



Diseño y validación de instrumentos de evaluación de pensamiento crítico en educación primaria

- Design and Validation of Assessment Tools for Critical Thinking in Primary Education
- Desenho e validação de ferramentas de avaliação do pensamento crítico no ensino básico

Resumen

El objetivo principal de este artículo de investigación es analizar la elaboración y validación de instrumentos de evaluación de habilidades de Pensamiento Crítico (PC) aplicables en el alumnado de entre 11 y 14 años, así como analizar una implementación de estos en un contexto educativo. En primer lugar, se realiza el diseño de una modificación del *Halpern Critical Thinking Assessment* (HCTA, por sus siglas en inglés). Seguidamente, se analizan las propiedades psicométricas con los resultados obtenidos de la aplicación del *test* en una muestra de 171 estudiantes. Los resultados obtenidos del análisis permiten determinar una estructura factorial similar a la del *test* original y una fiabilidad aceptable para su aplicación. En segundo lugar, se diseña una rúbrica cualitativa para evaluar actividades en el contexto de aula. En última instancia, se aplican ambos instrumentos en un aula integrada por diecisiete estudiantes de entre 11 y 12 años, siguiendo la estructura *pretest* —intervención (realización de actividades de trabajo explícito del PC)— *posttest*. Los resultados de la implementación, extraídos con la modificación del HCTA, permiten observar una mejora de las habilidades PC, que se corresponde con la mejora reflejada en los resultados obtenidos en la evaluación mediante la rúbrica de esas mismas habilidades en las actividades de aula.

Palabras clave

ciencia y la tecnología; educación; evaluación; pensamiento científico; pensamiento crítico

Vanessa Ortega-Quevedo*
Cristina Gil Puente**
Cristina Vallés Rapp***
María Antonia López-Luengo****

* Máster en investigación en Ciencias Sociales; Educación; contratada predoctoral FPU en Universidad de Valladolid, Departamento de Didáctica de las Ciencias Experimentales, Sociales y de la Matemática de la Universidad de Valladolid.
Orcid: <https://orcid.org/0000-0002-5742-4678>.
Correo electrónico: vanessa.ortega@uva.es

** Doctora en Ciencias Geológicas por la Universidad Complutense de Madrid; profesor contratado doctor en la Universidad de Valladolid, Departamento de Didáctica de las Ciencias Experimentales, Sociales y de la Matemática de la Universidad de Valladolid.
Orcid: <https://orcid.org/0000-0001-5794-5564>.
Correo electrónico: cristina.gil.puente@uva.es

*** Doctora en Ciencias por la Universidad Autónoma de Madrid; profesor contratado doctor en la Universidad de Valladolid, Departamento de Didáctica de las Ciencias Experimentales, Sociales y de la Matemática de la Universidad de Valladolid.
Orcid: <https://orcid.org/0000-0002-6772-0606>.
Correo electrónico: cvalles@dce.uva.es

**** Doctora en Ciencias de la Educación por la Universidad de Valladolid, profesor colaborador de la Universidad de Valladolid, Departamento de Didáctica de las Ciencias Experimentales, Sociales y de la Matemática de la Universidad de Valladolid.
Orcid: <https://orcid.org/0000-0001-7334-5278>.
Correo electrónico: mariaantonia.lopez@uva.es



Abstract

The main objective of this article is to analyze the design and validation of assessment instruments of Critical Thinking (CT) skills applicable to students between 11 and 14 years old; as well as to analyse an implementation of these in an educational context. First, a modification of the Halpern Critical Thinking Assessment (HCTA) is conducted. Next, the psychometric properties are analysed with the results obtained from the application of the test in a sample of 171 students. The results obtained from the analysis make it possible to determine a factorial structure like that of the original test and an acceptable reliability for its application. Secondly, a qualitative rubric is designed to evaluate activities in the classroom context. Ultimately, both instruments are applied in a classroom made up of 17 students aged between 11 and 12, following the pre-test-intervention (carrying out explicit work activities of the CT)-post-test structure. The implementation results extracted with the HCTA modification allow observing an improvement of the CT skills that corresponds to the improvement reflected in the results obtained in the evaluation through the rubric of those same skills in the classroom activities.

Keywords

science and technology; education; evaluation; scientific thinking; critical thinking

Resumo

O objetivo principal deste artigo de investigação é analisar a construção e validação de instrumentos de avaliação de competências de Pensamento Crítico (CP) aplicáveis a alunos entre os 11 e os 14 anos de idade, bem como analisar a sua implementação em contexto educativo. Primeiro, uma modificação da Avaliação de Pensamento Crítico Halpern (HCTA) é projetada. Em seguida, as propriedades psicométricas são analisadas com os resultados obtidos a partir da aplicação do teste em uma amostra de 171 alunos. Os resultados obtidos a partir da análise permitem determinar uma estrutura fatorial semelhante à do teste original e uma confiabilidade aceitável para sua aplicação. Em segundo lugar, uma rubrica qualitativa é concebida para avaliar as atividades no contexto da sala de aula. Em última análise, ambos instrumentos são aplicados numa sala de aula composta por 17 alunos com idades compreendidas entre os 11 e os 12 anos, seguindo a estrutura do pré-teste-intervenção (realização de actividades de trabalho explícito do CP)-pós-teste. Os resultados da implementação extraídos com a modificação da HCTA permitem observar uma melhoria das habilidades de CP que corresponde à melhoria refletida nos resultados obtidos na avaliação através da rubrica dessas mesmas habilidades nas atividades de sala de aula.

Palavras chave

ciência e tecnologia; ensino; avaliação; pensamento científico; pensamento crítico

Introducción

La sociedad digital facilita a la población gran cantidad de información acerca de distintas temáticas, entre las que se encuentran verdades y falsedades que los ciudadanos deben discernir para crear una opinión propia y fundamentada sobre estas o actuar y adaptarse a una situación compleja conforme a la información evaluada. Esta panorámica actual pone en manifiesto la necesidad del desarrollo de habilidades como el Pensamiento Crítico (PC), entendido como un proceso cognitivo que se ejecuta de forma deliberada con el fin de extraer conclusiones sobre distintas temáticas. En esta línea, estudios como el informe Work Skills 2020, elaborado por el instituto de investigación de la Universidad de Phoenix (Davies et al., 2011), prevén que las capacidades cognitivas como el PC serán necesarias para acceder a la mayor parte de los empleos.

El PC es una forma superior de razonamiento esencial para participar activa y críticamente en la sociedad, motivo por el cual debe contemplarse en los sistemas educativos como parte del desarrollo integral del alumnado (Vieira et al., 2014). Así, la normativa española (Ley Orgánica 8, 2013), determina que el PC es una competencia transversal que debe adquirirse y desarrollarse desde edades tempranas, dado que el proceso de aprendizaje se proyecta a lo largo de toda la vida, aunque no contempla su tratamiento explícito en el desarrollo curricular. Esta inclusión del PC como elemento transversal en los currículos educativos conlleva varias dificultades, entre las cuales destacamos: el trabajo eficaz de las habilidades de PC como fomento de su desarrollo y la evaluación de estas como forma de visualizar dicho desarrollo.

Según lo expuesto, los objetivos principales de este estudio son: 1) validar un instrumento cuantitativo que proporcione datos sobre la capacidad que tienen los estudiantes entre los 11 y los 14 años, de aplicar habilidades de PC en distintos contextos (transferencia); 2) contrastar los resultados del instrumento cuantitativo con la aplicación de instrumentos cualitativos (rúbricas), que permitan evaluar la aplicación de habilidades de PC en ejercicios concretos de aula.

Marco de referencia

¿Por qué trabajar el PC en el aula de Ciencias?

El pensamiento científico y el pensamiento crítico

Comúnmente a los científicos se les atribuye una característica forma de pensar y actuar propia de la construcción de conocimiento científico. Estos procesos de pensamiento son provisionales y están en constante revisión, como la propia ciencia. Vázquez-Alonso et al. (2018) relacionan los principales componentes del pensamiento científico y las destrezas de pensamiento que estos suponen (Tabla 1). Como señalan dichos autores, las destrezas de pensamiento se pueden aplicar en distintas configuraciones en función de la disciplina científica en la cual se esté implementando o según la problemática a resolver.

Tabla 1. Desarrollo de los aspectos del pensamiento científico y sus destrezas específicas asociadas elaborado a partir de las citas sobre este tema

Aspectos	Destrezas asociadas al pensamiento científico
1) Observar con los sentidos	Observar. Recoger datos. Recopilar información. Describir. Historiar
2) Categorizar lo observado	Definir supuestos conceptos y problemas. Clasificar, organizar, ordenar, entes. Analizar, comparar, contrastar
3) Reconocer patrones en objetos y sucesos	Identificar datos probatorios. Cuantificar medidas. Descubrir regularidades. Sintetizar. Generalizar empíricamente
4) Crear y comprobar hipótesis	Formular preguntas. Identificar problemas y supuestos. Formular hipótesis. Planificar y desarrollar investigaciones. Aplicar análisis estadísticos. Usar adecuadamente la tecnología. Aplicar pensamiento matemático y computacional
5) Pensar en causas y efectos	Controlar efectos de múltiples variables. Atribuir causalidad. Usar la lógica. Dar interpretaciones válidas y fiables. Resaltar datos sorprendentes o contradictorios. Plantear problemas prácticos

Aspectos	Destrezas asociadas al pensamiento científico
6) Apoyar eficazmente la teoría con pruebas	Emitir juicios críticos. Sacar conclusiones. Aceptar y rechazar teorías. Utilizar pruebas. Justificar y validar conocimientos. Argumentar. Sugerir solución o comparación. Construir explicaciones y teorías coherentes
7) Visualizar el pensamiento científico	Crear e imaginar investigaciones y experimentos. Simular modelos y datos en figuras, diagramas y gráficos. Construir tablas. Desarrollar tecnologías. Afrontar necesidades y problemas sociales. Informar políticas científicas
8) Ser consciente y controlar el pensamiento propio (meta-procesos)	Pensar correctamente. Tener una actitud escéptica. Mostrar apertura de mente. Desafiar el conocimiento con alternativas. Evaluar suposiciones. Predecir. Buscar con creatividad e imaginación nuevas ideas y conocimientos
9) Usar metáforas y analogías	Crear modelos, analogías y teorías. Resolver problemas. Buscar bases para el conocimiento. Replicar teorías, modelos, conocimientos
10) Usar la heurística de "confirmar rápido, revocar después"	Explorar literatura, hipótesis. Deducir conclusiones. Inferir conclusiones. Tomar decisiones. Abducir conclusiones
11) Comunicar, evaluar, compartir, colaborar y pensar la información.	Comunicar y compartir conocimiento. Publicar descubrimientos. Trabajar en equipo cooperativamente. Participar en congresos. Debatir con los colegas teorías y soluciones. Evaluar comunicaciones de otros para publicar. Asumir responsabilidades (personal y social)

Fuente: Vázquez et al. (2018, p. 314).

Por otra parte, desde el punto de vista de la psicología, se establece el constructo de PC, definido de diferentes formas y delimitado por distintas taxonomías, según los principales autores (Ennis, 1996; Facione, 1990; Halpern 2016; Lipman, 1991; Paul, 2005; Paul et al., 2002), pero que comparte una esencia común: el PC está conformado por un componente cognitivo (habilidades o destrezas delimitadas) y otro disposicional (disposiciones que ponen en manifiesto la motivación para poner en práctica estos procesos de pensamiento).

Existen múltiples definiciones de PC cuyo análisis permite extraer que el PC es un conjunto de procesos que se implementan de forma intencional, con el objetivo de establecer conclusiones sobre diferentes temas y determinar cómo al aplicar los procesos asociados al PC, los datos relacionados o problemáticas a resolver se descomponen, sintetizan y evalúan reflexivamente para llegar una conclusión o solución sobre estos. Dicha solución a su vez también es evaluada con el fin esclarecer si es mejorable (Ortega-Quevedo et al., 2019).

En conclusión, el constructo PC contempla cuestiones fundamentales del pensamiento científico desde una perspectiva general, pensada para ser aplicada en la vida cotidiana. Este planteamiento coincide con la necesidad de desarrollo de estas capacidades en la población, especialmente en los jóvenes, y de ahí la necesidad de su inclusión y promoción en los contextos educativos (Vieira et al., 2014). De todo ello se ve beneficiada la didáctica de las ciencias, no solo porque la mejora de estas habilidades permite comprender mejor los procesos del pensamiento científico, sino porque el desarrollo de estos procesos se advierte necesario para solventar algunas de las dificultades existentes a la hora

de que los estudiantes puedan comprender temas sobre naturaleza de la ciencia (Vázquez-Alonso et al., 2018). En consecuencia, desde este estudio se defiende el desarrollo del PC en contextos educativos y, por ende, la necesidad de creación de instrumentos para su evaluación en las distintas etapas educativas.

Pensamiento crítico, taxonomía de Halpern

Entre los distintos autores que estudian el PC, Halpern (2014) considera que el componente cognitivo del PC está dividido en cinco dimensiones (análisis argumental, comprobación de hipótesis, probabilidad e incertidumbre, resolución de problemas, y razonamiento verbal), que se pasan a analizar:

Análisis argumental: para realizar un análisis argumental hay que comprender los componentes de un argumento; los argumentos están formados por distintos componentes, de los que los fundamentales son las razones y las conclusiones. No es posible formar un argumento sin, al menos, una razón y una conclusión. Otros componentes son las suposiciones, los calificadores y los contraargumentos (Halpern, 2014).

Comprobación de hipótesis; una hipótesis es una afirmación, sobre el mundo que nos rodea, establecida según una creencia no comprobada que posteriormente se confirma o refuta. Es decir, comprobar hipótesis es una forma de esclarecer la verdad sobre algo. Es preciso emplear este tipo de confirmación del conocimiento para interactuar en el mundo actual. Este proceso de confirmación de hipótesis se puede realizar mediante métodos de evaluación de la información (Halpern, 2014).

Probabilidad e incertidumbre: el estudio de la probabilidad y la incertidumbre es importante en el desarrollo de la vida cotidiana, debido a que contempla las posibilidades de que un suceso ocurra (Halpern, 2014).

Resolución de problemas: para resolver un problema se debe seguir un modelo que permita comprender las cuestiones relacionadas con dicho problema y contemplar las

alternativas disponibles que permiten llegar a una solución (Halpern, 2014).

Razonamiento verbal: el razonamiento verbal contempla los conocimientos sobre el lenguaje cotidiano que permiten identificar ambigüedades o analogías en el discurso argumental (Halpern, 1998).

Evaluación del PC

Como se ha reflejado, el PC es un constructo de difícil delimitación y sobre el cual no hay acuerdos cerrados relativos a sus componentes. Estas dificultades encontradas en su definición conceptual implican también a su evaluación, lo cual hace que la creación de instrumentos de evaluación del PC sea una tarea muy compleja (Saiz et al., 2008). Un ejemplo de instrumento de evaluación del PC es el *Halpern Critical Thinking Assessment* (HCTA, por sus siglas en inglés), el cual se elaboró empleando situaciones de la vida cotidiana donde se precisa el uso del PC, para así poder medir su aplicación a contextos. Este tipo de planteamientos "consiguen que a la persona que cumplimenta la prueba le resulte familiar lo que se le propone; en consecuencia, se enfrentará a ello como acostumbra a hacerlo en su entorno" (Saiz et al., 2008, p. 39).

En concreto, el HCTA está formado por veinte situaciones cotidianas (veinticinco en sus versiones iniciales), sobre las que hay que resolver una serie de cuestiones planteadas. En primer lugar, se proponen preguntas abiertas sobre la situación, dando la oportunidad a la persona de crear su propio pensamiento sobre algo. Finalmente, se presenta una opción de respuestas cerradas que responden a cuestiones planteadas sobre la situación (Halpern, 2016).

Desde el punto de vista de este estudio, el HCTA es un instrumento de evaluación muy completo, pero también muy complejo, que presenta dos inconvenientes principales para su uso en contextos educativos: 1) está dirigido a un público adulto, lo cual limita su uso a etapas de educación superior; 2) aunque evalúa la transferencia de la aplicación del PC a contextos reales, algo muy útil como forma

de diagnóstico y de comprobación del desarrollo de las habilidades asociadas, no es posible aplicarlo para evaluar actividades de aula ajenas a las situaciones de la propia prueba.

Metodología

El proceso de generación y validación del *test* para contextos educativos (11-14 años), a partir de la adaptación de dos dimensiones del HCTA, así como la elaboración de rúbricas, surgió de la inexistencia de instrumentos que permitieran: 1) medir la capacidad del alumnado de aplicar habilidades de PC en diferentes contextos; y 2) evaluar la implementación de actividades que potencian las habilidades de PC en el desarrollo de tareas diarias en el aula. Tras el diagnóstico de estas necesidades se inició el proceso de elaboración y aplicación de los instrumentos que se refleja en la Figura 1 y que se detalla en los siguientes apartados.

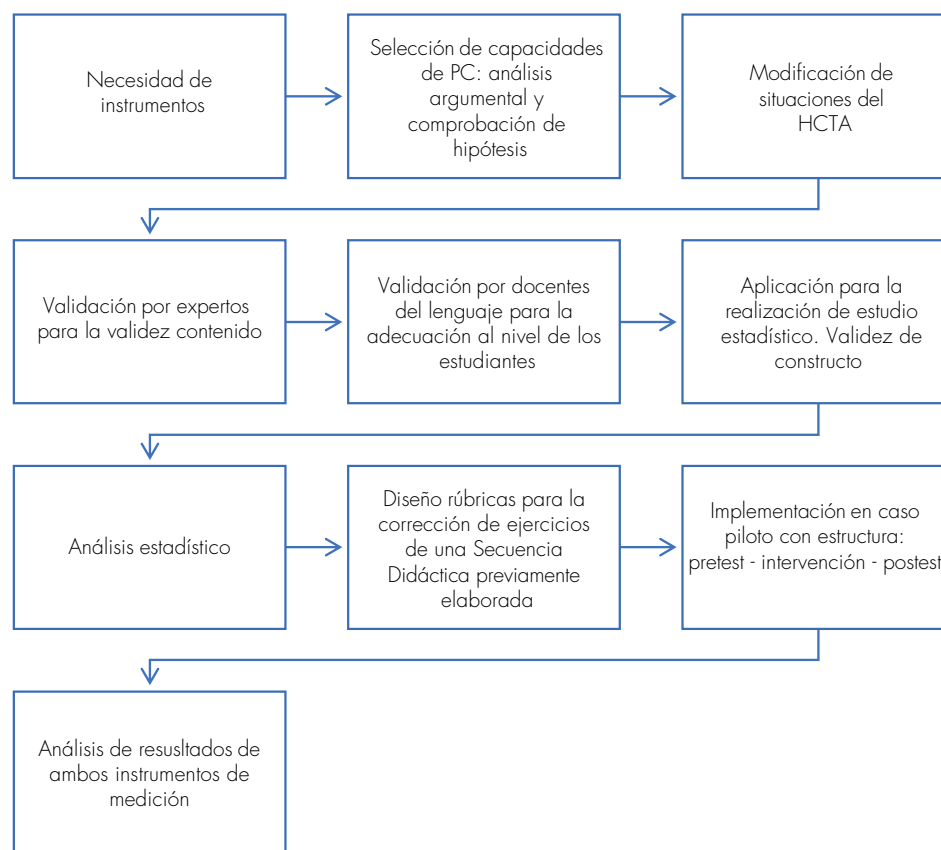


Figura 1. Procesos de diseño, validación e implementación de instrumentos de evaluación de PC

Fuente: elaboración propia.

Modificación del HCTA adaptada a estudiantes de entre 11 y 14 años

En primer lugar, tomando como referencia la taxonomía de Halpern, se seleccionaron las capacidades de PC que se trabajan de forma explícita en las diversas actividades de enseñanza-aprendizaje que conforman una secuencia didáctica determinada (ver sección Instrumentos didácticos. Secuencia de Enseñanza Aprendizaje SEA). Estas capacidades son análisis argumental y comprobación de hipótesis, cuya evaluación se persigue mediante los instrumentos diseñados y que se delimitan a continuación tomando como referente teórico la taxonomía de Halpern.

Para emitir una conclusión es preciso efectuar un análisis que evalúe las razones que apoyan o que no apoyan dicha conclusión, a este análisis le denominamos análisis argumental. Mediante las situaciones planteadas para esta dimensión se persigue medir las capacidades de identificación y elaboración de los componentes de un argumento (razón, contraargumento y conclusión).

La comprobación de hipótesis debe considerarse en contextos cotidianos y no solamente en investigación formal; así, las situaciones concebidas para esta dimensión evalúan si la información o muestra en la cual se fundamenta una hipótesis es válida para la confirmación o la refutación de esta.

En segundo lugar, se procede a la construcción y validación del instrumento. A continuación, se detallan las fases preliminares de elaboración del *test* y la fase de validación estadística.

Fases preliminares

1. Modificación de situaciones del HCTA. Se selecciona la estructura de las situaciones del *test* HCTA de las dimensiones de estudio y se decide limitar la extensión del *test* a dos situaciones por dimensión, con el fin de evitar la fatiga cognitiva de los estudiantes. Seguidamente, se escogen temas cercanos a la realidad del alumnado para realizar las modificaciones. Después, se reestructura la situación en función de la temática seleccionada;

finalmente, se adecúan las preguntas de cada situación a la temática y al nivel de los estudiantes, y se establece la métrica (ver a modo de ejemplo la situación 1 enmarcada en la Tabla 2 y el conjunto de las situaciones en el Anexo I). Cabe señalar que en este proceso de construcción se tiene en consideración el empleo de un lenguaje sencillo y comprensible para el alumnado del intervalo de edad.

Tabla 2. Métrica de la segunda situación de la modificación del HCTA para estudiantes de 11-14 años

Situación 2		Categoría: análisis argumental	
Hay muchas oportunidades para los especialistas en informática. La verdad es que deberías especializarte en esta ciencia. El trabajo es interesante, hay muchas posibilidades de empleo y los sueldos son buenos. Por supuesto, no es una buena especialidad si le temas a las matemáticas o te gusta trabajar al aire libre			
Parte Abierta, primera parte Puntuación máxima 1	1 punto	0 puntos	
	Extrae una conclusión del texto	Respuestas irrelevantes	
Parte Abierta, segunda parte Puntuación máxima 3	1 punto	0 puntos	
	1 punto por cada una de las razones expuestas en el texto	Respuestas irrelevantes	
Parte cerrada Puntuación máxima 4	1 punto si la respuesta a la primera opción es conclusión	1 punto si la respuesta a las opciones 2 y 3 es razón	1 punto si la respuesta a la tercera opción es contraargumento

Fuente: elaboración propia.

2. Validación por expertos. Se envía la modificación del instrumento a tres expertos en la materia, considerados como tal por su trayectoria de investigación en el campo de didáctica de las ciencias experimentales o en psicología. Este proceso de revisión propuso la realización de cambios mayores, como el planteamiento de una de las situaciones, y menores que contemplaban aspectos de adecuación de las respuestas cerradas, los cambios sugeridos fueron ejecutados y revisados. Con este proceso se pretende afianzar la validez de contenido, demostrando que la modificación se ajusta a la fundamentación teórica del constructo.

3. Validación del lenguaje por docentes. Se envía el *test* a docentes para comprobar que las modificaciones en la temática y la adaptación al lenguaje son adecuadas para la implementación con estudiantes de entre 11 y 14 años. En este caso no se sugirieron cambios.

Fase de validación estadística

Una vez aplicadas las modificaciones y validaciones anteriores se implementa el *test* modificado en una muestra de 171 estudiantes de la provincia de Segovia (España) para poder realizar el análisis estadístico empleando el software SPSS Statistics 24. En esta fase se realiza la prueba de fiabilidad de Alfa de Cronbach y un análisis factorial exploratorio para demostrar la validez de constructo.

Prueba de fiabilidad Alfa de Cronbach

Los resultados de la prueba Alfa se presentan en la Tabla 3 por dimensiones, junto con los resultados de esta misma prueba en el en su versión española HCTAes (Nieto et al., 2009). Para la interpretación de esta prueba y su comparación entre los distintos instrumentos es preciso no perder de vista el número de elementos del *test* por su influencia en la propia prueba alfa. Por ejemplo, en la dimensión comprobación de hipótesis se observa que el *test* HCTA en su versión en español obtiene un índice Alfa de fiabilidad de 0,635 con 10 ítems, mientras que el índice obtenido para la modificación realizada por este equipo es un poco más bajo (0,615), pero se obtiene con solamente cuatro ítems. Por último, se presenta la fiabilidad total del instrumento modificado mediante esta misma prueba 0,611; de estos resultados se comprueba que la fiabilidad tiene índices aceptables.

Tabla 3. Resultados de la prueba Alfa de Cronbach por dimensiones del HCTA

Dimensión	N.º de elementos del test		Alfa de Cronbach	
	HCTAes	HCTA modificado	HCTAes	HCTA modificado
Comprobación de hipótesis	10	4	0,635	0,615
Análisis de argumentos	10	4	0,465	0,447

Fuente: elaboración propia.

Análisis factorial exploratorio

Con el fin de analizar la validez de constructo se realiza un análisis factorial exploratorio. Una vez obtenidos los índices pertinentes de la prueba de Kaiser-Meyer-Olkin (0,69) y la prueba de Esfericidad de Bartlett ($\chi^2 = 212,226$; $p < 0,05$), se lleva a cabo un análisis de los componentes principales. De este análisis se extraen los valores recogidos en la Tabla 4 en la que se muestra la varianza explicada. En la tabla se puede observar cómo se obtienen dos factores que explican el 48,884 % de la varianza, el cual se considera suficiente en vista de la reducción de ítems aplicados y el nivel de la población a la que está dirigido el instrumento.

Tabla 4. Varianza total explicada

Componente	Sumas de extracción de cargas al cuadrado		
	Total	% de varianza	% acumulado
1	2,534	31,676	31,676
2	1,377	17,209	48,885

Nota: método de extracción: análisis de componentes principales.

Fuente: elaboración propia.

En la Tabla 5 se presentan los ítems que conforman cada uno de los factores. En este caso se puede observar que el componente 1 está formado por los ítems formulados para las situaciones de comprobación de hipótesis más el ítem correspondiente con las respuestas abiertas de la cuarta situación, destinada al análisis argumental. Además, el ítem de respuestas cerradas de la cuarta situación correlaciona en negativo con este primer componente. Por otra parte, el componente 2 está formado por todos los ítems relacionados con las situaciones destinadas al análisis argumental, a excepción del ítem de respuestas abiertas de la cuarta situación.

Tabla 5. Matriz de componente

		Componentes	
		1	2
Comprobación de hipótesis	Situación 1 Abierta	0,751	
	Situación 1 Cerrada	0,639	
	Situación 3 Abierta	0,694	
	Situación 3 Cerrada	0,511	
Análisis argumental	Situación 2 Abierta		0,655
	Situación 2 Cerrada		0,764
	Situación 4 Abierta	0,519	
	Situación 4 Cerrada	-0,491	0,646

Nota: método de extracción: análisis de componentes principales.
Nota: método de rotación: Promax con normalización Kaiser.

Fuente: elaboración propia.

Esta división demuestra que no existe unidimensionalidad y que la construcción del PC se compone de un conjunto de habilidades, como se presenta en la fundamentación teórica. Asimismo, los componentes resultantes son los esperados, pues se ajustan al modelo obtenido por Halpern (2016) en el HCTA, ya que se ha conseguido la misma cantidad de factores que dimensiones se pretende trabajar, lo que sostiene la validez de constructo del instrumento modificado.

Diseño de rúbricas para la evaluación de dimensiones de PC en actividades diarias del aula

El uso de rúbricas para evaluar el desarrollo de movimientos del pensamiento en educación primaria ha resultado eficaz anteriormente (García et al., 2017). En este caso, se evalúan la progresión en la aplicación de habilidades de PC relacionadas con las dos dimensiones señaladas (comprobación de hipótesis y análisis argumental). En esta línea, se diseña una rúbrica de tres niveles en los que se detallan los grados de progresión en las dimensiones de PC desde una perspectiva general (Tabla 6). Para concretar más esta rúbrica y adaptarla a cualquier ejercicio, con el fin de evaluar cuestiones sobre análisis argumental y comprobación de hipótesis, es necesario concretar el número de razones o contraargumentos que el alumnado puede elaborar o localizar por nivel y ejercicio; y de igual modo, detallar la información a evaluar, para asegurar su validez como argumento para poder aceptar o no una hipótesis.

Tabla 6. Rúbrica de evaluación de PC

	Máximo nivel de logro	Nivel de logro medio	Nivel del logro bajo
Estándar 1: conclusión	Expresa la conclusión completa y fundamentada en evidencias válidas	Expresa la conclusión incompleta o mal expresada pero fundamentada en evidencias válidas	No expresa la conclusión de forma coherente con las evidencias
Estándar 2: razón	Expresa varias razones fundamentadas en evidencias válidas	Expresa una razón fundamentada en evidencias válidas	No expresa razones fundamentadas en evidencias válidas
Estándar 3: contra argumento	Expresa varios contraargumentos fundamentados en evidencias válidas	Expresa un contraargumento fundamentado en evidencias válidas	No expresa contraargumentos fundamentados en evidencias válidas
Estándar 4: evaluación de la validez de la información	Contrasta la información y en función de la evaluación emite conclusiones adecuadas y correctamente expresadas sobre la misma	Contrasta la información y, en función de la evaluación, emite conclusiones adecuadas sobre la misma	Contrasta la información, pero no evalúa correctamente su validez antes de emitir conclusiones
Estándar 5: modifica sus hipótesis iniciales en función de las nuevas evidencias	Analiza los argumentos y evidencias presentados y comprueba la validez de sus hipótesis en función de las nuevas evidencias. Modifica esas hipótesis si es necesario	Analiza los argumentos y evidencias presentados y comprueba la validez de sus hipótesis en función de las nuevas evidencias	No contrasta sus hipótesis con las evidencias y argumentos presentados y no modifica sus hipótesis

Fuente: elaboración propia.

Instrumentos didácticos. Secuencia de Enseñanza Aprendizaje (SEA)

La SEA aplicada parte de una versión anterior ya validada (Ortega-Quevedo et al., 2019), a la cual se le realizan pequeñas modificaciones para la mejora de la misma (Tabla 7). Esta SEA se diseña para el desarrollo de dos concepciones sobre Naturaleza de la Ciencia y la Tecnología (NdCyT) (influencia de la ciencia y la tecnología en la sociedad y construcción social de la tecnología) conforme a la contextualización de la NdCyT propuesta por Manassero-Mas y Vázquez-Alonso (2019), así como para el desarrollo de dos capacidades de PC presentadas anteriormente (análisis argumental y comprobación de hipótesis).

Tabla 7. Desarrollo de contenidos por actividades de la SEA

Sesión		1	2	3	
Actividad	Inicial	Descripción	Rutina de pensamiento: generar, clasificar, conectar y elaborar	Rutina de pensamiento: compara contrasta	Rutina de pensamiento: veo, pienso, me pregunto
		Contenidos curriculares	Concepto de energía, propiedades de la energía y manifestaciones de la energía	Impactos ambientales relacionados con la energía	Fracturación hidráulica y sus consecuencias
		Contenidos transversales del estudio	Relaciones Ciencia-Tecnología-Sociedad. Iniciación a conceptos de PC	Argumentación y comprobación de hipótesis (CH)	Todos los elementos transversales
	Desarrollo	Descripción	Se genera un diálogo con el alumnado (escuchan preguntan)		
		Contenidos curriculares	Concepto de energía, propiedades de la energía y manifestaciones de la energía	Impactos ambientales relacionados con la energía y fuentes de energías renovables y no renovables	Fracturación hidráulica y sus consecuencias
		Contenidos transversales del estudio	Influencia de la Ciencia y la Tecnología en la Sociedad (ICTS), argumentación y CH	ICTS, argumentación y CH	Todos los elementos transversales
	Consolidación	Descripción	Se realiza una ficha por parejas que contiene preguntas que inducen a la reflexión. Seguidamente se ponen en común dichas respuestas generando debates		
		Contenidos curriculares	Concepto de energía, propiedades de la energía y manifestaciones de la energía	Impactos ambientales relacionados con la energía y fuentes de energía renovables y no renovables	Fracturación hidráulica y sus consecuencias
		Contenidos transversales del estudio	ICTS, razonamiento verbal, argumentación	ICTS, razonamiento verbal, argumentación y probabilidad	Todos los elementos transversales

Fuente: elaboración propia.

En concreto, la SEA está compuesta por tres sesiones de una hora aproximada de duración (el tiempo es variable en función de la capacidad de adaptación del grupo a la metodología comunicativa y de sus conocimientos previos sobre los contenidos curriculares). Cada sesión comparte la misma estructura formada por: 1) implementación de una rutina de pensamiento (Ritchhart et al., 2014) como forma de refrescar los contenidos previos e inducir al alumnado al empleo de procesos de pensamiento superior (es en estos momentos cuando se trabaja de forma explícita las dimensiones de PC a desarrollar); 2) presentación dialógica de los contenidos curriculares. Se persigue contextualizar y reflexionar sobre los contenidos curriculares con el alumnado (se presentan de forma explícita y contextualizada los contenidos curriculares, así como cuestiones sobre naturaleza de la ciencia y la tecnología); 3) realización de fichas de consolidación y sus tertulias asociadas, las cuales permiten poner en práctica los conocimientos y procesos trabajados.

Implementación de los instrumentos elaborados

Contextualización de la experiencia

La experiencia didáctica se implementa en un centro público de educación primaria de un pueblo de la provincia de Segovia (España), concretamente en el sexto curso de esta etapa correspondiente con estudiantes de 11-12 años. El total de discentes de dicho curso es de 18, sin embargo, la muestra empleada es 17 debido a la falta de uno de los participantes en una de las fases de la implementación.

Este centro es un centro rural cercano al núcleo urbano de Segovia. Como dato de interés sobre el centro cabe destacar que: el nivel sociocultural de las familias inherentes al centro es medio alto; y existe participación del centro en proyectos relacionados con las nuevas tecnologías y el medio ambiente, lo que muestra cierto interés por la ciencia y la tecnología. Sin embargo, desde el centro no conocían el constructo naturaleza de la ciencia y la tecnología, ni su papel en el desarrollo de la alfabetización científica (Vázquez-Alonso et al., 2019)

Diseño de la experiencia

Se presenta un diseño longitudinal con una estructura *pretest-intervención-postest* implementada en el transcurso de tres meses. Para la implementación de la experiencia el docente del grupo de participantes recibió una formación de tres horas de duración en las que se presentaron cuestiones fundamentales sobre los instrumentos a implementar, los contenidos transversales a trabajar y los materiales y estructura de la SEA.

Por último, cabe señalar que para el análisis de datos de la versión modificada del HCTA se utiliza el software IBM SPSS Statistics 24. En concreto, se emplea la prueba no paramétrica de Wilcoxon tomando un margen de error del 0,5%.

Resultados y análisis

Los resultados se presentan en función de las dimensiones de estudio: comprobación de hipótesis y análisis argumental. En consecuencia, se dividen los resultados de las situaciones de la modificación del HCTA y se relacionan con los resultados extraídos de la aplicación de las rúbricas en las actividades de consolidación de las distintas sesiones.

Comprobación de hipótesis

En primer lugar, se analizan los resultados de las partes abiertas de las situaciones 1 y 3, correspondientes con la habilidad de comprobación de hipótesis. Estas situaciones miden la capacidad del alumnado para evaluar la validez de una afirmación en función de la información disponible sin guía. La Figura 2a muestra los resultados de la parte abierta de la

primera situación en sus dos fases, en esta se observa cómo la distribución de las respuestas de los participantes varía de una fase a otra obteniendo resultados superiores en la fase de *postest*. La comparación del incremento de estos resultados se calcula mediante la prueba de Wilcoxon que determina un *p*-valor de 0,004, con base en el cual se rechaza la igualdad de medianas y se estima una mejora significativa en los resultados de la segunda fase. Por otra parte, la Figura 2c refleja los resultados de la parte abierta de la tercera situación, en la cual se observan escasos cambios, aunque positivos, en la distribución de la fase *postest*. Tomando como referencia los resultados de la prueba de Wilcoxon (*p*-valor = 0,166) se acepta la igualdad en las medianas, lo cual indica que los cambios en la distribución no son significativos.

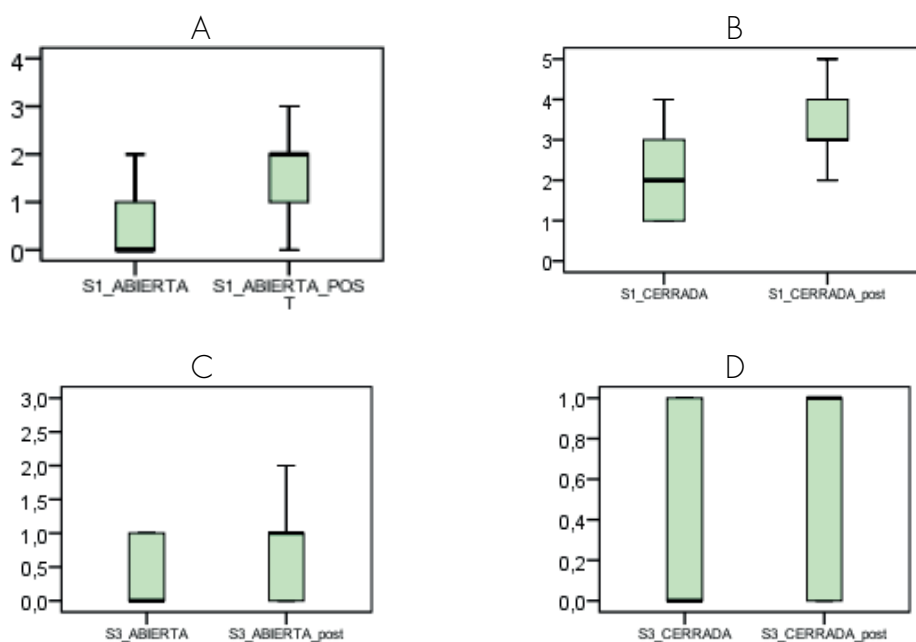


Figura 2. Resultados de las situaciones de comprobación de hipótesis en las dos fases de aplicación (2a: situación 1 abierta. 2b: situación 1 cerrada. 2c: situación 3 abierta. 2d: situación 3 cerrada)

Fuente: elaboración propia.

La parte cerrada de las situaciones 1 y 3 miden la capacidad de los estudiantes para seleccionar la evaluación que ellos estiman mejor sobre la validez de una afirmación y la información en la que se sustenta. En esta ocasión en la Figuras 2b y 2d se pueden apreciar cambios en la distribución que reflejan una mejora en la fase de *posttest*. Conforme los *p*-valor 0,025 y 0,025 obtenidos respectivamente de las pruebas Wilcoxon se determina que existe desigualdad en las medianas, de modo que la mejora percibida es significativa.

En segundo lugar, de los resultados obtenidos en las rúbricas se percibe una mejora en la cantidad de pares de estudiantes que obtienen un nivel medio en los estándares de “evaluación de la validez de la información” y “modifica sus hipótesis iniciales en función de las nuevas evidencias” en las sucesivas sesiones. Como se aprecia en las figuras 3 y 4 el nivel inicial es bajo, a pesar del trabajo explícito de las capacidades evaluadas en la primera sesión. Sin embargo, aumenta en cada una de las siguientes sesiones, permitiendo observar una mejora que alcanza el nivel medio sin llegar al alto en ningún caso.

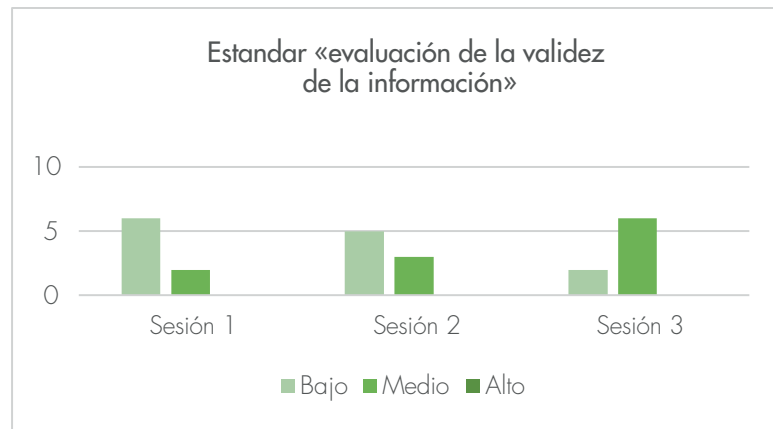


Figura 3. Resultados del estándar 4 de las rúbricas por parejas de estudiantes

Fuente: elaboración propia.

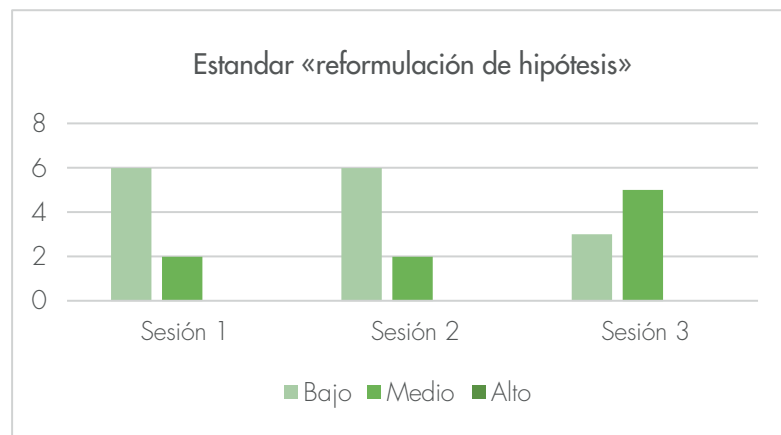


Figura 4. Resultados del estándar 5 de las rúbricas por parejas de estudiantes

Fuente: elaboración propia.

Análisis argumental

Primeramente, se analizan los resultados de las partes abiertas de las situaciones 2 y 4 correspondientes con la dimensión de análisis argumental. Estas situaciones miden la capacidad del alumnado para identificar y expresar los elementos de una argumentación (razón, contraargumento, conclusión y evidencias o hechos). Las Figuras 5a y 5c permiten observar cómo la distribución de las respuestas de

los participantes a la parte abierta de las situaciones mencionadas no sufre un incremento importante entre las distintas fases de aplicación, aunque sí se aprecia una leve mejora en las respuestas del *postest* de la situación 4. En estos casos la prueba Wilcoxon con *p*-valor de 0,083 y 0,070 respectivamente acepta la igualdad en las medianas, esclareciendo que no existe significatividad en la mejora de los resultados obtenidos en la fase de *postest*.

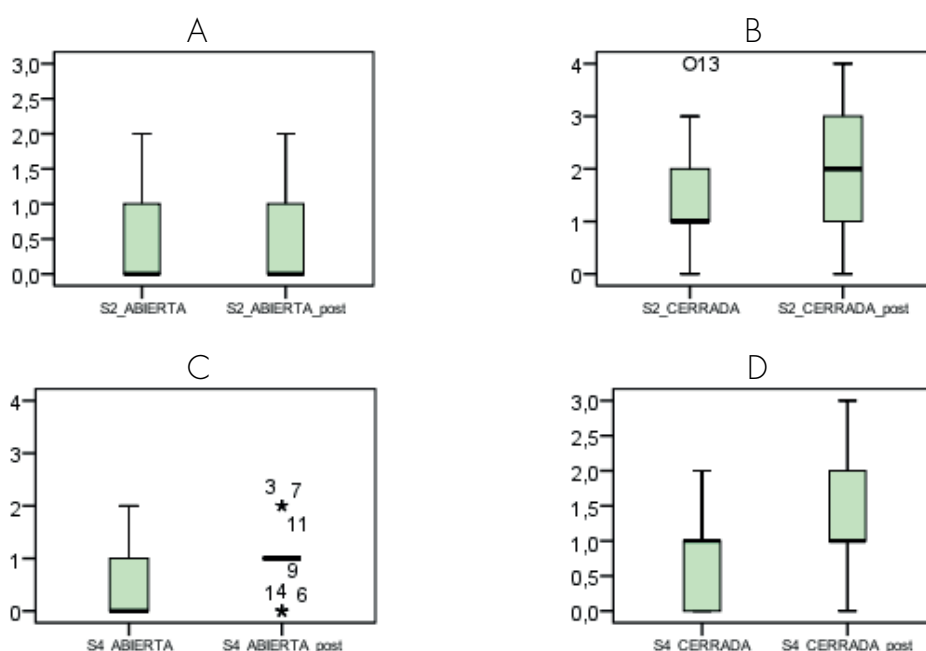


Figura 5. Resultados de las situaciones de análisis argumental en las dos fases de aplicación (3a: situación 2 abierta. 3b: situación 2 cerrada. 3c: situación 4 abierta. 3d: situación 4 cerrada)

Fuente: elaboración propia.

La parte cerrada de las situaciones 2 y 4 evalúan la capacidad de los estudiantes para reconocer los elementos de un argumento. En las figuras 5b y 5d se puede apreciar cambios en la distribución de los resultados de los estudiantes, siendo estos mejores en la fase de *postest*. Tomando como referencia los *p*-valores 0,027 y 0,005 obtenidos respectivamente de las pruebas Wilcoxon se determina que existe

desigualdad en las medianas, de modo que la mejora percibida es significativa.

Por último, en las rúbricas de evaluación se puede observar cómo al avanzar las sesiones de la SEA, la cantidad de pares de estudiantes que obtienen un nivel medio en los estándares de "Razón y contraargumento" y "conclusión" aumenta (figuras 6 y 7), aunque no se consigue alcanzar un nivel de logro alto.

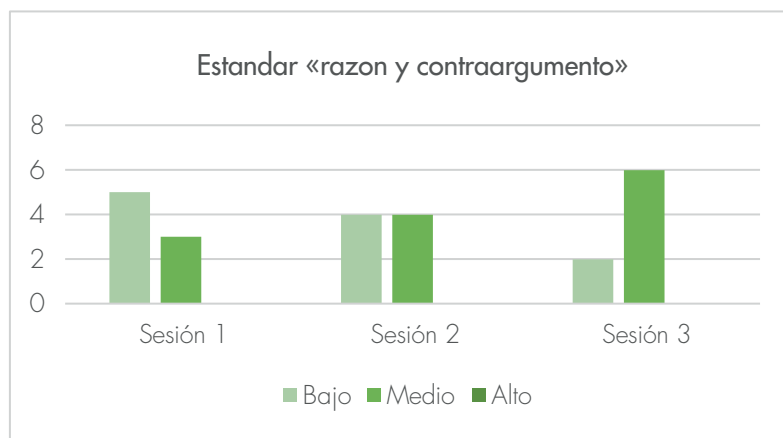


Figura 6. Resultados de los estándares 1 y 2 de las rúbricas por parejas de estudiantes

Fuente: elaboración propia.

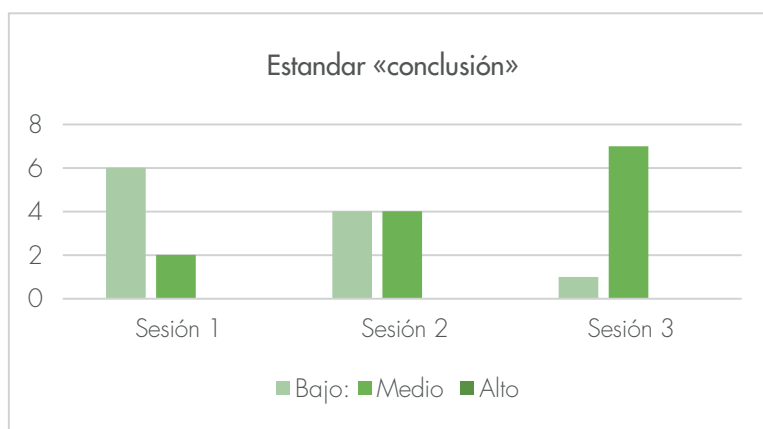


Figura 7. Resultados del estándar 3 de las rúbricas por parejas de estudiantes

Fuente: elaboración propia.

Consideraciones finales

Se concluye, tras la realización del análisis de componentes principales de la modificación del *test* HCTA, que esta mantiene una estructura factorial similar a la del *test* original y ajustada a las pretensiones establecidas en el diseño teórico; además de tener unos índices de fiabilidad aceptables. Sobre la base de estas pruebas se fundamenta la implementación del *test* en el contexto didáctico junto con la rúbrica diseñada.

Los resultados obtenidos de sendos instrumentos muestran una ligera evolución y marcan tendencias similares, lo cual permite concluir que ambas herramientas de evaluación miden las capacidades seleccionadas de forma complementaria. En consecuencia, se presentan dos instrumentos de evaluación de dos habilida-

des de PC, que dan respuesta a parte de la necesidad detectada de elaborar instrumentos que permitan evaluar habilidades de PC en estudiantes de estudios primarios (de 11 a 14 años); dichos instrumentos se ponen a disposición de la comunidad investigadora y docente. En esta misma línea, cabe destacar que existen tres dimensiones de PC, contempladas por Halpern que no se han tenido en cuenta en los instrumentos elaborados, no obstante, se continuará en esta línea de trabajo para crear nuevas adaptaciones.

Finalmente, se destaca la mejora de los discentes en los resultados obtenidos mediante los instrumentos de evaluación. El diseño de la implementación y la aplicación de instrumentos de evaluación, antes, durante y después de una intervención, donde se trabaja explícitamente el desarrollo del PC, permiten observar la evolución de los estudiantes en cada momento del proceso de desarrollo. En concreto, la evaluación final permite observar una mejora significativa por parte de los estudiantes en cinco de los ocho ítems a evaluar, lo cual con una intervención de tres sesiones es altamente satisfactorio.

Agradecimientos

La autoría agradece al proyecto EDU2015-64642-R (MINECO/FEDER) con financiación del Ministerio de Economía y Competitividad de España y el Fondo Europeo de Desarrollo Regional su colaboración.

Referencias

- Davies, A., Fidler, D. y Gorbis, M. (2011). *Future Work Skills 2020*. University of Phoenix Research Institute.
- Ennis, R. H. (1996). *Critical Thinking*. Prentice-Hall.
- Facione, P. (1989). Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction. *Research Findings and Recommendations*, 315.

García, N., Cañas, M. y Pinedo, R. (2017). Métodos de evaluación de rutinas de pensamiento: Aplicaciones en diferentes etapas educativas. En J. C. Núñez (Comps.). *Temas actuales de investigación en las áreas de la Salud y la Educación* (pp. 237-243). Scinfoper.

Halpern, D. F. (1998). Teaching critical thinking for transfer across domains. *American Psychologist*, 53(4), 449-455. <https://doi.org/10.1037/0003-066X.53.4.449>

Halpern, D. F. (2014). *Thought and knowledge. An Introduction to Critical Thinking*. Psychology Press. <https://doi.org/10.4324/9781315885278>

Halpern, D. F. (2016). *Manual Halpern Critical Thinking Assessment*. Test Label hcta. Schuhfried.

Lipman, M. (1991). *Thinking in Education*. Cambridge University Press.

Manassero-Mas, M. A. y Vázquez-Alonso, A. (2019). Conceptualización y taxonomía para estructurar los conocimientos acerca de la ciencia. *Revista Eureka sobre enseñanza y divulgación de las ciencias*, 16(3), 3104-17. https://doi.org/10.25267/Rev_Eureka_ensen_divulg_cienc.2019.v16.i3.3104

Nieto, A. M., Saiz, C. y Orgaz, B. (2009). Análisis de las propiedades psicométricas de la versión española del HCTAES-Test de Halpern para la evaluación del pensamiento crítico mediante situaciones cotidianas. *Revista Electrónica de Metodología Aplicada*, 14(1), 1-15. <https://doi.org/10.17811/rem.a.14.1.2009.1-15>

Ortega-Quevedo, V. y Gil, C. (2019). La naturaleza de la ciencia y la tecnología. Una experiencia para desarrollar el pensamiento crítico. *Revista Científica*, 35(2), 167-182. <https://doi.org/10.14483/23448350.14095>

Paul, R. (2005). The State of Critical Thinking Today. *New Directions for Community Colleges, Summer 2005*, 27-38. <https://doi.org/10.1002/cc.193>

- Paul, R. y Elder, L. (2002). *Critical Thinking: Tools for Taking Charge of your Professional and Personal Life*. Pearson Education.
- Reino de España. (2013, 9 de diciembre). *Ley Orgánica 8/2013, de 9 de diciembre, para la mejora de la calidad educativa*. Boletín Oficial del Estado, 295.
- Ritchhart, R., Church, M. y Morrison, K. (2014). *Hacer visible el pensamiento. Cómo promover el compromiso, la comprensión y la autonomía de los estudiantes*. Paidós.
- Saiz, C. y Rivas, S. F. (2008). Evaluación en pensamiento crítico: una propuesta para diferenciar formas de pensar. *Ergo, Nueva Época*, (22-23), 25-66.
- Vázquez-Alonso, A. 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. (2019). La educación de ciencias en contexto: Aportaciones a la formación del profesorado. *Tecné, Episteme y Didaxis: TED*, 46, 15-37.
- Vieira, R. M. y Tenreiro-Vieira, C. (2014). Investigaçao sobre o pensamento crítico na educaçao: contributos para a didáctica das ciências. En R. M. Vieira (Coord), *Pensamento Crítico na Educaçao: Perspectivas atuais no paranorama internacional* (pp. 29-40). Universidad de Aveiro.

Para citar este artículo:

- Ortega-Quevedo, V., Gil, C., Vallés, C., y López-Luengo, M. (2020). Diseño y validación de instrumentos de evaluación de pensamiento crítico en educación primaria. *Tecné, Episteme y Didaxis: TED*, (48), 91-110. <https://doi.org/10.17227/ted.num48-12383>

Anexo I

En el presente Anexo se recogen cuatro tablas (tablas 8, 9, 10 y 11), donde se especifica la métrica de cada una de las cuatro situaciones planteadas en el instrumento modificado a partir del HCTA.

Tabla 8. Enunciado y métrica de la primera situación de la modificación del HCTA

Situación 1		Categoría: comprobación de hipótesis	
<p>La dirección de un colegio, con muchos alumnos, ha observado que un elevado número de estudiantes comen grandes cantidades de bollería industrial durante el almuerzo. Para hacer que se reduzca el consumo de bollería industrial y conseguir aumentar el consumo de alimentos saludables, la dirección diseña un programa llamado "Almuerzo Saludable". La secretaria del centro puso un gran cartel con el nombre del programa y se envió al alumnado un folleto informativo con diferentes almuerzos saludables. Al cabo de un año se comprobó, de la misma forma, que el consumo medio de bollería industrial por los estudiantes que solían comer este tipo de dulces en el centro se redujo un 2%, en comparación con los datos del año anterior. El director del programa "Almuerzo Saludable" anunció que: "Este programa ha tenido un gran éxito como puede verse por la reducción del consumo de bollería industrial por parte de los estudiantes"</p>			
Parte Abierta, primera parte Puntuación máxima 2	2 puntos	1 punto	0 puntos
	Explica el descenso del consumo de bollería industrial	Menciona el descenso del consumo de bollería industrial	Respuestas irrelevantes
Parte Abierta, segunda parte Puntuación máxima 2	2 puntos	1 punto	0 puntos
	Se incluyen cuestiones relacionadas con el escaso descenso o la poca fiabilidad de las evidencias presentadas	Se hace referencia al escaso descenso	Respuestas irrelevantes
Parte cerrada Puntuación máxima 5		1 punto si la respuesta a la primera opción es "apoya"	1 punto si la respuesta a las opciones 2, 3, 4 y 5 es "debilita"

Fuente: elaboración propia.

Tabla 9. Enunciado y métrica de la segunda situación de la modificación del HCTA

Situación 3		Categoría: comprobación de hipótesis	
<p>Un informe reciente aparecido en una revista para padres y profesores muestra que los niños que comen chicle compulsivamente suelen obtener peores calificaciones en clase. A medida que aumenta el número de chicles por día, disminuyen sus notas. Una sugerencia que hace el informe es que podríamos mejorar el rendimiento escolar evitando que los niños coman chicles</p>			
Parte Abierta, primera parte Puntuación máxima 1	1 punto		0 puntos
	Si		No
Parte Abierta, segunda parte Puntuación máxima 2	2 puntos	1 punto	0 puntos
	Se expresan ideas que sostienen una relación sin causa-efecto	Se menciona que personas con malas notas no siempre comen chicle	Respuestas irrelevantes
Parte cerrada puntuación máxima 1		1 punto si seleccionan la tercera opción	

Fuente: elaboración propia.

Tabla 10. Enunciado y métrica de la tercera situación de la modificación del HCTA

Situación 2		Categoría: análisis argumental	
Hay muchas oportunidades para los especialistas en informática. La verdad es que deberías especializarte en esta ciencia. El trabajo es interesante, hay muchas posibilidades de empleo y los sueldos son buenos. Por supuesto, no es una buena especialidad si le temes a las matemáticas o te gusta trabajar al aire libre			
Parte Abierta, primera parte Puntuación máxima 1 Extrae una conclusión del texto	1 punto	0 puntos	
	Respuestas irrelevantes		
Parte Abierta, segunda parte Puntuación máxima 3 1 punto por cada una de las razones expuestas en el texto	1 punto	0 puntos	
	Respuestas irrelevantes		
Parte cerrada Puntuación máxima 4	1 punto si la respuesta a la primera opción es conclusión	1 punto si la respuesta a las opciones 2 y 3 es razón	1 punto si la respuesta a la tercera opción es contraargumento

Fuente: elaboración propia.

Tabla 11. Enunciado y métrica de la cuarta situación de la modificación del HCTA

Situación 4		Categoría: análisis argumental	
El alcalde ha propuesto que todos los edificios del centro de la ciudad se pinten con una pintura que permita limpiar fácilmente los grafitis			
Parte Abierta, primera parte Puntuación máxima 2	2 puntos	1 punto	0 puntos
	Si da una opinión coherente y bien expresada	Si da una opinión coherente	Respuestas irrelevantes
Parte Abierta, segunda parte Puntuación máxima 2	2 puntos	1 punto	0 puntos
	Un punto por conclusión completa y otro por razón completa	Un punto por conclusión parcial y otro por razón parcial	Respuestas irrelevantes
Parte cerrada Puntuación máxima 3	1 punto si la respuesta a la primera opción opinión	1 punto si la respuesta a la segunda opción hecho	1 punto si la respuesta a la tercera opción es argumento razonado

Fuente: elaboración propia.