos propuesto una definición y una taxonomía del pensamiento crítico (Manassero-Mas y Vázquez-Alonso, 2019). Nuestra aportación a la definición de pensamiento crítico es la siguiente: "Pensamiento crítico es el pensamiento claro y preciso en sus justificaciones y conclusiones, que además evalúa y juzga meticulosamente todos sus elementos".

La taxonomía elaborada pretende ofrecer un modelo simple para un constructo complejo como el pensamiento crítico, de modo que sea comprensible y útil a las personas no especialistas, especialmente los docentes encargados de enseñarlo, y está formado por las siguientes cuatro áreas y múltiples destrezas:

- Creatividad (generar ideas, conclusiones):
 - Plantear buenas preguntas.
 - Observación (comparar, clasificar, secuenciar).
 - Análisis y síntesis (partes-todo, analogías, modelos).
- Razonamiento y argumentación (justificar predicciones, implicaciones, conclusiones):
 - Lógico (deductivo).
 - Empírico (explicar con datos, informaciones, pruebas):
 - Inductivo (generalizaciones).
 - Argumentación (abductivo).
 - Estadístico (probabilístico).
 - Falacias y errores.
- Procesos complejos:
 - Toma de decisiones.
 - Resolución de problemas.
- Evaluación y juicio (valoración de la calidad del pensamiento):
 - Estándares intelectuales (claridad, precisión, relevancia, ...).
 - Razonamiento y argumentación.
 - Acciones (soluciones, decisiones, consecuencias, ...).
 - Credibilidad de fuentes.
 - Identificar supuestos.
 - Comunicación (clarificación de significados).
 - Metacognición:
 - Autorregulación y autorreflexión.
 - Actitudes y afectos (disposiciones).

En esta taxonomía, el concepto de pensamiento crítico es fundante, como constructo general que engloba a todas las demás destrezas de pensamiento, y que se desarrolla en cuatro grandes áreas: creatividad; razonamiento y argumentación; procesos complejos;

y evaluación y juicio. Cada una de estas incluye destrezas y subdestrezas, que pueden desarrollarse todavía en constituyentes más simples.

En resumen, la psicología encuadra el pensamiento crítico dentro del constructo habilidades de pensamiento de alto nivel, formado por un conjunto complejo de destrezas de pensamiento, como una forma de pensamiento autodirigido, autodisciplinado, autosupervisado y autocorrectivo que presupone ajuste y conformidad con rigurosos estándares de calidad y dominio consciente de su uso. Además, implica comunicación efectiva, capacidades de razonamiento, toma de decisiones, resolución de problemas y compromiso y actitudes para superar las tendencias naturales a la falacia y al sesgo (egocentrismo y sociocentrismo).

Por otro lado, es un lugar común de la psicología cognitiva el supuesto sobre la relación positiva entre el pensamiento y el aprendizaje, ya que ambos implican múltiples componentes y niveles, cognitivos y no cognitivos, interdependientes e inextricablemente relacionados entre sí. Además, la evidencia empírica actual indica que la dinámica interactiva entre pensamiento crítico y otras variables educativas puede contribuir a desarrollar mejor el aprendizaje (Phan, 2010).

Este estudio plantea el desarrollo de dos pruebas de evaluación específicas para un conjunto de destrezas de pensamiento crítico, educadas en el proyecto de enseñanza del pensamiento crítico desarrollado con los estudiantes participantes en este estudio piloto. Las hipótesis iniciales planteadas establecen la validez y fiabilidad de las pruebas de pensamiento crítico diseñadas y aplicadas a los estudiantes, tomando las calificaciones escolares de los estudiantes participantes como indicador de la validez de criterio de las pruebas y la investigación de la relación empírica positiva con el aprendizaje escolar.

Metodología

En este apartado se presenta la metodología general diseñada para este estudio piloto, donde se describe la muestra de los estudiantes participantes, los instrumentos de evaluación de pensamiento crítico y los procedimientos seguidos en la aplicación piloto.

Participantes

Los participantes en la aplicación piloto presentada en este estudio son dos grupos naturales de estudiantes en un colegio ubicado en una ciudad de tamaño medio, que aplica el programa de enseñanza del pensamiento crítico *teaching-based learning* (TBL) (Swartz et ál., 2013). Se trata, pues, de una muestra opinática, condicionada por la participación de los estudiantes en el programa TBL, que está formada por un grupo clase de 22 niña del grado 6 (curso final de la educación primaria) y 16 en el grupo del grado 8 (segundo curso de educación secundaria).

El pequeño tamaño de la muestra de conveniencia aquí empleada limita las posibilidades de algunos análisis sobre las pruebas, que serían importantes para los objetivos del estudio como los análisis factoriales exploratorios para la confirmación empírica de la validez factorial.

Tabla 1. Especificaciones generales de las pruebas piloto ensayadas en este estudio para evaluar ocho destrezas de pensamiento según los niveles educativos donde se han aplicado (grado 6, sexto curso de educación primaria, EP6, y grado 8, segundo curso de educación secundaria, ESO2)

| Destrezas de pensamiento | Fuente | Tipo | Número de reactivos | |
|--------------------------|--------------------|-------------|---------------------|-------------|
| | | | Prueba EP6 | Prueba ESO2 |
| Predicción | Cornell (Nicoma) | Verbal | 13 | - |
| Comparación y contraste | Cornell (Nicoma) | Verbal | 13 | - |
| Clasificación | Elaboración propia | Figurativo | 8 | - |
| Resolución de problemas | Elaboración propia | Figurativo | 10 | - |
| | Halpern | Situaciones | 1 | |
| Explicación causal | Cornell (Nicoma) | Verbal | - | 13 |
| Toma de decisiones | Cornell (Nicoma) + | Verbal | - | 7 |
| | Adaptación propias | Situaciones | | 4 |
| Secuenciación | Elaboración propia | Figurativo | - | 10 |
| Relación partes-todo | Elaboración propia | Figurativo | - | 4 |

Verbal: preguntas verbales de opción múltiple con un pie y diversas alternativas para elegir una.

Situaciones: plantean una situación real con información verbal y varias preguntas sobre ella.

Figurativo: plantean conjuntos de figuras sobre las que se plantean diversas preguntas.

Fuente: elaboración propia.

Instrumentos

Los instrumentos pilotados fueron diseñados a medida de las ocho destrezas de pensamiento crítico que los estudiantes estaban aprendiendo en cada uno de los cursos, de acuerdo con el programa TBL desarrollado en el colegio (tabla 1). Las destrezas evaluadas en el grado 6 de primaria (en adelante, EP6) son predicción, comparación y contraste, clasificación y resolución de problemas. Las destrezas evaluadas en el grado 8 de educación secundaria obligatoria (en adelante, ESO2) son secuenciación, relación de las partes con el todo, explicación causal y toma de decisiones.

Las cuestiones creadas para la evaluación de esas ocho destrezas forman los dos instrumentos de este estudio piloto de evaluación. El contenido de las cuestiones de las pruebas se diseñó con arreglo a los siguientes criterios generales: la demanda cognitiva de cada cuestión debe ir dirigida a la destreza para la que está diseñada, sus contenidos deben ser legibles, comprensibles, adecuados e inte-

resantes para la edad de la muestra a la que van dirigidas y deben plantear un reto cuyo logro motive y desafíe a los estudiantes. A partir estas premisas, las cuestiones que integran dos destrezas en cada curso fueron tomadas de un instrumento estandarizado de evaluación del pensamiento crítico (Ennis y Millman, 2005), denominado aquí test de Cornell (Nicoma), y las otras dos destrezas están formadas por cuestiones principalmente figurativas de elaboración propia, donde se ha incluido una cuestión del cuestionario de Halpern (resolución de problemas).

El test de pensamiento crítico de Cornell presenta verbalmente una historia ficticia y sencilla sobre unos exploradores en el planeta Nicoma; los estudiantes deben responder preguntas relacionadas con la información desplegada en la historia y que van dirigidas a evaluar cuatro destrezas de pensamiento crítico. La información de la historia y cuestiones referidas a las destrezas de comparación y predicción se han incluido en la prueba del grado 6 (EP6), y la información y cuestiones

referidas a las destrezas de explicación causal y toma de decisiones se han incluido en la prueba para el grado 8 (ESO2). Algunas preguntas de cada una de esas dimensiones se excluyeron porque se consideraron repetitivas o porque presentaban algún problema de comprensión (ver muestra en anexo).

Las otras cuatro dimensiones de las pruebas fueron desarrolladas a partir de materiales principalmente figurativos para facilitar la motivación, la comprensión de los estudiantes y la agilidad de sus respuestas; además, el diseño figurativo pretende que estas cuestiones sean aún más independientes de la cultura y los conocimientos escolares (ver muestra en anexo).

La prueba del grado 6 está formada por 45 reactivos (destrezas clasificación y resolución de problemas), mientras la prueba correspondiente al grado 8 (ESO2) está formada por 38 reactivos para las destrezas de secuenciación y relación partes y todo (tabla 1). En conjunto, todas las cuestiones incluidas en las pruebas están libres de cultura, puesto que los contenidos de las preguntas no están relacionados ni anclados en conocimientos curriculares de las materias escolares. Ningún conocimiento previo es un requisito para responder a la demanda cognitiva que exige lograr la respuesta correcta, de manera que las cuestiones plantean retos de pensamiento auténticos e independientes de conocimientos curriculares previos de ningún tipo para los estudiantes.

Los formatos de respuesta elegidos para las preguntas planteadas en las pruebas son formatos cerrados, porque facilitan la asignación de puntuaciones y hacen más rápida y fiable la obtención y análisis de los resultados. Este formato tiene el inconveniente de no evaluar el desarrollo de los procesos de pensamiento, pero tiene la ventaja de ofrecer una evaluación estandarizada y rápida de destrezas de pensamiento. Las ventajas metodológicas de un instrumento estandarizado son: facilitar el establecimiento de líneas base para comparar investigaciones, programas y metodologías de enseñanza y lograr que sus medidas ajusten la demanda cognitiva de las preguntas a la destreza que representan para constituir una evaluación válida y fiable de cada destreza.

Procedimiento

Las pruebas no se han aplicado como pruebas de velocidad, con un tiempo improrrogable para responder; los estudiantes podían disponer del tiempo necesario para concluir sus respuestas.

Fueron aplicadas por la profesora tutora de los estudiantes en su grupo natural, en el mes de mayo al final de curso, utilizando dispositivos digitales; se plantearon como una prueba más de evaluación dentro del curso, para incentivar el esfuerzo y motivación de los estudiantes.

Las respuestas correctas reciben un punto, y las respuestas incorrectas cero puntos y no se han aplicado correcciones por respuestas al azar. La base de datos resultante con las puntuaciones de los estudiantes se ha procesado con el paquete estadístico *SPSS 25*.

Las variables consideradas para los análisis son las puntuaciones alcanzadas en las diferentes destrezas evaluadas por las dos pruebas, como suma de los reactivos que forman cada destreza (tabla 1). La suma de las puntuaciones de cada una de las cuatro destrezas evaluadas en cada prueba ha generado una

variable de puntuación global (pensamiento) que sería la estimación del nivel de pensamiento crítico de los estudiantes con base en las cuatro destrezas evaluadas.

Aunque existen diversos tipos de validez psicométrica, estos no constituyen conceptos independientes entre sí, ya que mantienen una relación de influencia recíproca y conforman las diversas partes de un todo; en este pilotaje se han atendido especialmente las formas de validez compatibles con el contexto del estudio (el tamaño de la muestra impide el análisis correlacional de validez factorial). Por ello, la validez de contenido se ha basado en el análisis y evaluación de decenas de cuestiones e instrumentos de evaluación de pensamiento crítico por dos expertos, cuyo acuerdo ha seleccionado las cuestiones que se ajustaban mejor a la demanda cognitiva de cada destreza evaluada y las ha asignado a una de las dos pruebas pilotadas. La fiabilidad se ha calculado mediante el estadístico alfa de Cronbach.

Para evaluar empíricamente la validez convergente de las medidas de destrezas de

pensamiento crítico se han recogido para cada estudiante otras dos variables que actúan como criterios externos de validez de las pruebas. En primer lugar, el promedio de las calificaciones escolares finales (escala de 1 a 10) obtenidas por cada estudiante en las asignaturas del curso (criterio nota media); en segundo lugar, la calificación específica (escala de 1 a 7) otorgada por el profesor del programa TBL sobre la ejecución de las actividades realizadas por los estudiantes en este programa (variable TBL). Esta última calificación fue otorgada por el profesor de manera ciega respecto a la ejecución de las pruebas, ya que el profesor emitió su calificación sin saber que se aplicaban las pruebas, y, por supuesto, sin conocer sus resultados, de manera que se puede considerar una valoración totalmente ciega e independiente de la prueba.

Resultados

Los descriptores estadísticos de las puntuaciones de las variables consideradas en este estudio, obtenidas de la aplicación de las pruebas, están resumidos en la tabla 2.

Tabla 2. Estadística descriptiva básica de las variables empleadas en este estudio

| Prueba EP6 (n = 22) | Criterios externos | | Destrezas de pensamiento | | | | Pensamiento* |
|-------------------------|--------------------|-----------|--------------------------|--------------------|---------------|-----------------------|--------------|
| | Notas Medias | TBL (1-7) | Predicción | Comparación | Clasificación | Problema | |
| Media | 8,46 | 4,41 | 8,95 | 7,95 | 3,95 | 7,64 | 28,50 |
| Desviación estándar | 0,85 | 2,28 | 1,94 | 1,73 | 2,24 | 2,08 | 5,12 |
| Máximo | 9,6 | 7 | 13 | 11 | 8 | 11 | 40 |
| Mínimo | 6,6 | 1 | 4 | 5 | 0 | 3 | 21 |
| Mediana | 8,75 | 4,5 | 9 | 8 | 3,5 | 7 | 27,5 |
| Prueba ESO2 (n = 16) | Notas Medias | TBL (1-7) | Explicación causal | Toma de decisiones | Secuenciación | Relaciones parte-todo | Pensamiento* |
| N | 16 | 16 | 16 | 16 | 16 | 16 | 16 |
| Media | 800 | 4,38 | 7,88 | 4,94 | 7,56 | 1,00 | 21,38 |
| Desviación estándar | 1,14 | 1,86 | 2,33 | 1,39 | 1,71 | 1,63 | 4,76 |
| Máximo | 9,4 | 6 | 11 | 7 | 10 | 4 | 29 |
| Mínimo | 5,3 | 1 | 4 | 2 | 4 | 0 | 12 |
| Mediana | 8,15 | 5 | 8 | 4.5 | 8 | 0 | 22 |

* La puntuación de la variable pensamiento es la suma de las puntuaciones en las cuatro destrezas evaluadas. Fuente: elaboración propia.

La variable nota media es el promedio de las calificaciones finales obtenidas en las asignaturas del curso correspondiente; en ambos grupos se observa una calificación media situada en el intervalo del notable alto. La variable TBL es la calificación final otorgada por el profesor del aprendizaje realizado sobre las actividades de destrezas de pensamiento, en el marco del programa de enseñanza basada en el pensamiento (TBL); la calificación media y la mediana de esta variable también se sitúa por encima del punto medio del rango de estas calificaciones (4).

Las puntuaciones medias de las cuatro destrezas evaluadas mediante las pruebas aplicadas de retos de pensamiento se sitúan aproximadamente por encima de la media del rango de cada una de estas variables. Las puntuaciones de la variable pensamiento se obtienen, en cada caso, como suma de las cuatro puntuaciones obtenidas por cada estudiante en las cuatro destrezas de pensamiento evaluadas; las puntuaciones medias y las medianas de esta variable en cada curso se sitúan también aproximadamente por encima del punto medio del rango de esta variable. En conjunto, estos resultados sugieren que las pruebas están muy centradas en las capacidades de respuesta y asequibles a los niveles cognitivos de los estudiantes.

El análisis de correlaciones entre la variable pensamiento y las puntuaciones en las cuatro destrezas de pensamiento con la variable TBL (valoración por el profesor del nivel de aprendizaje de destrezas en clase) permite una estimación empírica de la validez convergente de las pruebas de retos de pensamiento. En el caso de la prueba de sexto curso de educación primaria EP6 la correlación obtenida entre la variable pensamiento y la variable TBL es positiva y moderadamente elevada (0,345), aunque su nivel de significación estadística ($p=0,116$) está próximo, pero ligeramente por encima del umbral de significación estadística ($p < 0,05$). En el caso de la prueba de segundo curso de educación secundaria obligatoria ESO2 la correlación obtenida entre la variable pensamiento y la variable TBL es positiva y muy alta (0,775), y su nivel de significación estadística es muy elevado, muy por debajo nivel de significación mínimo ($p < 0,000$). Sin duda, es el resultado de más impacto en este estudio piloto, ya que demuestra una profunda concomitancia entre ambas formas independientes de evaluar el pensamiento crítico, el profesor y la prueba retos de pensamiento.

Tabla 3a. Coeficientes de correlación de Pearson entre las variables de este estudio y su grado de significación estadística –en cursiva– para la prueba aplicada en sexto curso de educación primaria (los coeficientes de correlación de las destrezas con la variable pensamiento aparecen tachados porque ambas variables son linealmente dependientes entre sí, por definición)

| Prueba EP6 | Notas Medias | TBL (1-7) | Predicción | Comparación | Clasificación | Problema |
|--|--------------|--------------|----------------|----------------|----------------|----------------|
| Pensamiento <i>(significación)</i> | 0,337 | 0,345 | <i>0,795**</i> | <i>0,578**</i> | <i>0,635**</i> | <i>0,554**</i> |
| | <i>0,125</i> | <i>0,116</i> | <i>0,000</i> | <i>0,005</i> | <i>0,001</i> | <i>0,007</i> |
| Notas medias <i>(significación)</i> | 1 | 0,572** | 0,304 | 0,054 | 0,314 | 0,163 |
| | | <i>0,005</i> | <i>0,169</i> | <i>0,813</i> | <i>0,155</i> | <i>0,469</i> |
| TBL (1-7) <i>(significación)</i> | | 1 | 0,241 | 0,174 | 0,284 | 0,173 |
| | | | <i>0,280</i> | <i>0,440</i> | <i>0,200</i> | <i>0,441</i> |

| Prueba EP6 | Notas Medias | TBL (1-7) | Predicción | Comparación | Clasificación | Problema |
|----------------------------------|--------------|-----------|------------|-------------|---------------|----------|
| Predicción (significación) | | | 1 | 0,453* | 0,230 | 0,397 |
| | | | | 0,034 | 0,303 | 0,068 |
| Comparación (significación) | | | | 1 | 0,245 | -0,097 |
| | | | | | 0,271 | 0,667 |
| Clasificación (significación) | | | | | 1 | 0,068 |
| | | | | | | 0,764 |

* La correlación es significativa en el nivel 0.05 (bilateral).

** La correlación es significativa en el nivel 0.01 (bilateral).

Fuente: elaboración propia.

Las correlaciones entre las puntuaciones de las cuatro destrezas de pensamiento evaluadas por las pruebas y la calificación TBL son también positivas en las cuatro destrezas de las dos pruebas. En el caso de la prueba de sexto curso de educación primaria las correlaciones obtenidas entre las cuatro destrezas de pensamiento y la calificación TBL son positivas para las cuatro destrezas (0,173 a 0,284), aunque su nivel de significación no alcanza el umbral de significación estadística ($p < 0,05$). En el caso de la prueba de segundo curso de educación secundaria obligatoria las correlaciones obtenidas entre las cuatro variables de destrezas de pensamiento y la variable TBL son positivas y más altas (0,663 a 0,320), y su nivel es estadísticamente significativo, por debajo del umbral de significación, en la destreza explicación causal ($p = 0,015$) y muy significativo ($p = 0,005$) en el caso de la destreza secuenciación (tablas 3a, 3b).

El análisis de las correlaciones entre la variable de pensamiento, medida por las pruebas como suma de las cuatro puntuaciones de destrezas, y la variable de rendimiento escolar, medida por la nota media obtenida por los estudiantes entre las calificaciones finales de las asignaturas del curso, produce resultados empíricos similares, aunque no iguales en los dos cursos (tablas 3a, 3b). En el caso de la prueba de sexto curso de educación primaria la correlación entre pensamiento y nota media es positiva y moderadamente alta (0,337), y su nivel de significación estadística ($p=0,12$) también está próximo, pero ligeramente por encima del umbral de significación estadística ($p < 0,05$). En el caso de la prueba de segundo curso de educación secundaria obligatoria la correlación obtenida entre la variable pensamiento y la nota media del curso es positiva y más alta (0,522), y su nivel de significación estadística ($p=0,03$) es bueno, por debajo del umbral de significación mínimo ($p < 0,05$).

Tabla 3b. Coeficientes de correlación de Pearson entre las variables de este estudio y su grado de significación –en cursiva– para la prueba aplicada en segundo curso de educación secundaria (los coeficientes de correlación las destrezas con la variable pensamiento aparecen tachados porque son linealmente dependientes entre sí, por definición)

| Prueba ESO2 | Notas medias | TBL (1-7) | Explicación causal | Toma de decisiones | Secuenciación | Relaciones parte-todo |
|---------------------------------|--------------|-----------|--------------------|--------------------|--------------------|-----------------------|
| Pensamiento (significación) | 0,522* | 0,775** | 0,821** | 0,468 | 0,644** | 0,669** |
| | 0,038 | 0,000 | 0,000 | 0,068 | 0,007 | 0,005 |
| Notas medias (significación) | 1 | 0,683** | 0,468 | 0,021 | 0,535* | 0,273 |
| | | 0,004 | 0,068 | 0,938 | 0,033 | 0,306 |

| Prueba ESO2 | Notas medias | TBL (1-7) | Explicación causal | Toma de decisiones | Secuenciación | Relaciones parte-todo |
|---------------------------------------|--------------|-----------|--------------------|--------------------|---------------|-----------------------|
| TBL (1-7) (significación) | | 1 | 0,596* | 0,320 | 0,663** | 0,440 |
| | | | 0,015 | 0,227 | 0,005 | 0,088 |
| Explicación causal (significación) | | | 1 | 0,141 | 0,369 | 0,455 |
| | | | | 0,602 | 0,159 | 0,077 |
| Toma de decisiones (significación) | | | | 1 | 0,156 | 0,147 |
| | | | | | 0,564 | 0,587 |
| Secuenciación (significación) | | | | | 1 | 0,167 |
| | | | | | | 0,537 |

* La correlación es significativa en el nivel 0.05 (bilateral).

** La correlación es significativa en el nivel 0.01 (bilateral).

Fuente: elaboración propia.

Estos resultados avalan empíricamente la validez de criterio de las pruebas de pensamiento, ya que las puntuaciones obtenidas con las pruebas (la puntuación de pensamiento global y las puntuaciones de las cuatro destrezas) correlacionan positivamente con una medida de destrezas de pensamiento TBL otorgada por el profesor e independiente de las anteriores. Ambas medidas de pensamiento se pueden considerar independientes porque los profesores de ambos niveles fueron ciegos a esta experiencia, ya que emitieron sus calificaciones TBL sin información de las pruebas, y, por otro lado, las pruebas han sido diseñadas por personas ciegas a las experiencias y actividades de enseñanza concretas realizadas por los alumnos en la clase, en el marco del programa TBL.

El resultado se puede calificar de extraordinario en el caso de la prueba de retos de pensamiento para segundo curso de educación secundaria obligatoria debido a la correlación tan alta y estadísticamente significativa entre ambas medidas independientes. Los resultados también son positivos y muestran la misma tendencia para la relación entre destrezas de pensamiento y rendimiento escolar medido por la nota media de las calificaciones en las asignaturas de cada curso. No obstante, el pequeño tamaño de la muestra requiere confirmar estos prometedores resultados con nuevas medidas empíricas.

Fiabilidad de la prueba y las escalas

Otra propiedad importante de los instrumentos de medida es la fiabilidad del instrumento, como la consistencia interna de las puntuaciones emanadas del instrumento para cada una de las escalas que lo componen. En este caso, se ha calculado el estadístico alfa de Cronbach para la puntuación de pensamiento global y para las puntuaciones parciales en las cuatro destrezas evaluadas en cada prueba, obtenidas sumando las puntuaciones dicotómicas (0-1) de cada pregunta de las pruebas retos de pensamiento aplicadas (tabla 4). Para computar el estadístico alfa, cada variable de pensamiento se considera integrada por todas las cuestiones que la conforman; en el caso de la variable de pensamiento global serían todas las cuestiones de la prueba, y en el caso de las destrezas, serían las cuestiones pertenecientes a cada destreza.

Estos cálculos de fiabilidad tienen dos particularidades estadísticas que proyectan condicionantes sobre el resultado de estos parámetros: las cuestiones están medidas dicotómicamente (0-1) y el tamaño de la muestra de este estudio piloto es pequeño. Una amplia literatura está de acuerdo en que las estimaciones del parámetro alfa resultan minoradas cuando se usan variables dicotómicas y muestras de tamaño pequeño, el cual, además, puede condicionar la posibilidad de implementar algunos cálculos. En este último caso se encuentra el parámetro alfa para la variable pensamiento global de la prueba retos de pensamiento ESO2, que no pudo computarse por no satisfacer la muestra las condiciones de contorno requeridas para el cómputo.

Tabla 4. Resultados del coeficiente de consistencia interna (alfa de Cronbach) para las dos pruebas de retos de pensamiento (EP6 y ESO2)

| Prueba EP6 | Alfa Cronbach | Prueba ESO2 | Alfa de Cronbach |
|---------------|-------------------|-----------------------|------------------|
| Predicción | 0,455 | Explicación causal | 0,617 |
| Comparación | 0,185 | Toma de decisiones | 0,226 |
| Clasificación | 0,809 | Secuenciación | 0,479 |
| Problema | 0,545 (0,843*) | Relaciones parte-todo | 0,938 |
| Pensamiento | 0,700 | Pensamiento | No computable |

* Índice de fiabilidad si se elimina la cuestión verbal proveniente del test de Halpern de la escala de resolución de problemas (solo reactivos figurativos).

Fuente: elaboración propia.

El coeficiente alfa global para la prueba retos de pensamiento EP6 obtuvo un valor aceptable (0,70), pero las cuatro escalas que miden las destrezas de pensamiento obtuvieron valores más diversificados. La destreza clasificación (figurativa) obtuvo un valor bueno de su fiabilidad (0,809) y la destreza comparación (verbal) obtuvo el valor más bajo (0,185); las otras dos destrezas, comparación y resolución de problemas, obtuvieron valores intermedios, que podrían ser aceptables teniendo en cuenta, además, el pequeño número de cuestiones que las forman (el estadístico alfa depende críticamente del número de ítems que conforman la variable).

En el caso de la destreza resolución de problemas, formada por una cuestión procedente del test de Halpern (verbal) y otras cuestiones figurativas, se observó una gran diferencia de comportamiento de ambos grupos de cuestiones. Realizando un cómputo del parámetro alfa para los dos grupos de cuestiones por separado se obtuvo un valor de la fiabilidad

muy bueno (0,843) para las cuestiones figurativas (a pesar de su reducido número) y un valor menor de toda la escala, si se incluye la cuestión verbal del test de Halpern (0,545). Por tanto, las cuestiones figurativas aumentan la fiabilidad de las pruebas de pensamiento respecto a las cuestiones verbales.

El coeficiente de fiabilidad para la prueba global retos de pensamiento ESO2 no pudo calcularse por no satisfacer los requerimientos estadísticos; las diversas escalas que miden las cuatro destrezas de pensamiento obtuvieron también valores diversificados. Las cuestiones figurativas de la destreza relaciones entre las partes y el todo registraron un excelente resultado de fiabilidad (0,938), mientras la destreza de toma de decisiones alcanza un valor muy bajo; las otras destrezas, explicación causal y secuenciación logran valores intermedios.

En conjunto, estos resultados acerca de la fiabilidad de las pruebas y las escalas de destrezas muestran que las cuestiones y esca-

las figurativas tienden a tener resultados de fiabilidad mejores que las escalas verbales, representadas principalmente por las escalas del test de Cornell y la cuestión verbal sobre una situación cotidiana de resolución de problemas, tomada del test de Halpern, las cuales no superan el valor umbral usual ($\alpha > 0,70$).

Análisis de ítems

Se han calculado los índices de facilidad (porcentaje de respuestas correctas) y las correlaciones interítem, aunque en ambos casos no se exponen los resultados empíricos totales por su excesiva extensión.

El análisis de la prueba EP6 detecta seis cuestiones muy fáciles ($> 90\%$) y dos cuestiones muy difíciles ($< 10\%$); el análisis de la prueba ESO2 detecta cinco muy fáciles ($> 90\%$) y una muy difícil ($< 10\%$). Estas cuestiones muy fáciles y muy difíciles serán sometidas a una revisión cuidadosa para considerar su exclusión o reformulación en una futura prueba.

El análisis de las correlaciones interítems muestra la existencia de siete cuestiones en la prueba EP6 y siete en la ESO2 que presentan un número relevante de correlaciones negativas que indican una potencial disfunción estadística de estas cuestiones. Aquellas con correlaciones negativas también serán revisadas simultáneamente con su impacto en la fiabilidad de la prueba, para considerar su exclusión o reformulación en una futura prueba más sólida, para cuya mejora se ha realizado este estudio piloto.

Conclusiones

El objetivo de este estudio es validar empíricamente dos pruebas de evaluación de destrezas de pensamiento crítico para estudiantes de sexto curso de educación primaria (grado 6) y segundo curso educación secundaria obligatoria (grado 8), que siguen un curso de aprendizaje basado en pensamiento en su colegio. Como variables adicionales, que sirven como criterios de validación convergente, se han considerado las calificaciones escolares finales medias de las asignaturas del curso y la calificación específica de las actividades del programa de aprendizaje basado en pensamiento otorgadas por el profesor. El pequeño tamaño de la muestra empleada en este pilotaje limita las técnicas de validación, pues no pueden emplearse en este caso análisis factoriales y otros (por ejemplo, la percentilación de puntuaciones de las escalas o el análisis de la validez factorial de las destrezas de pensamiento).

Los resultados globales de este estudio piloto sobre las dos pruebas pensamiento crítico se pueden considerar positivos. En primer lugar, porque las correlaciones de las variables destrezas de pensamiento crítico con los criterios de validación (notas medias y nota TBL) son positivas en todos los casos. Para la prueba EP6 las correlaciones son positivas y próximas a la significación estadística, pero en la ESO2 esas correlaciones no solamente son positivas, sino que son muy altas y con una significación estadística alta. Cabe destacar el valor extraordinariamente alto de la correlación entre la variable global de pensamiento y la nota TBL (0,775) logrado en la prueba ESO2, que constituye un respaldo empí-

rico extraordinario como validación de criterio para la prueba (y también para el programa de aprendizaje TBL aplicado).

En segundo lugar, los resultados también son positivos porque la mayoría de los coeficientes de fiabilidad obtenidos son positivos y aceptables. La mitad de las cuestiones aplicadas provenían de un instrumento estandarizado construido hace años (test de Cornell, Nicoma, una cuestión del test de Halpern y otras); el resto son cuestiones de elaboración propia de tipo figurativo, para hacer las pruebas aún más libres de cultura. Pues bien, la tendencia observada en los índices de fiabilidad de las distintas destrezas es que las cuestiones figurativas tienden a presentar índices de fiabilidad buenos (0,479 a 0,938) y mejores que las cuestiones verbales, resultado que apoya una mejora de las pruebas, con énfasis en la inclusión de cuestiones figurativas.

Estos resultados también constituyen un respaldo empírico importante para el propio curso de enseñanza por el proyecto de aprendizaje basado en pensamiento (TBL) aplicado en el colegio (Swartz et ál., 2013): la correlación entre la nota TBL otorgada por el profesor del curso ESO2 y las puntuaciones alcanzadas en la prueba de pensamiento por los estudiantes es casi perfecta para una muestra tan pequeña, y en EP6, la correlación es positiva y próxima a la significación estadística. El resultado excepcional de la prueba ESO2 implica que las varianzas de ambas medidas son prácticamente idénticas y, en consecuencia, ambas medidas se realimentan en su validez: los resultados del curso apoyan la validez de la prueba y la prueba apoya la eficacia del curso. Este resultado es valioso porque identifica empíricamente efectos concretos y apoya la enseñanza del pensamiento, conforme propone Saiz (2017).

Como se ha expuesto, la mayoría de programas de enseñanza del pensamiento, que se aplican desde hace décadas en distintas instituciones educativas, tienen pocos datos que valoren sus efectos concretos y específicos, circunstancia que realza todavía más el

valor excepcional del resultado anterior. Desde esta perspectiva, estos resultados constituyen un acicate no solo para mejorar psicométricamente las pruebas, sino también para promover decididamente la evaluación de los efectos de la enseñanza del pensamiento crítico. Para ello, se necesita disponer de instrumentos de evaluación del pensamiento, válidos y fiables, cuya aplicación logre evaluaciones cuantitativas, empíricas y estandarizadas que aporten datos de los efectos positivos y beneficiosos, un aspecto vivamente abandonado de la investigación del pensamiento crítico (Saiz, 2017).

La estandarización de estos instrumentos de evaluación del pensamiento permite también convertirlos en herramientas de evaluación funcionales, a disposición del profesorado que enseña pensamiento crítico, para facilitar la evaluación educativa de las destrezas del pensamiento adquiridas por los estudiantes como también desarrollan otros autores (Gómez, Jiménez, De las Heras y Vázquez, 2018).

En resumen, este estudio presenta la validación piloto (tentativa) de instrumentos de medida de destrezas de pensamiento crítico con estudiantes adolescentes (12-14 años), compuestos por reactivos tomados de pruebas consolidadas (Cornell y Halpern) y otros reactivos figurativos, de nuevo diseño, como primera etapa para continuar su mejora en desarrollos futuros, y que demuestra la significativa relación entre pensamiento crítico y aprendizaje escolar. El estudio innova hacia un enfoque de instrumentos independientes de la cultura y de los currículos escolares, y además, flexibles, es decir, susceptibles de diseñarse a la medida, para satisfacer necesidades prácticas concretas, como adaptarse a las destrezas del programa de enseñanza, el curso o la edad de los estudiantes.

Agradecimientos

Proyecto EDU2015-64642-R (AEI/FEDER, UE) financiado por la Agencia Estatal de Investigación (AEI) de España y el Fondo Europeo de Desarrollo Regional (FEDER).

Referencias

- Bailin, S., Case, R., Coombs, J.R. y Daniels, L.B. (1999). Conceptualizing critical thinking. *Journal of Curriculum Studies*, 31, 285-302.
- Colom, R., García Moriyón, F., Magro, C. y Morilla, E. (2014). The Long-term Impact of Philosophy for Children: A Longitudinal Study (Preliminary Results). *Analytic Teaching and Philosophical Praxis*, 35, 50-55.
- Ennis, R.H. y Millman, J. (2005). *Cornell Critical Thinking Test Level X*. Seaside, CA: The Critical Thinking Company.
- Ennis, R.H. (2019). Long definition of critical thinking. *CriticalThinking.net*. Recuperado de <http://criticalthinking.net/definition/long-definition-of-critical-thinking/>
- Fisher, A. (2009). *Critical Thinking. An Introduction*. Cambridge: Cambridge University Press.
- Follmann, D., Mattos, K.R.C. y Güllich, R.I. da C. (2018). Estratégias de Ensino de Ciências e a Promoção de Pensamento Crítico em Portugal. *Tecné, Episteme y Didaxis, TED*, (número extraordinario). <https://revistas.pedagogica.edu.co/index.php/TED/article/view/8789>
- Gómez Pérez, J.E., Jiménez Pérez, R., De las Heras Pérez, Ángeles, M. y Vázquez Bernal, B. (2018). Avance de investigación: Validación de un cuestionario para medir el nivel de autorregulación en el tratamiento de cuestiones sociocientíficas en un aula de clase de séptimo grado. *Tecné, Episteme y Didaxis: TED*, (número extraordinario). <https://revistas.pedagogica.edu.co/index.php/TED/article/view/8862>
- Halpern, D.F. (2010). *Halpern Critical Thinking Assessment*. Vienna: Schuhfried. <http://www.schuhfried.com/vienna-test-system-vts/all-tests-from-a-z/test/hcta-halpern-critical-thinking-assessment-1/>
- Madison, J. (2004). *James Madison Critical Thinking Course*. Seaside, CA: The Critical Thinking Co. <https://www.criticalthinking.com/james-madison-critical-thinking-course.html>
- Manassero-Mas, M.A. y Vázquez-Alonso, A (2019, en prensa). Taxonomía de las destrezas de pensamiento: una herramienta clave para la alfabetización científica. En M.D. Maciel y E. Albrecht (org.), *Ciência, Tecnologia & Sociedade: Ensino, Pesquisa e Formação* (pp. 17-37). São Paulo: Terracota.
- Paul, R. y Nosich, G.M. (2019). A Model for the National Assessment of Higher Order Thinking. *CriticalThinking.org*. <http://www.criticalthinking.org/pages/a-model-for-the-national-assessment-of-higher-order-thinking/591>
- Phan, H.P. (2010). Critical thinking as a self-regulatory process component in teaching and learning. *Psicothema*, 22, 284-292.

- Rivas, S.F. y Saiz, C. (2012). Validación y propiedades psicométricas de la prueba de pensamiento crítico Penscrisal. *Revista Electrónica de Metodología Aplicada*, 17, 18-34.
- Saiz, C. (2017). *Pensamiento crítico y cambio*. Madrid: Pirámide.
- Swartz, R.J., Costa, A.L., Beyer, B.K., Reagan R. y Kallick, B. (2013). *El aprendizaje basado en el pensamiento*. Madrid: SM.
- Vázquez-Alonso, A. y Manassero-Mas, M.A. (2018). *Pensamiento científico y pensamiento crítico: dos caras de la misma moneda para la educación científica*. Comunicación presentada en el IV Simposio Internacional de Enseñanza de las Ciencias, Vigo, España. <http://congresos.educacioneditora.net/index.php/SIEC/SIEC2018/paper/view/219>
- Walton, D. y Macagno F. (2015). A Classification System for Argumentation Schemes. *Argument and Computation*, 6(3), 214-249.
- Watson, G. y Glaser, E.M. (2002). *Watson-Glaser Critical Thinking Appraisal-II Form E*. Londres: Pearson.

Para citar este artículo

- Manassero-Mas, M.A. y Vázquez-Alonso, A. (2020). Evaluación de destrezas de pensamiento crítico: Validación de instrumentos libres de cultura. *Revista Tecné, Episteme y Didaxis: TED*, (47), 15-32. <https://doi.org/10.17227/ted.num47-9801>

Anexo

Textos de dos reactivos de muestra de las pruebas; el primero, tomado del test de Cornell (Nicoma) es verbal, y el segundo, tomado de la destreza secuenciación, es figurativo.

óó. "Los exploradores no pueden escapar debido a que no pueden derribar las paredes de la cabaña de piedra."

¿Cuál de las siguientes afirmaciones probablemente se da por supuesta?

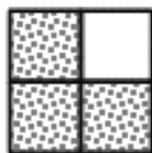
- Los exploradores pueden saltar por la ventana.
- Los centinelas están alerta.
- Todos los medios de escape son imposibles, excepto a través de las paredes.

Las figuras siguientes están construidas con unidades cuadradas y forman una secuencia regular de la cual se muestran las tres primeras etapas.

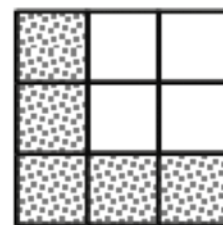
La etapa 3 de la secuencia está formada por nueve unidades cuadradas, de las cuales cinco están sombreadas y cuatro son blancas (sin sombra).



Etapa 1



Etapa 2



Etapa 3

Completa los números correspondientes a la etapa cuatro:

- Número de unidades cuadradas sombreadas correspondientes a la etapa cuatro: _____
- Número de unidades cuadradas blancas correspondientes a la etapa cuatro: _____
- ¿Cuántas unidades cuadradas sombreadas tendrá la figura de la etapa número once? _____
- ¿Cuántas unidades cuadradas blancas tendrá la figura de la etapa número once? _____
- ¿Cuántas unidades cuadradas sombreadas tendrá la figura de la etapa número catorce? _____