

Pautas para mejorar la evaluación del pensamiento crítico en ciencias de la actividad física y del deporte en estudiantes universitarios

Guidelines for improving critical thinking assessment in physical activity and sports sciences among university students

Mireia Vendrell-Moranco

Universidad Internacional de La Rioja (España)

Resumen. El Pensamiento Crítico (PC) se considera una competencia fundamental en la educación del siglo XXI debido a sus beneficios en los ámbitos académicos, personales, profesionales y sociales. Sin embargo, la evaluación del PC en estudiantes universitarios enfrenta desafíos en la conceptualización, diseño, criterios de enseñanza y evaluación, así como en el análisis de datos. Este artículo presenta pautas basadas en investigación para facilitar la promoción y evaluación efectiva del PC en el campo de las Ciencias de la Actividad Física y del Deporte. Las pautas abarcan cuatro fases: definición del constructo, elección o diseño de herramientas de evaluación, diseño de investigación y recopilación de datos, e interpretación de resultados. Se enfatiza la importancia de establecer objetivos de evaluación claros, seleccionar o crear herramientas de evaluación adecuadas, aplicar la evaluación de manera sistemática, brindar retroalimentación y monitorear el proceso. Además, se destaca la relevancia de integrar la evaluación del PC como parte esencial del proceso de enseñanza y aprendizaje. La implementación de estas pautas promoverá una evaluación más eficiente y efectiva del PC, fomentando el desarrollo de competencias críticas en los estudiantes universitarios y fortaleciendo la cultura físico-deportiva en general. Mediante la adopción de estas recomendaciones, se espera elevar la calidad educativa y potenciar el desarrollo integral de los estudiantes, preparándolos para enfrentar los desafíos actuales y futuros con éxito.

Palabras clave: Pensamiento crítico, evaluación educativa, evaluación de resultados universitarios, pautas, estudiantes universitarios.

Abstract. Critical Thinking (CT) is recognised as a pivotal competency in 21st-century education, offering benefits across academic, personal, professional, and social spheres. However, evaluating CT in university students poses challenges in terms of conceptualization, design, teaching and evaluation criteria, as well as data analysis. This article introduces evidence-based guidelines to facilitate the promotion and effective assessment of CT in the realm of Physical Activity and Sports Sciences. The guidelines encompass four key phases: construct definition, selection or design of assessment tools, application and data collection, and interpretation of results. Emphasis is placed on the importance of setting clear assessment objectives, selecting or creating appropriate assessment tools, applying systematic evaluation, providing constructive feedback, and monitoring the entire assessment process. Furthermore, the significance of integrating CT assessment as an integral part of the teaching and learning process is underscored. Implementing these guidelines is expected to foster a more efficient and effective CT assessment, thus promoting the development of critical competencies among university students and reinforcing the overall physical activity and sports culture. By adopting these recommendations, it is anticipated that the quality of education will be elevated, and students will be better equipped to confront present and future challenges with confidence and success.

Keywords: Critical thinking, Educational Assessment, College Outcomes Assessment, Guidelines, University students.

Fecha recepción: 03-08-23. Fecha de aceptación: 19-12-23

Mireia Vendrell-Moranco

mireia.vendrell@unir.net

Introducción

El concepto del Pensamiento Crítico (PC) es una noción multifacética ampliamente debatida y definida de diversas maneras en la literatura académica (Ericson, 2022), influenciada por varios marcos teóricos y adaptada a diferentes contextos (Moore, 2013). A pesar de esta diversidad, las definiciones del PC comparten elementos comunes. En esencia, el PC se comprende como un proceso cognitivo complejo y holístico cuyo propósito es explorar afirmaciones o problemas con el fin de llegar a una conclusión válida o seleccionar la alternativa con mayor probabilidad de éxito (Dwyer, 2017; Ennis, 1985; Halpern, 2014; McBride, 1991).

Para lograrlo, se requiere la aplicación organizada y disciplinada de habilidades cognitivas, como el análisis de evidencia (Cottrell, 2017), la detección de sesgos (Sokolnick, 2019; Tittle, 2011) y la reflexión autocrítica (Dwyer, 2017), siguiendo estándares racionales (Ennis, 2018). La dimensión disposicional del PC implica actitudes que contribuyen a superar sesgos cognitivos y a fomentar la ética y la

equidad en el razonamiento (Thomas & Lok, 2015), como el escepticismo (Danczak et al., 2020), la honestidad intelectual (Porter et al., 2021) y la apertura a nuevas ideas (Ennis, 2018). En última instancia, el desarrollo del PC exige la adquisición de habilidades específicas, conocimiento sustantivo en la materia de interés y la disposición personal para llevar a cabo este proceso de manera efectiva. Estas tres dimensiones son cruciales para fomentar y fortalecer el PC en entornos académicos y situaciones cotidianas (Paul & Elder, 2019).

Es importante destacar que "pensar críticamente" (habilidades + disposiciones + conocimiento) no equivale a "ser crítico" (PC + acción), es decir, a la criticidad (Davies, 2015; Davies & Barnett, 2015), aunque constituye una condición necesaria. En este contexto, es relevante señalar que la discusión propuesta sobre el PC busca precisamente la promoción de acciones basadas en este pensamiento. En otras palabras, pensar críticamente para actuar en consonancia con ese pensamiento.

En el contexto de la educación del siglo XXI, el PC se considera una competencia clave debido a sus beneficios en

el ámbito académico, personal, profesional y social (Prat-Sala & van Duuren, 2022; Thornhill-Miller et al., 2023). Investigaciones recientes han demostrado que el desarrollo del PC está asociado con un mayor éxito académico (Akpur, 2020; D'Alessio et al., 2019), una mayor resistencia a la desinformación (Joshi et al., 2022), la toma de decisiones más acertadas en situaciones personales y profesionales (Butler et al., 2017), mejores oportunidades laborales (Indrašienė et al., 2021) y la formación de ciudadanos éticos en una sociedad democrática (Aktoprak & Hursen, 2022).

Dada la importancia del PC para los individuos y sus comunidades, se han realizado numerosos esfuerzos para diseñar programas educativos centrados en su adquisición y transferencia (Butler, 2012; Gillespie & Culpan, 2000). En este proceso, la evaluación del PC juega un papel crucial para su promoción. La cuantificación del PC es necesaria para identificar las deficiencias y las áreas de desarrollo de los estudiantes (Everett et al., 2018), realizar un seguimiento de su evolución (Mueller et al., 2020), determinar la eficacia de las estrategias pedagógicas (Betancourth Zambrano et al., 2017; Braun et al., 2020) y orientar el diseño de planes de estudio que fomenten el progreso en esta tipología de pensamiento (Jacob et al., 2017).

A nivel internacional, existe un creciente interés por parte de profesores e instituciones de educación superior en evaluar el PC de los estudiantes universitarios (Liu et al., 2018). Sin embargo, se han señalado repetidamente las limitaciones de estos estudios, como la falta de claridad y rigor en la conceptualización del constructo, los principios de diseño, los criterios que sustentan la enseñanza y evaluación del PC, los informes estadísticos, la presentación de los datos y la discusión de los resultados (Cruz et al., 2017; Morais et al., 2019; Puig et al., 2019).

A nivel institucional, aunque la mayoría de los centros educativos superiores consideran el PC como un criterio de evaluación clave en sus programas de formación, rara vez se especifica cómo se materializa esta evaluación (Danczak et al., 2020; O'Leary et al., 2020; Reynders et al., 2020), especialmente en el campo de las Ciencias de la Actividad Física y del Deporte (AFyD). Una de las principales razones de esta falta de estudios y transparencia radica en la dificultad de definir y evaluar adecuadamente el PC en el contexto universitario. Entre las causas más citadas se encuentran la falta de consenso en cuanto a una definición operativa del constructo (Dwyer et al., 2014), la insuficiencia de instrumentos de evaluación diseñados específicamente para estudiantes universitarios (O'Leary et al., 2020), la dificultad para seleccionar o adaptar herramientas de evaluación sensibles y la necesidad de que los profesores comprendan y apliquen estrategias de evaluación científicamente fundamentadas (Hammonds & Schwarze, 2019).

Como solución parcial, algunos investigadores proponen el uso de herramientas estandarizadas disponibles comercialmente (Jacob et al., 2017). Sin embargo, se ha señalado que ninguna de estas herramientas está exenta de limitaciones (Liu et al., 2014), y su elección y aplicación pueden resultar desafiantes.

Con el objetivo de facilitar la promoción y evaluación efectiva del PC en el contexto universitario, este artículo presenta una aproximación detallada al proceso de evaluación del PC, particularmente en el campo de la AFyD. En este sentido, se ofrecen pautas respaldadas por estudios de investigación cuidadosamente seleccionados a través de una revisión sistemática de la literatura. El objetivo central es lograr una evaluación precisa y rigurosa del PC en estudiantes universitarios.

Es importante destacar que, aunque este artículo se centra en metodologías cuantitativas debido a restricciones de espacio, se reconoce la importancia de las metodologías cualitativas. Estas últimas, como la evaluación del PC a través de entrevistas, proporcionan una comprensión profunda de las habilidades de PC de los estudiantes, lo cual es valioso en un campo donde la interacción y la toma de decisiones en tiempo real son comunes.

Estas directrices abarcan diferentes aspectos del proceso de evaluación, desde la definición clara del constructo hasta la interpretación adecuada de los resultados obtenidos. Su propósito fundamental es mejorar significativamente la calidad y rigurosidad de la evaluación del PC, particularmente en el ámbito de la AFyD.

La relevancia de estas pautas radica en la naturaleza multidisciplinaria y compleja de la AFyD, donde el PC es esencial. La toma de decisiones en este campo abarca desde la planificación de entrenamientos deportivos hasta la gestión de lesiones y la formulación de estrategias deportivas (Cardozo et al., 2019; Pérez-Tejero et al., 2023). En este sentido, la evaluación del PC no solo contribuye a una formación académica sólida, sino que también es necesaria para el desarrollo de profesionales capaces de abordar los desafíos específicos y variados que enfrenta el sector deportivo.

Aproximación al proceso de evaluación del PC

Antes de abordar las directrices, es importante destacar que estas se han desarrollado a partir de un análisis crítico de artículos que se centran en la medición del PC en estudiantes universitarios. Con base en esta revisión, se presentan algunas directrices generales. Cabe mencionar que estas pautas no pretenden ser una lista exhaustiva de indicadores, sino más bien ofrecer sugerencias prácticas para llevar a cabo una evaluación del PC desde una perspectiva científica.

Con el fin de facilitar su comprensión, esta guía práctica se divide en cuatro fases: (1) Definición del constructo; (2) Selección o diseño de la herramienta de evaluación; (3) Diseño de investigación; y (4) Recopilación e interpretación de datos.

Fase 1. Definición del Constructo

El primer paso crucial en el proceso de evaluación del PC es establecer una definición clara y robusta del constructo, proporcionando una base teórica sólida para su estudio. Esto implica la identificación de habilidades, disposiciones y contenidos específicos que los estudiantes deben demostrar dentro del contexto de la AFyD.

En el ámbito de la AFyD, las habilidades del PC abarcan

la capacidad de analizar y evaluar argumentos en diversos contextos (Hatcher & Possin, 2020), como debates sobre técnicas de entrenamiento o interpretación de datos de rendimiento deportivo. Además, se espera que los estudiantes desarrollen habilidades para resolver problemas complejos relacionados con la AFyD deporte (Aktoprak & Hursen, 2022; Gillespie & Culpan, 2000; Southworth, 2022), como mejorar el rendimiento de un equipo o abordar desafíos en la gestión de lesiones deportivas.

Por otro lado, las disposiciones del PC son fundamentales para el desarrollo de un PC efectivo. Los estudiantes deben demostrar curiosidad intelectual (Liu et al., 2018), lo que implica un genuino interés por explorar nuevas ideas y perspectivas en el campo de la AFyD. Asimismo, deben desarrollar autonomía en el pensamiento, lo que significa pensar de manera independiente y cuestionar ideas preconcebidas (Alejo, 2017). Además, se espera que muestren flexibilidad cognitiva, es decir, estar dispuestos a revisar sus creencias y opiniones cuando se encuentre evidencia que las cuestione (Gul & Akcay, 2020). En cuanto a los contenidos del PC, los estudiantes deben adquirir conocimientos sobre

aspectos éticos relacionados con el deporte, tales como el dopaje, la justicia deportiva y la integridad en la competición (Hernández, 2004). Además, es crucial que comprendan la metodología de investigación en ciencias del deporte para poder evaluar críticamente estudios científicos y aplicar la evidencia en su práctica profesional (Negret, 2016). Asimismo, deben familiarizarse con diferentes estrategias de planificación deportiva y aprender a evaluar las implicaciones de cada enfoque, considerando factores como la seguridad de los atletas y los objetivos del equipo o individuo (Poblete-Valderrama et al., 2023).

Una vez que se ha establecido una definición clara del PC en el contexto de la AFyD, es posible diseñar instrumentos de evaluación que sean sensibles a las habilidades, disposiciones y contenidos mencionados anteriormente. Estos instrumentos deben medir de manera efectiva el desarrollo del PC en los estudiantes y reflejar los cambios esperados en el PC como resultado de la formación.

Con el objetivo de facilitar la evaluación del PC, es recomendable establecer objetivos y resultados de aprendizaje claros y estándares. Por ejemplo (Tabla 1):

Tabla 1.
Ejemplo de objetivo, resultado de aprendizaje y ejemplo

Objetivo	Resultado de aprendizaje	Ejemplo
Analizar y evaluar argumentos en distintos contextos	El estudiantado será capaz de identificar y evaluar argumentos presentes en debates sobre técnicas de entrenamiento o interpretación de datos de rendimiento deportivo	El estudiantado podría, por ejemplo, analizar y evaluar críticamente las afirmaciones hechas en un debate sobre la efectividad de dos métodos de entrenamiento de resistencia, considerando la evidencia científica disponible, el contexto específico del deporte y las implicaciones para el rendimiento de los atletas. Esto le permitirá tomar decisiones informadas en la planificación de programas de entrenamiento
Demostrar curiosidad intelectual en el campo de la AFyD.	El estudiantado mostrará un genuino interés por explorar nuevas ideas y perspectivas en su área de estudio	El estudiantado podría demostrar su curiosidad intelectual al buscar activamente investigaciones recientes o enfoques innovadores en la ciencia del deporte. Por ejemplo, podría explorar estudios sobre la aplicación de la tecnología de monitoreo del rendimiento en deportes específicos o investigar nuevas teorías sobre la motivación de los atletas, lo que reflejaría su compromiso en mantenerse actualizado/a con los avances en el campo
Familiarizarse con diferentes estrategias de planificación deportiva.	El estudiantado será capaz de evaluar las implicaciones de diversas estrategias de planificación, considerando la seguridad de los atletas y los objetivos del equipo o individuo	El estudiantado podría aplicar este conocimiento en un contexto práctico al evaluar y seleccionar estrategias de planificación para un equipo de deportes de equipo, como el voleibol. Esto implicaría considerar la edad y estado físico de los/as jugadores/as, los objetivos del equipo (por ejemplo, ganar el campeonato o desarrollar habilidades individuales), y asegurarse de que las estrategias seleccionadas minimicen el riesgo de lesiones y promuevan el rendimiento sostenible

Igualmente, es relevante señalar que si la medición del PC tiene como objetivo evaluar la eficacia de un modelo de enseñanza-aprendizaje, es fundamental que el instrumento de evaluación sea sensible a los cambios específicos en el PC que se espera lograr con la formación. En consecuencia, los resultados de aprendizaje deben reflejar dichos cambios para obtener una evaluación efectiva del desarrollo del PC en los estudiantes.

Fase 2. Elección o Diseño de la Herramienta

Decidir cuál es la herramienta más adecuada para medir el PC implica considerar la alineación entre las intenciones y posibilidades de los investigadores con las características de los instrumentos. Como se ha señalado anteriormente, los instrumentos difieren en cuanto a las facetas del constructo que pretende medir y al formato de sus ítems. Igualmente, también difieren en varios aspectos más, como el contenido de fondo, el número de preguntas, el tiempo de ejecución, la aplicación (en papel u online) o el precio, entre otros. El resultado de esta decisión es seleccionar y, si

es necesario, adaptar un instrumento disponible o, en el caso de que no se logre la alineación mencionada, diseñarlo.

La elección de la herramienta adecuada para medir el PC en el ámbito de la AFyD se vuelve crucial debido a la naturaleza específica y multidisciplinaria de este campo. Los desafíos que los profesionales y estudiantes de Ciencias de la AFyD enfrentan requieren una evaluación precisa de las habilidades y disposiciones de PC, ya que estas capacidades son fundamentales para la toma de decisiones en situaciones prácticas, como la planificación de entrenamientos, la gestión de lesiones deportivas y la formulación de estrategias deportivas.

Si la elección es utilizar un instrumento importado de otro contexto, antes de su aplicación el personal investigador debe asegurarse de que los estudiantes a examinar tienen la misma probabilidad de comprender cada ítem que la muestra para la que se validó el instrumento. Para ello, deben atenderse dos cuestiones. En primer lugar, si el instrumento está en una lengua extranjera, los investigadores deben considerar deliberadamente la traducción de los

instrumentos a la lengua local. De lo contrario, existe el riesgo potencial de confundir la capacidad de la lengua extranjera con la competencia en PC (Schendel & Tolmie, 2017).

En segundo lugar, incluso cuando la herramienta no se traduce, es necesario examinar cada ítem en busca de matices culturales. Si se detectan, las preguntas deben adaptarse para que sean pertinentes desde el punto de vista cultural. De lo contrario, las disparidades culturales entre los estudiantes podrían dar lugar a una variación que carece de relevancia en el marco del constructo que se pretende medir. Esto implica que la medición podría resultar inexacta o incluso engañosa debido a la falta de consideración de las diferencias culturales subyacentes.

Como ejemplo, consideremos un instrumento previamente validado en estudiantes estadounidenses que incluye un ítem que interroga acerca de la importancia de los deportes de equipo en la formación de los valores tanto individuales como colectivos de la juventud (por ejemplo: "¿Cuál es tu percepción acerca de la relevancia de los deportes de equipo en la forja de valores tanto individuales como colectivos entre los jóvenes?"). Este ítem podría generar respuestas variadas debido a las diferencias culturales en la percepción de la importancia de los deportes de equipo en el ámbito educativo y en la sociedad en general. De este modo, se pone de manifiesto cómo las diferencias culturales ejercen influencia sobre la percepción de la relación entre los deportes de equipo y los valores en la juventud, subrayando la trascendencia de tener en cuenta el contexto cultural al emplear instrumentos de evaluación en diversas regiones geográficas.

En Schendel y Tolmie (2017) se muestra una metodología bien fundamentada para adaptar un instrumento de PC basado en tareas de rendimiento. Esta metodología es particularmente útil para equipos de investigación pequeños que operan en un entorno de bajos recursos y que necesitan adaptar un instrumento en un período de tiempo relativamente corto. Se pueden encontrar más ejemplos de traducción, adaptación y validación de un instrumento existente en Liu et al. (2016) y Rodrigues Franco et al. (2018).

Por otro lado, si ninguno de los instrumentos disponibles se adapta a la realidad de los investigadores, una segunda opción es desarrollar una herramienta propia. El objetivo de los investigadores, en este caso, es redactar ítems lo suficientemente sensibles como para desencadenar el proceso de PC de los encuestados y que esté apoyado con evidencia empírica (Brückner & Pellegrino, 2016). Aunque este proceso es complejo, algunas prácticas de los artículos revisados allanan el camino para los nuevos diseñadores. Un buen ejemplo de cómo desarrollar un instrumento de PC se encuentra en Carter et al. (2016).

Como sugieren Brijmohan et al. (2018), el diseño de los instrumentos resulta significativamente mejor cuando existe colaboración entre especialistas en evaluación con formación en la redacción de ítems y expertos en el contenido del constructo en cuestión. En esta fase, los investigadores deben seleccionar el formato que mejor se adapte a

sus propósitos e identificar y minimizar sus limitaciones. A continuación, se presentan brevemente los formatos más utilizados: métodos indirectos, ítems de respuesta cerrada e ítems de respuesta abierta.

En primer lugar, los métodos de evaluación indirectos, es decir, los autoinformes sobre autopercepción y autoevaluaciones, no son considerados por varios autores como métodos idóneos para evaluar el PC (Hyytinen & Toom, 2019), particularmente debido a la dependencia de la puntuación al entendimiento, la precisión y la honestidad de los encuestados (Repo et al., 2017). Así, se detecta un discurso generalizado que considera que los autoinformes pueden no ser fiables ni válidos como medida del PC, ya que pueden no reflejar adecuadamente las habilidades y disposiciones reales del estudiante en cuestión.

En cuanto a los formatos de elección forzada (preguntas que requieren seleccionar una respuesta predefinida de un conjunto limitado de opciones), estos suelen considerarse la opción más rápida y económica para calificar. Además, pueden garantizar una mayor fiabilidad en tanto que no requieren que el evaluador interprete las respuestas de los evaluados. No obstante, en el ámbito del PC han sido duramente criticados por infrarrepresentar el PC. En concreto, según sus detractores, estos formatos no miden características significativas del PC, como las disposiciones o el razonamiento subyacente para elegir una determinada respuesta (Hyytinen & Toom, 2019; Rodrigues et al., 2018).

En respuesta a las limitaciones de los formatos de respuesta múltiple, varios evaluadores optan por captar las facetas del constructo mediante enfoques de evaluación menos restrictivos, es decir, a través de preguntas abiertas. Al respecto, aunque los instrumentos disponibles que requieren respuestas constructivas para medir el PC pueden adoptar formas y estilos diferentes, a menudo implican tareas escritas de respuesta construida y calificadas por una persona con la ayuda de una rúbrica. La principal ventaja de estas estrategias, ampliamente conocidas como "estrategias de evaluación auténticas" (Mueller et al., 2020), es que proporcionan un enfoque de mayor fidelidad para medir el PC. El motivo es que requieren que los examinados respondan de una manera más parecida a como lo harían al resolver problemas de la vida real (Braun et al., 2020; Kim et al., 2017). Así, a través de este formato parece que se puede obtener una representación del constructo superior en tanto que permiten obtener una visión más detallada de las dimensiones cognitivas y afectivas de los examinados (Re et al., 2019; Repo et al., 2017). Además, el diseño de estos ítems requiere menos tiempo y esfuerzo (Dominguez et al., 2015).

No obstante, al igual que el resto de los formatos, las tareas abiertas también han sido criticado en el área de la medición del PC. En este sentido, presentan importantes retos, especialmente relacionados con la varianza irrelevante al constructo. Así, las evaluaciones abiertas requieren de la interpretación de las respuestas de los examinados, por lo que no solo necesitan más tiempo y esfuerzo para ser corregidas, sino que también están más expuestas a sesgos por

parte del calificador (Re et al., 2019; Zlatkin-Troitschanskaia et al., 2019). En consecuencia, presentan más desafíos para asegurar la objetividad del proceso de medición (Ossa-Cornejo et al., 2018). Además, en el caso de que el PC sea evaluado únicamente a través de tareas

escritas, se corre el riesgo de que no se capture la totalidad del constructo (Hathcoat et al., 2016).

En este sentido, es importante tener en cuenta que todos los formatos tienen limitaciones y es necesario detectarlas y tratar de superarlas (Tabla 2).

Tabla 2.

Comparación de métodos de evaluación del PC

Método de Evaluación	Ventajas	Limitaciones y Formas de Superarlas
Autoinformes y Autoevaluaciones	Rápido y económico para calificar Percepción del progreso del estudiante Mejora de la enseñanza y el aprendizaje	Dependencia de la honestidad y precisión de los estudiantes No refleja habilidades y disposiciones reales Combinar con otros métodos de evaluación
Formatos de Elección Forzada	Rápido y económico Mayor fiabilidad al eliminar interpretación	Subrepresentación del PC Requiere de mucho esfuerzo para que los estudiantes usen el PC
Respuestas Abiertas	Mayor fidelidad en la medición Capta dimensiones cognitivas y afectivas Similitud con problemas de la vida real Estimula resolución de problemas y formación de argumentos	Sesgos de interpretación y corrección Uso de rúbricas para reducir sesgos Identificar y minimizar factores de varianza irrelevante Combinar con otros métodos para una evaluación completa

Por ejemplo, la autopercepción de los estudiantes puede ser una herramienta útil para evaluar la calidad de la enseñanza y el aprendizaje. Proporciona información sobre cómo los estudiantes perciben su propio progreso y el impacto de la enseñanza en su proceso de aprendizaje. Esta información puede ser útil para mejorar la experiencia de aprendizaje de los estudiantes a través del diseño y la planificación de la enseñanza y para tomar decisiones sobre cómo mejorar la enseñanza y el aprendizaje. Sin embargo, es importante tener en cuenta que la autopercepción de los estudiantes no debe ser el único factor en la evaluación.

Luego, si la evaluación del PC a través de formatos de elección forzada puede ser un desafío debido a la subrepresentación del constructo, si se decide utilizar este tipo de ítem, es necesario abordar este desafío haciendo que se requiera del uso del PC para identificar la respuesta correcta.

En relación con las preguntas abiertas, su principal reto, relativo a la varianza irrelevante, puede ser mitigado aplicando estrategias sistemáticas destinadas a minimizar el efecto de los factores que pueden actuar como fuente de error sistemático. Para ello, primero deben identificarse estos factores. Por ejemplo, las habilidades comunicativas de los alumnos, la longitud de las respuestas, los errores gramaticales y el sesgo de confirmación y de actitudes. Después, se puede proporcionar a los calificadores una rúbrica multidimensional claramente definida y estructurada. Es decir, una rúbrica que integre descripciones claramente justificadas, específicas y distinguidas de cada nivel de rendimiento y para cada criterio. Además, con el objetivo de mejorar la evaluación, las descripciones pueden estar ancladas a ejemplos concretos (Zlatkin-Troitschanskaia et al., 2019).

Otra práctica para minimizar los sesgos de interpretación es proporcionar formación a los calificadores sobre los protocolos de aplicación adecuados. Esto puede incluir ayudar a los evaluadores a concienciarse sobre el potencial de sesgo (sesgo de expectativas, de simpatía, de familiaridad, de grupo, entre otros) y animarlos a reconocer y superar los que les son inherentes. En este sentido, varios autores se esfuerzan por obtener puntuaciones en el PC que dependan, en la medida de lo posible, del contenido factual. Ejemplos de ello se encuentran en Buur et al. (2012) y Zlatkin-

Troitschanskaia et al. (2019). En el primer estudio, los autores trabajan para construir rúbricas detalladas alineadas con las prácticas de desarrollo de los estudiantes. En el segundo, se utiliza un enfoque holístico para construir la evaluación y se complementa con un enfoque analítico para calificar las respuestas escritas del estudiantado.

Por su parte, en el artículo de González-de Paz et al. (2016) siguen un método de puntuación de alta precisión a través de una rúbrica. En su artículo, los autores argumentan que la media ponderada de las puntuaciones de cada evaluador no tiene en cuenta los elementos de los que puede depender la calificación final del estudiantado. Es decir, la subjetividad de los evaluadores, la dificultad de los ítems y la capacidad de los estudiantes. Por ello, examinan los factores mencionados que influyen en las puntuaciones del estudiantado con un modelo Rasch multifacético. Además, una vez que las puntuaciones alcanzan un nivel aceptable de acuerdo con las puntuaciones humanas, utilizan la puntuación automatizada como solución para reducir los costes de las medidas abiertas. Por último, las percepciones de los estudiantes pueden utilizarse para añadir valor a la evaluación del PC (Cargas et al., 2017).

Asimismo, como la mayoría de los artículos han justificado, el PC debe ser medido a través de situaciones o problemas auténticos de la vida real y relevantes para los participantes. Un ejemplo de esto se encuentra en Schendel y Tolmie (2017). Por un lado, el PC es un constructo cognitivamente exigente, por lo que si el estudiante está familiarizado con el contenido de las preguntas, el préstamo cognitivo de los participantes no se reducirá al interpretar las preguntas (Schendel & Tolmie, 2017). Además, se recomienda que la medición del PC se realice a través de contenido actual y socialmente polémico (Halpern, 2014; Santos et al., 2021; Velez & Power, 2020).

Actualmente, el discurso hegemónico es que, evaluar el PC mediante contenido actual y polémicamente relevante puede ser beneficioso ya que promueve el aprendizaje basado en la indagación (Archila et al., 2022) y estimula a los estudiantes a ejercitar habilidades de resolución de problemas, toma de decisiones, formación de perspectivas y argumentación. También puede ayudar a los estudiantes a

distinguir sus creencias de la evidencia (Pnevmatikos et al., 2019), formar una mente basada en la evidencia (Ku et al., 2017), entender la importancia de la evaluación de la evidencia en el avance de la ciencia y en sociedades democráticas (Archila et al., 2022), enfrentarse a controversias sociales, reconocer que no se puede sugerir fácilmente una solución sin conocer las circunstancias concretas del problema (Pnevmatikos et al., 2019), promover estrategias no violentas para la gestión de conflictos (Su et al., 2016), aumentar la apreciación de la diversidad cultural, concienciar de su rol como agentes sociales activos (Maravé-Vivas et al., 2022) y contribuir a ser ciudadanos más informados y activos. En el caso de que el PC se evalúe en disciplinas específicas, la argumentación sobre cuestiones socialmente controvertidas también dependerá de la consideración de los valores éticos (Archila et al., 2022).

Fase 3. Diseño de investigación

Para asegurar la validez científica, es esencial que la investigación se lleve a cabo con un diseño cuidadosamente planificado y que las puntuaciones de los estudiantes se interpreten en función de las limitaciones del diseño aplicado. Un ejemplo de un diseño comúnmente utilizado en la medición del PC es el pretest-postest con un grupo de control. Sin embargo, es importante señalar que este enfoque, al carecer de asignación aleatoria, no permite realizar inferencias causales.

En esta etapa, se debe integrar y especificar claramente la definición operativa del PC, el diseño de la investigación, el desarrollo de la evaluación y su implementación (Braun et al., 2020). Esto abarca aspectos como quiénes serán los participantes en la evaluación, cómo se recopilarán y analizarán los datos, y cómo se proporcionará retroalimentación a los estudiantes acerca de sus puntos fuertes y áreas de mejora. Además, se pueden considerar varios diseños de investigación para abordar de manera efectiva las preguntas de investigación. A continuación, se presentan ejemplos de algunos diseños comunes que podrían ser adecuados:

a. **Pretest-postest con grupo de control y asignación aleatoria:** En este diseño, se asigna aleatoriamente a los participantes a un grupo de control y un grupo experimental. Se realiza una evaluación inicial (pretest) antes de la intervención, seguida de una evaluación posterior (postest) después de la intervención. Este diseño permite inferencias causales más sólidas.

b. **Diseño cuasiexperimental:** En situaciones en las que la asignación aleatoria no es factible, este diseño involucra un grupo experimental y un grupo de control, pero la asignación no es aleatoria. Puede utilizarse cuando se necesita un mayor control sobre ciertas variables.

c. **Estudio longitudinal:** Este diseño implica la recopilación de datos de la misma muestra de participantes a lo largo de un período de tiempo más extenso. Permite observar cambios y tendencias en el PC con el tiempo.

Cada uno de estos diseños tiene sus ventajas y desventajas, y la elección depende de las preguntas de investigación específicas y las limitaciones del contexto.

La selección y descripción adecuada del diseño de la investigación es esencial para garantizar la validez y la utilidad de los resultados.

Fase 4. Recogida e Interpretación de los Datos

A la hora de recoger los datos es necesario proporcionar instrucciones claras y detalladas a los estudiantes para asegurar que entiendan lo que se espera de ellos y respondan de manera adecuada, garantizar la confidencialidad de los resultados de la evaluación para proteger la privacidad de los participantes y fomentar su confianza, y monitorear y evaluar el proceso de evaluación para asegurar su efectividad y detectar cualquier posible mejora.

Luego, independientemente del instrumento empleado, es decir, una herramienta validada en otro contexto o construida por el investigador, de elección múltiple o abierta, las puntuaciones deben someterse a un análisis psicométrico. Ejemplos al respecto se encuentran en Liu et al. (2018), O'Leary et al. (2020) y Sánchez et al. (2019). En este sentido, las afirmaciones de que el instrumento mide el PC y si esa medición puede aplicarse al contexto deben ser validadas empíricamente cada vez que se utilice el instrumento.

El nivel de fiabilidad requerido depende de las consecuencias derivadas del uso del puntaje. Por ejemplo, cuando se toman decisiones que implican consecuencias relevantes, la fiabilidad debe ser mayor que la que se puede requerir para describir las diferencias individuales a nivel de grupo o para el uso en propósitos formativos (Braun et al., 2020). Sin embargo, esta indicación general debe tomarse con cautela, ya que el grado de confianza que se puede otorgar a los niveles de fiabilidad depende del paradigma utilizado para el diseño de la prueba y el método de puntuación utilizado. Por ejemplo, si las puntuaciones se utilizan para asignar a los estudiantes a una categoría de diagnóstico o de rendimiento (ej.: aceptación o rechazo de la universidad), se podría confiar más en el punto de partida que guía la clasificación si el instrumento ha sido diseñado y puntuado por la Teoría de Respuesta al Ítem (TRI) que por la Teoría Clásica de los Test (TCT).

La TCT representa un enfoque tradicional en la medición de puntuaciones, basado en principios estadísticos clásicos que se enfoca en la consistencia y la confiabilidad de las puntuaciones mediante correlaciones entre los ítems de la prueba. En contraste, la TRI es un enfoque más contemporáneo que se apoya en modelos matemáticos para evaluar la probabilidad de que un individuo responda correctamente a un ítem específico, considerando la relación entre el nivel latente de la habilidad del individuo y la dificultad de los ítems. En resumen, la TRI proporciona una base más sólida para la evaluación y la toma de decisiones, particularmente en situaciones cruciales.

Con respecto a los estudios que pretenden analizar las diferencias entre grupos, es necesario informar sobre la significatividad y la magnitud de la diferencia. Algunas medidas para informar del tamaño del efecto son la *d* de

Cohen, la g de Hedges y la d de Glass¹.

Por último, es crucial reconocer las limitaciones al interpretar las puntuaciones del estudiantado y utilizar esta información como base para implementar mejoras. Igualmente, proporcionar un feedback regular y constructivo al estudiantado es esencial para ayudarle a comprender su progreso, identificar áreas de mejora y motivarle a esforzarse más. Este enfoque no solo permite al estudiantado descubrir sus fortalezas, sino que también fomenta el aprendizaje autónomo y contribuye a crear un ambiente de aprendizaje colaborativo y de apoyo mutuo. Implementar mejoras en este sentido podría incluir la incorporación de herramientas tecnológicas para agilizar la retroalimentación y hacerla más personalizada, así como la realización de encuestas periódicas al estudiantado para recopilar sus comentarios y adaptar las estrategias de enseñanza en consecuencia.

Conclusiones

Captar el PC de los estudiantes de educación superior en el ámbito de las Ciencias de la AFyD es una tarea compleja y desafiante. Es fundamental que el profesorado y el personal investigador cuenten con una comprensión clara de qué se entiende por PC y cómo se puede medir desde una perspectiva científica.

En este artículo, se presentan recomendaciones respaldadas por estudios de investigación para evaluar de manera eficiente y efectiva el PC en estudiantes universitarios en el contexto de las Ciencias de la AFyD. Estas directrices tienen como objetivo apoyar el proceso de evaluación del PC en el ámbito académico e incluyen la definición operativa del constructo, el establecimiento de objetivos de evaluación claros, la selección o creación de herramientas de evaluación adecuadas, el diseño de un proceso de evaluación efectivo, la aplicación sistemática de la evaluación, el suministro de feedback a los estudiantes y el seguimiento y evaluación continua del proceso de evaluación.

Es crucial definir operativamente el PC, especialmente cuando se trata de constructos complejos y abstractos. Establecer objetivos de evaluación precisos y seleccionar las herramientas adecuadas son aspectos fundamentales para una evaluación sólida. La elección del instrumento de evaluación puede impactar significativamente la relevancia de los resultados obtenidos. Además, es necesario tener en cuenta cuidadosamente el diseño de investigación y sus limitaciones al interpretar los resultados de la evaluación del PC en los estudiantes. La recopilación e interpretación de datos deben garantizar precisión y consistencia, minimizando sesgos como el de las expectativas o la memoria. Asimismo, es esencial seleccionar el análisis adecuado para los datos recopilados e interpretarlos con cautela, considerando todas las implicaciones y limitaciones.

En cuanto a las propiedades psicométricas, diversos

estudios en la evaluación del PC sugieren que este es un proceso complejo y multidimensional que puede ser difícil de cuantificar con técnicas estadísticas (Hyytinen & Toom, 2019; Repo et al., 2017; Rodrigues et al., 2018). El PC abarca diversas dimensiones y aspectos, como el razonamiento lógico, las consideraciones éticas y la toma de decisiones, y puede variar en diferentes situaciones. No obstante, estas dificultades no deben ser utilizadas como excusa para evitar la evaluación del PC. Por el contrario, deben servir como incentivo para seguir mejorando la evaluación de este constructo y encontrar formas de superar estos desafíos. Es esencial continuar evaluando el PC de los estudiantes para comprender mejor cómo pueden mejorar sus habilidades y tener éxito en sus estudios, así como en su vida personal y profesional. La evaluación del PC es un proceso integral que debe involucrar a los estudiantes y adaptarse a sus necesidades y características individuales. Es importante dedicar suficiente tiempo y oportunidades para desarrollar las habilidades de PC, y recordar que la evaluación es un proceso continuo que requiere revisión y mejora constantes. Esto implica evaluar la efectividad de las herramientas de evaluación y adaptarlas en consecuencia, así como revisar el proceso de evaluación en su conjunto para asegurar que contribuye de manera efectiva al desarrollo del PC de los estudiantes.

En cuanto a las limitaciones, es relevante reconocer que este artículo se ha centrado principalmente en la evaluación del PC a través de una metodología cuantitativa. No obstante, es importante destacar que las metodologías cualitativas y mixtas poseen un gran potencial para una comprensión más completa del PC. Para futuras investigaciones, sería interesante explorar el PC desde una perspectiva cualitativa o mixta, ya que esta variedad de enfoques permitiría captar las dimensiones subyacentes y las experiencias de los estudiantes en un nivel más profundo. Igualmente, para que esta investigación tenga un impacto práctico significativo, es esencial llevar a cabo una revisión más profunda de las fortalezas y debilidades de estas recomendaciones en situaciones reales de enseñanza y aprendizaje en el ámbito de la AFyD. Esto implica un análisis más detallado de cómo estas directrices pueden ser implementadas en entornos educativos y cómo podrían influir en la mejora de la enseñanza y el desarrollo del PC del estudiantado. Además, sería beneficioso explorar la retroalimentación de personal docente y estudiantado que han aplicado estas recomendaciones y considerar sus experiencias y perspectivas. Esta reflexión adicional podría arrojar luz sobre la viabilidad y la eficacia práctica de las pautas presentadas en este artículo, ayudando así a cerrar la brecha entre la teoría y la práctica en la evaluación del PC en el ámbito de la AFyD.

En síntesis, la evaluación del PC en el estudiantado universitario, particularmente en el contexto de la AFyD, plantea una serie de desafíos y complejidades que no pueden

¹ Para investigadores con muestras pequeñas, el uso de análisis psicométricos sofisticados puede no ser factible. En estos casos, es necesario recurrir a métodos científicos alternativos. Las metodologías cualitativas pueden ser igualmente rigurosas y válidas

como las cuantitativas (Bensley y Murtagh, 2012). Ya sea que los datos sean cualitativos o cuantitativos, se requiere una medición y análisis sistemáticos adecuados de los datos.

pasarse por alto. Este artículo ha ofrecido recomendaciones fundamentadas en investigaciones previas para llevar a cabo evaluaciones efectivas del PC. Sin embargo, la mejora continua y la refinación de este proceso son imperativas. El PC es un constructo intrincado y polifacético, manifestándose de diversas maneras en diferentes situaciones. Por lo tanto, la evaluación del PC debe concebirse como un esfuerzo en constante evolución que implica la colaboración activa de estudiantado, personal docente y expertos y expertas en la materia. Promover una cultura de evaluación en las instituciones de Educación Superior y aprovechar los hallazgos de la investigación para impulsar mejoras en la enseñanza y el aprendizaje son pilares esenciales para el progreso continuo en el campo de la AFyD.

Referencias

- Akpur, U. (2020). Critical, Reflective, Creative Thinking and Their Reflections on Academic Achievement. *Thinking Skills and Creativity*, 37(July). <https://doi.org/10.1016/j.tsc.2020.100683>
- Aktoprak, A., & Hursen, C. (2022). A bibliometric and content analysis of critical thinking in primary education. *Thinking Skills and Creativity*, 44, 101029. <https://doi.org/10.1016/J.TSC.2022.101029>
- Alejo, L. (2017). *El Pensamiento Crítico en estudiantes del grado en Maestro/a en Educación Primaria desde la Didáctica de las Ciencias Sociales*. Universidad de Málaga.
- Archila, P. A., Molina, J., Danies, G., Truscott De Mejía, A.-M., & Restrepo, S. (2022). Using the Controversy over Human Race to Introduce Students to the Identification and the Evaluation of Arguments. *Science & Education*, 31, 861–892. <https://doi.org/10.1007/s11191-021-00299-8>
- Bensley, D. A., & Murtagh, M. P. (2012). Guidelines for a Scientific Approach to Critical Thinking Assessment. *Teaching of Psychology*, 39(1), 5–16. <https://doi.org/10.1177/0098628311430642>
- Betancourth Zambrano, S., Muñoz Moran, K. T., & Rosas Lagos, T. J. (2017). Evaluación del pensamiento crítico en estudiantes de educación superior de la región de Atacama-Chile. *Prospectiva*, 23, 199. <https://doi.org/10.25100/prts.v0i23.4594>
- Braun, H. I., Shavelson, R. J., Zlatkin-Troitschanskaia, O., & Borowiec, K. (2020). Performance Assessment of Critical Thinking: Conceptualization, Design, and Implementation. *Frontiers in Education*, 5(September), 1–10. <https://doi.org/10.3389/feduc.2020.00156>
- Brijmohan, A., Khan, G. A., Orpwood, G., Brown, E. S., & Childs, R. A. (2018). Collaboration Between Content Experts and Assessment Specialists: Using a Validity Argument Framework to Develop a College Mathematics Assessment. *Canadian Journal of Education/Revue Canadienne de l'Éducation*, 41(2), 584–600. <https://www.jstor.org/stable/90025226>
- Brückner, S., & Pellegrino, J. W. (2016). *Integrating the Analysis of Mental Operations Into Multilevel Models to Validate an Assessment of Higher Education Students' Competency* Sebastian Brückner. 53(3), 293–312.
- Butler, H. A. (2012). *Halpern Critical Thinking Assessment Predicts Real-World Outcomes of Critical Thinking*. <https://doi.org/10.1002/acp.2851>
- Butler, H. A., Pentoney, C., & Bong, M. P. (2017). Predicting real-world outcomes: Critical thinking ability is a better predictor of life decisions than intelligence. *Thinking Skills and Creativity*, 25, 38–46. <https://doi.org/10.1016/J.TSC.2017.06.005>
- Buur, J. L., Schmidt, P., Smylie, D., Irizarry, K., Crocker, C., Tyler, J., & Barr, M. (2012). Validation of a scenario-based assessment of critical thinking using an externally validated tool. *Journal of Veterinary Medical Education*, 39(3), 276–282. <https://doi.org/10.3138/jvme.0112-009R>
- Cardozo, J. M. C., Velasco, A. D., Domínguez, S. L., & Reyes, J. A. M. (2019). ¿Qué se investiga en formación docente en educación física y en recreación? *Retos*, 2041(36), 3–8. <https://doi.org/https://doi.org/10.47197/retos.v36i36.51456>
- Cargas, S., Williams, S., & Rosenberg, M. (2017). An approach to teaching critical thinking across disciplines using performance tasks with a common rubric. *Thinking Skills and Creativity*, 26(November 2016), 24–37. <https://doi.org/10.1016/j.tsc.2017.05.005>
- Carter, A. G., Creedy, D. K., & Sidebotham, M. (2016). Efficacy of teaching methods used to develop critical thinking in nursing and midwifery undergraduate students: A systematic review of the literature. *Nurse Education Today*, 40, 209–218. <https://doi.org/10.1016/j.nedt.2016.03.010>
- Cottrell, S. (2017). *Critical Thinking Skills: Developing Effective Analysis and Argument* (Vol. 13, Issue 1). Palgrave Macmillan.
- Cruz, G., Payan-Carreira, R., & Dominguez, C. (2017). L'éducation de la pensée critique dans l'enseignement supérieur au Portugal: Une révision systématique des pratiques éducatives. *Revista Lusófona de Educação*, 38(38), 43–61. <https://doi.org/10.24140/issn.1645-7250.rle38.03>
- D'Alessio, F. A., Avolio, B. E., & Charles, V. (2019). Studying the impact of critical thinking on the academic performance of executive MBA students. *Thinking Skills and Creativity*, 31, 275–283. <https://doi.org/10.1016/j.tsc.2019.02.002>
- Danczak, S. M., Thompson, C. D., & Overton, T. L. (2020). Development and validation of an instrument to measure undergraduate chemistry students' critical thinking skills. *Chemistry Education Research and Practice*, 21(1), 62–78. <https://doi.org/10.1039/c8rp00130h>
- Davies, M. (2015). A Model of Critical Thinking in Higher Education. In M. B. Paulsen (Ed.), *Higher Education: Handbook of Theory and Research Volume 30* (pp. 41–92). Springer Cham. <https://doi.org/https://doi.org/10.1007/978-3->

319-12835-1

- Davies, M., & Barnett, R. (2015). The palgrave handbook of critical thinking in higher education. In M. Davies & R. Barnett (Eds.), *The Palgrave Handbook of Critical Thinking in Higher Education*. Palgrave Macmillan New York. <https://doi.org/10.1057/9781137378057>
- Dominguez, C., Nascimento, M. M., Payan-Carreira, R., Cruz, G., Silva, H., Lopes, J., Morais, M. da F. A., & Morais, E. (2015). Adding value to the learning process by online peer review activities: towards the elaboration of a methodology to promote critical thinking in future engineers. *European Journal of Engineering Education*, 40(5), 573–591. <https://doi.org/10.1080/03043797.2014.987649>
- Dwyer, C. P. (2017). *Critical thinking: Conceptual perspectives and practical guidelines*. Cambridge University Press.
- Dwyer, C. P., Hogan, M. J., & Stewart, I. (2014). An integrated critical thinking framework for the 21st century. *Thinking Skills and Creativity*, 12, 43–52. <https://doi.org/10.1016/j.tsc.2013.12.004>
- Ennis, R. H. (1985). A Logical Basis for Measuring Critical Thinking Skills. *Educational Leadership*, 43(2), 44–48. <https://pdfs.semanticscholar.org/80a7/c7d4a98987590751df4b1bd9adf747fd7aaa.pdf>
- Ennis, R. H. (2018). Critical Thinking Across the Curriculum: A Vision. *Topoi*, 37(1), 165–184. <https://doi.org/10.1007/s11245-016-9401-4>
- Ericson, J. D. (2022). Mapping the Relationship Between Critical Thinking and Design Thinking. *Journal of the Knowledge Economy*, 13(1), 406–429. <https://doi.org/10.1007/s13132-021-00733-w>
- Everett, A. C., Anderson, O. S., Wright, M. C., & Fontana, M. (2018). Longitudinal Assessment of Critical Thinking Skills Across a Dental Curriculum. *Journal of Dental Education*, 82(9), 921–928. <https://doi.org/10.21815/jde.018.088>
- Fadhullah, A., & Ahmad, N. (2017). Thinking Outside of the Box: Determining Students' Level of Critical Thinking Skills in Teaching and Learning. *Asian Journal of University Education*, 13(2), 51–70.
- Gillespie, L., & Culpan, I. (2000). Critical thinking: ensuring the “education” aspect is evident in Physical Education. *Journal of Physical Education New Zealand*, 33(3), 84–96. <http://dx.doi.org/10.1016/j.jaci.2012.05.050>
- González-de Paz, L., Elorduy-Vaquero, M., Virumbrales, M., Real, J., Sureda, X., Borrás, A., & Martínez-Sánchez, J. M. (2016). Learning to perform research in the degree of medicine: analysis and evaluation of the student ratings of the final degree dissertation. *Anales de Psicología*, 32(2), 484. <https://doi.org/10.6018/analeps.32.2.212261>
- Goodsett, M. (2020). Best practices for teaching and assessing critical thinking in information literacy online learning objects. *The Journal of Academic Librarianship*, 46(5), 102163. <https://doi.org/10.1016/j.acalib.2020.102163>
- Gul, M. G., & Akcay, H. (2020). Structuring a new socioscientific issues (SSI) based instruction model: Impacts on pre-service science teachers' (PSTs) critical thinking skills and dispositions. *International Journal of Research in Education and Science*, 6(1), 141–159. <https://doi.org/10.46328/ijres.v6i1.785>
- Halpern, D. F. (2014). *Thought and Knowledge: An Introduction to Critical Thinking* (5th ed.). Psychology Press.
- Hammonds, D. S., & Schwarze, M. J. (2019). Toward Oral Assessment in Counselor Education: Practice-Based Measurement of Competence in Pre-Practicum Experiences. *Journal of Creativity in Mental Health*, 14(4), 532–548. <https://doi.org/10.1080/15401383.2019.1632770>
- Hatcher, K. ., & Possin, D. (2020). Chapter 16 Commentary. In F. Fasko, Daniel; Fair (Ed.), *Critical Thinking and Reasoning. Theory, Development, Instruction, and Assessment* (pp. 298–322). Brill/ Sense Publishing. <https://doi.org/10.4135/9781412991445.d102>
- Hathcoat, J. D., Penn, J. D., Barnes, L. L. B., & Comer, J. C. (2016). A Second Dystopia in Education: Validity Issues in Authentic Assessment Practices. *Research in Higher Education*, 57(7), 892–912. <https://doi.org/10.1007/s11162-016-9407-1>
- Hernández, J. L. (2004). *La Propuesta de Transformación del Título de Licenciado en Ciencias de la Actividad Física y del Deporte en el Marco del Espacio Europeo de Educación Superior: Análisis del Contexto y del Contenido*.
- Hyytinen, H., & Toom, A. (2019). Developing a performance assessment task in the Finnish higher education context: Conceptual and empirical insights. *British Journal of Educational Psychology*, 89(3), 551–563. <https://doi.org/10.1111/bjep.12283>
- Indrašienė, V., Jegelevičienė, V., Merfeldaitė, O., Penkauskienė, D., Pivorienė, J., Railienė, A., Sadauskas, J., & Valavičienė, N. (2021). *Critical Thinking in Higher Education and Labour Market*. Peter Lang International Academic Publishers. <https://doi.org/10.3726/18636>
- Jacob, E., Duffield, C., & Jacob, D. (2017). A protocol for the development of a critical thinking assessment tool for nurses using a Delphi technique. *Journal of Advanced Nursing*, 73(8), 1982–1988. <https://doi.org/10.1111/jan.13306>
- Joshi, S. C., Gupta, K., & Manektala, S. (2022). Misinformation, Public Opinion, and the Role of Critical Thinking. *International Journal of Management and Humanities*, 8(9), 15–18. <https://doi.org/10.35940/ijmh.i1483.058922>
- Kim, K. J., Meir, E., Pope, D. S., & Wendel, D. (2017). WordBytes: Exploring an Intermediate Constraint Format for Rapid Classification of Student Answers on Constructed Response Assessments. *Journal of Educational Data Mining*, 9(2), 45–71.
- Ku, K. Y. L., Lee, V. S. L., & Ellis, J. W. (2017). Using artwork as problem context in generic critical thinking

- instruction: A strategy for thoughts. *Thinking Skills and Creativity*, 25, 53–59. <https://doi.org/10.1016/J.TSC.2017.07.001>
- Liu, J., McBride, R. E., Xiang, P., & Scarmardo-Rhodes, M. (2018). Physical Education Pre-service Teachers' Understanding, Application, and Development of Critical Thinking. *Quest*, 70(1), 12–27. <https://doi.org/10.1080/00336297.2017.1330218>
- Liu, O. L., Frankel, L., & Roohr, K. C. (2014). *Assessing Critical Thinking in Higher Education: Current State and Directions for Next-Generation Assessment*. <https://doi.org/10.1002/ets2.12009>
- Liu, O. L., Mao, L., Zhao, T., Yang, Y., Xu, J., & Wang, Z. (2016). Pilot Testing the Chinese Version of the ETS® Proficiency Profile Critical Thinking Test. *ETS Research Report Series*, 2016(2), 1–11. <https://doi.org/10.1002/ets2.12123>
- Liu, O. L., Shaw, A., Gu, L., Li, G., Hu, S., Yu, N., Ma, L., Xu, C., Guo, F., Su, Q., Kardanovaj, E., Chirikov, I., Shi, J., Shi, Z., Wang, H., & Loyalka, P. (2018). Assessing college critical thinking: preliminary results from the Chinese HEIghten® Critical Thinking assessment. *Higher Education Research and Development*, 37(5), 999–1014. <https://doi.org/10.1080/07294360.2018.1467381>
- Maravé-Vivas, M., Salvador-García, C., Gil-Gómez, J., Valverde-Esteve, T., & Martín-Moya, R. (2022). How Can Service-Learning Shape the Political Perspectives of Pre-Service Teachers? A Program in the Field of Physical Education. *International Journal of Environmental Research and Public Health*, 19(15). <https://doi.org/10.3390/ijerph19159175>
- McBride, R. E. (1991). Critical Thinking—An Overview with Implications for Physical Education. *Journal of Teaching in Physical Education*, 11(2), 112–125. <https://doi.org/10.1123/jtpe.11.2.112>
- McComas, M. J., Wright, R. A., Mann, N. K., Cooper, M. D., & Jacks, M. E. (2013). M-OSCE as a Method to Measure Dental Hygiene Students' Critical Thinking: A Pilot Study. *Journal of Dental Education*, 77(4), 485–493. <https://doi.org/10.1002/j.0022-0337.2013.77.4.tb05494.x>
- Moore, T. (2013). Critical thinking: Seven definitions in search of a concept. *Studies in Higher Education*, 38(4), 506–522. <https://doi.org/10.1080/03075079.2011.586995>
- Morais, F., Silva, H., Cruz, G., Pedrosa, D., Payan-Carreira, R., Dominguez, C., & Nascimento, M. M. (2019). Perceptions of portuguese university teachers about critical thinking educational practices. *Communications in Computer and Information Science*, 993, 223–239. https://doi.org/10.1007/978-3-030-20954-4_17
- Mueller, J. F., Taylor, H. K., Brakke, K., Drysdale, M., Kelly, K., Levine, G. M., & Ronquillo-Adachi, J. (2020). Assessment of Scientific Inquiry and Critical Thinking: Measuring APA Goal 2 Student Learning Outcomes. *Teaching of Psychology*, 47(4), 274–284. <https://doi.org/10.1177/0098628320945114>
- Negret, J. (2016). Formación ciudadana, cultura física y deporte: estrategia para una formación de calidad. *Revista Cubana de Educación Superior*, 2016(1), 4–17.
- O'Leary, M., Reynolds, K., Ling, G., Liu, O. L., Belton, S., O'Reilly, N., & McKenna, J. (2020). Assessing Critical Thining in Higher Education: Validity Evidence for the Use of the HEIghten Critical Thinking in Ireland. *Journal of Higher Education Theory and Practice*, 20(12), 115–130. <https://doi.org/10.33423/jhetp.v20i12.3783>
- Ossa-Cornejo, C., Palma-Luengo, M., Martín, N. L. S., & Díaz-Larenas, C. (2018). Evaluación del pensamiento crítico y científico en estudiantes de pedagogía de una universidad chilena. *Revista Electronica Educare*, 22(2), 1–18. <https://doi.org/10.15359/ree.22-2.12>
- Paul, R., & Elder, L. (2019). *A Guide For Educators to Critical Thinking Competency Standards: Standards, Principles, Performance Indicators, and Outcomes With a Critical Thinking Master Rubric*. The Foundation for Critical Thinking. www.criticalthinking.org
- Pérez-Tejero, J., Casas-Carmona, M., & Chavez Hernandez, S. F. (2023). Competencias laborales en el mercado deportivo en España: la opinión de los empleadores. *Retos*, 50, 799–807. <https://doi.org/10.47197/retos.v50.99425>
- Pnevmatikos, D., Christodoulou, P., & Georgiadou, T. (2019). Promoting critical thinking in higher education through the values and knowledge education (VaKE) method. *Studies in Higher Education*, 44(5), 892–901. <https://doi.org/10.1080/03075079.2019.1586340>
- Poblete-Valderrama, F., Garrido-méndez, A., Castillo, C. M., Cáceres, R. C., Toro-, A., Aburto, J. C., Aguilar, L. I., Cenzano-castillo, L., Rodríguez, K. H., & Rivera, C. F. (2023). Formación Inicial Docente en Educación Física, paradigmas de la formación desde la percepción del formador de formadores, lineamientos ministeriales y mallas curriculares: estudio de caso. *Retos: Nuevas Tendencias En Educación Física, Deporte y Recreación*, 49, 552–563.
- Porter, T., Baldwin, C. R., Warren, M. T., Murray, E. D., Cotton Bronk, K., Forgeard, M. J. C., Snow, N. E., & Jayawickreme, E. (2021). Clarifying the Content of Intellectual Humility: A Systematic Review and Integrative Framework. *Journal of Personality Assessment*. <https://doi.org/10.1080/00223891.2021.1975725>
- Prat-Sala, M., & van Duuren, M. (2022). Critical Thinking Performance Increases in Psychology Undergraduates Measured Using a Workplace-Recognized Test. *Teaching of Psychology*, 49(2), 153–163. <https://doi.org/10.1177/0098628320957981>
- Prieto, G., & Delgado, A. R. (2010). Fiabilidad y Validez. *Papeles Del Psicólogo*, 3(1), 67–74. <https://www.redalyc.org/pdf/778/77812441007.pdf>
- Puig, B., Blanco-Anaya, P., Bargiela, I. M., & Crujeiras-

- Pérez, B. (2019). A systematic review on critical thinking intervention studies in higher education across professional fields. *Studies in Higher Education*, 44(5), 860–869. <https://doi.org/10.1080/03075079.2019.1586333>
- Re, M. R., Amenduni, F., De Medio, C., & Valente, M. (2019). How to use assessment data collected through writing activities to identify participants' critical thinking levels. *Journal of E-Learning and Knowledge Society*, 15(3), 117–132. <https://doi.org/10.20368/1971-8829/1135051>
- Repo, S., Lehtinen, T., Rusanen, E., & Hyytinen, H. (2017). Prior education of Open University students contributes to their capability in critical thinking. *Journal of Adult and Continuing Education*, 23(1), 61–77. <https://doi.org/10.1177/1477971417693416>
- Reynders, G., Lantz, L., Ruder, S. M., Stanford, C. L., & Cole, R. S. (2020). Rubrics to assess critical thinking and information processing in undergraduate STEM courses. *International Journal of STEM Education*, 7(1), 1–15. <https://doi.org/https://doi.org/10.1186/s40594-020-00208-5>
- Rodrigues, A., Soares, P., & Da Silva, L. (2018). Translation, adaptation, and validation of the Halpern Critical Thinking Assessment to Portugal: Effect of disciplinary area and academic level on critical thinking. *Anales de Psicología*, 34(2), 292. <https://doi.org/10.6018/analesps.34.2.272401>
- Santos, L. F., Gyenes, A., Fujisaki, S., & Gay, S. (2021). *Critical Thinking*. viii(1).
- Schendel, R., & Tolmie, A. (2017). Beyond translation: adapting a performance-task-based assessment of critical thinking ability for use in Rwanda. *Assessment and Evaluation in Higher Education*, 42(5), 673–689. <https://doi.org/10.1080/02602938.2016.1177484>
- Serrano, J. L., Carrera, X., Brescó, E., & Suárez-Guerrero, C. (2019). Tratamiento crítico de la información de estudiantes universitarios desde los entornos personales de aprendizaje. *Educação e Pesquisa*, 45, 1–21. <https://doi.org/10.1590/s1678-4634201945193355>
- Simper, N. (2018). Rubric authoring tool supporting cognitive skills assessment across an institution. *Teaching and Learning Inquiry*, 6(1), 10–24. <https://doi.org/10.20343/teachlearninq.6.1.3>
- Smith, G. F. (2014). *Assessing Business Student Thinking Skills*. <https://doi.org/10.1177/1052562913489028>
- Sokolnicky, A. C. (2019). On loaded language. *CURRENTS. A Journal of Young English Philology Thought and Review*, 5, 9–25.
- Southworth, J. (2022). Bridging critical thinking and transformative learning: The role of perspective-taking. *Theory and Research in Education*, 20(1), 44–63. <https://doi.org/10.1177/14778785221090853>
- Su, H. F. H., Ricci, F. A., & Mnatsakanian, M. (2016). Mathematical teaching strategies: Pathways to critical thinking and metacognition. *International Journal of Research in Education and Science*, 2(1), 190–200. <https://doi.org/10.21890/ijres.57796>
- Thomas, K., & Lok, B. (2015). Teaching Critical Thinking: An Operational Framework. In M. Davies & R. Barnett (Eds.), *The Palgrave Handbook of Critical Thinking in Higher Education*. Palgrave Macmillan.
- Thornhill-Miller, B., Camarda, A., Mercier, M., Burkhardt, J. M., Morisseau, T., Bourgeois-Bougrine, S., Vinchon, F., El Hayek, S., Augereau-Landais, M., Mourey, F., Feybesse, C., Sundquist, D., & Lubart, T. (2023). Creativity, Critical Thinking, Communication, and Collaboration: Assessment, Certification, and Promotion of 21st Century Skills for the Future of Work and Education. *Journal of Intelligence*, 11(3). <https://doi.org/10.3390/jintelligence11030054>
- Tittle, P. (2011). *Critical thinking: An appeal to reason*. Routledge.
- Velez, G., & Power, S. A. (2020). Teaching students how to think, not what to think: Pedagogy and political psychology. *Journal of Social and Political Psychology*, 8(1), 388–403. <https://doi.org/10.5964/jspp.v8i1.1284>
- Zlatkin-Troitschanskaia, O., Shavelson, R. J., Schmidt, S., & Beck, K. (2019). On the complementarity of holistic and analytic approaches to performance assessment scoring. *British Journal of Educational Psychology*, 89(3), 468–484. <https://doi.org/10.1111/bjep.12286>