



## LOS RETOS DE LA EVALUACIÓN DEL PENSAMIENTO CRÍTICO EN LA EDUCACIÓN SUPERIOR

Silvia Fernández Rivas<sup>1</sup>

Carlos Saiz Sánchez<sup>2</sup>

### RESUMEN

La enseñanza superior debería adaptarse a los nuevos retos del siglo actual, pero no lo está logrando. Las transformaciones acontecidas en los últimos tiempos no han ido parejas con innovaciones en la preparación de nuestros futuros ciudadanos y profesionales. A nuestro juicio, este desajuste se debe en parte al hecho de que no se realiza un diagnóstico del funcionamiento del sistema de formación de los estudiantes, con el fin de saber qué funciona bien y qué no. Sin esta visión no es posible convencer a los profesionales de la educación para que muten sus objetivos y sus prácticas. Se necesita saber qué hará mejores a nuestros futuros ciudadanos y profesionales en estos tiempos tan exigentes en todos los ámbitos. Desde nuestro grupo de pensamiento crítico, proponemos enfocar la educación de un modo diferente al aún más extendido en la actualidad y, sobre todo, planteamos que sin una evaluación completa del proceso de aprendizaje no es posible el cambio y la mejora en el mismo. Llevamos tiempo desarrollando procedimientos de mejora del aprendizaje y diseñando sistemas de valoración del mismo. Ambas actividades son inseparables. En este trabajo presentaremos el núcleo de estas líneas de investigación.

**Palabras-clave:** Pensamiento crítico; Evaluación; Formación; Educación superior.

## OS DESAFIOS DA AVALIAÇÃO DO PENSAMENTO CRÍTICO NO ENSINO SUPERIOR

### RESUMO

O ensino superior deveria adaptar-se aos novos desafios do presente século, mas não está a conseguir fazê-lo. As transformações ocorridas nos últimos tempos não foram acompanhadas de inovações na preparação dos nossos futuros cidadãos e profissionais. Consideramos que este desajuste se deve, em parte, ao facto de não ser efectuado um diagnóstico do funcionamento do sistema de formação dos estudantes, para saber o que funciona bem e o que não funciona. Sem essa perspetiva, não é possível estimular os profissionais da educação a mudar os seus objetivos e as suas práticas. É necessário descobrir o que tornará os nossos futuros cidadãos e profissionais mais capacitados, em toda as áreas, nestes tempos difíceis. No quadro do nosso grupo de pensamento crítico, propomo-nos a abordar a educação de uma forma diferente daquela que é, ainda, a mais difundida na atualidade e, sobretudo, defendemos que sem uma avaliação completa do processo de aprendizagem não é possível mudá-lo e aperfeiçoá-lo. Desde há algum tempo que temos vindo a desenvolver procedimentos de melhoria da aprendizagem e concebemos sistemas de avaliação da mesma. Ambas as atividades são inseparáveis. No presente trabalho, apresentaremos o cerne dessas linhas de investigação.

**Palavras-chave:** Pensamento crítico; Avaliação; Formação; Ensino superior.

<sup>1</sup> Profesora de Pensamiento crítico Departamento de Psicología Básica, Psicobiología y Metodología de CC Facultad de Psicología. Universidad de Salamanca, España. E-mail: <[silviaferivas@usal.es](mailto:silviaferivas@usal.es)>. ORCID ID.: <http://orcid.org/0000-0001-6871-5246>.

<sup>2</sup> Profesor de Pensamiento crítico Departamento de Psicología Básica, Psicobiología y Metodología de CC Facultad de Psicología. Universidad de Salamanca, España. E-mail: <[csaiz@usal.es](mailto:csaiz@usal.es)> . ORCID ID.: <http://orcid.org/0000-0002-5243-958X>.



## THE CHALLENGES OF EVALUATING CRITICAL THINKING IN HIGHER EDUCATION

### ABSTRACT

Higher education should adapt to the new challenges of the current century, but it is not succeeding. The transformations that have taken place in recent times have not gone hand in hand with innovations in the preparation of our future citizens and professionals. In our opinion, this mismatch is due in part to the fact that a diagnosis of the functioning of the student training system is not carried out, in order to know what works well and what does not. Without this vision, it is not possible to convince education professionals to change their goals and practices. We need to know what will make our future citizens and professionals better in all areas in these demanding times. From our critical thinking group, we propose to approach education in a different way from the one that is even more widespread today and, above all, we propose that without a complete evaluation of the learning process, change and improvement in it is not possible. We have been developing learning improvement procedures for some time and designing learning assessment systems. Both activities are inseparable. In this work we will present the core of these lines of research.

**Keywords:** Critical thinking; Assessment; Instruction; Higher education.

### Introducción

En las últimas décadas, se han producido grandes cambios sociales y tecnológicos a los que no nos estamos adaptando todo lo bien que podríamos. El entorno se transforma tan rápido que nos cuesta seguirlo. Esto lo ha descrito muy bien José Antonio Marina al proponer la *Ley universal del aprendizaje* (LUA). Con ella expone lo siguiente: toda persona, institución o sociedad debe aprender a la misma velocidad con la que cambia su entorno si quiere sobrevivir y, si desea progresar, debe hacerlo a más velocidad (Marina; Rambaud, 2018). En los tiempos actuales, padecemos un preocupante enlentecimiento en la asimilación de esos cambios o la adaptación a los mismos. La lentitud del aprendizaje es evidente; y si se mantuviera por mucho tiempo, podría amenazar nuestra supervivencia. Afortunadamente, se puede evitar.

La ventaja enorme de esta revolución tecnológica como es internet, que permite disponer instantáneamente de cualquier información, supone un peaje del cual poco a poco vamos siendo conscientes. La sobrestimulación está dificultando nuestro sistema de procesamiento, porque se fracasa al seleccionar, discriminar o filtrar dicha información. Esto ocurre básicamente porque la atención funciona de un modo cada vez más limitado. Como ejemplo, a nuestros estudiantes, de un tiempo a esta parte, cada vez les cuesta más mantener su atención más allá de quince o veinte minutos, un hecho bastante preocupante. Esta dificultad de concentración impide que se pueda realizar un procesamiento profundo de la información, lo que significa que solo se produce un pensamiento superficial. De nuevo, esto

se puede ilustrar desde nuestra labor docente. Una vez más, hace tiempo que observamos que nuestros alumnos no son capaces de realizar tres inferencias seguidas, a la segunda relación ya no son capaces de continuar. Estas dos deficiencias graves dan cuenta de una parte del problema de un modo bastante obvio. El exceso de información o estimulación ocasiona dificultad de concentración que, a su vez, impide el procesamiento profundo y fomenta el superficial, de modo que no podremos manejar convenientemente nuestro entorno. Claramente, desde estos hechos podemos apreciar que la LUA no se cumple, pues cada vez más observamos una preocupante lentitud del aprendizaje, en general.

Otra consecuencia preocupante de este “nuevo mundo” en el que nos encontramos es la falta de conciencia de que lo que era importante ayer, hoy ya no lo es. Todavía no nos hemos percatado de que la administración de saberes no es la mejor forma de progresar o, simplemente, sobrevivir. Una vez más, los economistas señalan un hecho muy revelador. Un país que aumente veinticinco puntos en la prueba de PISA (**Programme for International Student Assessment**), incrementa en un 3 % su PIB (Saiz; Rivas; Almeida, 2020). Con todas las reservas que podamos poner a este sistema de evaluación, y se las ponemos sin lugar a duda, sin embargo, el hecho de que la riqueza de un país cambie en función de su nivel educativo, debe decirnos algo. Este dato, junto con otros estudios, hace tiempo que han puesto de manifiesto la necesidad de lo que han llamado *learnability* (LE) - Dumont; Istance; Benavides, 2010-. En la era de internet se sigue primando la administración de saberes y no la gestión del aprendizaje. Todavía se atiende principalmente a la transmisión del conocimiento y no a su adquisición, no al mejor modo de lograrlo. Aún no hemos entendido que la sociedad del conocimiento no debe ser predominante en la actualidad, pues debe dejar paso a la sociedad del aprendizaje, del LE. (Para un desarrollo más detallado de estos análisis, ver Saiz, Rivas y Almeida, 2020).

El incumplimiento de estos dos principios (LUA, LE) contribuye a otro problema preocupante para el futuro. Siguen siendo los economistas los que indirectamente nos llaman la atención sobre lo que no estamos haciendo del todo bien respecto a la educación o formación. Otro hecho que nos ponen delante y que nos debería invitar a pensar mucho es el siguiente: ellos ya saben que alguien que termine hoy día sus estudios de grado y postgrado, por ejemplo, en nuestra comunidad o país, a lo largo de su vida laboral, cambiará una media de doce a catorce veces de puesto de trabajo o empleo (Saiz; Rivas; Almeida, 2020). Sobre lo que nos llaman la atención es sobre el hecho de que la formación debe

preparar a futuros *knowmad* (KM) -Moravec, 2013-. Esta expresión nos está indicando que el conocimiento (*know*) debe ser, nómada (*mad*, abreviatura de *nomad*). El modo en el que debemos formar o formarnos debe ser para “nómadas del conocimiento”, puesto que las demandas sociales exigirán que apliquemos a diferentes contextos nuestra formación, que desarrollemos más nuestras competencias horizontales o transversales, que las verticales (ver Saiz; Rivas; Almeida, 2020). Y en este punto el resultado es claro, debemos decidir si queremos estar dentro o fuera del mundo laboral, y solo hay un modo. Este planteamiento se ha aplicado a otro ámbito fundamental como el bienestar personal, en otro trabajo del grupo (Saiz; Rivas, 2020), que por razones de espacio no podemos tratar aquí.

Estos tres principios (LUA, LE, KM) nos sirven para describir los cambios que están aconteciendo y sus consecuencias en el ámbito de la educación, que es nuestra preocupación. La *lentitud* de la formación, la predominancia de la *administración* de saberes, y la aún superior *verticalidad* de las enseñanzas, creemos, definen muy bien la situación actual en lo referente a los problemas esenciales de la formación. Las investigaciones que nuestro grupo lleva realizando desde hace ya más de una década avalan inequívocamente el diagnóstico que acabamos de realizar, guiados por estos tres ejes centrales. El pensamiento crítico (PC) es el campo desde el que venimos aportando diferentes soluciones a los problemas de formación actuales. Y desde aquí es de donde vamos a ofrecer un modo de resolver parte de las limitaciones actuales con las que nos encontramos ante los grandes retos del siglo actual. Esta propuesta de solución es el trabajo que presentamos en este artículo.

Desde diferentes ámbitos se viene insistiendo en la importancia de pensar críticamente como ciudadanos y como profesionales, y se reitera que sin él no es posible construir una sociedad mejor o ser un experto resolutivo. Sin embargo, una cosa es la teoría y otra muy diferente la práctica. Ciertamente, el PC tiene algo que decir en educación, quizás mucho más que otros enfoques u orientaciones, aunque solo sea porque desde su origen nace como una apuesta aplicada a este ámbito. Por ello, los investigadores que lo conforman llevan décadas tratando de mejorar el proceso de aprendizaje y reclamando a las instituciones responsables un cambio de mentalidad y de maneras de actuar. Y lo que plantean desde hace tiempo es sencillo, efectuar un buen diagnóstico para proponer la mejor intervención. Parte del diagnóstico lo acabamos de hacer (*lentitud*, *administración* y *verticalidad*). El PC ofrece para cada caso una forma de proceder diferente. El aprendizaje se puede mejorar o acelerar con un *cambio de metodología*. La formación debe ser esencialmente procedimental, esto es,

basada en mucha práctica intra e inter-dominio. Además, se debe trabajar con proyectos reales, en donde debe haber un mayor esfuerzo aplicado. Por último, cada vez es más necesario trabajar como una comunidad de indagación. Desde la investigación en el campo del PC y en nuestro grupo, la conclusión es clara, solo se mejorará el aprendizaje desde la *indagación*, la *práctica* y la *aplicación*. De este modo, según la LUA, no solo sobreviviremos, sino que además progresaremos (para un tratamiento detallado ver Saiz, 2017).

En lo que se refiere a LE, hemos dicho que lo que debe orientar la formación es la gestión del aprendizaje y no la administración de saberes. Además de lo ya propuesto, lo que el PC tiene que decir aquí es fomentar las competencias de manera ecléctica. Por un lado, desde los dominios específicos y, por otro, a través de la enseñanza directa. Debemos fortalecer desde todos los puntos dichas habilidades, en definitiva, centrar el esfuerzo en el proceso de adquisición, más que en el producto (Saiz; Rivas; Almeida, 2020). Pero al mismo tiempo debemos ser conscientes de la naturaleza doble de las competencias, cognitiva y no cognitiva. En realidad, su esencia nace de la suma de habilidades, más actitudes, más conocimiento (Campirán, 2016). Lo actitudinal, disposicional y motivacional son los elementos imprescindibles para el buen funcionamiento de nuestras capacidades. Lo actitudinal hace aparecer la perseverancia, la resistencia o el esfuerzo, tan necesarios para que nuestra maquinaria intelectual funcione con facilidad. Sin embargo, aún necesitamos un refuerzo adicional, pues lo que permite un buen manejo de los procesos de aprendizaje y su consolidación es el *darnos cuenta* de lo que mentalmente ha sucedido, lo que ha operado bien y lo que no. La *metacognición* siempre es el aliado esencial en cualquier mejora del aprendizaje, el que posibilita, junto con todo lo anterior, la gestión correcta del mismo (para más detalle ver Saiz, 2020).

Finalmente, en el KM, es donde el PC habla con mayor claridad. La verticalidad es lo menos indicado en un “nómada del conocimiento” y lo secundario en el PC. Las competencias aquí son esencialmente transversales o generalizables. La horizontalidad es la que nos permitirá ser versátiles y poder desempeñar con eficacia los diferentes trabajos por los que se pasará a lo largo de una vida laboral. La habilidad fundamental de esta clase es nuestra capacidad de solucionar problemas. Este tipo de pericia es la que nos permite producir cambios y mejoras, o modificar nuestro entorno para el bien personal, de una organización o de la sociedad. Pero esta competencia descansa en otras también muy importantes, como es nuestra capacidad de argumentar correctamente, de tomar buenas

decisiones o de explicar la realidad de manera cierta. Estos son los pilares de nuestro potencial resolutivo, en cualquier ámbito, en el personal o en el profesional. En realidad, la *adquisición del conocimiento basada en la explicación* es lo que nos proporciona la mayor transversalidad en cualquier contexto, pues nos llevará siempre a la mejor decisión y posterior solución idónea. Lo importante de esta maquinaria cognitiva está en que es la que verdaderamente consigue los resultados, la que produce el cambio, la que realmente permite obtener los logros o metas deseadas. Este conjunto de habilidades es lo que hace que las ideas sean tangibles, las que ejecutan y hacen realidad un plan, las que convierten la representación en acción, en definitiva, las que transforman nuestro entorno. Nada tendría sentido si no se es resolutivo, si no se obtienen resultados, cambio, si no se es eficaz (para una profundización ver Saiz, 2020).

Los ejes empleados para justificar lo que el PC tiene que decir respecto a la educación o formación, se puede resumir del modo siguiente, respondiendo a las siguientes cuestiones: ¿Por qué o para qué? ¿Cómo? y ¿Qué? El *porqué* de esas competencias se entiende porque son las que ofrecen la mejor solución a los problemas derivados de esos cambios y demandas sociales nuevas, porque permiten una formación procedimental imprescindible para sobrevivir y progresar. El *cómo* lo alcanzamos resolviendo problemas significativos y trabajando como una comunidad de indagación. Este modo de actuación permite fomentar al máximo el aprendizaje activo. Finalmente, el *qué* se alcanza enfocando siempre la formación por resultados, fijándose como meta el cambio y la eficacia.

Nuestro grupo lleva años trabajando en la dirección expuesta anteriormente. En este sentido, hemos desarrollado una metodología de enseñanza verificada y publicada, en particular, los instrumentos ARDESOS y DIAPROVE. En la web de nuestro grupo (<https://www.pensamiento-critico.com/>), se puede acceder a todas estas investigaciones publicadas en libros, capítulos de libros y revistas especializadas. Con la experiencia de todos estos años en la formación del pensamiento crítico, podemos acreditar qué deficiencias, dificultades y limitaciones son propias de nuestro tiempo, qué problemas son más urgentes, y cuál es el modo más conveniente de afrontarlos.

En nuestra exposición anterior, hemos descrito los problemas fundamentales en formación y educación. Llevamos tiempo aplicando soluciones eficientes a una buena parte de esos problemas. Sin embargo, como es evidente, antes de cualquier tratamiento se debe efectuar un diagnóstico. Para poder intervenir en cualquier ámbito, como el que aquí nos

importa, el educativo, debemos identificar las carencias y los obstáculos que dificultan el proceso de aprendizaje. Pero, para ello, necesitamos disponer de instrumentos de medida válidos. En el terreno de la evaluación, ya lo hemos mencionado, nuestro grupo tiene también una trayectoria de años. Hemos diseñado pruebas de diferente naturaleza: de rendimiento, y psicométricas o estandarizadas. Cada una por sí sola es insuficiente para captar la mejora de nuestro pensamiento. Las primeras son una seguridad para comenzar con las segundas. Sin un rendimiento específico en tareas propias de la intervención no podemos tener garantía de que haya cambios en lo que enseñamos. Después, una vez que estemos seguros de que hay mejora específica, podemos aplicar pruebas estandarizadas que nos permitan ir más allá del contexto de aprendizaje.

Las pruebas de rendimiento asociadas a un programa de instrucción concreto facilitan información inapreciable sobre la eficacia de la intervención. Nos aseguramos de saber qué competencias se dominan y cuáles no, con el fin de actuar en consecuencia, desde el punto de vista de la formación (Saiz; Rivas 2008b). Pero esta no es la única manera de evaluación que se necesita. Además, conviene emplear pruebas estandarizadas, en combinación con las de rendimiento. Con estas, además de replicar dichas medidas, nos permite compararlas con las realizadas por otros investigadores, con independencia de los métodos de formación y evaluación empleados, y esta es su gran virtud (Saiz, 2008b). Digamos que, si disponemos de una medida estandarizada de PC, podremos compararla con cualquier otra que mida lo mismo. Así pues, si comparo los valores de la prueba de PC con otras del mismo tipo, puedo enriquecer cualquier investigación de manera exponencial. Por esta razón, son tan importantes esta clase de medidas.

Una prueba estandarizada tiene más valor que otras medidas de otra clase, por las razones expuestas anteriormente, aunque no son excluyentes; y al ser más idóneas para la evaluación, también lo son como herramienta de diagnóstico, una de las carencias preocupantes, por generalizadas, en formación y educación. Hasta donde nosotros conocemos, en nuestra comunidad y en nuestro país, no existen estudios de evaluación de las competencias de PC. Esto impide saber el nivel de dominio sobre el mismo que, a su vez, imposibilita actuar desde el punto de vista de la formación, con el fin de mejorar las deficiencias que hubiera, simplemente, porque no las conocemos.

Por todo lo dicho hasta aquí, el hecho de que no exista un diagnóstico de las competencias transversales de nuestros estudiantes es una limitación importante, desde el



punto de vista de la mejora de la educación. Creemos que esta laguna es un problema preocupante y que debe acometerse su solución lo antes posible. En el trabajo que presentamos ofrecemos un modo de resolverlo. Llevamos unos años enfrentándonos a este problema con relativo acierto. Hace ya varios años, nuestro equipo de investigación comenzó la construcción de una prueba de evaluación del pensamiento crítico en nuestro idioma, como consecuencia del fracaso en la adaptación y validación de uno de los mejores test de PC que conocemos. Esta prueba se llamó primero el Halpern Critical Thinking Assessment Using Everyday Situations (HCTAES), de Halpern (2006), y después la comercialización se denominó el Halpern Critical Thinking Assessment (HCTA) de Halpern (2010, 2012); ambos son el mismo test, lo conforman los mismos ítems, y puesto que nosotros hemos trabajado la prueba HCTAES, de ahora en adelante, nos referiremos a esta, sin olvidar que es la misma que la comercial HCTA, de la que existen ya varias versiones en otros idiomas, menos en el nuestro.

Pues bien, intentamos adaptar y validar el HCTAES. Después de publicar un estudio sobre el mismo (Nieto; Saiz; Orgaz, 2009), fracasamos en su validación, a pesar de emplear una muestra considerable (más de 700 personas, frente a los 120 que empleó la autora, en la validación inglesa) -ver Nieto y Saiz, 2008-. En otros proyectos posteriores, estudiamos con mayor profundidad esta herramienta, concluyendo que adolecía de algunas limitaciones importantes, que no nos permitiría emplearlo como instrumento de diagnóstico. Finalmente, desarrollamos una prueba propia de evaluación del PC, el PENCRISAL, que describiremos más adelante. El test PENCRISAL (Rivas y Saiz, 2012; Saiz y Rivas, 2008b) lleva empleándose varios años, con el fin de diagnosticar a nuestros estudiantes del grado de psicología. Además, desde su publicación, ofrecemos un servicio de evaluación que ha permitido que se aplicara en diferentes países latinoamericanos en decenas a estudiantes de centros universitarios, de secundaria, y a profesionales de diferentes especialidades. De hecho, existe una adaptación peruana (Rivas; Morales; Saiz, 2014), que proporciona unos baremos más ajustados para esos países. Además, disponemos una versión en portugués que se aplica en Brasil, principalmente, y en la actualidad, se está empleando también dentro de un proyecto de investigación en la Universidade de Aveiro (Portugal).

### **Antecedentes y situación actual de la investigación**

A pesar de todas las dificultades expuestas anteriormente, en los últimos



tiempos, hay un creciente interés en la educación superior por la mejora de las competencias transversales de PC (Saiz, 2017). El nuevo enfoque por competencias de la enseñanza universitaria le otorga un papel central en los nuevos planes de estudio, y aunque todavía son escasos, está aumentado el desarrollo de programas que mejoren estas habilidades dentro de diferentes áreas de los estudios universitarios. Pero de todos es sabido, que a la implementación de cualquier metodología le debe seguir necesariamente la evaluación de su eficacia, pues ésta es la que nos permite mejorarla y cambiarla, y esto no siempre se hace. Y para ello no debe valernos cualquier instrumento. La evaluación debe caminar de la mano de la intervención. Necesitamos sistemas de evaluación eficaces y válidos que nos permitan evaluar de la manera más precisa las habilidades de PC y que se ajusten a los objetivos de nuestras intervenciones (Saiz; Rivas, 2008b). Para ello se debe comprobar si el rendimiento de los alumnos mejora después de recibir la instrucción en comparación al momento antes de la intervención, es decir, estamos valorando la eficacia en función de la magnitud del efecto de la intervención. Las investigaciones en este campo se centran fundamentalmente en comprobar si los programas de instrucción son eficaces y, por tanto, si los alumnos mejoran después de los mismos. Son muchas las iniciativas de instrucción que ponen de manifiesto resultados de eficacia (ver en Saiz, 2017). Por citar los trabajos más clásicos, Perkins y Grotzer (1997) hacen una extensa revisión sobre iniciativas de instrucción de programas generales como “Filosofía para niños” de Lipman (1976, 2003), el de “Pensamiento productivo” de Covington, Crutchfield, Davies y Olton (1974), o el de “Enriquecimiento instrumental” de Feuerstein (1980), con resultados positivos que muestran la eficacia de estos programas en la mejora de las habilidades de PC. Al respecto existen revisiones extensas sobre los programas de enseñar a pensar, que el lector interesado puede consultar (Maclure, 1994; Maclure y Davies, 1994; Nickerson, Perkins y Smith, 1987; Saiz, 2002; Segal, Chapman y Glaser, 1985). Por otra parte, Halpern (2014) y Saiz (2017), presentan una revisión más actual de estudios que abordan los mecanismos específicos del PC, todos ellos con resultados también positivos después de una instrucción. Un tratamiento amplio y detallado de los programas del PC en psicología se puede encontrar en el libro editado por Sternberg y Halpern (2020). En todos estos estudios, estamos valorando la eficacia de los programas en función de la magnitud del efecto de la intervención y en muchos de ellos también la transferencia de las habilidades a nuevos contextos y dominios de conocimiento. Sin embargo, uno de los mayores inconvenientes en este tipo de estudios sigue siendo la falta de instrumentos confiables y

válidos que permitan evaluar con certeza los logros en las distintas propuestas de intervención.

De nada nos sirve una buena iniciativa sin una adecuada herramienta de evaluación. Es en este punto donde existe un gran vacío, sobre todo en nuestro idioma. Por esta razón, nuestros esfuerzos en investigación van encaminados en llenar esta laguna en el ámbito de la evaluación del PC.

### **La evaluación del pensamiento crítico**

En cuanto a la evaluación del PC, se han desarrollado diversidad de test (Ennis, 2003), pero con sus limitaciones y ninguno en castellano. Pasamos a hacer un breve recorrido por los más importantes. Uno de los instrumentos más conocidos para evaluar habilidades de PC es el *Watson-Glaser Critical Thinking Appraisal* (WGCTA) -Watson y Glaser, 1984-. Este test es el que se ha utilizado más ampliamente. Sin embargo, existen fuertes críticas en relación a su validez para ser usado en la actualidad (Hau *et.al.*, 2006). Por ejemplo, se ha señalado que presenta sesgos lingüísticos y verbales, ya que correlaciona fuertemente con pruebas estandarizadas de lectura. El formato no permite, que los que realizan el test elaboren su propio pensamiento y análisis, dado que carece de preguntas de respuesta abierta (Halpern, 2006; Hau *et.al.*, 2006 y).

Otros dos instrumentos también muy utilizados son el *California Critical Thinking Skills Test* (CCTST) de Facione (1990), que tiene dos formas paralelas (A y B) y se construyó tomando en consideración el informe Delphi; el segundo, el *Cornell Critical Thinking Test* (CCTT) -Ennis; Millman; Tornko, 1985-, que posee dos niveles, uno para estudiantes de enseñanza media y otro para estudiantes de enseñanza superior. Las críticas que han recibido estos dos instrumentos se refieren a que se pone demasiado énfasis sobre la inteligencia verbal, y que poseen bajos valores en su consistencia interna y validez de constructo (Halpern, 2006; Hau *et al.*, 2006). Este tipo de formato tiene como consecuencias las tres limitaciones fundamentales de la mayoría de las pruebas publicadas, esto es, las que afectan a la validez, a la producción y a su complejidad (Saiz; Rivas, 2008b).

Los tres instrumentos mencionados anteriormente, se presentan en formato de respuesta múltiple, lo que ha originado dificultades acerca de su validez como indicadores de las habilidades de PC, debido a que no se da la oportunidad a los examinados para que

expresen su propio criterio. Una alternativa a estos instrumentos es el test *Ennis-Weir Critical Thinking Essay Test* (Ennis y Weir, 1985), el cual está constituido por 14 sub-escalas en las que se requiere que los estudiantes lean un ensayo sobre un tema cotidiano, que contiene numerosos errores de razonamiento y, construyan su propia respuesta. La principal limitación para su uso se encuentra en el tiempo y costo implicado para calificar las evaluaciones de las respuestas abiertas y la necesidad de tener evaluadores expertos para ello. Este test evalúa bien la capacidad de argumentación. Pero necesitamos medir todas las habilidades de pensamiento, no solamente una.

En este sentido, comprobamos que el pensamiento no es fácil de evaluar mediante preguntas que se refieren a situaciones poco realistas, en las que además se pide al sujeto que elija entre un conjunto de alternativas, sin que pueda hacer aportaciones, reflexiones, justificaciones, inferencias, etc. más allá de lo presentado en el material. Govier (1987) señalaba que no es fácil producir PC mediante pruebas con alternativas de respuesta múltiple, en los que este pensamiento se trata a través de una serie de temas sencillos y delimitados, que la mayoría de las veces suelen ser artificiales, expresados en frases breves y lo más neutras posibles para que no sean susceptibles de distintas interpretaciones. Esto implica que hay muchos aspectos del pensamiento crítico que no pueden ser evaluados mediante este formato. Sugiere que este pensamiento debería ser evaluado a través de preguntas abiertas, argumentos, o entrevistas. Pero en este tipo de formatos el coste es mucho mayor. Ennis (2003), en la misma línea, proponía que posiblemente el mejor formato para la evaluación del pensamiento crítico sería aquel en el que el sujeto tuviera que elegir entre una serie de alternativas de respuesta junto con la justificación de esta.

Como hemos expuesto anteriormente, la mayoría de las pruebas estandarizadas publicadas presentan tres problemas: 1) que no se sabe bien lo que miden, 2) que no activan las habilidades fundamentales y, 3) que los problemas son artificiales y alejados del funcionamiento cotidiano de las personas. En esta línea, y como ya hemos mencionado antes, el test HCTAES de Halpern intenta subsanar algunos de los problemas señalados anteriormente. Este instrumento fue innovador en el campo ya que presenta unas características diferentes a las existentes hasta el momento. En primer lugar, utiliza situaciones cotidianas y similares a las encontradas en la vida real, por lo que los ítems presentan una muy buena validez ecológica, al ser representativos de lo que puede encontrarse en nuestro funcionamiento diario. En segundo lugar, utiliza un doble formato de

pregunta. Se plantea una situación o problema sobre la que se formula una pregunta abierta y, después, se pide que elija la mejor alternativa que resuelva la situación o problema planteado. Por lo tanto, cada uno de los 25 escenarios que constituyen el HCTAES es evaluado mediante una pregunta de formato abierto y una pregunta de formato cerrado. Según Halpern (2006) este doble formato de preguntas permite conocer, por un lado, si el que responde al HCTAES manifiesta un uso espontáneo de la habilidad y, por otro, si es capaz de usarla cuando se le señala que es necesaria para esa situación. La autora recomienda una medición del pensamiento crítico, basada en habilidades claramente definibles, las cuales serían: 1) comprobación de hipótesis, 2) razonamiento verbal, 3) análisis de argumentos, 4) probabilidad e incertidumbre, y 5) toma de decisiones y resolución de problemas.

Las investigaciones de nuestro grupo, como ya se ha mencionado, permitieron corroborar las virtudes del test de Halpern, principalmente porque las situaciones cotidianas planteadas a través de los ítems permiten recoger bien las habilidades fundamentales de pensamiento. Sin embargo, se identificó una limitación importante en los criterios de corrección, que afectaba directamente a la validez de la prueba. Por ello, como ya se ha expuesto, hemos construido el PENCRISAL (Rivas y Saiz, 2012, Saiz y Rivas, 2008b), con el fin de superar las limitaciones que presenta el HCTAES. Pasamos a describir los fundamentos de este instrumento.

Recordemos algunos aspectos de la necesidad de valorar el PC. Una de las razones para medir esta clase de procesos es estimar si una iniciativa de intervención es eficaz o no. Además de que nos sirva, como ya hemos señalado, para el diagnóstico de competencias. Desde hace algún tiempo, llevamos desarrollando un programa de intervención que intentamos mejorar (Saiz y Rivas, 2008a, 2011). Un programa es eficaz cuando: a) produce un cambio, b) este cambio perdura y, c) además se trasfiere o generaliza (Saiz, 2002). La permanencia del cambio sólo impone una medida después de transcurrido un tiempo. La eficacia exige emplear una prueba que cumpla con las condiciones anteriormente establecidas en nuestros análisis. Y esto es lo que hemos desarrollado, un test que corrija las deficiencias que otros no han podido resolver. A esta prueba la hemos denominado PENCRISAL (**P**ensamiento, **C**rítico, **S**alamanca). Finalmente, la transferencia es la medida de eficacia más importante, porque evalúa capacidad de generalizar habilidades a diferentes dominios, una vez que esas habilidades se han adquirido o desarrollado, pues no podemos transferir lo que no existe. En nuestra prueba, se emplean diferentes dominios para medir este índice (para un

estudio en profundidad, ver Rivas y Saiz, 2012; Saiz y Rivas, 2008b).

La evaluación de la eficacia del PC debe ser *cuantitativa* fundamentalmente, por varias razones que iremos exponiendo a continuación. Nuestra forma de entender la evaluación, como ya se ha dicho, incorpora *pruebas de rendimiento específicas y pruebas estandarizadas*. Aquí nos vamos a ocupar de las segundas, y el PENCRISAL es un ejemplo de estas. Las razones para utilizar pruebas estandarizadas para evaluar el PC, además de las de rendimiento, son sobradamente conocidas. Recordemos las principales. Su uso nos permite *evaluar el constructo de PC*, nos proporciona un indicador fiable de validez, esto es, que estamos midiendo lo que deseamos y no otra cosa. Esta es la razón fundamental, la que realmente importa, el disponer de una *medida fiable y válida* de pensamiento. Hay otras razones que deben tenerse en cuenta. Una de estas es que nos proporciona *indicadores de diferencias individuales*, entre personas y poblaciones. La tercera razón es que una medida de esta clase nos permite *compartir el conocimiento* con otros investigadores, al emplear medidas semejantes. Una penúltima razón es que nos posibilita el *comparar estudios* de diferentes investigaciones y poder aumentar nuestro conocimiento y capacidad de explicación. Y finalmente, con un instrumento de esta clase, podemos evaluar las competencias del PC y programas de instrucción.

Por todo lo expuesto, la construcción de nuestras pruebas surge, por una parte, de la necesidad diagnóstica en educación y de la de verificar la eficacia del programa de instrucción que se lleva desarrollando hace algún tiempo (Saiz y Rivas, 2008a, 2011) y, por otra, de la urgencia de disponer de pruebas de medida del PC, realmente válidas. Así pues, para superar los obstáculos anteriormente mencionados, y a la vez evaluar la eficacia de nuestra instrucción en su doble vertiente, magnitud del efecto y transferencia, desarrollamos este instrumento específico, el PENCRISAL.

Con este instrumento podemos valorar si la instrucción ha producido los cambios esperados, o lo que es lo mismo, si hay una mejora del PCy, si se utiliza de un modo generalizado, es decir, si se produce la transferencia de esas habilidades a otros dominios.

La prueba PENCRISAL, como ya se ha dicho, tiene su origen en el HCTAES. Este instrumento es realmente innovador, de modo que hemos mantenido parte de sus principios en nuestras pruebas y hemos modificado otros que no resultan muy apropiados (ver Saiz y Rivas, 2008b). Por ello, para la construcción de esta prueba nos hemos guiado por las conclusiones derivadas de nuestro análisis previo, y que nos llevan a seguir cuatro principios

fundamentales: 1) *la utilización de ítems que sean problemas de situaciones cotidianas*, 2) *el uso de diferentes temáticas de conocimiento o dominios*, 3) *un formato de respuesta abierto*, y 4) *el empleo de problemas con respuestas únicas*. El primero mantiene una de las aportaciones más importantes del HCTAES. El segundo y el cuarto son la aportación de nuestra prueba, en especial, el uso de problemas de única respuesta. El formato abierto de las pruebas es la forma de permitir que los procesos de pensamiento se puedan expresar. Finalmente, el uso de diferentes dominios nos permite valorar el grado de generalización de las habilidades, otro de los aspectos fundamentales, la transferencia.

Los ítems formados por situaciones cotidianas o situaciones problema intentan evaluar el PC en la vida diaria y hacer que la tarea sea más interesante, motive más. Al usar *problemas cotidianos*, intentamos simular el comportamiento diario de las personas, de modo que podamos apreciar hasta qué punto las habilidades son de uso generalizado en nuestras tareas del día a día. Por otra parte, afrontar problemas que simulen la realidad le añade un plus de motivación importante a la hora de realizar la prueba. Se busca, con las situaciones cotidianas, conseguir la mayor proximidad entre el contexto de la evaluación y el funcionamiento natural y cotidiano de las personas.

Un segundo rasgo de nuestra prueba es el uso de *diferentes dominios* (se emplean diferentes contextos como salud, deportes, educación, ocio...). El propósito de usar diferentes dominios en la intervención y en la evaluación es doble, por una parte, acortar la distancia entre lo formal y lo aplicado y, por otra, poder medir la transferencia. Recordemos que esta hace referencia a la capacidad de generalización de las habilidades aprendidas a cualquier otro contexto.

La tercera característica de nuestras pruebas es el uso de un *formato de respuesta abierta*. Este rasgo tiene más importancia de lo que pueda parecer. Un formato cerrado es solo una prueba de *comprensión* del problema que se nos plantea con el fin de discriminar entre las opciones que se nos proponen. Tan solo tenemos que reconocer o identificar la opción que consideramos correcta, pero solo señalar, sin más explicación. No tenemos que justificar por qué elegimos esa opción y no otra, y esto nos impide averiguar qué procesos de pensamiento se están empleando en cada situación. El formato cerrado ofrece muy poca información, y lo que es más problemático, nos imposibilita para indagar qué mecanismo se pone en marcha con cada situación-problema. En cambio, un formato de respuesta abierto obliga a mostrar los pasos seguidos para resolver el problema, en realidad hace que la prueba

sea de *producción*, no de comprensión. Obliga a generar o producir respuestas que justifiquen la forma de resolver cada situación. Cada problema cotidiano que planteamos activa un proceso de pensamiento determinado. Y esto nos permite saber lo que más importa, cómo se resuelve ese problema, esto es, con qué proceso de pensamiento concreto. De este modo, podemos disponer de la información que importa, a saber, la forma de pensar en cada problema o situación. Esto, no es posible obtenerlo con un formato de respuesta cerrado. Cuando se da una justificación de la respuesta podemos ver fácilmente si está resolviendo el problema usando el mecanismo que pretendemos medir. Además de decir si es correcto lo que se afirma en la situación problema, debe explicarlo. Al usar el formato de respuesta abierta podemos saber si realmente estamos midiendo lo que pretendemos medir y no otra cosa. Nuestros mecanismos de pensamiento son demasiados complejos como para captarlos con una simple marca en una casilla.

El último rasgo de nuestra prueba consiste en plantear situaciones-problema que solo se puedan responder de una manera, es decir, diseñar ítems que sean *problemas de respuesta única*. Recordemos, que esta propuesta es la que permite resolver el problema serio de validez de la mayoría de las pruebas, al conseguir que frente a una tarea-problema se activen procesos concretos que llevan a respuestas concretas. El test predecesor de nuestra prueba, el HCTA, consta de situaciones cotidianas que se pueden responder de diferentes formas, no plantea un problema para cada proceso. Esto plantea un problema semejante al generado por las respuestas cerradas, en las que no sabemos qué está haciendo mentalmente quien responde el problema. Para solucionar esta grave limitación, en nuestra prueba hemos incorporado la metodología del *análisis de tareas* (ver Saiz y Rivas, 2008b). Con este procedimiento, conseguimos fijar qué proceso se usa para cada situación problema, de modo que quien lo resuelve bien, sabemos con certeza cómo lo ha hecho, qué operaciones mentales a empleado. La idea es diseñar problemas que sólo se puedan resolver de una manera, con el fin de identificar el mecanismo de pensamiento correspondiente. Esta técnica nos proporciona ítems altamente informativos sobre los procesos y facilita la cuantificación de los ítems, al permitir solo una forma de respuesta. En realidad, con esta técnica, al puntuar la prueba, conseguimos una precisión muy próxima a la de las respuestas cerradas.

Para más detalle sobre las características de la prueba el lector puede consultar: Rivas y Saiz (2010, 2012), y Saiz y Rivas (2008b).



## Resultados y estado actual de la investigación

La descripción anterior de nuestra propuesta de evaluación ofrece las mejoras que hemos planteado como solución a los problemas de medida que poseen muchas de las pruebas estandarizadas publicadas. A través de nuestro método podemos tener la certeza de impulsar los mecanismos de pensamientos que deseamos, de modo que cuando evaluemos esta actividad podamos estar seguros de lo que medimos. Esta garantía arroja luz sobre las deficiencias más importantes en el terreno de la valoración de los procesos de pensamiento. Para poder abordar la complejidad de las habilidades de esta clase debemos ser capaces de cuantificar cada una de ellas por separado. Debemos medir destrezas concretas, con el fin de tener una visión más precisa de PC. Nuestro modo de evaluar contribuye mucho a esta clarificación conceptual.

Como se ha dicho anteriormente, el PC es un proceso multicomponencial (Saiz y Rivas, 2008a). Un modo de evaluar como el propuesto aquí, hace que esos componentes del pensamiento sean adecuadamente cuantificados. Los factores que componen las pruebas de medida, desarrolladas por nosotros, dejan a las claras esas diferentes facetas de lo que es reflexionar. Como también queda razonablemente establecido que esos componentes no son independientes unos de otros. Podemos hacer que se activen unos y no otros, pero en un contexto amplio, como es el de nuestra actividad diaria, todos los procesos interactúan entre sí. Así, por ejemplo, en el ámbito del razonamiento práctico podemos encontrar cualquier forma posible de inferencia. Los procedimientos que debemos seguir para tomar buenas decisiones requieren así mismo de nuestros procesos de razonamiento, que nos permiten lograr con mayor eficacia los resultados deseados. Por último, cuando solucionamos un problema, también estamos haciendo uso de nuestros procesos de razonamiento y de nuestra capacidad de decidir bien. Para alcanzar una solución adecuada frente a un problema, a veces, debemos razonar, otras decidir. Nuestro método de evaluación de estas destrezas es capaz de aislar cualquiera de los procesos fundamentales de pensamiento.

Hasta ahora nos hemos ocupado de justificar nuestro sistema de evaluación en PC, en contextos cotidianos. Las dificultades que acompañan a una tarea tan compleja como esta, son enormes, por lo que cualquier propuesta debe recorrer un camino de mucho tiempo hasta que se perfila como estable. Nosotros llevamos trabajando ya tiempo en este sistema de evaluación. De momento, hemos conseguido disponer de un instrumento fiable y válido

para medir PC, de manera estandarizada. El trabajo es mucho y los recursos pocos, por esta razón, vamos avanzando con más parsimonia de la deseada. Lo importante, para nosotros, es comprobar, a partir de los resultados que vamos obteniendo, que nuestras propuestas van por el buen camino, aunque lo recorramos con lentitud. Creemos que hemos tomado una buena dirección en nuestro trabajo por mejorar y constatar el desarrollo del PC.

### **Bibliografía**

Campirán, A. **Habilidades de pensamiento crítico y creativo. Toma de decisiones y resolución de problemas. Lecturas y ejercicios para el nivel universitario.** Xalapa: Universidad Veracruzana, 2016.

Covington, M. V., Crutchfield, R. S., Davies, L. y Olton, R. M. **The productive thinking program: A course of learning to think.** Columbus, OH: Merrill, 1974.

Dumont, H., Istance, D. y Benavides, F. (Eds.). **The nature of learning. Using research to inspire practice.** <https://doi.org/10.1787/9789264086487-en>, 2010.

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

Ennis, R., Millman, J. y Tornko, T. **Manual, Cornell critical thinking test.** Pacific Grove, CA: Midwest Publications, 1985.

Ennis, R. H. y Weir, E. **The Ennis-Weir Critical Thinking Essay Test.** Pacific Grove: Midwest Publications, 1985

Facione, P. A. **The California Critical Thinking Skills Test (CCTST): Forms A and B; and the CCTST test manual.** Millbrae. CA: The California Academic Press, 1990.

Feuerstein, R. **Instrumental enrichment: An intervention program for cognitive modifiability.** Baltimore: University Park Press, 1980.

Govier, T. **Problems in argument analysis and evaluation.** Dordrecht, Holland: Foris, 1987.

Halpern, D.F. **Halpern Critical Thinking Assessment Using Everyday Situations: Background and scoring standards (2º Report).** Unpublished manuscript. Claremont, CA: Claremont McKenna College, 2006.

Halpern, D.F. **Halpern Critical Thinking Assessment. SCHUHFRIED (Vienna Test System).** <http://www.schuhfried.com/vienna-test-system-vts/alltests-from-a-z/test/hcta-halpern-critical-thinking-assessment-1/>, 2010.

Halpern, D.F. **Halpern critical thinking assessment manual.** Austria: Schuhfried, 2012.

Halpern, D.F. **Thought and knowledge. An introduction to critical thinking (5th ed.)**. New York: Psychology Press, 2014.

Hau, K., Halpern, D.F., Marin-Burkhart, L, Ho, I., Ku, K., Chan, N. y Miu-Chi Lun, V. **Chinese and United States student's critical thinking: cross-cultural construct validation of a critical thinking assessment**. En American Educational Research Association Annual Meeting, San Francisco, EE. UU, 2006.

Lipman, M. Philosophy for Children. **Metaphilosophy**, 1976, 7 (1), p. 17-33.

Lipman, M. **Thinking in education (2nd Edition)**. Cambridge, MA: Cambridge University Press, 2003.

Maclure, S. Introducción: Panorama general. In S. Maclure y P. Davies (Eds.), **Aprender a pensar, pensar en aprender**. Barcelona: Gedisa, 1994, p. 11-32.

Maclure, S. y Davies, P. (Eds.). **Aprender a pensar, pensar en aprender**. Barcelona: Gedisa, 1994.

Marina, J. A. y Rambaud, J. **Biografía de la humanidad: Historia de la evolución de las culturas**. Barcelona: Ariel, 2018.

Moravec, J. W. (Ed.). **Knowmad society**. Minneapolis: Education Futures LLC, 2013.

Nickerson, R. S., Perkins, D. N. y Smith, E. E. **Enseñar a pensar. Aspectos de la aptitud intelectual**. Madrid: Paidós/M.E.C., 1987.

Nieto, A.M. y Saiz, C. Evaluation of Halpern's "Structural Component" for Improving Critical Thinking. **The Spanish Journal of Psychology**, 2008, 11 (1), 266-274.

Nieto, A.M., Saiz, C. y Orgaz, B. 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**, 2009, 14 (1), p. 1-15.

Perkins, D. N. y Grotzer, T. A. Teaching intelligence. **American Psychologist**, 1997, 52, p. 1125-1133.

Rivas, S.F. Morales, P. y Saiz, C. Propiedades psicométricas de la adaptación peruana de la prueba de pensamiento crítico PENCRISAL. **Avaliação Psicológica**, 2014, 13 (2), pág. 257-268.

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

Rivas, S. F. y Saiz, C. Validación y propiedades psicométricas de la prueba de pensamiento crítico PENCRISAL. **Revista Electrónica de Metodología Aplicada**, 2012, 17, p. 18-34.

Saiz, C. Enseñar o aprender a pensar. **Escritos de Psicología**, 2002, 6, p. 53-72.

Saiz, C. **Pensamiento crítico y cambio**. Madrid: Pirámide, 2017.

Saiz, C. **Pensamiento crítico y eficacia (2ª ed.)**. Madrid: Pirámide, 2020.

Saiz, C. y Rivas, S.F. Intervenir para transferir en pensamiento crítico. **Praxis**, 2008a, 10 (13), p. 129-149.

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

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

Saiz, C. y Rivas, S.F. Pensamiento crítico y bienestar como prevención del abandono de los estudios. En Bernardo, A.B., Tuero, E., Almeida, L.S. y Nuñez, J.C. (Eds.). **Motivos y factores explicativos del abandono de los estudios: claves y estrategias para superarlo** (p. 201-219). Madrid: Pirámide, 2020

Saiz, C., Rivas, S.F. y Almeida, L.S. Los cambios necesarios en la enseñanza superior que seguro mejorarían la calidad de la educación. **E-Psi. Revista Eletrónica de Psicologia, Educação e Saúde**, 2020, 9 (1), 9-26.

Segal, J. W., Chipman, S. F. y Glaser, R. (Eds.). **Thinking and learning skills**. Vol. 1: Relating instruction to research. New York: Erlbaum, 1985.

Sternberg, R. J. y Halpern, D. F. (Eds.). **Critical thinking in psychology. (Second ed.)**. New York: Cambridge, 2020.

Watson, G. y Glaser, E. M. **The Watson-Glaser Critical Thinking Appraisal**. San Antonio, TX: The Psychological Corporation, 1984.

---

**Revisão gramatical realizada pelos próprios autores.**

**RECEBIDO 22 DE AGOSTO DE 2020.**

**APROVADO 13 DE NOVEMBRO DE 2020.**