

TED

Tecné, Episteme y Didaxis: TED

ISSN: 2665-3184

ISSN: 2323-0126

Universidad Pedagógica Nacional; Facultad de Ciencia y Tecnología

Manassero-Mas, María Antonia; Vázquez-Alonso, Ángel
Las destrezas de pensamiento y las calificaciones escolares en educación secundaria: validación de un instrumento de evaluación libre de cultura¹
Tecné, Episteme y Didaxis: TED, núm. 48, 2020, pp. 33-54
Universidad Pedagógica Nacional; Facultad de Ciencia y Tecnología

DOI: <https://doi.org/10.7440/res64.2018.03>

Disponible en: <https://www.redalyc.org/articulo.oa?id=614270272003>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en [redalyc.org](https://www.redalyc.org)

UAEM [redalyc.org](https://www.redalyc.org)

Sistema de Información Científica Redalyc
Red de Revistas Científicas de América Latina y el Caribe, España y Portugal
Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto



Las destrezas de pensamiento y las calificaciones escolares en educación secundaria: validación de un instrumento de evaluación libre de cultura

- Thinking Skills and School Grades in Secondary Education: Validation of A Culture-Free Assessment Instrument
- Habilidades de pensamento e notas escolares no ensino médio: validação de um instrumento de avaliação sem cultura

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

Resumen

Los estudios sobre pensamiento crítico se han centrado y preocupado más por crear programas de enseñanza que evaluar el impacto de los programas y, por ello, este artículo de investigación plantea la evaluación de las destrezas de pensamiento de estudiantes españoles al final de la educación secundaria (15-16 años), su relación con las calificaciones del aprendizaje y la validación del instrumento de evaluación. Este se plantea como un instrumento libre de cultura, con un componente figurativo importante y cuyo propósito es evaluar cuatro destrezas: deducción, asunciones, secuenciación y toma de decisiones. Los resultados permiten un diagnóstico de las destrezas de los estudiantes, así como resaltan el hallazgo más importante: la correlación significativa entre las calificaciones escolares de los estudiantes y las puntuaciones de las destrezas de pensamiento, que tienden a ser más altas en las asignaturas matemáticas y científicas. Los resultados psicométricos de validez y fiabilidad son buenos para las cuestiones figurativas y más bajos para las cuestiones verbales, a pesar de que estas últimas se extrajeron de instrumentos estandarizados de evaluación de pensamiento crítico contruidos por otros autores. Finalmente, se discute la importancia de los resultados para la educación, puesto que confirman la relevancia transversal que pueden tener las destrezas de pensamiento para los aprendizajes escolares, y se propone revisar algunas cuestiones del instrumento de evaluación para mejorar su validez y fiabilidad.

Palabras clave

Pensamiento crítico; evaluación de destrezas; aprendizaje competencial; calificaciones escolares; validez; fiabilidad

* Universidad de las Islas Baleares, España.
Departamento de Psicología, Facultad de Psicología,
Universidad de las Islas Baleares, Edificio Guillem
Cifre de Colonya.

Orcid: <https://orcid.org/0000-0002-7804-7779>.
Correo: ma.manassero@uib.es

** Universidad de las Islas Baleares, España.
Orcid: <https://orcid.org/0000-0001-5830-7062>
Correo: angel.vazquez@uib.es

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



Abstract

Both the focus and concern of critical thinking have been placed upon the development of teaching programs rather than upon evaluating the impact of the programs; therefore, this research article study aims at the evaluation of the thinking skills of Spanish students at the end of secondary education (15-16 years), its relationship with learning grades and the validation of the assessment instrument. The latter is planned as a culture-free instrument with an important figurative component, and is designed to evaluate four skills: deduction, assumptions, sequencing and decision making. The results facilitate a diagnosis of students' skills, and they highlight the most important finding —the significant and high correlation between the students' school grades and the thinking skills scores, which tend to be higher in mathematical and scientific subjects. The psychometric results on the instrument validity and reliability are good for figurative issues and lower for verbal issues, in spite of the fact that the latter are drawn from standardized critical thinking assessment instruments constructed by other authors. Finally, the importance of the results for education is discussed —as they confirm the cross-cutting relevance that thinking skills might have for school learning— and some issues related to the assessment instrument are pondered in order to improve its validity and reliability.

Keywords

Critical thinking; skill assessment; competence learning; school grades; validity; reliability

Resumo

Os estudos sobre pensamento crítico têm se concentrado e mais preocupado em criar programas de ensino do que em avaliar o impacto dos programas e, portanto, este artigo de pesquisa estudo propõe a avaliação das habilidades de pensamento de estudantes espanhóis no final do ensino médio (15-16 anos), sua relação com as qualificações de aprendizagem e a validação do instrumento de avaliação utilizado. Ele é proposto como um instrumento livre de cultura, com um importante componente figurativo, e é projetado para avaliar quatro habilidades: dedução, suposições, sequenciamento e tomada de decisão. Os resultados permitem um diagnóstico das habilidades dos alunos, onde a descoberta mais importante é a correlação significativa e alta entre as notas escolares dos alunos e as pontuações das habilidades de pensamento, tendendo a ser maior nas disciplinas matemáticas e científicas. Os resultados psicométricos de validade e confiabilidade são bons para questões figurativas e mais baixos para questões verbais, embora os últimos sejam extraídos de instrumentos padronizados de avaliação do pensamento crítico construídos por outros autores. Por fim, discute-se a importância dos resultados para a educação, pois confirmam a importância transversal que as habilidades de pensamento podem ter para a aprendizagem escolar e propõe-se revisar algumas questões do instrumento de avaliação para melhorar sua validade e confiabilidade.

Palavras-chave

Pensamento crítico; avaliação de habilidades; aprendizado de competências; notas escolares; validade; confiabilidade

Introducción

Para afrontar los desafíos de las actuales sociedades del conocimiento, tales como el creciente impacto científico y tecnológico, la acelerada innovación digital e informativa, la globalización y la emergencia ecológica, se solicita a los sistemas educativos enseñar las denominadas habilidades del siglo XXI, que incluyen pensamiento crítico (PC en adelante), comunicación, colaboración, creatividad, emprendimiento, resolución de problemas, habilidades de investigación, argumentación, análisis, interpretación, creatividad, innovación y toma de decisiones. Ciertamente, esta formación de la ciudadanía es la más adecuada para participar y contribuir a la sociedad, tomar decisiones críticas y garantizar la empleabilidad (Fullan y Scott, 2014; International Society for Technology Education, 2003; National Research Council, 2012; European Union, 2014, Ananiadou y Claro, 2009).

Por otra parte, la investigación sobre aprendizaje sugiere que muchas de las anteriores destrezas de pensamiento se corresponden con las categorías superiores de la taxonomía de Bloom (analizar, juzgar y crear), también denominadas *destrezas de pensamiento de alto nivel* (Krathwohl, 2002). Además, son consideradas factores clave para los aprendizajes no rutinarios o memorísticos, como los aprendizajes en las áreas de ciencias, matemáticas y tecnologías (STEM), que son aprendizajes profundos y sensibles al dominio de destrezas de pensamiento. Los estudios pioneros de Piaget (Piaget y Inhelder, 1997), continuados por los programas de aceleración (Shayer y Adey, 2002), han desarrollado estas ideas. Paralelamente, desde hace algunos años se ha denunciado que la formación de los científicos adolece de "falta de entrenamiento y estímulo del pensamiento crítico" (Popper, 1975, p. 150); más recientemente, muchas investigaciones sostienen que las destrezas del PC son fundamentales para la educación STEM (Ford y Yore, 2014; McDonald y McRobbie, 2012; Simonneaux, 2014; Tamayo, 2017; Tenreiro-Vieira y Vieira, 2014; Vázquez-Alonso y Manassero-Mas, 2018), tesis

justificable por la identidad constitutiva entre PC y pensamiento científico (Vázquez-Alonso y Manassero-Mas, 2019).

Desde la perspectiva de la investigación educativa general, también se han aportado pruebas empíricas del impacto de las variables cognitivas sobre el aprendizaje. El metanálisis del aprendizaje visible de Hattie (2009, 2012) informa que el tamaño del efecto de los programas piagetianos es muy alto ($d = 1,28$), segundo en magnitud; además, otras variables de pensamiento (estrategias metacognitivas, creatividad, resolución de problemas, etc.) alcanzan también impactos relevantes sobre el aprendizaje ($d > .40$). En consecuencia, todos estos resultados demuestran que las destrezas de pensamiento, devaluadas por modas y etiquetas didácticas, siguen siendo un factor altamente relevante para el aprendizaje profundo y justifican una atención innovadora hacia ellas.

Marco teórico

Evaluación del pensamiento crítico

La investigación sobre PC se ha centrado en la conceptualización del constructo, su enseñanza y su evaluación, aunque el desarrollo de cada área ha sido desigual (Saiz, 2017). La conceptualización encuadra el PC como habilidades de alto nivel, a pesar de que es notoria la ausencia de consenso acerca de su definición. Con todo, se suele caracterizar como una forma de pensamiento que supone un dominio consciente de múltiples destrezas, su conformidad con estándares de calidad y las disposiciones para superar las tendencias de la condición humana a la falacia y al sesgo (egocentrismo y sociocentrismo). Por la ausencia de consenso en su definición, muchos prefieren definirlo por extensión, es decir, por medio de las destrezas constituyentes (por ejemplo, Fisher, 2009), aunque tampoco existe consenso completo en este caso (al respecto, se pueden consultar los siguientes sitios web: <http://www.criticalthinking.org> y <http://criticalthinking.net>).

El plan nacional para la evaluación del PC de Paul y Nosich (1993) contiene la lista de destrezas de PC más extensa (en total, 88), agrupadas en las siguientes dimensiones: 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 de los elementos de pensamiento y dimensiones afectivas. Así mismo, otro planteamiento global sobre el PC ha sido diseñado y reelaborado a lo largo de los años por Robert Ennis quien propone las siguientes destrezas: clarificación básica de las ideas (centrarse en una pregunta, analizar argumentos, preguntas de aclaración y comprender y usar gráficas), bases de una decisión (juzgar la credibilidad de una fuente, los informes de observación, utilizar el conocimiento existente), inferencias (deducir y juzgar deducciones, hacer y juzgar inferencias y argumentos inductivos, inducción enumerativa, argumento e inferencia a la mejor explicación, hacer y juzgar juicios de valor), clarificación avanzada (definir términos y juzgar definiciones, manejar el error, atribuir y juzgar suposiciones, pensamiento supuesto, hacer frente a falacias y calidad del pensamiento propio) y metacognición. Evidentemente, los contenidos y la nomenclatura resaltan el parecido con el pensamiento científico, de modo que, por ejemplo, todas las destrezas de inferencia son homologables directamente como destrezas de pensamiento científico.

Una variedad de programas para la enseñanza del PC con variadas orientaciones y prácticas se han creado desde los años ochenta, cuya hipótesis básica es que el pensamiento puede ser mejorado mediante programas educativos adecuados. Así, pues, la diversidad de programas es tan amplia que resulta también difícil de conseguir una síntesis virtuosa entre la variedad de programas (Follmann, Mattos y Güllich, 2018; Saiz, 2017; Swartz, Costa, Beyer, Reagan y Kallick, 2013).

La investigación y los programas de enseñanza necesitan estudiar los niveles, los cambios y las mejoras del pensamiento, para lo cual la evaluación del PC es crucial; en otras palabras, constituye una necesidad ineludible para la investigación y un apoyo significativo para la mejora de los programas de enseñanza, que requieren la aplicación de instrumentos de evaluación apropiados para lograr medidas válidas y fiables de esas variables. Sin embargo, los estudios de evaluación que acrediten empíricamente los efectos de los programas de enseñanza, son la excepción más que la regla (Saiz, 2017), pues solo el programa de filosofía para niños desarrollado por Lipman ha sido evaluado repetidamente (Colom, García, Magro y Morilla, 2014), pero otros, como el aprendizaje basado en pensamiento —TBL— (Swartz et al., 2013), solo han sido evaluados ocasionalmente, y aun otros, como el programa canadiense Razonamiento, argumentación y retórica (Walton y Macagno, 2015), carecen de evaluaciones de sus efectos.

La investigación ha desarrollado diversos instrumentos para evaluar PC, cuyo núcleo común es definir el conjunto de destrezas que valora cada uno. En ese sentido, el cuestionario de situaciones cotidianas de Halpern (2010) evalúa las siguientes destrezas: resolución de problemas, razonamiento verbal, probabilidad e incertidumbre, prueba de hipótesis, análisis de argumentos y toma de decisiones. El instrumento de Watson y Glaser (2002), por su parte, propone evaluar

la deducción, el reconocimiento de supuestos, la interpretación, la inferencia y la evaluación de argumentos. De manera similar, el cuestionario de PC de la universidad de Salamanca (PENCRISAL) evalúa la solución de problemas, el razonamiento inductivo, el razonamiento deductivo, el razonamiento práctico o informal y la toma de decisiones (Rivas y Saiz, 2012).

En cuanto a la prueba de evaluación de PC de James Madison (2004), esta incluye las siguientes destrezas: reconocer y clarificar problemas, conclusiones, argumentos y explicaciones; distinguir condiciones necesarias y suficientes; identificar y evitar errores y falacias; evaluar argumentos; distinguir conclusiones, premisas (razones), argumentos, explicaciones, supuestos (declarados/no declarados), problemas, conclusiones tácitas, premisas no declaradas e implicaciones, reivindicaciones, 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 relevancia de las declaraciones para otras; evaluar argumentos en términos de criterios y, por último, discernir si ciertos pares de afirmaciones son consistentes, contrarias, contradictorias o paradójicas.

La conceptualización del pensamiento crítico

Las destrezas constitutivas del PC, a partir de los instrumentos de evaluación, se han comparado cualitativamente por Manassero y Vázquez (2019), estableciendo equivalencias entre las distintas denominaciones y calculando las frecuencias relativas de aparición como orientación para determinar la relevancia relativa de las distintas destrezas, con apoyo en la literatura de investigación. Este análisis produjo la siguiente lista de categorías (en orden decreciente de frecuencias): comunicación, resolución de problemas, razonamiento, creatividad, argumentación, toma de decisiones, clarificación y precisión de ideas, falacias y

errores, identificación de supuestos, evaluación, etc. A partir de ese análisis empírico, se propone una taxonomía donde el PC se considera el constructo estructurante de las diversas habilidades y destrezas de pensamiento y de nivel superior a cualquiera de ellas.

De este modo, la taxonomía propuesta está estructurada en cuatro grandes dimensiones: la dimensión de creatividad, que engloba las acciones y operaciones cognitivas dirigidas a generar preguntas, ideas y conclusiones, guiar las observaciones, elaborar modelos, analizar y sintetizar; la dimensión del razonamiento y argumentación, que comprende el conjunto de las operaciones mentales y cognitivas dirigidas a justificar la validez de una conclusión —en esta dimensión se ubican las cuatro formas básicas de razonamiento: deductivo, inductivo, abductivo y estadístico—. La dimensión de procesos complejos (resolución de problemas y toma de decisiones) aborda las tareas más complejas del pensamiento porque implican el uso de otras destrezas para encontrar solución a un problema o tomar la decisión más adecuada; finalmente, la dimensión de evaluación y juicio incluye aquellas operaciones dirigidas a valorar la calidad y el valor de los procesos de pensamiento, propios y ajenos, y en todos sus elementos (información, supuestos, conclusiones y consecuencias); en resumen juzgar el pensamiento significa, entre otras cosas, justificar la adecuación, justicia, validez y fiabilidad de las conclusiones alcanzadas (tabla 1).

Tabla 1. Esquema de la taxonomía del pensamiento crítico

<p>CREATIVIDAD (generar ideas, conclusiones) Plantear buenas preguntas Observación (comparar, clasificar) Análisis y síntesis (partes-todo, analogías, modelos)</p>	<p>PROCESOS COMPLEJOS (destrezas orientadas a buscar soluciones o decisiones) Toma de decisiones Resolución de problemas</p>
<p>RAZONAMIENTO Y ARGUMENTACIÓN (justificar predicción, implicación, conclusión) Lógico (deductivo) Empírico (explicar con datos, informaciones, pruebas) Inductivo (generalizaciones) Argumentación (abductivo) Estadístico (probabilístico) Falacias y errores</p>	<p>EVALUACIÓN Y JUICIO (valoración de la calidad del pensamiento) Razonamientos Estándares intelectuales (claridad, precisión, relevancia, etc.) Acciones (soluciones, decisiones, consecuencias, etc.) Credibilidad de fuentes Identificar supuestos Comunicación (clarificación de significados) Metacognición Autorregulación y autorreflexión Actitudes y afectos (disposiciones)</p>

Fuente: Elaboración propia

Esta taxonomía propone el constructo PC como concepto universal y estructurante de las destrezas de pensamiento de alto nivel que se desarrolla en cuatro dimensiones básicas constituyentes. Adicionalmente, esta clasificación sirve para poner de manifiesto las relaciones entre el PC, como concepto organizador, y el resto de las destrezas, dispuestas en dimensiones y categorías del PC. La taxonomía desarrolla cada dimensión, a su vez, en categorías y subcategorías, las cuales pueden contener múltiples destrezas específicas. De esta manera, la dimensión de creatividad pone en juego las categorías de plantear buenas preguntas (con la contraparte de proponer buenas respuestas), observar, analizar y sintetizar, así como se ejemplifican algunas destrezas concretas de dicha dimensión, tales como comparar, clasificar, analizar, sintetizar, relacionar las partes y el todo y crear modelos y analogías; igualmente, la categoría de razonamiento empírico, que comporta la explicación con base en datos, se divide a su vez en las subcategorías de pensamiento inductivo, abductivo y estadístico.

Ahora bien, el objetivo de la taxonomía es ordenar el campo del PC con una finalidad doble: proveer un marco de referencia organizado para la investigación y facilitar la comprensión de este campo a los no especialistas —por ejemplo, los profesores—. Sin embargo, esta estructuración no debe perder de vista el hecho fundamental de que todas las dimensiones, categorías, subcategorías y destrezas elementales que forman parte del constructo general, no constituyen elementos independientes o disjuntos, sino que están relacionadas entre sí, ya que el pensamiento tiene un carácter integral y globalizado en cada persona; este hecho básico también implica que no se pueden establecer niveles de precedencia o superioridad de unas dimensiones, categorías o destrezas sobre las otras, aunque

algunas puedan considerarse cognitivamente más simples (por ejemplo, hacer preguntas) que otras, consideradas más complejas (la resolución de problemas).

Finalmente, la declaración de consenso de los expertos en PC para fines de evaluación educativa sostiene que

es esencial el desarrollo de estrategias de evaluación válidas y confiables a partir de las cuales los profesores puedan sacar inferencias razonables sobre el pensamiento crítico de los estudiantes, en contraste con su conocimiento específico y dominio de otras habilidades académicas (como leer o escribir). (Facione, 1998, p. 16).

Este estudio recomienda que, para la aceptabilidad de una estrategia o instrumento de evaluación de PC, se debe considerar la validez de contenido, la validez de constructo, la confiabilidad y la equidad. Además, la recomendación número 13 propone evaluar con frecuencia, tanto de forma diagnóstica como sumativa, empleando diferentes tipos de instrumentos. En últimas, la evaluación de PC debe ser explícita con el fin de reforzar su valor ante los estudiantes, sus familias y el público, apoyar los objetivos de mejora de los educadores e informar adecuadamente sobre las políticas educativas.

Por otro lado, es un lugar común de la psicología cognitiva el supuesto sobre la relación positiva entre el pensamiento y el aprendizaje, puesto que ambos comportan múltiples componentes y niveles, cognitivos y no cognitivos, interdependientes e inextricablemente relacionados entre sí. Además, diversas evidencias empíricas actuales apuntan a la dinámica interactiva entre PC y otras variables educativas que pueden contribuir a desarrollar mejor el aprendizaje (Phan, 2010). Por ello, este estudio plantea la validación de una prueba de evaluación para un conjunto específico de destrezas de PC. Las hipótesis iniciales dan cuenta de la validez y la fiabilidad de las pruebas de PC diseñadas y aplicadas a los estudiantes,

pues se tomaron 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

A continuación, se presentan los aspectos generales de la metodología adoptada en este estudio, que comprende la muestra de los estudiantes participantes, los reactivos de evaluación de PC elaborados y los procedimientos seguidos en su aplicación.

Participantes

Los participantes son tres grupos naturales de estudiantes pertenecientes a tres centros educativos diferentes —uno privado concertado y dos públicos, situados en dos ciudades pequeñas—. Se trata de una muestra opinática, seleccionada por su favorable disposición a participar en el estudio y formada por 88 estudiantes de los grados 9.º y 10.º (tercer y cuarto cursos de educación secundaria obligatoria, ESO); son 48 hombres y 40 mujeres con edades comprendidas entre los 14 y 17 años (un promedio de 15.4 años).

Instrumento

Este estudio es continuación de un banco de ítems y de un piloto previo desarrollado por los autores (Manassero y Vázquez, 2020); los ítems del instrumento actual son el producto de la revisión de las cuestiones cuyos resultados psicométricos fueron disfuncionales en el piloto, la adición de otras nuevas y la adaptación de las restantes al nivel educativo de los participantes. Las destrezas evaluadas por este instrumento son: deducción (como explicación causal, no como conclusión lógica), asunciones, secuenciación y toma de decisiones (tabla 2).

Los reactivos para evaluar esas destrezas y su contenido se diseñaron de acuerdo con los criterios generales para construcción de pruebas (representatividad, relevancia, diver-

sidad, claridad, sencillez y comprensibilidad) y otros específicos —la demanda cognitiva de cada reactivo va dirigida a la destreza para la que está diseñada y plantean un reto cuyo logro motiva y desafía a los estudiantes—. Es preciso añadir que el diseño equilibra el número de cuestiones por destreza y su diversidad permite también comparar reactivos estandarizados con otros figurativos y libres de cultura.

Tabla 2. Especificaciones del instrumento de evaluación aplicado en este estudio para evaluar cuatro destrezas de pensamiento en los grados 9.º y 10.º (educación secundaria).

Destrezas de pensamiento	Fuente	Tipo	Número de cuestiones
Deducción	Cornell (Nicoma)	Verbal	6
Asunción	Cornell (Nicoma)	Verbal	5
Secuenciación	Elaboración propia	Figurativo	6
Decisiones	Elaboración propia	Situaciones	6

Fuente: Elaboración propia

El test de Cornell presenta verbalmente una historia ficticia sobre unos exploradores que llegan al planeta Nicoma; los estudiantes responden preguntas sencillas relacionadas con la información desplegada en la historia y tal información, junto con algunas cuestiones referidas a las destrezas de deducción y asunción, se han incluido en este estudio—. Por otra parte, se excluyeron preguntas repetitivas o que presentaban problemas de comprensión (Ennis y Millman, 2005)

Un reactivo verbal sobre toma de decisiones fue tomado del test de Halpern y plantea una situación cotidiana que ofrece distintas alternativas de decisiones sobre la situación; los estudiantes valoran la calidad de esas alternativas para la situación planteada. En cuanto a los reactivos destreza de secuenciación y de decisiones, estos emplean materiales figurativos de elaboración propia para facilitar en los estudiantes motivación, comprensión y agilidad de respuestas; en ese orden de ideas, el diseño figurativo pretende que estas cuestiones sean aún más independientes de la cultura y los conocimientos curriculares de las materias escolares (muestras en anexo).

Así, pues, el instrumento está formado por 23 cuestiones libres de cultura, de las cuales 9 corresponden a contenidos enteramente figurativos (esto es, ejercicio del pensamiento basado en figuras). Conviene señalar que no hay conocimientos previos que sean requisitos para responder a la demanda cognitiva que exige lograr la respuesta correcta, de manera que las cuestiones suponen retos de pensamiento auténticos e independientes de todo tipo de conocimientos curriculares de los estudiantes. Adicionalmente, los formatos de respuesta son cerrados, pero de diversos tipos, porque facilitan la asignación de puntuaciones y hacen más rápidos y fiables la obtención y el análisis de los resultados. Aunque este formato cerrado no permite evaluar procesos de pensamiento, ofrece la ventaja de una evaluación estandarizada y rápida de las destrezas; en términos metodológicos, favorece el establecimiento de líneas base para comparar investigaciones, programas y metodologías de enseñanza, así como lograr que sus medidas ajusten la demanda cognitiva de las preguntas a la destreza que representan, esto con el fin de obtener una evaluación más válida y fiable de cada destreza.

Procedimiento

El instrumento fue aplicado por la profesora tutora de los estudiantes en su grupo natural al final de curso, utilizando dispositivos digitales y siguiendo directrices estandarizadas; para incentivar el esfuerzo y la motivación, el ejercicio se planteó como una prueba de evaluación de los aprendizajes del curso. Los estudiantes dispusieron del tiempo necesario y razonable para concluir sus respuestas sin límite fijo (usualmente, un periodo de clase).

Las respuestas correctas reciben un punto, las incorrectas cero puntos, y no se aplicaron correcciones por respuestas al azar. A su vez, las variables consideradas en los análisis son las puntuaciones alcanzadas en las diferentes destrezas evaluadas, obtenidas como suma de las puntuaciones en las preguntas que forman cada destreza (tabla 2). La suma de las puntuaciones de las cuatro destrezas evaluadas generó una de puntuación global de pensamiento, que puede considerarse un indicador que estima el nivel de PC de los estudiantes, con base en las cuatro destrezas evaluadas. Previo control de calidad y depuración de los datos por los autores, la base de datos con las puntuaciones de los estudiantes se procesó con SPSS 25.

Puesto que los diversos tipos de validez psicométrica no son independientes entre sí —pues conforman diversas partes de un todo y mantienen una influencia recíproca—, en este estudio se han atendido especialmente las formas de validez compatibles con el contexto, mediante análisis correlacional factorial de componentes principales (CP). La validez de contenido se basa en el análisis y evaluación inicial de un banco de cuestiones e instrumentos de PC por expertos, cuyo acuerdo seleccionó las cuestiones que se ajustaban mejor a la demanda cognitiva de cada destreza evaluada y al nivel educativo de los estudiantes, desde las pruebas pilotadas previamente. Por último, la fiabilidad se calculó mediante el estadístico alfa de Cronbach.

Para evaluar empíricamente la validez convergente de las destrezas de PC, se han recogido para cada estudiante sus calificaciones escolares finales obtenidas por cada estudiante en las nueve asignaturas del curso (escala de 1 a 10), en razón de que constituyen variables que actúan como criterios externos de validez.

Resultados y discusiones

Los descriptores estadísticos de las puntuaciones de las 23 variables consideradas en este estudio, obtenidos a partir las respuestas de los estudiantes en la aplicación de las pruebas, están resumidos en la tabla 3. Las puntuaciones de aciertos muestran una distribución equilibrada entre cuestiones fáciles y difíciles, con la gran mayoría de cuestiones (16) que logran un promedio de respuestas correctas intermedio (.70 y .30), una minoría de cuestiones (4) muy fáciles (aciertos por encima de .70) y otro reducido grupo de cuestiones (3) muy difíciles (aciertos por debajo de .30). Por otro lado, el promedio de aciertos global muy próximo al 50 por ciento (0.499) confirma la dificultad media del instrumento de evaluación para los grados 9.º y 10.º (tercer y cuarto cursos de educación secundaria), un valor deseable en general para cualquier prueba de evaluación.

Tabla 3. Proporción de aciertos en las 23 variables evaluadas con el instrumento de pensamiento crítico para los grados 9.º y 10.º, formado por cuatro destrezas (deducción, asunciones, secuenciación y toma de decisiones)

Ítems / Variables	Aciertos (0-1)	Desviación estándar
DEDUC1	.8864	.31919
DEDUC2	.5682	.49817
DEDUC3	.4205	.49646
DEDUC4	.5000	.50287
DEDUC5	.2727	.44791
DEDUC6	.8750	.33261
ASUNC1	.3409	.47673
ASUNC2	.3977	.49223
ASUNC3	.5455	.50078
ASUNC4	.3636	.48380
ASUNC5	.4545	.50078
SECUE1	.9318	.25350
SECUE2	.5568	.49961
SECUE3	.7273	.44791
SECUE4	.5795	.49646
SECUE5	.4205	.49646
SECUE6	.4659	.50170
DECIS1	.4205	.49646
DECIS2	.5568	.49961
DECIS3	.4886	.50274
DECIS4	.3523	.48042
DECIS5	.1705	.37819
DECIS6	.1818	.38790

Fuente: Elaboración propia

Las puntuaciones medias de las cuatro destrezas de PC calculadas sumando las puntuaciones obtenidas en las cuestiones que las forman, están resumidas en la tabla 4. En general, están próximas al punto medio del rango total de cada variable, pero por debajo del punto y medio en los casos de las destrezas asunciones y toma de decisiones, y por encima del punto medio del rango en deducción y secuenciación. A su vez, la puntuación máxima del rango se alcanza en todas, excepto en la toma de decisiones. En conjunto, este grupo de parámetros estadísticos para las destrezas de PC del instrumento de evaluación muestran también valores muy centrados y sin desviaciones notables.

Tabla 4. Estadística básica descriptiva de las destrezas pensamiento crítico evaluadas a partir del instrumento de evaluación aplicado

Destrezas de pensamiento	Número de cuestiones	Rango	Mínimo	Máximo	Media	Desviación estándar
Deducción	6	0 – 6	1	6	3.5227	1.11394
Asunción	5	0 – 5	0	5	2.1023	1.24143
Secuenciación	6	0 – 6	0	6	3.6818	1.58691
Decisiones	6	0 – 6	0	4	2.1705	1.34951
Global	23	0 – 23	5	19	11.4773	3.60707

Fuente: Elaboración propia

De igual modo, los parámetros estadísticos básicos de las calificaciones escolares obtenidas por los estudiantes en las asignaturas curriculares de los grados 9.º y 10.º, pueden apreciarse en la tabla 5. Como puede observarse, el rasgo más notable que presenta la distribución de las calificaciones escolares es la asimetría entre las puntuaciones máximas y mínimas, pues mientras las calificaciones máximas son alcanzadas en todas las asignaturas, las calificaciones mínimas se distribuyen

más irregularmente. La puntuación mínima de 1 punto solo se alcanza en lengua catalana, mientras geografía-historia presenta como calificación mínima la puntuación 5, y el resto de las asignaturas presentan como calificación mínima las puntuaciones 2 y 3. La asignatura física y química, por su parte, tiene la puntuación media más alta de la muestra, en tanto que la puntuación media más baja corresponde a la asignatura lengua catalana.

Tabla 5. Estadística básica descriptiva de las calificaciones obtenidas por los estudiantes participantes en este estudio en las asignaturas escolares (n = 88)

Asignaturas curriculares	Mínimo	Máximo	Media	Desviación estándar
Biología y geología	3	10	7.01	1.568
Educación física	3	10	7.38	1.549
Física y química	3	10	7.53	1.735
Geografía e historia	5	10	7.19	1.388
Lengua castellana	3	10	6.70	1.669
Lengua catalana	1	10	6.34	1.929
Matemáticas	2	10	6.59	1.779
Religión	2	10	7.13	1.622
Segunda lengua (SAL)	2	10	7.02	1.849
NOTA MEDIA			6.955	1.312

Fuente: Elaboración propia

Correlaciones entre las destrezas de pensamiento crítico y las calificaciones escolares

Uno de los argumentos más generalizados en favor de la educación de las destrezas de PC es su esperado impacto transversal en las competencias, los aprendizajes y la formación general de las personas, con independencia del debate entre el contexto general y específico de su enseñanza. Para confirmar empíricamente este impacto entre destrezas de pensamiento y aprendizaje, se presenta el análisis empírico de las correlaciones entre las variables que representan ambos elementos, pensamiento (destrezas) y aprendizaje (calificaciones escolares), como parámetros que justifican su relación mutua.

En ese sentido, el análisis de las correlaciones entre las destrezas indica que la única destreza de PC que presenta correlaciones significativas con las otras destrezas es la secuenciación, evaluada mediante cuestiones figurativas, y que es especialmente más alta si se la vincula con la destreza toma de decisiones. Por su lado, la destreza de asunciones solo se correlaciona significativamente con la destreza secuenciación (tabla 6).

Tabla 6. Correlaciones de Pearson entre las destrezas de pensamiento crítico (deducción, asunciones, secuenciación y toma de decisiones)

Destrezas	ASUNCIONES	SECUENCIA	DECISIONES
DEDUCCIÓN Correlación	.160	.225*	.269*
Sig. (bilateral)	.136	.035	.011
ASUNCIONES Correlación		.361**	.134
Sig. (bilateral)		.001	.215
SECUENCIA Correlación			.460**
Sig. (bilateral)			.000

**. La correlación es significativa en el nivel 0,01 (bilateral).

*. La correlación es significativa en el nivel 0,05 (bilateral).

Fuente: Elaboración propia

Puede decirse, entonces, que el análisis de las correlaciones entre las destrezas del PC y el aprendizaje escolar tiene varios aspectos notables. La puntuación global de pensamiento y la nota media de las calificaciones escolares, las dos variables globalizadoras de ambos constructos son las que exhiben valores de las correlaciones más elevados y significativos. Así, la puntuación global de PC presenta las correlaciones más altas con las calificaciones de las asignaturas SAL, lengua catalana y matemáticas; en contraste, las correlaciones más bajas se establecen con la asignatura religión. Por otro lado, la nota media de las calificaciones escolares muestra el índice de correlación más alto con la destreza toma de decisiones y la puntuación global de PC, aunque no son significativas con las destrezas y secuencias.

Las correlaciones de las cuatro destrezas con las calificaciones de las distintas asignaturas muestran un patrón diferencial claro entre ellas: las destrezas asunciones y secuencias no tienen correlaciones significativas (aunque son positivas) con ninguna de las asignaturas, excepto una (SAL); la destreza de deducción se correlaciona solo con tres asignaturas (lengua castellana, lengua catalana y SAL), y, por último, la destreza toma de decisiones correlaciona significativamente con todas las asignaturas,

de las que se destaca educación física. Es importante anotar que las correlaciones de las distintas asignaturas con las cuatro destrezas de PC son principalmente consecuencias del patrón diferencial anterior, donde asunciones y secuencias no se correlacionan con ninguna asignatura, en tanto que la toma de decisiones se vincula con todas y solo la destreza de deducción marca algunas diferencias en el grupo de tres asignaturas correlacionadas.

Tabla 7. Correlaciones entre las puntuaciones de las cuatro destrezas de pensamiento crítico, la puntuación global de pensamiento, las calificaciones las asignaturas escolares y la nota media de estas calificaciones escolares

Calificaciones	Pensamiento				
	GLOBAL	DEDUCCIÓN	ASUNCIONES	SECUENCIA	DECISIONES
Nota media	.350**	.225*	.162	.144	.430**
Biología y geología	.250*	.226	.096	.078	.320**
Educación física	.241*	.205	.004	.007	.464**
Física y química	.258*	.098	.140	.083	.383**
Geografía e historia	.232*	.068	.128	.112	.313**
Lengua castellana	.241*	.232*	.120	.025	.314**
Lengua catalana	.331**	.269*	.163	.148	.339**
Matemáticas	.321**	.196	.149	.198	.326**
Religión	.200	.087	.109	.094	.252*
Segunda lengua (SAL)	.372**	.225*	.135	.247*	.393**

** . La correlación es significativa en el nivel 0,01 (bilateral).

* . La correlación es significativa en el nivel 0,05 (bilateral).

Fuente: Elaboración propia

Las correlaciones entre las destrezas de PC y las calificaciones de aprendizajes escolares indican un patrón diferencial, donde la destreza de toma de decisiones se correlaciona significativamente con todos los aprendizajes escolares; además, asunciones y secuencias establecen vínculos positiva pero no significativamente con los aprendizajes, mientras que la destreza de deducción correlaciona significativamente con los aprendizajes lingüísticos. En suma, estos resultados sugieren una importancia positiva de las destrezas del PC en relación con los diversos aprendizajes escolares, pero a la vez diferencial, pues su mayor o menor magnitud depende de la destreza y el aprendizaje.

Fiabilidad del instrumento de evaluación de pensamiento crítico

La fiabilidad del instrumento de evaluación de PC para los grados 9.º y 10.º (tercer y cuarto cursos de educación secundaria), se ha calculado mediante el coeficiente de consistencia interna de Cronbach para la escala global de PC y cada una de las subescalas que corresponden a las cuatro destrezas evaluadas (tabla 8).

Tabla 8. Coeficientes de consistencia interna de la escala y las subescalas de destrezas de pensamiento crítico medidos con el estadístico alfa de Cronbach

Destrezas	N.º de ítems	Alfa de Cronbach (estandarizada)
Deducción	6	.114
Asunción	5	.272
Secuenciación	6	.609
Decisiones	6	.316
Global	23	.652

Fuente: Elaboración propia

Como puede apreciarse, el coeficiente de consistencia interna para la escala global de PC es moderado pero aceptable; en cambio, la fiabilidad de las cuatro subescalas que representan las cuatro destrezas de PC evaluadas es muy diferente: la deducción, las asunciones y la toma de decisiones presentan bajos coeficientes, que, por el contrario, podrían ser aceptables para la destreza secuencias. No obstante, debe tenerse en cuenta el efecto mecánico de la reducción del número de ítems sobre la fiabilidad, ya que todas las sub-escalas tienen muy pocos ítems. En conjunto, estos resultados indican la necesidad de una revisión del instrumento global y, especialmente, de las escalas que miden las destrezas de deducción, asunciones y toma de decisiones para mejorar su consistencia interna.

Validez del instrumento: análisis factorial de componentes principales

Por construcción, las cuatro destrezas de PC evaluadas con el instrumento son constructos independientes entre sí, pues están articuladas con cuestiones diferentes que representan las distintas destrezas de pensamiento. Sin embargo, todas las correlaciones entre ellas son positivas y algunas significativas, que indican distintos grados de una cierta relación empírica entre ellas (tabla 6). Así, pues, en este apartado se intenta verificar la validez

del instrumento contrastando, en primer lugar, el grado en que las cuestiones que representan las destrezas constituyen un constructo unidimensional y, en segundo lugar, el grado en que todas las cuestiones representan las cuatro destrezas definidas teóricamente en el instrumento. Para ello, las cuestiones que representan a cada destreza se sometieron, de manera separada, a un análisis factorial exploratorio de componentes principales (CP) con un modelo de factor único, esto con el propósito de verificar la hipótesis unidimensional en cada destreza (tabla 9).

Tabla 9. Análisis factorial exploratorio de componentes principales de las cuestiones correspondientes a cada una de las destrezas sometidas a un modelo de factor único

	DEDUCCIÓN		ASUNCIONES		SECUENCIA		DECISIONES	
Estadísticos	Variables	Cargas	Variables	Cargas	Variables	Cargas	Variables	Cargas
	DEDUC1	0,656	ASUNC4	0,591	SECUE3	0,687	DECIS1	0,838
	DEDUC5	-0,507	ASUNC5	0,557	SECUE4	0,666	DECIS2	0,821
	DEDUC6	0,471	ASUNC3	0,556	SECUE5	0,666	DECIS3	0,749
	DEDUC4	0,464	ASUNC1	0,489	SECUE6	0,517	DECIS6	-0,462
	DEDUC3	0,436	ASUNC2	0,290	SECUE1	0,503	DECIS5	0,19
	DEDUC2	0,394					DECIS4	-0,163
KMO*	.423		.516		.509		.679	
Bartlett** (Sig.)	.049		.626		.000		.000	
% varianza	24.50		25.84		34.38		36.90	

*. KMO: Prueba Kaiser-Meyer-Olkin de adecuación de muestreo

** . Bartlett: Prueba de esfericidad de Bartlett

Fuente: Elaboración propia

Los valores de prueba Kaiser-Meyer-Olkin de adecuación de muestreo son, entonces, relativamente bajos en las cuatro destrezas de pensamiento —siendo el más alto en toma de decisiones—. Sin embargo, la prueba de esfericidad de Bartlett es significativa en tres de las destrezas, excepto para la destreza asunciones. Análogamente, el porcentaje de varianza explicado por cada uno de los factores únicos extraídos en cada una de las cuatro destrezas de pensamiento sigue el mismo patrón definido por la prueba KMO, a saber, muy alto para toma de decisiones (37 %), un poco menor en secuenciación (34 %), y más bajo en asunciones (26 %) y deducción (25 %).

Por su parte, el análisis de las cargas de cada una de las variables sobre el factor único de cada destreza permite identificar las cuestiones del instrumento que aportan menos a la varianza explicada. Así, la destreza de secuenciación, formada toda ella por cuestiones figurativas, es la que presenta un perfil de cargas más altas y positivas, lo que indica que todas las cuestiones aportan una carga apreciable sobre el factor único; de manera similar, las cinco cuestiones de la destreza de asunciones presentan también cargas altas y positivas respecto al factor único que representan.

En contraste, la destreza de deducción, formada por cuestiones verbales, presenta una carga negativa sobre el factor único de un ítem (DEDUC5), que sería susceptible de revisión. Análogamente, la destreza toma de decisiones, formada por cuestiones de elaboración propia sobre situaciones decisivas, presenta cargas muy altas respecto al factor único en tres ítems, baja en otro (DECIS5) y negativa en otros dos (DECIS6 y DECIS4), que deberían asimismo revisarse.

En segundo lugar, y en vista de que el instrumento total de evaluación de PC está formado teóricamente por cuestiones correspondientes a cuatro destrezas, se ensayó una solución de cuatro factores con todas las cuestiones del instrumento de evaluación para el análisis factorial de CP. En ese sentido, la medida Kaiser-Meyer-Olkin de adecuación de muestreo es moderada (.600), aunque la prueba de esfericidad de Bartlett es favorable (significación < .000). El modelo de cuatro factores con una rotación Varimax converge en 6 iteraciones y explica el 31,1 % de la varianza generada por las 23 variables del cuestionario.

Tabla 10. Matriz de componentes rotados resultante de un análisis factorial exploratorio de componentes principales realizado sobre todas las cuestiones del instrumento de evaluación de pensamiento crítico, sometidas a un modelo de cuatro factores y una rotación Varimax

Variables	Factor 1	Factor 2	Factor 3	Factor 4
DECIS2	0,773			
DECIS1	0,704	0,256		
DECIS3	0,697			
SECUE4	0,531	0,207	0,241	
SECUE5	0,523	0,224	0,371	

Variables	Factor 1	Factor 2	Factor 3	Factor 4
DECIS6	-0,498		0,333	
SECUE1	0,452		0,204	
SECUE2		0,745		
ASUNC5		0,621		
SECUE3	0,347	0,468		
DEDUC5		0,463		
DECIS5		0,445	0,407	
DECIS4				
DEDUC6			0,719	
DEDUC1			0,515	
ASUNC1			0,449	-0,230
SECUE6	0,300	0,272	0,410	
DEDUC2				0,575
DEDUC3				0,511
ASUNC3	0,277			-0,461
ASUNC2		0,309	0,400	0,455
ASUNC4				-0,390
DEDUC4	0,220		0,255	0,380

Rotación Varimax con normalización Kaiser (converge en 6 iteraciones).

Cargas < .20 eliminadas y tachadas las cargas cruzadas sobre factores diferentes al teórico

Fuente: Elaboración propia

Ahora bien, los resultados para la matriz de cargas del análisis factorial exploratorio con cuatro factores sometidos a una rotación Varimax (tabla 10), indican que la constitución de los factores empíricos por la mayoría de las variables se ajusta a un modelo de cuatro factores basado en las cuatro destrezas del diseño teórico de la prueba, ya que las variables con las cargas principales en los cuatro factores empíricos se corresponden con varias cuestiones de las cuatro destrezas evaluadas.

Empero, la reproducción del modelo teórico por los factores latentes obtenidos en el análisis exploratorio de CP, presenta cargas cruzadas entre factores que sugieren aspectos que pueden ser mejorables.

En principio, vale la pena señalar que la matriz de componentes para los cuatro factores obtenidos (tabla 10) muestra que los dos factores latentes primeros y los dos últimos deberían corresponder a las destrezas

toma de decisiones, secuenciación, deducción y asunciones, respectivamente, porque contienen la mayoría de las variables correspondientes a las destrezas mencionadas y las cargas altas de las variables que lideran cada factor. De esta forma, el factor primero se hace corresponder con la destreza toma de decisiones, puesto que las tres primeras variables con las cargas altas corresponden a esa destreza clasificación e incluye las seis variables de esta destreza, aunque una con carga negativa y otras dos (DECIS5 y DECIS4) con cargas bajas. El factor segundo se corresponde significativamente con la destreza de secuenciación, pues las variables SECUE2 y SECUE3 tienen cargas altas y el resto de variables de esta destreza responden también a ese factor, aunque muestran cargas cruzadas sobre otros factores.

Por otro lado, los factores tercero y cuarto contienen la mayoría de las variables correspondientes a las destrezas de deducción y asunciones, a pesar de que se presentan mezcladas entre ambos factores. El tercer factor se corresponde apreciablemente con la destreza deducción, ya que las variables con mayores cargas (DEDUCÓ y DEDUC1) se relacionan con esta destreza y el resto de las variables de esta destreza tienen también cargas más bajas en este factor, si bien está contaminado con otros elementos de asunción. El cuarto factor corresponde a la destreza de asunción, pues las cargas mayores de este factor se vinculan con tres variables de asunción, aunque con dos que poseen cargas negativas, y el resto tienen cargas menores sobre este factor, que está contaminado también con tres variables procedentes de la destreza deducción.

En resumen, el análisis factorial exploratorio de CP indica que la estructura del conjunto de cuestiones correspondientes a cada destreza es unidimensional, lo cual apoyaría la validez de cada una de las subescalas como constructos con identidad propia coherente para representar cada destreza, aunque se identifican también algunas variables con cargas bajas sobre el factor único que podrían ser objeto de revisión. Además, la estructura del instrumento completo no ajusta bien el modelo teórico de cuatro destrezas, pues, aunque los cuatro factores latentes corresponden con las destrezas, aparecen contaminados con variables pertenecientes a otras destrezas. Por ello, los resultados configuran la validez del instrumento, aunque ofrecen datos para la revisión del instrumento sobre las cuestiones que contaminan los diferentes factores con cargas cruzadas. Con todo, estas últimas no son determinantes de modificación o supresión, dado que el PC tiene un sustrato común subyacente a todas las destrezas que justificaría las cargas cruzadas en los cuatro factores.

Conclusiones

El fin principal de esta investigación fue validar un instrumento para evaluar cuatro destrezas de PC (deducción, asunciones, secuenciación y toma de decisiones), adaptado según las necesidades de la enseñanza y la edad de estudiantes de grados 9.º y 10.º (15 y 16 años); así mismo, uno de sus aportes constituyó la novedad de evaluar la secuenciación, una destreza poco habitual en los instrumentos conocidos, lo que permite afirmar que se debe ir más allá de los instrumentos estandarizados, generalmente dirigidos a una población amplia para un uso general y, por ello, no adaptados a las necesidades específicas de edad o programas educativos concretos. No obstante, resulta necesario mantener algu-

nas cuestiones estandarizadas como referencia para las figurativas libres de cultura creadas por los autores y cuya comparación también es fuente de interesantes conclusiones (Ennis y Millman, 2005; Halpern, 2010).

De igual modo, los resultados muestran la validez del instrumento global, pues se correlaciona bien con el promedio de la calificación escolar y tiene un factor principal dominante; también es fiable, pues alcanza un valor aceptable del coeficiente de fiabilidad. Los resultados de las subescalas que representan las cuatro destrezas muestran una validez aceptable, pues cada destreza es unidimensional, aunque respecto de la prueba global exhiben numerosas cargas cruzadas entre ellas; las correlaciones con las calificaciones son todas positivas y más altas para la toma de decisiones. Por el contrario, la fiabilidad de las destrezas es más diversa, pues en todas está aminorada por el impacto a la baja de la reducción de su longitud (5-6 ítems), sin desconocer que la escala figurativa de secuenciación logra aun una fiabilidad aceptable. Estos resultados confirman la validez y fiabilidad de la prueba global, aunque los resultados para las escalas de destrezas deben tomarse con precaución, sugiriéndose aumentar el número de ítems y eliminar aquellos disfuncionales. No obstante, la muestra es baja para este tipo de análisis, que debe ser incrementada en una nueva aplicación (Muñiz y Fonseca-Pedrero, 2019).

En consecuencia, el hallazgo más novedoso derivado de este estudio es la posibilidad de desarrollar instrumentos de evaluación de PC válidos y fiables, adaptados a los perfiles y necesidades del programa de enseñanza de destrezas y a la edad de los alumnos. En particular, la correlación positiva entre las puntuaciones del instrumento y las calificaciones escolares es consistente con el carácter transversal del pensamiento respecto de los aprendizajes escolares, predicado por algunos (European Union, 2014), y constituye una cierta verificación empírica del mismo, aunque se requieren nuevos estudios para confirmar esta hipótesis.

Adicionalmente, los reactivos figurativos libres de cultura muestran una fiabilidad y validez mejor que otros más verbales, lo que apunta a una línea investigativa prometedora para el desarrollo de pruebas de pensamiento, especialmente, para edades de los alumnos donde la verbalidad y los conocimientos específicos pueden mediar más la elaboración de las respuestas. La línea de valoración del dominio de destrezas con reactivos libres de cultura sigue la tendencia actual de controlar las interferencias de los conocimientos, pero su elaboración y desarrollo requiere comparaciones adicionales con nuevas cuestiones y aplicaciones. Finalmente, la importancia de las destrezas de pensamiento para el aprendizaje en general, y para los aprendizajes más dependientes de las destrezas de pensamiento —tales como los aprendizajes STEM (ciencia, tecnología, ingeniería y matemáticas)—, hace esperable la promoción de futuras investigaciones empíricas que elaboren con mayor extensión los diversos aspectos pendientes en esta área relativamente inexplorada de la investigación didáctica en las diversas áreas específicas de la enseñanza (Tamayo, 2017; Tenreiro-Vieira y Vieira, 2014).

Referencias

- Ananiadou, K. y Claro, M. (2009). 21st Century Skills and Competences for New Millennium Learners in OECD Countries. *OECD Education Working Papers*, 41. <https://doi.org/10.1787/19939019>
- Colom, R., García, 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(1), 50-55. <https://journal.viterbo.edu/index.php/atpp/article/view/1129>
- European Union (2014). *Key Competence Development in School Education in Europe: KeyCoNet's review of the literature. A summary*. http://keyconet.eun.org/c/document_library/get_file?uuid=3a7a093c-4c8f-473c-8702-f38ed86b-b730&groupId=11028

- Ennis, R. H. y Millman, J. (2005). *Cornell Critical Thinking Test, Level x*. The Critical Thinking Company.
- Facione, P. A. (1998). Insight Assessment. <https://www.insightassessment.com/article/delphi-consensus-report-executive-summary-pdf>
- Fisher, A. (2009). *Critical Thinking: An Introduction*. 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*, (extraordin). <https://revistas.pedagogica.edu.co/index.php/TED/article/view/8789>
- Ford C. L. y Yore, L. D. (2014). Toward Convergence of Critical Thinking, Metacognition, and Reflection: Illustrations from Natural and Social Sciences, Teacher Education, and Classroom Practice. En A. Zohar y Y. J. Dori (eds.), *Metacognition in Science Education* (pp. 251-271). Springer.
- Fullan, M. y Scott, G. (2014). *Education Plus*. Collaborative Impact SPC.
- Halpern, D. F. (2010). *Halpern Critical Thinking Assessment*. SCHUHFRID.
- Hattie, J. (2009). *Visible learning: A Synthesis of over 800 Meta-analyses Relating to Achievement*. Routledge.
- Hattie, J. (2012). *Visible learning for teachers: Maximizing Impact on Learning*. Routledge.
- International Society for Technology Education. (2003). *National Educational Technology Standards for Teachers: Preparing Teachers to Use Technology*. Eugene, Oregon: International Society for Technology Education.
- Krathwohl, D. (2002). A Revision of Bloom's Taxonomy. *Theory into Practice*, 41(4), 212-218. <https://www.depauw.edu/files/resources/krathwohl.pdf>
- Madison, J. (2004). *James Madison Critical Thinking Course*. The Critical Thinking Co.
- Manassero-Mas, M. A. y Vázquez-Alonso, A (2019). 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. 17-38). UNICSUL.
- Manassero-Mas, M. A. y Vázquez-Alonso, A (2020, en prensa). Evaluación de destrezas de PC: Validación de instrumentos libres de cultura.
- McDonald, C. V., y McRobbie, C. J. (2012). Utilising Argumentation to Teach Nature of Science. En B. J. Fraser, K. G. Tobin y C. J. McRobbie (eds.), *Second International Handbook of Science Education* (pp. 969-986). Springer.
- Muñiz, J. y Fonseca-Pedrero, E. (2019). Diez pasos para la construcción de un test. *Psicothema*, 31(1), 7-16. <http://www.psicothema.com/pdf/4508.pdf>
- National Research Council (2012). *Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century*. The National Academies Press.
- Paul, R., y Nosich, G. M. (1993). A Model for the National Assessment of Higher Order Thinking. En R. Paul (ed.), *Critical Thinking: What Every Student Needs to Survive in a Rapidly Changing World* (pp. 78-123). The Foundation for Critical Thinking.

- Piaget, J. e Inhelder, B. (1997). *Psicología del niño*. Morata.
- Popper, K. R. (1975). La ciencia normal y sus peligros. En I. Lakatos y A. Musgrave (eds.), *La crítica y el desarrollo del conocimiento científico* (pp. 149-158). Grijalbo.
- Saiz, C. (2017). *Pensamiento crítico y cambio*. Pirámide.
- Shayer, M. y Adey, P.S. (eds.) (2002). *Learning Intelligence: Cognitive Acceleration across the Curriculum from 5 to 15 Years*. Open University Press.
- Simonneaux, L. (2014). From Promoting the Techno-sciences to Activism: A Variety of Objectives Involved in the Teaching of ssis. En L. Bencze y S. Alsop (eds.), *Activist Science and Technology Education* (pp. 99-112). Springer.
- Swartz, R. J., Costa, A. L., Beyer, B. K., Reagan R. y Kallick, B. (2013). *El aprendizaje basado en el pensamiento*. SM.
- Tamayo, O. (2017). Interacciones entre naturaleza de la ciencia y pensamiento crítico en dominios específicos del conocimiento. *Enseñanza de las ciencias*, (n.º extraordinario), 521-526. <https://core.ac.uk/download/pdf/158654749.pdf>
- Tenreiro-Vieira, C. y Vieira, R. M. (2014). *Construyendo prácticas didático-pedagógicas promotoras da literacia científica e do pensamento crítico*. Iberciencia.
- Vázquez-Alonso, Á. y Manassero-Mas, M. A. (2018). Más allá de la comprensión científica: educación científica para desarrollar el pensamiento. *Revista Electrónica de Enseñanza de las Ciencias*, 17(2), 309-336.
- Vázquez-Alonso, A. y Manassero-Mas, M. A. (2019). Scientific Thinking and Critical Thinking: A Keystone for History, Philosophy and Sociology of Science Teaching. En F. Seroglou y V. Kolountzos (eds.), *Conference Proceedings 15th International History, Philosophy and Science Teaching Conference*. Grafima Publications.
- Walton, D. y Macagno F. (2015). A Classification System for Argumentation Schemes. *Argument and Computation*, 6(3), 219-245.
- Watson, G. y Glaser, E. M. (2002). *Watson-Glaser Critical Thinking Appraisal-ii: Form e*. Pearson.

Para citar este artículo

Manassero-Mas, M. A. y Vázquez-Alonso, A. (2020). Las destrezas de pensamiento y las calificaciones escolares en educación secundaria: validación de un instrumento de evaluación libre de cultura. *Revista Tecné, Episteme y Didaxis: TED*, (48), 33-54. <https://doi.org/10.17227/ted.num48-12375>

Anexo

La siguiente es una muestra de reactivos de la prueba; el primero es verbal, tomado del test de Cornell (Nicoma), y el segundo es figurativo, representativo de la destreza secuenciación.

Test de Cornell (Nicoma)

Después de observar la aldea durante una hora, usted toma su grupo para volver al campamento. Ordena al sargento elaborar un informe para presentar al capitán. Para hacer el informe, el sargento toma como ciertas algunas ideas, aunque sin decirlo abiertamente. Estas ideas servirán de base para su razonamiento.

Su trabajo ahora consiste en seleccionar las ideas que el sargento probablemente toma como supuestamente ciertas para justificar esos razonamientos, sin expresarlas realmente.

74. "Los aldeanos no tienen centinelas en el exterior. Puedo garantizarlo porque miramos con mucha atención y no hemos visto ninguno".

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

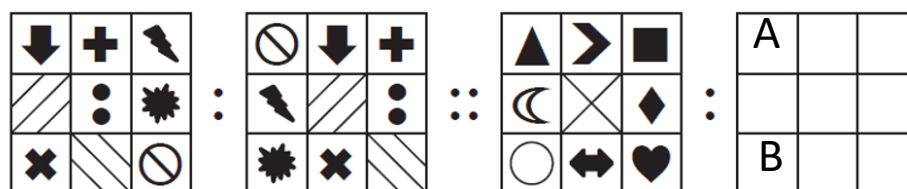
A. Los centinelas sólo son empleados por personas que quieren que alguien vigile por ellas.

B. Los centinelas pueden ser vistos por personas atentos a ellos.

C. Si se viese un centinela, entonces éste no sería muy importante.

Destreza secuenciación

27. Observa cuidadosamente las secuencias de figuras contenidas en las nueve casillas de los dos primeros tableros, comparándolas, para intentar comprender la relación entre ellos.



Tu tarea consiste en señalar las imágenes que deberían contener las casillas A y B del cuarto tablero, si éste se relacionase con el tercero de la misma manera que el segundo está relacionado con el primero.

¿Qué imagen debería tener la casilla A del cuarto tablero?

¿Qué imagen debería tener la casilla B del cuarto tablero?

