

EVALUACIÓN EN PENSAMIENTO CRÍTICO: UNA PROPUESTA PARA DIFERENCIAR FORMAS DE PENSAR

Carlos Saiz
Universidad de Salamanca
csaiz@usal.es
Silvia F. Rivas
U. de Salamanca e IE Universidad Segovia
silvia.fernandez@ie.edu

Resumen: En este trabajo, estudiaremos las formas de evaluación del pensamiento crítico (PC) que existen en la actualidad. El interés por medir las competencias intelectuales proviene de la necesidad de comprobar la eficacia de las iniciativas de intervención que se aplican. Es importante saber si un programa de instrucción consigue mejorar las habilidades de pensamiento que se propone. Pero la evaluación del PC se enfrenta a una dificultad previa, que es delimitar qué se entiende por ello. Hay muchas formas de definir lo que es el PC, y según entendamos lo que es, así debemos proceder en su evaluación. El modo de valorar las destrezas intelectuales depende mucho de cómo las entendamos. Las diferentes iniciativas de evaluación que se han desarrollado, la mayoría, han mostrado problemas graves de validez, que han puesto en tela de juicio su viabilidad. El objetivo de este trabajo, precisamente, es proponer una forma de evaluación del PC que solucione los problemas más importantes que poseen la mayoría de esas iniciativas. En nuestro estudio, vamos a proceder del siguiente modo: primero, vamos a justificar la importancia de la evaluación del PC, después trataremos los problemas existentes en las diferentes formas de evaluación, para pasar después a sopesar la viabilidad de una medida de esta naturaleza; finalmente, expondremos nuestra propuesta de evaluación.

Términos clave: pensamiento crítico, evaluación, habilidades de pensamiento.

1. La necesidad de valorar la buena reflexión

¿Por qué evaluar nuestra capacidad intelectual?, o ¿para qué valorar nuestras aptitudes? Cuando se trata de estas habilidades, parece que el interés mayor siempre proviene del ámbito de la educación. Un objetivo importante de la enseñanza es evaluar sus resultados. En lo referente al desarrollo de las habilidades de pensamiento, también es necesario medir de qué modo una forma de enseñanza funciona o no. El beneficio estaría en ver si el rendimiento de las personas mejora después de recibir un curso

para tal fin, en comparación al momento anterior a recibirlo. El propósito sería saber si una intervención educativa es eficaz. La clase de instrucción que nos interesa a nosotros es la dirigida al desarrollo del pensamiento crítico.

La necesidad más inmediata para evaluar las habilidades de pensamiento surge de saber si una intervención, un programa de enseñar a pensar, funciona o no. Si se instruye para mejorar determinadas habilidades, es imprescindible saber si esta enseñanza tiene algún efecto. Para ver esta influencia debemos comparar el rendimiento después de la instrucción con el obtenido antes de la misma. Si dicho rendimiento es mejor después de la intervención que antes de ella, estamos en el camino adecuado para mostrar que, probablemente, esa ganancia posterior se deba a nuestras enseñanzas. El origen de la mayoría de los proyectos de evaluación del pensamiento es: demostrar la eficacia de una intervención. Pero ello no es o, al menos, no debería ser la única necesidad de una medida del pensamiento.

El pensamiento crítico está cobrando especial interés en la enseñanza superior en países como Estados Unidos. Allí, se ha creado hace poco una comisión de expertos para estudiar el “futuro de la educación superior” y esta comisión ha sugerido el uso generalizado, en el ámbito federal, de pruebas de pensamiento crítico en los estudiantes universitarios (Ennis, 2008). El objetivo de esta iniciativa se recogía en el informe titulado: *A Test of Leadership*. El deseo era valorar la competencia de los futuros dirigentes de la sociedad (U.S. Department of Education, 2006). Pero ¿cómo realizar esto? Parece razonable pensar que los ciudadanos con responsabilidades en el futuro en la sociedad deberían poseer un suficiente desarrollo de sus dotes de argumentación o de tomar buenas decisiones, en definitiva, deberían manifestar un buen rendimiento en las habilidades que definen al pensamiento crítico (PC, en adelante). La preocupación social para que los dirigentes de la misma sean personas capaces de tomar buenas decisiones o de resolver problemas parece perfectamente razonable. De este modo, podemos observar que el

empeño por medir la capacidad de reflexión crítica no es exclusivo de los estudiosos de las “aulas”, sino una preocupación social seria.

Como lo que se mueve en Estados Unidos influye más de lo que uno se espera, esa preocupación se está extendiendo a muchos países de su área de influencia, de modo que se están tomando iniciativas para evaluar las competencias intelectuales de los estudiantes universitarios en otros lugares. Una de las propuestas más generalizadas es la de evaluar las habilidades fundamentales de pensamiento, esto es, lo que se entiende por PC, como capacidad de argumentar, de plantear hipótesis, de emitir juicios de probabilidad, de decidir o resolver bien problemas complejos. El interés que muestran los responsables de la enseñanza superior en varios países radica en asegurarse que la enseñanza universitaria, aparte de ofrecer una buena formación para una profesión, dispense un adecuado desarrollo del PC. Desde esta perspectiva, los tests de PC serían una herramienta de diagnóstico y de pronóstico. En el primer caso, nos permitirían saber si un sistema educativo logra lo que se espera de él, formar buenos pensadores. En el caso de no lograrse este objetivo, se pueden tomar medidas para que el sistema lo consiga en el futuro. La evaluación, pues, sirve para mejorar nuestro sistema educativo.

En cuanto a la función de pronóstico, las pruebas nos posibilitan saber quién posee ciertas capacidades necesarias, para asumir determinadas responsabilidades en la sociedad. El buen juicio y la capacidad de decidir de manera razonada parecen muy convenientes, cuando se tiene el compromiso de resolver problemas que afectan a muchos ciudadanos.

Parece, pues, necesario fomentar el buen pensar en los ciudadanos en general y en nuestros estudiantes en particular. Este deseo obliga a valorar esas aptitudes deseables, para saber si se poseen de manera razonable o para ver qué programa educativo las desarrolla suficientemente. En definitiva, el deseo es conocer qué sistema educativo, institucional o no, funciona, qué programa de intervención fomenta las habilidades que forman parte de lo que se considera reflexión crítica. Éste parece ser el punto de partida

de la evaluación del PC, la necesidad de saber en qué medida un colectivo posee estas capacidades.

2. El modo de valorar el pensamiento eficaz

Ahora bien, saber hasta qué punto alguien piensa de manera crítica exige al menos dos requisitos: un concepto claro de lo que es PC y un instrumento que lo mida. La realidad es bien diferente, porque ni tenemos una idea precisa de lo que es PC ni una herramienta válida. La pregunta que se deriva de esta situación es, pues, si resulta viable una evaluación de esta naturaleza en tales circunstancias. Dejemos su contestación para el siguiente apartado. Ahora ocupémonos de relativizar el problema planteado, esto es, la falta de claridad conceptual. Y decimos problema, porque no puede haber un instrumento que mida aquello que no podemos definir con precisión, dado que no sabemos exactamente en qué consiste. Una prueba se crea para medir algo que sabemos lo que es, ahí descansa su propiedad más importante, la validez, saber lo que mide. Dedicuémonos pues, antes de nada, a clarificar lo que queremos medir. Nuestro esfuerzo va a dirigirse más a exponer lo que entendemos por pensamiento crítico y menos a discutir las diferentes ideas que existen al respecto (una buena discusión al respecto se puede encontrar en Johnson, 2008).

Como siempre sucede, el grado de precisión conceptual es el reflejo de nuestro desconocimiento. Cuando se dice, por ejemplo, que: “El pensamiento crítico es la actividad intelectual que nos permite conseguir nuestros fines de la manera más eficaz” (Saiz y Nieto, 2002, p. 16), no estamos detallando la actividad intelectual, los procesos propiamente dichos. En realidad, indicamos que no sabemos muy bien en qué consiste el pensamiento. La definición anterior sirve con modelo, como idea ampliamente utilizada en el campo. Hagamos un intento por precisar más el concepto y digamos que el pensamiento crítico *es un proceso de búsqueda de conocimiento, a través de las habilidades de razonamiento, solución de problemas y toma de decisiones, que nos permite lograr, con la mayor eficacia, los resultados deseados.* En esta

segunda definición, hemos especificado la actividad intelectual con una finalidad intrínseca a todo proceso mental, a saber, buscar conocimiento. Conseguir nuestros fines no depende sólo de la dimensión intelectual, podemos necesitar nuestra actividad motora o perceptiva, de modo que aporta poco afirmar que el PC permite conseguir nuestros objetivos, puesto que también lo podemos conseguir moviendo nuestras piernas, por ejemplo. Es importante esforzarnos por identificar los procesos mentales responsables del pensar y no otras cosas.

El pensamiento si con algo tiene que ver, más que con otras cosas, es con derivar algo de algo, inferir, razonar. El proceso fundamental consiste en extraer información nueva, en buscar saber más. En definitiva, se trata de un mecanismo esencial de adquisición de conocimiento. Entendemos que lo fundamental de este proceso de aprendizaje es inferir o razonar. El razonamiento, pues, constituye lo nuclear del pensamiento. Pero no sólo el razonamiento. Y aquí ya debemos referirnos a la parte de la definición para conseguir nuestros fines. Buscamos algo cuando no lo tenemos o, teniéndolo, lo queremos en mayor cantidad o calidad. Lo que plantea esta carencia es una situación de “solución de problemas”.

Normalmente, pensamos para resolver nuestras dificultades. Ésta es la segunda actividad importante de pensamiento. Un problema se puede resolver razonando, pero también planificando cursos de acción o seleccionando la estrategia más adecuada para la situación. Así, además de razonar, debemos tomar decisiones para resolver los problemas. Elegir es una de las actividades más frecuentes e importantes que realizamos al solucionar un problema. Por esta razón, preferimos darle el protagonismo que se merecen en una definición de pensamiento. Resolver problemas exige mucha actividad intelectual, como razonar, decidir, planificar... Esta última característica va más allá de los mecanismos propios de inferencia. Lo que se observa a la hora de acotar en qué consiste pensar de manera eficaz es que se aglutinan conceptos que van más allá de lo nuclear, de lo que tiene que ver con inferir o razonar.

Por esta razón, nos parece oportuno ahora hablar de los componentes del PC. Observemos a continuación la figura 1, inspirada en la descripción que hace Halpern (2003) de un modelo de pensamiento crítico:

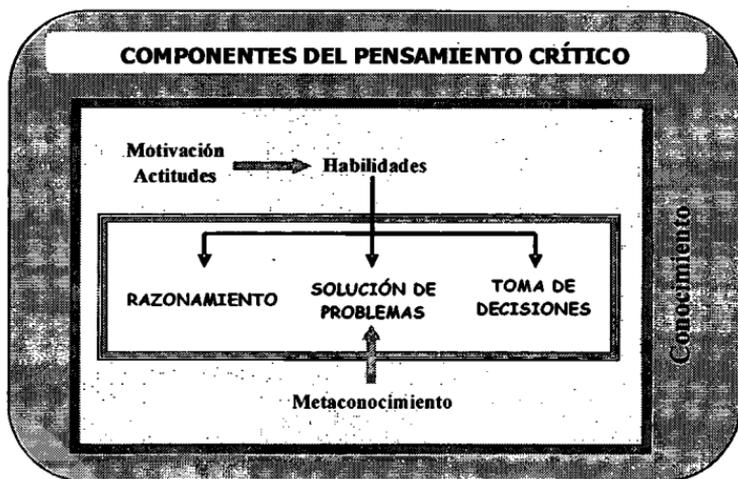


Figura 1: Componentes del pensamiento crítico

En ella, podemos encontrar todos los conceptos de la definición anterior menos dos: motivación y metaconocimiento (las actitudes suelen entenderse como disposiciones, inclinaciones... algo próximo a los motivos, pero también al metaconocimiento). El núcleo fundamental del PC sigue siendo lo que tiene que ver con las habilidades, en nuestro caso, razonamiento, solución de problemas y toma de decisiones. Pero ¿por qué incorporar conceptos de otra naturaleza, como la motivación en lo que es una descripción del pensamiento crítico? Hace ya algunos años se ha comprobado que, a la hora de hablar de PC, el hecho de centrarse sólo en las habilidades no permite recoger toda la complejidad del mismo. El propósito del esquema de la figura 1 es aportar claridad conceptual al adjetivo "crítico" en la expresión "pensamiento crítico".

Hay quien puede considerar innecesario este término, por entender que es redundante, ya que el carácter crítico es

consustancial a la esencia misma del pensamiento, forma parte integrante de ella. Sin distraernos por estos matices, lo que sugerimos es que si entendemos que *crítico* se refiere a *eficaz*, debemos apreciar también que esto no se logra sólo con las habilidades. Es necesario que se unan a ellas otros protagonistas y lo hagan en momentos distintos. Las capacidades intelectuales solas no consiguen la eficacia que se asume en crítico. Primero, para que se pongan en marcha (para que pensemos) debemos quererlo (“el saber comienza por el querer”, dicen). La motivación pues entra en juego antes que las habilidades, las pone en funcionamiento. Por su parte, el metaconocimiento nos permite una dirección, organización y planificación de nuestras habilidades de manera rentable, y actúa una vez que las capacidades han empezado a funcionar. La motivación, pues, activa nuestras destrezas y el metaconocimiento consigue que sean más eficaces. El final siempre debería ser un deseable conocimiento de la realidad, una mayor sabiduría.

Nos parece que el hecho de referirnos a componentes del pensamiento crítico, diferenciando a su vez habilidades de motivación y metaconocimiento, ayuda en la clarificación conceptual que buscamos. Por un lado, especificamos de qué destrezas hablamos y, por otro, qué otros componentes (diferentes del *pensar*) se relacionan, incluso se solapan, con ellas. Debemos ser conscientes de lo vana que resulta la ilusión de encontrar procesos mentales “puros”. Planificar un curso de acción, un rasgo esencial del metaconocimiento, exige reflexionar, pronosticar, elegir, comparar y evaluar... ¿Acaso esto no es pensar? Los diferentes niveles o dimensiones de nuestra actividad mental deben relacionarse, integrarse. Nuestra indagación, creemos, saldrá así mejor parada. Por esto, nuestro esfuerzo de clarificación conceptual va dirigido a lograr esa integración de componentes del pensamiento. Nuestra apuesta es poder identificar lo que es sustancial en el pensar para saber qué es lo que podemos evaluar.

Hay poca duda de que, para nuestro funcionamiento personal y profesional, pensar bien es rentable. Cuando deseamos conseguir algo o cambiar una situación, se pone en marcha (o debería) toda

nuestra mente. Percibimos la situación, recordamos aspectos relevantes para el problema, analizamos toda la información disponible junto con la conocida, valoramos todo lo analizado, esto es, emitimos juicios acerca de los asuntos más relevantes, decidimos sobre las opciones o vías de solución, ejecutamos el plan, obtenemos unos resultados, los evaluamos, estimamos si hemos alcanzado nuestro propósito y según sea el grado de satisfacción posterior a esta estimación, damos por bueno o no nuestro curso de acción. La pregunta que debemos hacernos frente a esta descripción es ¿dónde está el proceso de *pensar*? ¿En todo ello o en una parte?

Nuestra respuesta es sencilla: el *pensar* lo encontramos una vez que tenemos sobre el tablero, de la partida que constituye la vida, toda la información que necesitamos; cuando comenzamos a analizarla, a partir de ese momento ya podemos continuar diciendo que “seguimos” pensando. Para concretar, en aras de la claridad, inferir después del análisis, decidir, planificar, usar estrategias... Todo esto sería *pensar*. Como mencionamos antes, razonar, decidir y resolver problemas son procesos que consideramos como habilidades fundamentales de pensamiento crítico.

Hemos definido qué entendemos por PC y hemos especificado lo que es para nosotros (razonar, decidir y resolver). Nuestra concepción puede ser discutible, pero no imprecisa; puede ser más o menos acertada, pero no vaga ni falta de concreción. Nuestro empeño por precisar qué es el PC viene impuesto por el *modo de valorarlo*. No podemos evaluar desde la inconcreción. Saber *qué medimos* es lo que legitima una prueba diseñada para tal fin. A nadie se le escapa que evaluar cualquier actividad mental es una empresa difícil y frustrante, y más una conducta tan multicomponencial como es pensar. Una vez que sabemos lo que debemos evaluar todo parece más sencillo, pero sigue sin serlo. Sólo hay que recurrir a algunas fuentes para observar lo dicho. No es nuestro propósito aquí efectuar ninguna revisión sobre el particular, aunque, obviamente, recurriremos a las más interesantes. En un libro editado por Daniel Fasko, hay varios capítulos sobre evaluación del pensamiento crítico que muestran a

las claras la difícil empresa de medir lo mental (ver Ennis, 2003; Halpern, 2003b).

Desde los trabajos pioneros de Donders (1868/1969) hasta los más recientes, (ver, por ejemplo Possin, 2008), el investigador siempre ha tenido la tentación de cuantificar, incluso lo más escurridizo, como es la actividad mental. Para ello, como nosotros ya hemos señalado, se ha indagado en la dirección del *cómo* y el *qué*: qué medir y cómo hacerlo. El *qué* ya lo hemos expuesto, nos falta el *cómo* medir. Quizás, el autor que más sepa sobre evaluación del pensamiento crítico sea Robert H. Ennis. Es quien más tests ha publicado y quien ha realizado mejores estudios sobre el particular (Ennis, 2003, 2008). Ennis (2003), por ejemplo, nos ofrece hasta siete razones para medir el PC, y nos presenta y valora todas las pruebas publicadas. De las veintiuna que cita, cuatro (un veinte por ciento) fueron elaboradas por él y colaboradores. Con tanta oferta de medida podríamos imaginar que el modo en el que se evalúa el PC hoy es acertado. Nada más lejos de la realidad. La mayoría de las pruebas fracasan por caer en la tentación de la trampa de la objetividad. Casi todas consisten en preguntas, excesivamente formales, a las que hay que responder marcando una opción de las tres o cuatro propuestas. Esta clase de tests tienen una virtud a la que es difícil resistirse: la rapidez en la corrección. Pero una prueba con un formato de elección múltiple y un contenido artificial difícilmente pueden medir habilidades fundamentales de pensamiento. Un conjunto de ítems frente a los que sólo hay que señalar el que nos parece correcto exige poco pensamiento y mucha percepción; obliga a discriminar, pero no a justificar; impone elegir, pero no valorar. En definitiva, impide que se exprese lo que es consustancial a la naturaleza del pensamiento. Esta forma de proceder tiene como consecuencia las tres limitaciones esenciales de la mayoría de las pruebas publicadas, a saber: las que afectan a la *validez*, a la *producción* y a la *complejidad* de las mismas.

Para medir la capacidad de razonamiento –por señalar la más significativa del PC– de la persona que se enfrenta a una prueba,

ésta debe hacer explícito su proceso mental para que, a partir de sus manifestaciones pueda inferirse el modo como está razonando. Si frente a un problema sólo es necesario decidir si la solución que se nos ofrece es correcta o no, o bien elegir una de las respuestas que se presentan ya formuladas, nunca se podrá saber qué reflexión se ha realizado para responder, lo más que se puede hacer es especular sobre ello. Ante un silogismo del tipo: “Todo hombre es mortal, Juan es hombre, luego Juan es mortal”, cuando sólo tenemos que decir si es correcto o no, no hay modo de deducir de qué forma se llega a esa respuesta. Varios estudios muestran que guiándose por los cuantificadores, se pueden desarrollar estrategias que garantizan hasta un ochenta por ciento de eficacia sin conocimiento alguno de lógica. El *efecto de atmósfera* es un truco que permite dicha eficacia (Anderson, 1990). Quien responde a estos problemas sigue reglas prácticas sencillas como ver qué cuantificador se repite en las premisas y en la conclusión. De este modo, se pueden conseguir un número elevado de respuestas correctas, acertando sin pensar bien el problema. La imposibilidad de conocer el proceso de pensamiento seguido por quien da la respuesta es quizás el problema más serio que presentan la mayoría de las pruebas estandarizadas existentes. Y lo es porque les mina la validez.

La segunda limitación grave de los tests publicados tiene que ver con el tipo de respuesta que se pide. Si utilizamos un formato de elección múltiple, además del problema de validez señalado, nos encontramos con otro importante: sólo se pide *comprensión* de la tarea, pero no *producción* de pensamiento. Evaluar el pensamiento consiste en valorar lo que se produce, no lo que se supone que puede haberse producido. Con una prueba de alternativas, asumimos que se genera pensamiento, pero no tenemos constancia de ello. No podemos cuantificar el proceso, porque no se ha hecho explícito. Paradójicamente, se mide algo que no se ha expresado. Esto es, a nuestro juicio, otro problema serio. En un test de esta clase se debe pedir que se exprese *por qué* se responde de una manera determinada, se debe exigir justificación de todas y cada una de las respuestas que se pide que se den. El problema, se

aduce, es el tiempo de corrección. La réplica es muy sencilla ¿de qué sirve un test que se corrige mecánicamente pero que no sabemos si mide lo que queremos?

La última limitación de la que nos vamos a ocupar se refiere a la de la *complejidad*. El ejemplo de antes “Todo hombre es mortal, Juan es...” ilustra bien este problema. Una tarea trivial, excesivamente formal, artificial, desmotiva y no activa los mecanismos de pensamiento habituales que empleamos en nuestra vida diaria. Y si esto sucede, estamos valorando unas respuestas a un test que refleja poca habilidad, porque lo que puede estar sucediendo es que la prueba no active convenientemente los pertinentes procesos de pensamiento. La mayoría de los tests publicados padecen también esta tercera dolencia. Para verificar en qué medida se reflexiona frente a una tarea, es importante proponer problemas semejantes a los cotidianos, es recomendable “simular” las dificultades diarias en el test, es conveniente proponer situaciones problema como los que abordamos habitualmente. Los problemas de los test estandarizados no plantean situaciones cotidianas que haya que resolver. Nuevamente, nos encontramos con una deficiencia importante, que impide valorar realmente lo que se pretende. Si quien responde a un test no se siente en absoluto atraído por los ejercicios que se le planteen, responderá sin implicarse intelectualmente, sin poner en marcha la maquinaria de reflexión que deseamos captar. Si no se dedica a pensar con determinación, nunca podremos evaluarlo.

El modo en el que la mayoría de los tests evalúan el pensamiento crítico es inadecuado. Salvo algunas excepciones que mencionaremos después, los tests con formato de elección múltiple no pueden justificar conceptualmente lo que se pretende medir, porque el tipo de respuesta que se demanda lo impide. Podemos jugar, y mucho, con la estadística y el análisis factorial y observar agrupamientos más o menos cercanos a los planteamientos de partida, pero nunca podremos decir que esos “factores”, esos agrupamientos, se producen porque se ha puesto en marcha el proceso de pensamiento adecuado. No lo podemos afirmar, sencillamente porque el tipo de respuestas no pueden mostrarlo.

El pedir a quien se enfrenta a una prueba estandarizada que “marque” lo que le parece bueno, impide que realmente piense sobre el problema planteado. Si el formato de prueba limita lo que se desea medir, parece razonable afirmar que algo se está haciendo mal. No se puede pretender medir pensamiento y al mismo tiempo imposibilitar, o limitar, que se piense. Finalmente, utilizar en las pruebas problemas estereotipados o demasiado “escolares” invita al distanciamiento de la tarea por parte del que responde. Esta lejanía de los problemas propuestos no estimula precisamente a la reflexión profunda, que se supone que es lo que se desea. Decíamos antes que hay algunos tests que son la excepción, porque no poseen las limitaciones que acabamos de señalar. El problema de estos instrumentos es que miden sólo un aspecto del pensamiento crítico. Por ejemplo, el *The Ennis-Weir Critical Thinking Essay Test* (Ennis & Weir, 1985), evalúa bien, pero sólo capacidad de argumentación. Hay otras pruebas unidimensionales que se pueden consultar en Ennis (2003). Ahora bien, lo que necesitamos es medir todas las habilidades de pensamiento crítico, no una solamente.

3. Viabilidad de la evaluación en el pensamiento crítico

La mayoría de las pruebas estandarizadas que hemos valorado, con sus deficiencias importantes, dibujan un panorama bastante pesimista sobre la posibilidad de evaluar el PC. La viabilidad de un proyecto de medida de esta clase parece encontrar problemas tan serios, de tanto calado, que puede disuadir al más optimista de proponer una empresa de esta índole. Sí es cierto que la evaluación de procesos complejos, como los de PC, plantea retos desalentadores, incluso para grupos de investigación sobrados de recursos materiales y humanos. Sin embargo, las dificultades mayores no suelen radicar en la falta de medios, sino en la carencia de ideas. Cuando alguien abre la puerta de la solución con la estrategia adecuada, desaparecen de inmediato muchas dificultades que hace nada parecían insalvables. Un proyecto de medida, desarrollado por Halpern (2006), ha conseguido precisamente esto, abrir la puerta a muchos de los callejones sin salida de la

evaluación del pensamiento crítico. La prueba HCTAES (*Halpern Critical Thinking Assessment Using Everyday Situations*), en nuestro parecer, marca un *antes y un después* en la evaluación del PC. Su aportación no consiste tanto en la calidad de la prueba, como en el modelo de evaluación que se propone en ella. La prueba ha sido adaptada al castellano por nosotros, y estudiada con suficiente detenimiento como para poder afirmar que debe mejorarse en aspectos fundamentales (Nieto, Saiz y Orgaz, 2009, en prensa). Son precisamente estas deficiencias del test lo que nos ha llevado a desarrollar instrumentos alternativos, que describiremos más adelante. Antes de esto, lo que perseguimos es estudiar si es posible evaluar el pensamiento crítico con cierta eficacia. A lo que estamos obligados antes de nada es a demostrar la *viabilidad* de un proyecto de valoración de las habilidades de pensamiento.

Tal como hemos justificado anteriormente, existen al menos tres problemas graves en la mayoría de las pruebas estandarizadas que se han publicado: el primero, que no se sabe bien lo que miden, el segundo, que no activan las habilidades fundamentales objeto de la medida y, finalmente, que los problemas son artificiales y alejados de la actividad cotidiana de la gente. Pues bien, el proyecto HCTAES intenta salvar estos escollos, aunque el resultado no lo haya acompañado. Su aportación no es tanto el producto final, que se puede mejorar, como la novedad de ofrecer una propuesta decisiva que, ahora sí, hace viable la valoración del PC. Este es su enorme mérito. Y esto es lo que nos ha guiado a nosotros en nuestro proyecto de evaluación. Comenzaremos por describir cómo se resuelven los problemas de validez, producción y complejidad en esta prueba, analizaremos después qué es lo que consigue medir, para terminar exponiendo cómo intenta resolver nuestro proyecto aquello en lo que fallan. En nuestra evaluación asumimos el modelo de medida del HCTAES, pero corrigiendo sus limitaciones e incorporando una propuesta nueva que, a nuestro juicio, resuelve el problema más grave de los tres, la validez. Para solventar esta limitación fundamental de cualquier medida, planteamos un diseño

de ítems basado en un *análisis de tarea*. Pero, para entender esta propuesta, es necesario adentrarse antes en otros terrenos.

Tomemos un ítem del test HCTAES, para ayudarnos en el análisis que vamos a realizar a continuación.

Ítem 21, parte 1 del Test HCTAES

Supón que eres un estudiante de primer curso (año) en una escuela de Odontología (dentista). Te das cuenta de que tu nueva amiga, que también es estudiante de primer curso en la misma escuela, se emborracha varias veces por semana. Tú no observas ninguna señal de su problema con la bebida en la escuela, pero te afecta porque los dos empezáis a ver pacientes en la clínica dental de la escuela dentro de un mes. Ella no ha respondido a tus insinuaciones acerca de su problema con la bebida. Por lo que tú sabes, nadie más se ha enterado de que bebe.

- (A) Expresa el problema de dos formas distintas.
- (B) Para cada aspecto del problema, propón dos posibles soluciones.

Este ítem del test ilustra muy bien las bondades del planteamiento de la prueba, como veremos a continuación. La prueba tiene dos partes: una abierta (como la del ejemplo) y otra cerrada, en la que la misma situación se vuelve a presentar con varias opciones, de las que hay que elegir una. Lo importante del planteamiento de la prueba se encuentra en la parte abierta, por lo que vamos a ocuparnos de ella (para un conocimiento del instrumento ver Halpern, 2006; y Nieto, Saiz, y Orgaz, 2009 en prensa). Esta herramienta es original en casi todos sus aspectos, tanto en la naturaleza de los ítems como en la forma de corregirlos. En las 25 situaciones de las que consta se plantean problemas de las habilidades fundamentales de pensamiento; en concreto, se formulan situaciones de: 1) comprobación de hipótesis, 2) razonamiento verbal, 3) argumentación, 4) probabilidad e incertidumbre, y 5) toma de decisiones y solución de problemas. La situación 21 sería del último tipo, decisión y resolución. Los ítems

de la prueba son situaciones cotidianas, que se responden, al menos una vez, de forma abierta, y proponen problemas que se deben resolver mediante esas cinco habilidades fundamentales de pensamiento. Si se plantea un problema que se debe resolver por un mecanismo de pensamiento concreto, se asegura una validez robusta de la prueba. Si, además, se pide que se explique la respuesta que se da a cada ítem, se obliga a producir pensamiento, y finalmente, si se plantean problemas que son situaciones cotidianas, se consigue una proximidad y *complejidad* primordial con las tareas y problemas que se abordan a diario en la vida. Sirva como ilustración la situación 21 de antes. En ella, debemos responder empleando alguna estrategia de solución de problemas, además tenemos que explicar la solución que proponemos, y finalmente estamos resolviendo un problema cotidiano, común. Esta prueba es muy ingeniosa porque de un modo sencillo aborda los tres problemas fundamentales a la vez. Estudiemos uno por uno.

Una situación que plantea un problema cotidiano consigue que a la persona que cumplimenta la prueba le resulte familiar lo que se le propone; en consecuencia, se enfrentará a ello como lo hace normalmente, pensando o reflexionando como acostumbra a hacerlo en su entorno. Esto posee al menos dos ventajas claras, una, que la persona se implique en la tarea y otra, que emplee las destrezas que mejor conoce. Esta clase de tareas propuestas motivan más que otras, por ser semejantes a lo que resolvemos en nuestra vida diaria, en cuanto a *complejidad* e importancia. No se plantean problemas artificiales, ajenos a los intereses de la gente, ni tan simples, que puedan distraer al sujeto buscando modos de resolverlo que no tienen por costumbre. Con estos ítems, al menos conseguimos una mayor motivación y actividad intelectual por parte de quien los resuelve. Hay una importante implicación en la tarea. Sobre esto, tanto la autora como nosotros disponemos de datos de las aplicaciones del test que lo avalan. Es importante, creemos, emplear tareas que resulten familiares, que permitan a las personas desenvolverse con comodidad e interés. La validez ecológica de una prueba es importante. Pero también lo es la *generalización o transferencia* de las destrezas. Este concepto no se ha

mencionado hasta ahora porque es patrimonio, tanto de la intervención como de la evaluación. Y aquí nos ocupamos sobre todo de lo segundo. Sin embargo, ahora es el momento de considerarlo.

Si uno plantea un problema muy diferente a los cotidianos en la evaluación, corremos el riesgo de debilitar la validez de la prueba. Si en un ítem buscamos que la persona utilice una habilidad determinada, debemos facilitar al máximo que emplee esa habilidad y no otra. Pero si el ítem propone un problema muy alejado de los que normalmente abordamos, se plantean dos posibilidades al resolverlo: una, que se resuelva sin que se disponga de la destreza que se necesita, y dos, que teniéndola no se sepa aplicar, por las diferencias que hay entre las situaciones cotidianas y las de la tarea. Esta segunda posibilidad, si se da, impide que una prueba mida lo que pretende, sencillamente porque no activa los mecanismos de pensamiento que desea captar y para lo que fue diseñada. Debe entenderse que la validez ecológica refuerza la validez teórica, también llamada de constructo. No es sólo una cuestión de que la prueba resulte más entretenida para quien la realiza, que también, sino que además consiga poner en funcionamiento los procesos mentales que quiere medir. Dificilmente mediremos lo que no aparece. Este aspecto de la transferencia es mucho más importante de lo que aquí podemos explicar, pero puede encontrarse un tratamiento extenso del mismo en Saiz y Rivas (2008). La primera gran virtud de la prueba de Halpern es que acierta de lleno en plantear problemas próximos a la gente. Si analizamos la situación 21, observaremos que es una especie de “simulación” de problemas comunes en nuestra vida diaria. Se describe un acontecimiento para el que nos piden que nos pongamos en una situación y que propongamos soluciones. Como no podemos crear problemas reales, los simulamos y pedimos soluciones. Permitásenos una comparación. Se procede de forma parecida aquí al igual que con los futuros pilotos; estos antes de subirles a un avión, han hecho muchas horas de simulador. Éste no es un avión, pero se le parece. Algo semejante hacemos en la evaluación. No medimos cómo se resuelven problemas reales

(tememos que no nos permitirían entrar en sus vidas, obviamente), sino problemas simulados. Las situaciones cotidianas son tareas-problema que estimulan todo lo anteriormente señalado y esta aportación es tan importante que es recogida en nuestra propuesta, la cual será descrita más adelante.

En la situación 21 se pide:

- (A) Expresa el problema de dos formas distintas.
- (B) Para cada aspecto del problema, propón dos posibles soluciones.

En esta prueba se utiliza un formato de respuesta abierta, obviamente vinculada al enunciado de los ítems, como se recoge en la situación 21. Quien resuelve el test, es necesario que explique lo que hace. En el ítem 21, se pide que se defina el problema y que se ofrezcan soluciones al mismo. Se solicita que se piense sobre la situación, que se produzca “pensamiento”. Las situaciones-problema obligan a reflexionar y, especialmente importante para la evaluación, a expresar esa reflexión. El test de Halpern opta por utilizar dos formatos de respuesta, uno abierto, en la parte 1, y otro cerrado, en la parte 2. Sobre formato abierto debemos decir que es necesario para la evaluación del PC, ya que no es posible medir unas habilidades tan complejas de otro modo. Un formato de respuesta cerrado (de elección múltiple, por ejemplo) impide que se exprese el pensamiento, lo que incapacita para su medida. Este aspecto tiene poco que discutir: sólo podemos medir lo que aparece. El test HCTAES es especialmente sensible a este problema, a pesar del coste alto en tiempo, que tiene la corrección de respuestas abiertas. Los tests estandarizados, por el contrario, sacrifican validez al ahorro de tiempo. Esto, sencillamente, es un sinsentido. Una ventaja adicional, y no menor, ofrecen los formatos de respuesta abiertos: cuando una tarea es de producción, además de lo ya apuntado, obliga a un pensamiento más completo que cuando es sólo de comprensión. Un test de respuestas cerradas tan sólo mide comprensión. En un test de producción, sin embargo, debemos elaborar lo que hemos pensado para que sea inteligible antes de expresarlo. Necesitamos repensar lo pensado para expresarlo. Todos sabemos que no es lo mismo tener una idea que

comunicarla. Cuando hacemos esto, en muchos casos comprobamos que esa idea no estaba tan clara. Al exteriorizarla, aparecen sombras que no detectamos cuando está en nuestra mente, al abrigo del análisis. Este rasgo de la tarea (producción vs. comprensión) es importante en otros ámbitos, porque hace que, según sea de un tipo u otro, puedan aparecer o no determinados sesgos o distorsiones en nuestro pensamiento. Por citar sólo un dato y no distraernos de nuestro análisis principal, el efecto de atmósfera mencionado antes sólo aparece con tareas de comprensión y nunca de producción (Shaw & Johnson-Laird, 1998). Esta digresión tiene como fin apuntalar más si cabe la importancia de permitir que en una tarea de evaluación se exprese lo que se piensa. Es esencial, desde todos los puntos de vista. De nuevo, la prueba de Halpern cumple bien con este requisito.

Dos de los tres problemas señalados en el análisis anterior son resueltos bien por el test de Halpern. El problema más central, el de la validez, no lo es. Veamos por qué. La prueba dice que mide los cinco factores señalados antes (comprobación de hipótesis, razonamiento verbal, argumentación, probabilidad e incertidumbre, toma de decisiones y solución de problemas). El ítem 21 es un ejemplo del último factor. La situación recoge un problema cotidiano que se adecua bien al factor propuesto, pero ¿cómo se corrige ese ítem? En el protocolo de respuesta se establece lo siguiente:

PROTOCOLO DE RESPUESTA	
6 puntos son posibles: 1 por cada definición del problema y 1 por cada solución. Para conseguir los 6 puntos, se tiene que haber definido el problema de forma distinta y las soluciones también han de ser distintas.	
Forma 1:	
1 punto: Ej: Ella tiene un problema con el alcohol que se manifestará en su relación con los pacientes.	Ej: Tiene un problema con el alcohol
Solución 1: 1 punto: Debe reducir gradualmente su consumo de alcohol.	
Solución 2: 1 punto: No debe relacionarse con los pacientes	
Forma 2:	
1 punto: Ej: Ella tiene un problema con el alcohol pero no da señales de que esté afectando a su habilidad.	Ej: Los pacientes podrían sufrir algún tipo de daño debido a su problema.
Solución 1: 1 punto: Demostrarle cómo su problema podría afectar a su habilidad. Alejarla de la clínica.	
Solución 2: 1 punto: Convencerla de que está poniendo en peligro a la gente.	

Si leemos con detenimiento el protocolo de respuesta, observaremos que se están puntuando expresiones redundantes del problema, como en los ejemplos. Dar un punto por decir que tiene problemas con el alcohol y que afecta a sus pacientes se nos antoja demasiado generoso, sobre todo cuando estamos valorando pensamiento, no percepción. Y lo mismo podemos decir de las soluciones. El sugerir que debe reducir el consumo de alcohol y darle un punto sigue la misma tónica. ¿Qué estamos valorando? ¿Definir bien un problema y solucionarlo con eficacia? Nada de lo que se propone como puntuable se adecua a esto. Este problema se encuentra en la mayoría de las 25 situaciones del test. La prueba ofrece tareas que pueden recoger bien las habilidades fundamentales de pensamiento, pero fracasa al valorar formas de respuesta que, en muchos casos, nada tienen que ver con esas destrezas. Contempla respuestas alejadas de lo que se quiere medir y no se proponen respuestas correctas únicas. Al no ofrecer maneras exclusivas de solución, se deja que se diluya la validez de la prueba.

Govier (1987) ya recogía esta limitación de los test de pensamiento al indicar que las respuestas correctas y las respuestas de quien contesta la prueba deben ser las mismas en un alto porcentaje para poder afirmar que se posee la capacidad de pensamiento esgrimida. Si no diseñamos un problema que exija una única respuesta, nuestro empeño de medir una habilidad determinada fracasa rotundamente. La limitación más importante de la prueba de referencia descansa en la forma inadecuada de corrección, aunque no sólo en ella, pues muchas de las situaciones de la misma exigen una especificación más cuidada de los problemas. A nuestro juicio, la limitación fundamental de este test es el que no haya diseñado las situaciones cotidianas para que se puedan responder de una única manera. Esto tiene consecuencias graves para la validez de la prueba que ahora pasamos a demostrar, a la vez que ofrecemos una propuesta de solución.

Analicemos este asunto de un modo sencillo. Se desea medir, por ejemplo, razonamiento causal. Construyamos para ello un problema de esta naturaleza que se responda de una única forma

correcta posible, porque así, cuando la respuesta sea correcta, podemos saber que se está procediendo mentalmente de un modo imaginable para nosotros. Esa forma de proceder sería la que se necesita para solucionar problemas de causalidad y no sería posible resolverlos de otro modo. En consecuencia, cuando se resuelve ese problema, se espera que se describa ese procedimiento, porque no hay otro posible. Necesitamos situaciones cuyas estrategias de solución se excluyan entre sí, de modo que la que empleemos para responder al primero no pueda ser la misma que para el segundo. De este modo, podremos estar seguros de que el ítem de causalidad está midiendo únicamente causalidad. Solamente si logramos esta especificidad de las situaciones problema, conseguiremos medir lo que queremos y lograremos que nuestra prueba sea válida. La prueba de Halpern fracasa en los términos expuestos porque no realiza un análisis de las tareas que propone de manera correcta. Veamos esto con más detenimiento.

Como mencionábamos antes, siempre se ha tenido interés en medir los procesos mentales. Donders afirmaba hace tiempo algo muy interesante: “¿No debería el pensamiento tener una duración finita, y no sería posible determinar el tiempo requerido para formar un concepto o expresar los deseos propios?” (Donders, 1868/1969, p. 147). Este autor inaugura lo que después se conocerá como *cronometría mental*. Para nuestros propósitos de evaluar el pensamiento, lo que resulta especialmente relevante no es tanto la medida del tiempo de los procesos mentales, como el análisis del procedimiento previo que se debe realizar para poder medirlo. La metodología del *análisis de tarea* nace de este esfuerzo por anticipar qué “pasos mentales” se deben dar para poder realizar una determinada operación con éxito.

Con el fin de entender mejor esta técnica, resumamos las reflexiones de Donders. Comentemos la tarea que él creaba. El experimentador pronunciaba sílabas y el sujeto respondía con esas mismas sílabas. La tarea (a) consistía en pronunciar sólo una sílaba, que la otra persona debía repetir. En la tarea (b), se pronunciaba una sílaba de entre seis posibles y la otra persona la repetía. En la tarea (c), se fija una sílaba que es la única que se

ha de responder, mientras que las otras han de quedar sin respuesta. La realización correcta de cada una de las tres tareas exige, al menos, tres procesos: 1) codificación del estímulo, 2) decisión de respuesta, y 3) ejecución de la misma. Sin embargo, para la tarea (b) también es preciso discriminar entre estímulos y decidir qué respuesta emitir. En cambio para la tarea (c) sólo es necesario discriminar entre estímulos, pero no decidir qué responder. Como se puede apreciar, la tarea (a) no incorpora ni discriminación ni decisión de respuesta, procesos ambos que sí están presentes en la tarea (b), mientras que la tarea (c) tan solo exige discriminación. Lo que propone Donders es realmente ingenioso: utiliza tareas concretas para poner en marcha procesos concretos. De este modo, por comparación, se puede permitir evaluar procesos específicos de discriminación o de decisión motora.

Lo que interesa señalar aquí de esta técnica de análisis de tarea son las posibilidades que tiene para solucionar el problema de validez. La prueba de Halpern se queda a medio camino de resolver esta dificultad, porque no aplica este método. Como observábamos, las situaciones cotidianas pueden elicitar un único proceso de pensamiento pero, en la práctica, la prueba no lo consigue, porque en su diseño se ha olvidado establecer la respuesta que se debe dar. En todas las tareas de Donders, se fijan los estímulos y las respuestas que se deben presentar y emitir. Por el contrario, en la prueba de Halpern, se determinan sólo los primeros, esto es, las situaciones-problema. No se fijan las respuestas correctas para cada una de esas situaciones, de modo que no podremos saber cuál es la estrategia adecuada para responder bien y tampoco cómo corregir las respuestas de modo consistente.

Lo que nosotros planteamos es diseñar una prueba con situaciones cotidianas que definan problemas específicos para los que sólo haya una respuesta correcta. De este modo, al responder bien a cada tarea-problema, sabremos de manera inequívoca qué proceso mental se está empleando, de forma que podremos cuantificarlo de modo fiable. Con esta especificidad en el desencadenamiento de mecanismos de pensamiento concretos, se soluciona el grave problema de la validez, con lo que se puede

disponer de un instrumento que mide las habilidades de pensamiento que se desea medir, y no otra cosa.

Si nos retrotraemos al ejemplo de la situación 21 ya citada, para mejorar esa tarea, habría que diseñar el problema de manera más clara, con el propósito de que exista una única forma correcta de resolverlo. Sería necesario plantear una tarea que se aborde sólo de un modo y se resuelva también de manera singular. Esto permitiría cuantificar las respuestas como si fueran respuestas dicotómicas, correctas/incorrectas. Además, se resuelve otro problema no menor, el de la falta de fiabilidad de las respuestas abiertas, pues en ellas, la puntuación depende del criterio subjetivo del corrector. En nuestro caso, este problema no se suscita, porque sólo hay un modo de puntuar las respuestas, por muy abiertas que sean: se consideran como correctas las que resuelven el problema y se rechazan como incorrectas las que no lo resuelven. Pero esto se verá claramente al exponer nuestra propuesta.

4. Propuesta de evaluación de habilidades fundamentales de pensamiento

Nuestro estudio e investigación sobre la evaluación del pensamiento crítico nos ha permitido identificar algunos problemas de medida fundamentales y vislumbrar formas de solucionarlos. El empeño por emplear formatos de respuesta cerrados y tareas artificiales en la mayoría de los tests estandarizados publicados, los incapacita para medir aquello para lo que fueron diseñados. Si a una persona se le pide que resuelva problemas de razonamiento solamente marcando las opciones de respuesta que se le dan, se le impide generar o producir sus propias respuestas: se le induce a señalar respuestas ya hechas. Este tipo de pruebas no estimulan ni permiten expresar lo que pensamos, de este modo, imposibilitan su evaluación. Si las tareas que empleamos para dicha evaluación, además, están muy alejadas de las que resolvemos a diario, no resultarán un acicate para la reflexión, sino más bien una invitación a la pasividad y al distanciamiento. En definitiva, si no se ofrecen situaciones-problema que activen y exterioricen los procesos que se desean medir, cualquier proyecto de evaluación del pensamiento es un fracaso ya de partida.

El proyecto HCTAES es una propuesta que intenta solventar esas dificultades. Sus resultados no se han correspondido con la originalidad de su propuesta. Sin embargo, sus planteamientos han inspirado la nuestra. La prueba resuelve bien el problema de la artificialidad y de la producción. Ofrece tareas cercanas que implica a quien las realiza y le permite enfrentarse a ellas, exteriorizando las operaciones que se siguen para cumplimentarlas. Quien realiza la prueba tiene libertad para expresarse sobre problemas semejantes a los que resuelve a diario. Sin embargo, el test no cierra la manera de responder en una única dirección. Las situaciones-problema que se proponen no fijan una única forma de responder, porque no inducen hacia la utilización de un único proceso. Como hemos visto, si cada tarea estimula más de un proceso de pensamiento y más de una forma de responder, perdemos la oportunidad de saber qué proceso concreto lleva a una respuesta concreta. Bajo estas circunstancias, ya no es posible saber qué evaluamos, sencillamente, porque se puede pensar de varias formas frente a una tarea determinada. No hay ya posibilidad de medir lo que queremos, por tanto, no hay validez.

Un modo de subsanar este inconveniente es diseñar problemas únicos con una única respuesta correcta; en esto consiste nuestra propuesta. Aprovechamos todas las virtudes de la propuesta de Halpern y corregimos sus deficiencias. Para ello, incorporamos a su planteamiento un análisis de tareas. Las situaciones cotidianas del tipo de las presentadas en el test HCTAES se pueden rediseñar de tal manera que propongan problemas, que no admitan más que una respuesta. Siguiendo con los sabios planteamientos de los fundadores de la "medición mental", diseñemos tareas que podamos imaginar de antemano cómo se pueden resolver. Si para ejecutarlas hay que seguir el paso A, después el B para terminar con el C, por ejemplo, cuando la persona resuelve bien esa tarea podemos tener la certeza de que ha procedido así y no de otro modo. De este modo, al menos, podemos saber que, cuando se resuelve bien un problema de una prueba, es porque se han realizado esos procesos mentales A, B y C. Cuando medimos el rendimiento o la eficacia sabremos que lo hacemos sobre esos

mecanismos solamente. De esta manera, queda resuelto el gran problema de la validez.

Al analizar las características de una tarea, podemos imaginar las operaciones que se deben realizar para resolverla y esto nos garantiza saber qué estamos midiendo, pero además, nos permite proceder a la inversa. Con este método, no sólo proponemos un problema para el que podemos imaginar los pasos que se seguirían a la hora de resolverlo, sino que imaginamos qué pasos nos interesaría que se dieran y entonces diseñamos una situación para que se ejecuten los mismos. La ventaja de esta reversibilidad del análisis de tareas la encontramos a la hora de elaborar medidas de pensamiento, pues esta característica nos permite adecuar las tareas-problema a los procesos que deseamos medir. En nuestro caso, las habilidades fundamentales de pensamiento, ya señaladas al principio de este trabajo, consisten a su vez en otras más específicas. Estas habilidades de pensamiento más concretas son las que realmente evaluamos y las tareas se diseñan específicamente para medirlas. Una de las habilidades fundamentales es la de razonamiento, pero formas de razonar existen muchas, de modo que debemos seleccionar con buen criterio las que mejor representan esa capacidad y medirlas (Saiz y Rivas, 2008). Al proceder así, lo que en realidad hacemos es fijar la estructura de la prueba de evaluación, que será descrita más adelante. Ahora importa establecer cómo operativizamos dicha estructura, porque es otra de las aportaciones de nuestro trabajo: elegir procesos de pensamiento relevantes y crear tareas específicas para cada uno de ellos. Nuestra propuesta de evaluación garantiza la validez de las medidas por medio de este método de análisis.

La necesidad de valorar el pensamiento ya ha sido justificada al comienzo de nuestro estudio. Ahí señalábamos que una de las razones para medir esta clase de procesos es estimar si una iniciativa de intervención es eficaz o no. Desde hace algún tiempo, llevamos desarrollando un programa de intervención que intentamos mejorar (Nieto y Saiz, 2008; Saiz y Rivas, 2008). 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 vamos a desarrollar, un test que corrija las deficiencias que otros no han podido resolver. A esta prueba la vamos a denominar **PENCRISAL** (Pensamiento, Crítico, Salamanca). 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. Por este motivo es la más importante, porque presupone cambio previo en nuestros procesos de pensamiento. Esta es la segunda prueba que vamos a construir. La vamos a llamar **PENTRASAL** (Pensamiento, Transferencia, Salamanca).

Antes de describir nuestras dos pruebas, conviene aclarar lo que se entiende aquí por transferencia. El planteamiento de evaluación que nosotros defendemos consiste en emplear situaciones cotidianas que garanticen la validez teórica y ecológica de la prueba. Cuando alguien se enfrenta a una prueba de evaluación del pensamiento, se busca que utilice las habilidades que emplea a diario, para ello se diseñan tareas próximas a las actividades corrientes. Este planteamiento es el que mantenemos en la prueba de evaluación del efecto, de modo que la pregunta es: ¿Cuál es la diferencia entre ambas pruebas? ¿Por qué una mide magnitud del efecto y la otra transferencia, cuando el diseño de los ítems es el mismo? La respuesta está en uno de los objetivos de la evaluación, a saber, medir la eficacia de un programa de intervención. Nuestro planteamiento de la intervención condiciona nuestra medida de transferencia (para una discusión más detallada, ver Saiz y Rivas, 2008). Aquí nos basta con aclarar lo siguiente. Una intervención que se apoya en tareas también cotidianas no puede buscar transferencia en esos contextos, pues son los que previamente ha trabajado. Vamos a medir la transferencia a dominios cotidianos, esto es, ámbitos que cubran todas las esferas de nuestra actividad diaria. Aquí radica la diferencia fundamental

entre ambas pruebas. En la de evaluación de la magnitud del efecto, las situaciones plantean problemas personales, mientras que en la medida de la transferencia, las tareas versan sobre problemas de diferentes contextos, como, de educación, de salud, laborales... Más adelante se detallará todo esto. La idea, pues, es ver si se es capaz de resolver problemas con la misma eficacia en todas las esferas de nuestra vida personal. Ésta sería nuestra medida de transferencia. Su fundamentación se encuentra en Saiz y Rivas (2008). Pasemos ya a exponer nuestra propuesta de evaluación. A continuación, se expondrá nuestro método de evaluación.

Como decíamos, la construcción de nuestras pruebas surge, por una parte, de la necesidad de verificar la eficacia del programa de instrucción que se lleva desarrollando hace algún tiempo (Nieto y Saiz, 2008; Saiz y Rivas, 2008), y por otra, de la urgencia de disponer de pruebas de medida del pensamiento crítico, realmente válidas. Por ello, 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, hemos desarrollado dos medidas específicas, como ya hemos apuntado antes: **PENCRISAL** y **PENTRASAL**.

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

Para la construcción de estas pruebas nos hemos guiado por las conclusiones derivadas de nuestro análisis previo, y que nos llevan a seguir cuatro principios fundamentales: 1) plantear ítems que sean problemas de situaciones cotidianas, 2) utilizar un formato de respuesta abierta, 3) proponer problemas de diferentes temáticas de conocimiento o dominios y 4) plantear problemas con respuestas únicas.

Uno de los aspectos centrales en nuestro planteamiento es el que se refiere al uso de situaciones-problema que sean lo más

semejantes posibles a las de la vida diaria. La utilización de estas situaciones, que recogen problemas reales, nos permite por una parte, evaluar nuestras destrezas de pensamiento en un contexto semejante al real y por otra, introduce un factor motivacional importante, consiguiendo un mayor interés por la tarea que se propone. Consideremos los siguientes ejemplos:

Ejemplo 1:

Ítem 16. Prueba de Habilidad en Pensamiento Crítico de California

"Si Alex quiere a alguien, él quiere a Bárbara. Hay muchas personas a las que Bárbara no quiere, y Alex es una de ellas. Pero todo el mundo quiere a alguien".

Ejemplo 2:

Situación 1. PENCRISAL

"El entrenador personal de Juan le ha dicho que sólo si entrena dos horas al día aprobará las pruebas del cuerpo de bomberos. Juan está preocupado porque, al no haber entrenado las dos horas requeridas, cree que no va a superar ese examen".

Un problema deductivo como el del ejemplo 1, presenta una estructura demasiado formal, lo que la aleja de las situaciones de nuestro quehacer diario donde pudiéramos necesitar conocer este tipo de principio. Sin embargo, empleando tareas como las del ejemplo 2, nos alejamos de los planteamientos más formales en pensamiento para acercarnos a contextos más ecológicos, que nos van a permitir una mayor aplicación de las capacidades a las situaciones cotidianas.

En este contexto, la pregunta que se puede plantear es: ¿Cuáles son los temas sobre los que versan los problemas? Es aquí donde debemos hacer una diferenciación entre las dos pruebas que presentamos, y que ya hemos justificado antes, pero que ahora concretamos en cuanto a su contenido. Los ítems del PENCRISAL

que, como ya hemos comentado, es la medida de la magnitud del efecto, fundamentalmente se refieren a un único ámbito de conocimiento: el de las *relaciones personales*, sin embargo, los del PENTRASAL, la medida de la transferencia, hacen referencia a todas las esferas que consideramos importantes en nuestro desenvolvimiento diario, como son: salud, deporte, ocio, tráfico, educación, medio ambiente, política, laboral y consumo.

Recordemos la justificación que hemos hecho de esta distinción, para mayor claridad, en relación al concepto de transferencia. Si la dificultad para generalizar nuestra capacidad de razonar o de solucionar problemas descansa en la diferencia que existe entre el dominio de aprendizaje y la vida cotidiana, y planteamos una intervención con situaciones cotidianas para reducir esta distancia, lo que se desprende de esto es una reconceptualización de la idea de transferencia. Recordemos que transferir es la capacidad de generalizar una habilidad de un dominio a otro. Ahora nos encontramos que el dominio es el mismo, lo que introducimos en esta prueba son los ámbitos de conocimiento de los problemas, de manera que si somos capaces de aplicar nuestras habilidades en diferentes esferas de conocimiento, entonces podemos seguir hablando de transferencia.

Otro de los aspectos relevantes en la construcción de estos instrumentos es el empleo de un formato de respuesta abierto, cuyas ventajas ya han sido discutidas en relación con el de respuesta cerrada. No obstante, recordemos algunas de ellas. Estos últimos, los formatos de respuesta cerrada, nos proporcionan información insuficiente sobre los procesos de pensamiento que la persona, supuestamente, tiene que poner en marcha para resolver el problema. Son respuestas que exigen reconocimiento de opciones, no producción de las mismas. Los ítems de nuestras pruebas, por el contrario, al presentar un formato de respuesta abierto, permiten que se produzca pensamiento y nos capacitan para conocer los mecanismos o procesos concretos del mismo. Los ítems de este tipo obligan a generar la respuesta y a explicarla, por lo que la persona no sólo debe identificar aquellas situaciones en las que es preciso activar procesos de pensamiento sino que también

debe decidir cuál de los procesos es necesario para resolver de forma más adecuada la situación. A continuación se incluyen unos ejemplos aclaratorios:

ÍTEMES DE RAZONAMIENTO CATEGÓRICO

Ejemplo 3: Formato respuesta cerrada

Test de Pensamiento Crítico de Cornell: Ítem 6

El señor García propone que admitamos a todos los extranjeros que quieran entrar a nuestro amado país. Pero los extranjeros siempre han ocasionado problemas y siempre los darán. La mayoría de ellos ni siquiera pueden hablar castellano. Como todo grupo que ocasiona problemas es malo, de esto se sigue que los extranjeros son un colectivo perjudicial.

Contesta al ítem de la forma siguiente:

Si la conclusión *se sigue necesariamente* de las afirmaciones, marca A

Si la conclusión *contradice* las afirmaciones expuestas, marca B

Si la conclusión *ni se sigue ni contradice* las afirmaciones, marca C

FORMULACIÓN		RESPUESTA
Todos los extranjeros (E) ocasionan problemas (P)	E → P	Opción A
Todo grupo que ocasiona problemas (P) es malo (M)	P → M	
Todos los extranjeros (E) son malos (M)	E → M	

Ejemplo 4: Formato respuesta abierta

PENCRISAL: Situación 5

La función de comprender envuelve el lenguaje, y sólo los hombres lo usan, de modo que podemos concluir que sólo ellos comprenden.

¿Es correcto el razonamiento? Explica el porqué.

FORMULACIÓN		RESPUESTA
Comprender (C) envuelve el lenguaje (L)	C → L	Razonamiento correcto
Sólo los hombres (H) usan el lenguaje (L)	L → H	<i>Explicación:</i> Principio transitividad
Los hombres (H) comprenden (C)	C → H	

En estos dos ítems, las situaciones están igualmente bien planteadas. Como se observa en las tablas, es posible extraer adecuadamente la estructura formal de ambos argumentos. Sin embargo lo que queremos resaltar es cómo, aun así, lo que aporta un sistema de formato cerrado, como es el caso del ítem del test Cornell, cuando tan solo tiene que identificar que la respuesta correcta es la "A", es una información muy opaca, ya que el hecho de que alguien haya sido capaz de reconocer esa opción como válida no nos ofrece información sobre los procesos de pensamiento que ha seguido. Sólo está poniendo en funcionamiento estrategias de reconocimiento e identificación. Sin embargo, en el ítem de respuesta abierta (ejemplo 4), por una parte, identifica adecuadamente que el argumento es correcto, y después, pone en funcionamiento los procesos de producción al tener que hacer explícita la explicación de la misma y es aquí donde, para extraer que se debe a un principio de transitividad, debe recurrir a procesos concretos de pensamiento.

El último de los cuatro principios básicos, empleado en nuestras pruebas, es el diseño de problemas con respuestas únicas. Recordemos, que esta propuesta es la que permite resolver el problema serio de validez de la mayoría de los tests, al conseguir que frente a una tarea-problema se activen procesos concretos que llevan a respuestas concretas. Elaboramos tareas que eliciten procesos de pensamiento específicos que llevan a responder siempre de la misma manera. Dentro de este esquema, cuando esa respuesta única aparece sabemos qué mecanismo concreto se ha desencadenado. De esta manera, nos aseguramos que estamos midiendo lo que pretendemos, y además simplificamos y aportamos rigor a la corrección del test. Quizás se comprenda mejor nuestro

planteamiento, si comparamos dos ítems, uno que no garantiza la validez y el otro que sí, uno es de la prueba de Halpern y el otro de la nuestra.

Veamos unos ejemplos concretos del HCTAES y del PENCRIAL:

ÍTEMES DE SOLUCIÓN DE PROBLEMAS

Ejemplo 5: HCTAES (Halpern): Ítem 24

Supón que estás cuidando del perro de tu vecino y una de las tareas que tienes que hacer es darle una pastilla voluminosa y aparentemente amarga. Se trata de un perro de presa grande que mordió a un niño el año pasado. ¿Cómo te las arreglarías para darle la medicina?

Propón dos buenas soluciones para este problema.

CORRECCIÓN

2 puntos: Si hace uso de métodos creativos.

Se puntúan con 2, respuestas del tipo: *"Pedir ayuda, consultar a un experto, mezclar la pastilla con algo, etc."*

Ejemplo 6: PENCRIAL: Situación 22

Eres propietaria de un bar familiar donde trabaja tu madre en la cocina y tu hija ayudando como camarera de forma ocasional. Ahora necesitas contratar a un nuevo trabajador para poder mantener el bar abierto ya que han ingresado a tu madre en el hospital y los médicos te han dicho que la estancia en él será prolongada. Además, tu hija ha empezado las clases en una universidad que está fuera de la ciudad. Tienes que hacer una selección entre varios candidatos, pero no tienes tiempo y no quieres pedirle a tu hija que te ayude en esto pues tendría que desplazarse. Una amiga te habla de su hija que está en paro y necesita trabajar, pero tienes tus dudas, ya que no tiene experiencia laboral en la hostelería.

¿Qué estrategias o pasos debería seguir ella para alcanzar una buena solución? Indica la solución resultante.

CORRECCIÓN

2 puntos: Si aplica bien un método.

Se puntúan con 2 aquellas respuestas que incluyan:

- *La identificación del problema:* madre e hija no atienden el bar y necesita alguien que trabaje en el bar.
- *Los subproblemas o submetas:*
 - ✓ Hacer una selección de candidatos para elegir al trabajador adecuado, que le llevará tiempo y tendrá posiblemente que cerrar el bar.
 - ✓ Contratar a la hija de su amiga provisionalmente mientras busca un candidato adecuado.

En el ejemplo 5, se asignan 2 puntos a soluciones que, según el criterio de la autora son métodos creativos de resolución, pero a nuestro modo de ver, si realmente estamos evaluando la capacidad de una persona para resolver problemas, no consideramos que “pedir ayuda” o “consultar a un experto” suponga que está empleando algún tipo de procedimiento, ni de solución de problemas, ni de pensamiento. Este tipo de respuestas más bien podríamos valorarlas como “irracionales”, puesto que no activan ningún proceso de pensamiento. Sin embargo, en el ejemplo 6, perteneciente a nuestra prueba, se puede apreciar una primera diferencia con respecto al anterior en cuanto a la pregunta que se plantea sobre la situación. Al decir: “¿Qué estrategias o pasos debería seguir ella para alcanzar una buena solución? Indica la solución resultante” estamos concretando tanto las exigencias de la tarea como el procedimiento mediante el cual se llega a esa solución.

Otra diferencia evidente es la delimitación tan precisa que se hace de los tipos de respuestas a los que vamos a asignar, en este caso, los 2 puntos. Como ya hemos dicho, las situaciones-problema

han sido diseñadas, siguiendo el método de análisis de tarea, mediante el cual los problemas se plantean de tal manera que podemos anticipar las operaciones que las personas van a necesitar aplicar para resolver la situación. Y esto es posible porque sólo existe, como hemos dicho ya, un camino para resolverlo.

Ahora, vamos a dedicar un espacio a describir, con más detalle, las características más relevantes de nuestras dos pruebas y lo hacemos de manera conjunta, puesto que parten de los mismos fundamentos y presentan una estructura semejante.

El inicio de la construcción de las pruebas consistió en confeccionar un banco de ítems, lo suficientemente amplio, como para poder hacer una buena selección, atendiendo al nivel de dificultad de los mismos y lo más heterogéneo posible para evaluar de modo preciso las diferentes habilidades. Asimismo, este conjunto de ítems nos garantiza, en la medida de lo posible, la validez de contenido, reflejando los principales aspectos que representan las habilidades de pensamiento crítico.

En la elección de las situaciones-problema se tuvieron en cuenta dos aspectos: primero, que estuvieran bien representados los procesos u operaciones de pensamiento que se querían evaluar, y segundo, que la redacción de las mismas fuera clara y empleara un lenguaje coloquial, alejado del empleo de tecnicismos.

La primera versión de las pruebas se aplicó a una muestra de 469 estudiantes universitarios de diferente procedencia, con el objetivo de realizar el correspondiente análisis psicométrico de los ítems. Del resultado de esta aplicación, se eliminaron algunos de ellos que han sido sustituidos por otros nuevos y se depuraron aquellos que presentaban ligeras deficiencias. Todo ello iba encaminado a conseguir una mejora de la calidad de los instrumentos. A partir de nuestros resultados, podemos inferir que las pruebas demuestran una fiabilidad suficiente y una validez ajustada a los planteamientos teóricos. A la hora de redactar estas líneas, se está empezando a realizar la aplicación empírica de esta segunda versión, con la confianza de que en ella se demostrarán psicométricamente las mejoras introducidas.

En este momento, el PENCRIAL y PENTRASAL constan de 35 ítems (situaciones-problema) de respuestas abiertas. Los enunciados de los problemas se han planteado de tal manera que no requieren que la respuesta se elabore y se enuncie en términos técnicos, sino que se pueden expresar en lenguaje cotidiano. En el ejemplo siguiente se ilustra esta característica:

Este ítem nos enfrenta a una tarea deductiva, concretamente un razonamiento proposicional. La formulación de este problema sería la que se expone a continuación:

Ejemplo 7: PENCRIAL: Situación 20

Juan necesita utilizar el transporte público todos los días para ir a trabajar y tarda aproximadamente unas dos horas. Estos últimos días, con la huelga de autobuses, ha habido problemas de tráfico, por lo que siempre ha llegado tarde. Hoy tiene una reunión muy importante y su jefe está intranquilo por si llegará a tiempo. Le pregunta a un compañero por Juan y éste le dice que no se preocupe que hoy no hay huelga, luego no tendrá problemas de tráfico, así que llegará a tiempo para la reunión.

Este ítem nos enfrenta a una tarea deductiva, concretamente un razonamiento proposicional. La formulación de este problema sería la que se expone a continuación:

H: hay huelga de autobuses; T: llegar tarde al trabajo

$$H \rightarrow T, \neg H \Rightarrow \neg T$$

Dicho argumento pone de manifiesto una falacia formal del razonamiento proposicional: la *Negación del Antecedente* (NA) que tiene semejanza con el argumento válido conocido como *Modus Tollens* o *Negación del Consecuente* (NC). Ahora bien, como hemos comentado anteriormente, no es necesario que la persona conteste al ítem en estos términos, sino que puede justificar su respuesta indicando, por ejemplo, que “el

razonamiento no es válido porque el hecho de que no haya huelga no implica que no pueda llegar tarde, pues pueden darse otras circunstancias para que se retrase”.

En cuanto a la estructura de las pruebas, los ítems de ambas se configuran en torno a cinco factores: Razonamiento Deductivo (RD), Razonamiento Inductivo (RI) y Razonamiento Práctico (RP), y Toma de Decisiones (TD) y Solución de Problemas (SP). De todas las manifestaciones posibles que incluyen estas habilidades, se han seleccionado las estructuras más representativas de cada una de ellas por su uso más frecuente en nuestro funcionamiento cotidiano. Cada uno de estos factores se compone de 7 ítems.

Entre las situaciones problema que evalúan el Razonamiento Práctico, cuatro miden *argumentación*, que quizás sea el más común de los modos de razonar (integra todas las demás formas de razonamiento) ya que en nuestro funcionamiento cotidiano son continuas las situaciones que nos exigen de valoraciones o producciones de buenos argumentos para defender puntos de vista, opiniones, posiciones, etc. Las otras tres situaciones evalúan la identificación de *falacias*: errores de razonamiento que o bien por la fuerza persuasiva del argumento empleado, o bien por la ambigüedad del lenguaje utilizado, en la mayoría de las ocasiones resultan difíciles de identificar, son engañosas y nos pueden inducir a cometer sesgos en nuestras valoraciones de los argumentos. Tienen un uso frecuente, por ejemplo en los medios de comunicación y en la política.

Los ítems que conforman el factor de Razonamiento Deductivo evalúan las formas de razonamiento más importantes: el *razonamiento proposicional* (cuatro ítems), y el *razonamiento categórico* (tres ítems). El razonamiento formal es menos frecuente que el práctico y el inductivo, pero se emplea en cierta medida.

El factor de Razonamiento Inductivo incluye: 1) el razonamiento causal, tres ítems; 2) el razonamiento analógico, dos ítems; 3) el razonamiento hipotético, un ítem; y 4) las generalizaciones inductivas, un ítem.

La escala de Toma de Decisiones evalúa el uso de procedimientos generales de decisión, lo cual exige elaborar juicios precisos de probabilidad y utilizar los heurísticos adecuados para adoptar decisiones sólidas. Aquí, se incluyen dos situaciones generales, en las que se tendrá que proceder de determinada manera, para alcanzar la decisión más acertada, y las otras cinco, que requieren de la utilización de los principales procedimientos que seguimos a la hora de optar por una decisión, a saber, los heurísticos.

Por último, los ítems de Solución de Problemas, al igual que los de Toma de Decisiones se dividen en problemas generales (cuatro ítems) y específicos (tres ítems) que son los que van a requerir la puesta en marcha de estrategias específicas de solución.

Tanto en el factor de Toma de Decisiones como en el de Solución de Problemas se promueve el desempeño de procedimientos generales de decisión y solución, con la finalidad de estimular el uso necesario de estrategias para la planificación de un problema. El metaconocimiento y la toma de conciencia de los procesos de pensamiento es donde se planifica la acción, se dirige, se organiza y se regula. Una vez más, en esto seguimos la propuesta de Halpern (1998, 2003). A continuación, presentamos una tabla que resume la descripción anterior sobre la configuración de los ítems de acuerdo a los cinco factores ya dichos:

Tabla 1: Factores del PENCRIASL y PENTRASAL

DEDUCCIÓN	INDUCCIÓN	PRÁCTICO	TOMA DE DECISIONES	SOLUCIÓN DE PROBLEMAS
RPR=4	RC=3	ARG=4	GRAL=2	GRAL=4
RCT=3	CH=1	FAL=3	PRB=2	RGL=2
TOT=7	RA=2	TOT=7	CI=1	MF=1
	GI=1		REP=1	TOT=7
	TOT=7		DIS=1	
			TOT=7	

El orden de presentación de los ítems ha sido aleatorizado, pero evitando que apareciesen de forma consecutiva aquellos que pertenecen a un mismo factor.

La forma de corrección se ha ideado de tal manera que resuelva la limitación encontrada en el sistema de respuesta del HCTAES que, debido a sus ambigüedades, llevaba en ocasiones a confusión. La dificultad se resuelve dotando a cada ítem de una solución única. De esta manera se ha simplificado el procedimiento de valoración de las respuestas pudiendo establecerse tres valores estándar:

- 0 puntos: cuando la respuesta dada como solución del problema es incorrecta;
- 1 punto: cuando solamente la solución es correcta, pero no se argumenta adecuadamente, lo que pone de manifiesto que sólo identifica y demuestra la comprensión de los conceptos fundamentales;
- 2 puntos: cuando, además de dar la respuesta correcta, se justifica o explica adecuadamente el porqué, haciendo, por tanto, uso de procesos más complejos que implican verdaderos mecanismos de producción.

De esta manera se está utilizando un sistema de escalamiento cuantitativo, cuyo rango de valores se sitúa entre 0 y 72 puntos como límite máximo, para la puntuación global de las pruebas y entre 0-14 para cada una de las cinco escalas.

Con el ejemplo de uno de los ítems del PENCRISAL que presentamos a continuación, ilustramos los criterios de corrección:

Ejemplo 8: PENTRASAL

Situación 33: Toma de decisiones: representatividad

En un estudio realizado en dos institutos de secundaria, uno con 2000 alumnos y otro de 200, se encuentra que las calificaciones medias de los chicos y chicas son muy semejantes. Además en ambos centros hay aproximadamente el mismo número de varones que de mujeres. Sin embargo, en una evaluación, el 60% de las chicas sacan mejores notas que los chicos ¿dónde es más probable que suceda este resultado?

La valoración de este ítem, a partir de respuestas reales, sería la siguiente:

0 puntos, si escribe: *"En ambos por igual, porque las probabilidades para ambos son las mismas independientemente del tamaño de las muestras"*. Puntuamos con cero, porque no indica la respuesta correcta, que sería en el hospital pequeño.

1 punto, para respuestas del tipo: *"Es más fácil en el instituto pequeño, porque habrá menos chicas"*. Obtiene solo un punto porque ofrece la respuesta correcta, pero la justificación es vaga.

2 puntos, para respuestas como: *"En el pequeño, el de 200, porque la probabilidad será más real en las muestras grandes que en las pequeñas"*. Se asignan dos puntos ya que además de dar la respuesta correcta, la justificación de la misma es adecuada.

En cuanto al tiempo de administración, nuestras pruebas se definen como tests psicométricos de potencia, es decir, sin limitación de tiempo. Pero la duración promedio estimada para su resolución completa por una persona de nivel intelectual normal, se estima entre 50 y 60 minutos. Esto arroja una nueva ventaja con respecto al test HCTAES de respuesta abierta (Halpern, 2003), en el que aproximadamente el tiempo de realización es el doble (120 minutos), llegando incluso a acercarse al tiempo de aplicación de pruebas de elección múltiple, como es el caso del Test de Pensamiento Crítico de Cornell (50 minutos), o de la Prueba de Habilidad en Pensamiento Crítico de California (PHPC) de Faccione (46 minutos).

La forma de administración de las pruebas puede realizarse en formato lápiz y papel o mediante ordenador, vía Internet. Nosotros hemos optado por esta última por ser la que más ventajas ofrece tanto para el corrector, facilitando toda la tarea tediosa de volcado de datos, como para quien lo realiza, ya que el sistema de programación permite la realización de la prueba en varias sesiones, reduciendo así los posibles efectos que el cansancio puede ocasionar en especial sobre el rendimiento de los últimos ítems. Asimismo, este sistema permite controlar todos los aspectos relevantes de la prueba, como impedir que se dejen ítems sin contestar, puesto que el sistema no pasa al ítem siguiente hasta que no se da una respuesta; o que se retroceda para corregir las

respuestas, así como impedir que acuda de nuevo al test, una vez finalizado. La versión electrónica permite además una aplicación tanto individual como colectiva de ambas pruebas.

Los resultados que finalmente se obtienen con estas pruebas son: una puntuación global de la capacidad de pensamiento crítico, junto a las puntuaciones de las cinco subescalas (RD, RP, RI, TD y SP), que son las que hacen referencia a las habilidades específicas del mismo.

Los primeros datos que hemos obtenido con estas pruebas, muestran unas propiedades psicométricas aceptables, pero mejorables. En estos momentos, estamos realizando los cambios pertinentes para conseguir una mayor calidad de las mismas.

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 del PC. Nuestro modo de evaluar contribuye mucho a esta clarificación conceptual.

El PC es un proceso multicomponencial, como hemos descrito al comienzo de nuestro trabajo. 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 inferencia. Los procedimientos que debemos seguir para tomar buenas decisiones requieren asimismo 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.

5. Conclusiones y sugerencias

A lo largo de este trabajo, hemos estudiado las razones que justifican la evaluación del pensamiento crítico, la forma de abordar esta valoración, su viabilidad y la propuesta de medida que ofrecemos. Con frecuencia, la necesidad de evaluar las habilidades de pensamiento arranca de una iniciativa de intervención. Cuando se busca mejorar estas destrezas se impone el conocer si la instrucción produce algún cambio. Este es el origen principal de la evaluación. Hay ocasiones donde surgen iniciativas institucionales o administrativas encaminadas a valorar el rendimiento de los estudiantes en ciertas capacidades intelectuales. En ambos casos, el modo de estimar estas competencias es muy semejante. Desde hace tiempo, se vienen desarrollando proyectos de evaluación del pensamiento con estos fines, y con métodos psicométricos muy semejantes. Se construyen pruebas que plantean problemas frente a los que básicamente sólo hay que responder eligiendo opciones de respuesta ya dadas. Esta manera de valorar habilidades de pensamiento resulta poco útil, ya que no se puede saber qué destreza está empleando quien responde a la prueba; incluso es difícil de asegurar que esté utilizando alguna competencia de esa clase. La discusión que hemos planteado tenía como finalidad demostrar la incapacidad de medir pensamiento por parte de la mayoría de las pruebas estandarizadas publicadas. El modo de evaluar dichas competencias manifiesta graves problemas de validez.

A raíz del estudio que venimos realizando, desde hace tiempo, sobre la evaluación del pensamiento, hemos podido proponer un modo alternativo de solucionar este problema y otros también importantes. Nuestro planteamiento consistía en emplear el análisis de tareas para las situaciones-problema. Este método se ha mostrado muy eficaz a la hora de saber qué proceso de pensamiento se está empleando para cada tarea-problema. De este modo, podemos saber lo que estamos midiendo y conseguir así diagnosticar el nivel de competencia alcanzado por quien recibe una instrucción, y diseñar pautas de intervención futuras.

Faltan aún muchos aspectos por mejorar de nuestro método, pero hemos sentado unas bases sólidas, que nos permiten avanzar con ciertas garantías de éxito. La labor que nos falta por hacer consiste en obtener buenos índices psicométricos para nuestras pruebas. Ya disponemos de unos primeros resultados aceptables. El tiempo nos dará la oportunidad de mejorarlos.

Bibliografía

- Anderson, J. R. (1990). *Cognitive psychology and its implications* (Third edition). San Francisco, CA: W.H. Freeman.
- Donders, F. C. (1969). "On the speed of mental processes". *Acta Psychologica*, 30, pp. 412-431. Original: 1868.
- Ennis, R. H. (2003). "Critical thinking assessment". En D. Fasko (ed.), *Critical thinking and reasoning. Current research, theory, and practice*, pp. 293-313. Cresskill, NJ: Hampton Press.
- Ennis, R. H. (2008). "Nationwide testing of critical thinking for higher education: Vigilance required". *Teaching Philosophy*, 31 (1), pp. 1-26.
- Ennis, R. H., & Weir, E. (1985). *The Ennis-Weir-critical thinking essay test*. Pacific Grove: CA: Midwest Publications.
- Govier, T. (1987). *Problems in Argument Analysis and Evaluation*. Dordrecht, Holland: Foris Publications.
- Halpern, D. F. (1998). "Teaching critical thinking for transfer across domains - Dispositions, skills, structure training, and metacognitive monitoring". *American Psychologist*, 53 (4), pp. 449-455.

- Halpern, D. F. (2003a). *Thought and knowledge: An introduction to critical thinking* (Fourth edition). Hillsdale, NJ: Erlbaum.
- Halpern, D.F. (2003b). "The "How" and "Why" of critical thinking assessment". En D. Fasko (ed.), *Critical thinking and reasoning. Current research, theory, and practice*, pp. 355-366. NY: Hampton press.
- Johnson, R. H. (2008, 8-11 de enero). "Critical thinking, logic and argumentation". Ponencia presentada en la: Conferencia Internacional: Lógica, Argumentación y Pensamiento Crítico, Santiago de Chile.
- Nieto, A.M. y Saiz, C. (2008). "Evaluation of Halpern's "Structural Component" for Improving Critical Thinking". *The Spanish Journal of Psychology*, 11 (1), pp. 266-274.
- Nieto, A.M., Saiz, C. y Orgaz, B. (2009, en prensa). "Análisis de la 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*, n° 1 de 2009.
- Possin, K. (2008). "A field guide to critical-thinking assesment". *Teaching Philosophy*, 31 (3), pp. 201-228.
- Saiz, C. (2002). "Enseñar o aprender a pensar". *Escritos de Psicología*, 6, pp. 53-72.
- Saiz, C. y Nieto, A. M. (2002). "Pensamiento crítico: capacidades y desarrollo". En C. Saiz (ed.), *Pensamiento crítico: conceptos básicos y actividades prácticas*, pp. 15-19. Madrid: Pirámide.
- Saiz, C. y Rivas, S.F. (2008). "Intervenir para transferir en pensamiento crítico". *Praxis*. 10 (13), pp. 129-149.
- Shaw, V. F. & Johnson-Laird, P. N. (1998). "Dispelling the "atmosphere" effect on reasoning". En A. C. Quelhas & F. Pereira (eds.), *Cognition and context*, pp. 169-199. Lisboa: Instituto Superior de Psicologia Aplicada.
- U.S. Department of Education (2006), *A Test of Leadership: Charting the Future of U.S. Higher Education*. Washington, D.C. En: <http://www.ed.gov/about/bdscomm/list/hiedfuture/index.html>.

Recibido el 1 de agosto, revisado el 10 de agosto,
aprobado el 10 de septiembre de 2008.