# III. CONCLUSIONES Y SUGERENCIAS

Hemos desarrollado nuestra propuesta de intervención, con el fin de indagar si el resultado de la misma puede hacer crecer personalmente a los jóvenes. La manera de lograrlo, en nuestra opinión, pasa por el desarrollo de una instrucción que simule la realidad, que promueva la utilidad, la conciencia de nuestras deficiencias y sesgos, y que se fomente la producción de situaciones cotidianas.

Pero ¿esta iniciativa de intervención funciona? ¿Logra lo que pretende? La evaluación de un programa de instrucción es un tema centrar, que se debe abordar con el máximo rigor. Pero esto será el objeto de la próxima comunicación, no de ésta.

Son muchos los esfuerzos que todavía habrá que realizar hasta lograr la eficacia deseada, como es la transferencia, pero el camino, creemos, pasa por diseñar situaciones cotidianas, que permitan simular problemas cercanos al funcionamiento diario.

# CAPÍTULO<sup>3</sup>

# ¿ES POSIBLE EVALUAR LA CAPA CIDAD DE PENSAR CRÍTICAMENTE EN LA VIDA COTIDIANA?

sidad Pontificia de Salamanca y (Universersidad de Salamanca, España) Univ

#### I. INTRODUCCIÓN

En el presente trabajo, vamos a intentar resposible evaluar capacidad de pensar en la vida dia quívocamente sí, pero nuestro esfuerzo se dirigirá lector de la conveniencia de valorar nuestras con la finalidad de mejorar personal y socialmente. Siempre son complejos y problemáticos, y máximalgo tan amplio como es la actividad de pensar. Lo de esta comunicación será la defensa de que es de medida de esta clase, nuestra conclusión es que blemente factible y muy útil para las personas y que tenemos que ocuparnos es de las necesidades

ponder a la cuestión de si es ria. Nuestra respuesta es inea justificarlo y a persuadir al apetencias intelectuales, con Los sistemas de evaluación ne cuando se trata de medir que expondremos a lo largo posible construir un sistema le este proyecto es razonala sociedad. De lo primero y dificultados de la propia

evaluación, con el fin de proponer una, que cumpla con los objetivos que se buscan y que además sea rigurosa. A continuación, debemos tratar el origen de toda evaluación, los programas de intervención, con el propósito de dejar muy claro que toda medida depende de ellos, o que sin ellos no tiene sentido. Una medida de aprendizaje, cuando no la hay, es un contrasentido obvio. Todo sistema de evaluación, dirigido a medir lo que se ha adquirido, siempre tropieza con el ingrato problema de la transferencia. Es pues, imprescindible que nos detengamos en él, pues es el núcleo de toda evaluación de competencias intelectuales. Una vez que hayamos aclarado cómo vamos a tratar nosotros la transferencia, entonces, ya sí, podemos ocuparnos en exponer nuestro sistema de evaluación. Y finalmente, después de describir nuestras pruebas, mencionaremos los primeros resultados que hemos obtenido con ellas, para terminar señalando el estado actual de la evaluación en pensamiento crítico.

# II. NECESIDAD, IMPORTANCIA Y DIFICULTADES DE LA EVALUACIÓN

La necesidad e importancia de la evaluación del pensamiento crítico en la vida diaria proviene de si social o personalmente se desea que estas competencias se mejoren. Sin esta condición, nada importa. Reflexionar de un modo eficaz exige mejorar dicha capacidad, y saber si existe dicha mejora impone cuantificar o cualificarla de algún modo. En definitiva, tener como objetivo el desarrollo de nuestras habilidades de pensamiento obliga a medir o valorarlas, con el fin de constatar algún cambio o progreso en esas competencias. Pero la importancia de pensar críticamente supone que esto nos hace más eficaces a la hora de lograr nuestras metas o simplemente que nos adaptemos bien a nuestro entorno. Por su parte, su necesidad asume que no pensamos tan bien como pudiéramos. Hay pues dos ideas que conviene separar, una, que pensar bien nos beneficia, y otra, que podemos hacerlo mejor. No es el lugar aquí para justificar con datos o razones, de los muchos de los que se disponen, lo necesario e importante que es pensar críticamente. Sí procede tratar las dificultades a la hora de evaluarlo.

Las dificultades en la evaluación del pensamiento crítico son *conceptuales y metodológicas*. Las primeras, provienen de las diferentes formas que

hay de entender lo que es pensar críticamente. Para no distraernos de los objetivos de nuestro trabajo en valorar las ideas que existen, diremos solo lo que nosotros entendemos como tal. 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. Esta es la idea con la que estamos trabajando en intervención y evaluación. Esta definición intenta recoger lo común a todo proceso de pensar, buscar conocimiento¹, junto con sus mecanismos fundamentales, encaminados al logro de nuestras metas. Para nuestros fines, consideramos suficiente hacer explícita nuestra concepción sobre lo que es pensar críticamente.

Las dificultades metodológicas² tienen su origen principalmente en el uso de pruebas con un formato de respuesta cerrado. Una prueba que solo pida a quien la responda que marce unas casillas que se dan como opciones de respuesta impide captar lo esencial del pensamiento, esto es, los procesos que llevamos a cabo para realizar una buena reflexión. Se sacrifica el conocimiento a la rapidez en la corrección. Ya se sabe que es muy costoso, en tiempo, corregir los formatos de respuesta abiertos, pero es el único modo de conseguir adentrase en la mente de quien responde. Si realmente deseamos saber, si estamos midiendo pensamiento, debemos dar la oportunidad de que se exprese. Una marca en una casilla difícilmente puede lograr esto. Son pocas las iniciativas de evaluación en esta dirección. Una de las más recientes es la de Halpern. Esta autora ha desarrollado el test HCTAES3. Esta prueba se centra precisamente en la eficacia del aprendizaje, en los procesos. En ella, se plantean problemas que se deben resolver y explicar. Se pide que se expresen los pasos que se siguen para afrontar esos problemas, que se detallen las estrategias o procesos seguidos. Esta medida, muy innovadora, junto con la que hemos desarrollado y que describiremos después, son las únicas pruebas de

<sup>&</sup>lt;sup>1</sup> Ver D. F. Halpern, *Thought and knowledge: An introduction to critical thinking*, Hillsdale, NJ: Erlbaum (4<sup>th</sup> edition), 2003.

<sup>&</sup>lt;sup>2</sup> Ver R. H. Ennis, «Critical thinking assessment», in D. Fasko (ed.), Critical thinking and reasoning. Current research, theory, and practice, Cresskill, NJ: Hampton Press, 2003, pp. 293-313.

<sup>&</sup>lt;sup>3</sup> D. F. Halpern, Halpern Critical Thinking Assessment Using Everyday Situations: Background and scoring standards, 2006 (unpublished report).

pensamiento crítico, enfocadas hacia los procesos de pensamiento. La razón obvia ya se ha dicho, el coste tan grande de tiempo en su corrección.

#### III. EVALUACIÓN DE PROGRAMAS

Mejorar nuestra capacidad de pensar, como cualquier otra habilidad, se aprende. Y los procesos de aprendizaje necesitan evaluarse para conocer si se está consiguiendo desarrollar las competencias adiestradas. Como decíamos al principio, pensar es nuestra herramienta esencial de adaptación. Su importancia y necesidad son difíciles de cuestionar. La razón de ser, pues, de la evaluación del pensamiento crítico es cuantificar el proceso de aprendizaje, sea este reglado o no (académico o no). En nuestro caso, llevamos un tiempo desarrollando el programa de instrucción ARDESOS4, lo que nos ha obligado a desarrollar una evaluación del mismo. Sin embargo, conviene saber que este proceso de evaluación es de naturaleza conceptual y empírica. En el presente trabajo se requiere de nosotros una mayor atención para la segunda, puesto que la conceptual está muy vinculada a la construcción o selección de un programa de intervención y es, sobre todo, cualitativa. Además, la evaluación conceptual se realiza durante el proceso de elaboración de un programa, algo ya descrito en la comunicación previamente citada. La evaluación empírica es la que exige una mayor atención por nuestra parte. Ésta es de naturaleza cuantitativa y va dirigida a probar la eficacia de la instrucción, esto es, si después de la enseñanza se ha conseguido algún cambio. Es, pues, cuantitativa y se realiza después de que la intervención haya terminado. Dediquemos algo de tiempo a su descripción.

Una evaluación empírica necesita de un instrumento que nos permita cuantificar los cambios que una instrucción ha producido. Estos cambios son los que nos permitirán afirmar si una iniciativa de intervención como ARDES-OS<sup>5</sup>, u otras, son eficaces (hacen que pensemos mejor), después de ese proceso de aprendizaje. Pero para llegar a esta conclusión final, de si un programa ha funcionado o no, hay un largo camino que recorrer. El punto de partida de una

evaluación de esta naturaleza, empírica y del aprendizaje, impone un diseño singular. Lo primero que debemos tener presente es que aprender o adquirir habilidades mediante un método, temporalmente, nos exige que conozcamos el nivel de destreza antes del aprendizaje, de lo contrario nada podríamos decir sobre la eficacia. Una vez que dispongamos de una medida anterior a la instrucción, ya sí podremos cuantificar el cambio. Sin embargo, aún no estamos en condiciones de poder concluir si ese cambio se debe al programa. Para esto, debemos disponer de un nivel de pericia frente a una misma prueba en personas que hayan pasado por la instrucción y en personas que no. Esta medida, que se obtiene *después* de la enseñanza en el grupo de la instrucción, y en otro (que llamaremos *control*) que no ha pasado por ella, es lo que ya nos va a permitir concluir que, de haber una mejora en el primer grupo y no en el control, será fruto de esa instrucción. La organización de la medida de evaluación que conviene realizar aquí, se denomina *diseño antes-después con grupo control* (al menos uno).

El diseño que hemos descrito es nuestro campo de acción, pero en él conviene introducir «piezas» que nos permitan cuantificar la eficacia del aprendizaje. Estas piezas deben recoger todos los aspectos esenciales del proceso de adquisición de las habilidades instruidas. Téngase en cuenta que no estamos tratando de destrezas sencillas como puede ser «atornillar» o «apretar tuercas». Se trata de competencias que son la base del funcionamiento humano más singular, como lo es conocer o explicar el mundo o a nosotros mismos. Estas capacidades están presentes en casi todo lo que hacemos, por lo tanto, debemos ser capaces de ver si ese cambio que produzca el aprendizaje se aplica en todos los ámbitos. Pensar no es dependiente de un dominio, es independiente, de modo que es deseable que se refleje en cualquiera de ellos. Pero, hasta llegar a esta constatación, hay que andar un buen trecho. Cuando uno se esfuerza por mejorar sus capacidades, lo primero que debemos saber es que se ha dado un cambio, que después de un aprendizaje se posee una mayor competencia que antes. A este cambio se le denomina magnitud del efecto. Este es uno de los tres indicadores de eficacia que tenemos que conocer y obtener. Si hacemos las cosas mejor después del aprendizaje que antes, en principio, algo ha cambiado. Y si ese cambio sucede en quien ha pasado por la instrucción y no se observa en los que no la han recibido, ya podemos concluir que esa mag-

<sup>&</sup>lt;sup>4</sup> Ver C. Saiz y S. F. Rivas, «¿Mejorar el pensamiento crítico contribuye al desarrollo personal de los jóvenes?» (cap. 2)

<sup>&</sup>lt;sup>5</sup> Ver *idem*, «Evaluation of the ARDESOS program: an initiative to improve critical thinking skills», in *Journal of the Scholarship of Teaching and Learning* (2009, en revisión).

lar mejor e pone en procesos an flor de <sup>u</sup>mpo posibio; cuancesitamos nos grado juevo esas o. Si obteatisfechos. consiste en de obtener, ealizar por as, después porque muitivas como xperimental

a una buena
l'último es el
licia. Cuando
que está preluo criterio de
lotros ámbilese ha venido
ampoco este
trabajo es exlicia.

opropuestos Arolladas para tal fin: PENCRISAL y PENTRASAL?. La magnitud del efecto y el grado de persistencia los medimos con el primer instrumento, y la transferencia, con el segundo. Pero antes de pasar a la exposición de nuestro sistema de evaluación conviene reflexionar sobre el problema central de la evaluación: *la transferencia*. Si tuviéramos que quedarnos con un índice de eficacia, de los tres que se necesitan, sin lugar a dudas, sería la transferencia. Lograrla nos permitiría olvidarnos de la magnitud del efecto y de hasta la persistencia del mismo. Es tal su naturaleza, que recoge muy bien todo lo relevante de la evaluación, entre otras razones, por una cuestión de lógica, no es posible generalizar lo que no se tiene. Si fuéramos radicales, hablaríamos solo de este criterio. Sin embargo, la sensatez debe moderarnos, puesto que lograr la tan descada transferencia es un reto difícil de lograr y, si fracasamos, nos quedamos con las manos vacías, de modo que mejor medir la magnitud del efecto y después apostar por el reto de la generalización ¿Por qué es tan difícil conseguir que pensemos en cualquier ámbito o dominio? Intentemos responder a esta pregunta.

#### IV. TRANSFERENCIA DE LAS HABILIDADES

La generalización del conocimiento y de las habilidades a otros dominios distintos del de aprendizaje es uno de los problemas centrales de la educación. Obtener conocimiento o comprender la realidad no es suficiente, es imprescindible que seamos capaces de *utilizarlo* en cualquier circunstancia, en toda situación<sup>8</sup>. Pero aplicar el conocimiento o una destreza exige mucho más que la mera comprensión de un problema o situación. Y esto es así, porque debe funcionar bien todo nuestro sistema cognitivo, y esto no es fácil conseguirlo. Normalmente se fracasa en la transferencia por problemas de *identificación*,

ologist, 52 (10).

<sup>&</sup>lt;sup>7</sup> Ver C. Saiz y S. F. Rivas, «Evaluación en pensamiento crítico: una propuesta para diferenciar formas de pensar», in *Ergo, Nueva Época*, 22-23, 2008, pp. 25-66. En: <a href="http://www.pensamiento-critico.com/pensalen/evaluarpcergodf.pdf">http://www.pensamiento-critico.com/pensalen/evaluarpcergodf.pdf</a> Versión inglesa: «Assessment in critical thinking: a proposal for differentiating ways of thinking», en: <a href="http://www.pensamiento-critico.com/pensalen/evaluationCTergoENGLSH.pdf">http://www.pensamiento-critico.com/pensalen/evaluationCTergoENGLSH.pdf</a>

<sup>\*</sup> Ver D. N. Perkins, «¿Qué es la comprensión?», in M. S. Wiske (ed.), La enseñanza para la comprensión. Vinculación entre la investigación y la práctica, Buenos Aires: Paidós, 1999, pp. 69-92; y R. C. Schank, Lessons in Learning, e-Learning, and Training: Perspectives and Guidance for the Enlightened Trainer, San Francisco, CA: Pfeiffer/Wiley, 2005.

recuperación o aplicación<sup>9</sup>. Un conocimiento, como una determinada estrategia, para poder usarlo, debe estar accesible, debemos saber que disponemos de ella. Pero también necesitamos recuperarla, y finalmente ser capaz de aplicarla a un contexto o situación concreto. La transferencia exige que funcionen bien los mecanismos de atención y memoria primero, para poder pensar (en este ejemplo, inductivamente) cómo usar esa estrategia. El problema de la educación, al menos uno importante, es que apenas se ocupa de las limitaciones de nuestro sistema de procesamiento. Nuestra mente funciona razonablemente bien hasta que una parte de la misma deja de hacerlo. Al estar tan interrelacionados los mecanismos mentales, cualquier problema en uno, hace que el sistema deje de funcionar de manera idónea. Y esto es lo que sucede cuando intentamos ser eficaces empleando nuestro conocimiento. Deben ponerse en funcionamiento demasiadas partes del sistema para conseguir dicha eficacia.

Sin embargo, sabemos algo del funcionamiento mental como para poder evitar deficiencias importantes. En el caso que nos ocupa, sabemos que un procedimiento, recurso o estrategia, para poder usarlo (una vez adquirido, aprendido), debe estar accesible y ser fácilmente recuperable, para su uso. Por lo que sabemos, eso se puede reforzar (accesibilidad y recuperación) incorporando en el proceso de aprendizaje muchas actividades prácticas, que hagan que una habilidad se consolide en nuestro sistema con ayuda de la acción. Es necesario aprender haciendo, no sirve con la representación o imaginación mental solo, hay que poner esa estrategia o recursos a trabajar. De esta forma, conseguimos alta disponibilidad y fácil recuperación de nuestro conocimiento. Una vez que hemos conseguido disponer de un recurso o procedimiento, debemos aplicarlo, y esto es otro problema, pues debemos ser capaces de identificar las similitudes entre contextos, para poder aplicar la misma estrategia en todos ellos. Esto exige que nuestro sistema sea capaz de identificar regularidades, que pueda abstraer lo común, que pueda relacionar, que logre identificar patrones semejantes. Una vez conseguido esto, entonces se puede aplicar un procedimiento o habilidad con independencia del contexto o situación. Durante el aprendizaje o adquisición de habilidades es muy

importante, para fomentar esos mecanismos inductivos, realizar muchas prácticas en *diferentes dominios o contextos*, con el fin de facilitar la aplicación de las mismas. Muchas actividades prácticas y mucha práctica interdominios es parte de la clave, para conseguir una buena transferencia de habilidades. Si a esto le añadimos que las situaciones de aprendizaje sean problemas que se deben resolver, entonces garantizaremos aún más el uso de habilidades en cualquier dominio o situación. El *aprendizaje basado en problemas* fomenta aprender haciendo, dirige el aprendizaje hacia la adquisición de conocimiento que permita resolver problemas. El conocimiento que se aplica y es útil, adquiere un interés mayor que cualquier otro, motiva más que lo que no se etiqueta como aplicable.

Para conseguir que cuando se enseñe a pensar, se fomente la transferencia, debemos incorporar tareas que sean problemas que debamos resolver, pero no cualquier tipo de problema, deben ser cercanos, «ecológicos», problemas cotidianos. El aprendizaje basado en problemas lo debe ser de problemas comunes, a los que normalmente nos enfrentamos en la vida diaria, no valen acertijos, «sudokus» o similares, deben ser problemas relevantes, importantes para nosotros. Si no vemos un problema como importante en nuestra vida diaria, nuestra dedicación y esfuerzo a ellos serán más bien escasos y, una vez más, la transferencia fracasará. Para conseguir generalizar nuestras habilidades debemos incorporar en la instrucción de las mismas muchas cosas, la falta de cualquiera de ellas es un billete hacia ninguna parte. Actividades prácticas, prácticas interdominio, y problemas cotidianos deben ser los fundamentos de toda enseñanza que busque la transferencia del conocimiento. Nuestra iniciativa de enseñar a pensar, como ya se ha expuesto en la comunicación de Saiz y Rivas, se guía por esos fundamentos. Pero no olvidemos que en esta comunicación estamos discutiendo y proponiendo métodos de evaluación, no de intervención ¿por qué pues tanta atención a los aspectos de la intervención y la transferencia? Sencillamente porque la evaluación debe ser una especie de clon de la intervención. Solo podemos evaluar lo que enseñamos y en el modo que lo hacemos. Pasemos ahora a tratar nuestro sistema de evaluación del pensamiento crítico.

<sup>&</sup>lt;sup>9</sup> D. F. Halpern, «Teaching critical thinking for transfer across domains – *Dispositions, skills, structure training, and metacognitive monitoring*», in *American Psychologist*, 53 (4), 1998, pp. 449-455.

<sup>&</sup>lt;sup>10</sup> Ver C. E. Hmelo-Silver, «Problem-based learning: What and how do students learn?», in *Educational Psychology Review*, 16 (3), 2004, pp. 235-266.

#### V. PRUEBAS DE EVALUACIÓN

Cualquier iniciativa para aprender o enseñar a pensar no puede ser *convencional* ni en la instrucción ni en la evaluación, porque intentamos que cambien los procesos más complejos y sofisticados de la mente humana. Ya hemos hecho referencia a los requisitos de la intervención, abordemos ahora los de la evaluación.

Como ya señalamos, no se puede pretender medir mecanismos tan complejos como los del pensamiento con unas preguntas y unas simples alternativas de respuesta de las que solo tenemos que elegir una. Esta forma generalizada de medir pensamiento crítico, mediante pruebas cerradas, sencillamente, no sirve. Sobre todo medimos capacidad de discriminar entre opciones a partir de lo más familiar, pero no medimos procesos de pensamiento. No nos detendremos más en este problema de las pruebas de pensamiento crítico que ha sido tratado ampliamente en otro lugar<sup>11</sup>. El clon de la instrucción que debe ser la evaluación, en nuestro caso, rechaza pruebas convencionales como las ya citadas, pero además, nos obliga a adaptar la evaluación a los objetivos específicos de la intervención y a su naturaleza. En nuestro proyecto, la evaluación está vinculada a nuestro programa de intervención ARDESOS, pero no sólo. Recordemos que el objetivo de nuestra comunicación es contestar a la pregunta de si es posible evaluar la capacidad de pensar en la vida diaria. Nuestro sistema de evaluación comienza en nuestro programa pero termina en el pensamiento crítico de la vida cotidiana. Por lo tanto, se debe ver nuestra propuesta de medida de competencias de pensamiento, como susceptible de utilizar fuera del contexto de aprendizaje. Puede ser aplicable en cualquier momento o lugar.

La evaluación de la eficacia del pensamiento crítico debes ser *cuantita-tiva* fundamentalmente, por varias razones que iremos exponiendo a lo largo de este apartado. Nuestra forma de entender la evaluación pasa por incorporar *pruebas de rendimiento específicas y pruebas estandarizadas*. Cada una por sí sola es insuficiente para captar la mejora de nuestro pensamiento. Las primeras son un salvoconducto 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 intentar aplicar pruebas estandarizadas que nos permitan ir más allá del contexto de aprendizaje. En nuestro caso, hemos utilizado una tarea de rendimiento acumulativa y dos pruebas estandarizadas, PENCRISAL/PENTRASAL¹².

La tarea específica de rendimiento acumulativa mide aspectos esenciales de la instrucción como la capacidad de producción e integración de habilidades, así como un rendimiento por contenidos o temáticas. Precisando algo más nuestra concepción, entendemos el pensar como razonar y decidir para resolver. Desde nuestro punto de vista, el pensamiento es más acción para alcanzar nuestros fines o resolver problemas, que meditación. Y para resolver problemas, nuestras estrategias fundamentales son razonar y decidir bien. Lo que deseamos resaltar es que nuestra enseñanza del pensamiento crítico lo es integrando todas las habilidades o procesos. Pero además es una enseñanza desde el error. Esto es, aprender desde nuestros límites, sesgos o deficiencias. Y esto solo lo podemos observar cuando planteamos problemas que se deben resolver sin ningún tipo de instrucción. Al plantear situaciones de esta naturaleza, nos vemos obligados a generar o producir soluciones, sin más apoyo que el sentido común. Al pedir que se realicen tareas de producción, posibilitamos que aparezcan esas deficiencias o sesgo, que nos permitirán asentar mucho mejor las destrezas enseñadas, ya que se progresa desde la conciencia de nuestras propias limitaciones. Pero como decíamos antes, la evaluación debe ser una especie de clon de la intervención. Por esta razón, la tarea de rendimiento acumulativa valora de un modo integrado todas las habilidades de pensamiento. Describamos una tarea de esta naturaleza.

Se plantea un problema que hay que resolver, razonando y tomando decisiones. El problema puede ser el siguiente: ¿la adopción por parte de parejas homosexuales es positiva o negativa? A continuación, se les dice que se imaginen siendo un psicólogo que tiene que valorar la conveniencia de la adopción por parte de este colectivo. Se les pide que emitan un informe técnico que de respuesta a ese problema. Para realizar esta tarea, deben proceder indagando, argumentando, valorando los estudios científicos, y decidiendo en función de esa reflexión previa. Se van realizando tareas acumulativas y de

<sup>11</sup> C. Saiz y S. F. Rivas, art. cit. (cf. nota 7).

<sup>12</sup> Ver idem, ibidem

manera integrada, consistente en producir argumentos o explicaciones, y sopesando todas las opciones de parejas. Con esta tarea, valoramos bloques de competencias como: argumentar de manera integrada, ponderar alternativas y decidir, con el fin de resolver el problema planteado. La producción de argumentos integrados consiste en la elaboración de una argumentación en la que existan subargumentos condicionales, causales y analógicos. De este modo, evaluamos los aspectos esenciales de la instrucción, como ya hemos señalado. En ella, trabajamos por bloques de habilidades (argumentación, deducción, inducción y solución de problemas y toma de decisiones). Dentro de cada bloque se elaboran argumentos de cada tipo, se generan y aplican estrategias de solución de problemas y de toma de decisiones. Al usar una tarea acumulativa, valoramos las diferentes formas de razonamiento de un modo integrado y acumulativo, así como las distintas maneras en las que se resuelve y decide. Démonos cuenta, que la naturaleza de esta tarea permite valorar las competencias aplicadas a contextos cotidianos, esto es, nos proporciona un buen indicador del grado de dominio de las mismas, puesto que nos piden aplicarlas y a contextos diferentes de los del aprendizaje. Un reto que nos puede proporcionar información adicional para discutir el problema de la transferencia.

La tarea de rendimiento acumulativa obliga, pues, a utilizar todas las habilidades adquiridas. Como prueba de rendimiento, facilita información inapreciable sobre la eficacia de la intervención. Pero esta no es la única forma de evaluación que se necesita. Además, conviene emplear pruebas estandarizadas, en combinación con las de rendimiento. En nuestro caso, empleamos la pruebas PENCRISAL/PENTRASAL, que pasamos a describir. Las razones para utilizar pruebas estandarizadas para evaluar el pensamiento crítico, además de las de rendimiento, son sobradamente conocidas, pero conviene recordarlas. Una de las razones fundamentales para su uso es que nos permite *evaluar el constructo de pensamiento crítico*, nos proporciona un indicador fiable de validez, esto es, de 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 importantes, aunque no tanto. 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 pensamiento crítico y programas de instrucción. Démonos cuenta, que esta última razón hace posible evaluar el pensamiento crítico en la vida cotidiana. Permite valorar el nivel de competencias intelectuales de cualquier persona, pudiendo diagnosticar y corregir posibles deficiencias.

Las pruebas estandarizadas PENCRISAL y PENTRASAL tienen su origen en el test HCTAES<sup>13</sup>. 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<sup>14</sup>. Los principios que sustentan nuestras pruebas son: 1) la utilización de ítems que sean situaciones cotidianas, 2) el uso de diferentes dominios, 3) un formato de respuesta abierto, y 4) el empleo de problemas de respuesta única. El primero mantiene una de las aportaciones más importantes del test de Halpern. El segundo y el cuarto son la aportación de nuestras pruebas, 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 expresen. Finalmente, el uso de diferentes dominios nos permite valorar el grado de generalización de las habilidades. Este rasgo y el uso de problemas únicos es lo singular de nuestras pruebas.

Los ítems formados por situaciones cotidianas o situaciones problema intentan evaluar el pensamiento crítico 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.

<sup>13</sup> Ver D. F. Halpern, op. cit. (cf. nota 3)

<sup>14</sup> Ver C. Saiz y S. F. Rivas, art. cit.

#### Ejemplo 1

#### DEDUCCIÓN

## Item 16: Prueba de Habilidad de PC 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.

#### Situación 1: PENCRISAL

El entrenador personal de Juan le ha dicho que solo 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.

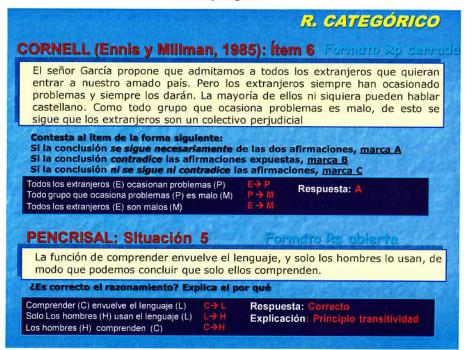
En el ejemplo 1, se muestran dos ítems, uno del test de California y otro del PENCRISAL. Se puede observa que la situación 1 de nuestra prueba plantea un problema de deducción de manera menos estereotipada y natural que el ítem 16. Aquí, además, se responde marcando una casilla, y en nuestra prueba se debe explicar si es correcta la reflexión del protagonista de la situación problema. Recordemos que en la instrucción se trabaja con situaciones problemas que simulan la realidad, con el fin de conseguir la mayor proximidad entre el contexto de aprendizaje y el cotidiano. En la evaluación debemos proceder del mismo modo.

Un segundo rasgo de nuestras pruebas es el uso de *diferentes dominios*. En el caso del PENCRISAL, el ámbito de los problemas son las relaciones personales, y en el PENTRASAL, se emplean diferentes contextos como salud, deportes, educación, ocio... El propósito de usar diferentes dominios en la intervención y en la evaluación es poder medir la transferencia. Recordemos que esta hace referencia a la capacidad de generalización de las habilidades aprendidas en la esfera del aprendizaje a otras. Sin embargo, en nuestro

planteamiento de la intervención, nos surge un problema, si en el aprendizaje trabajan con problemas cotidianos ¿a dónde deben generalizar esas competencias? Lo que proponemos es una *reconceptualización* de la transferencia, entendiéndola como el uso de las habilidades aprendidas en diferentes dominios o temáticas. Puesto que en la instrucción, quien la recibe, ya se enfrenta a problemas cotidianos, solo queda que generalice sus destrezas a la mayoría de los dominios en su desenvolvimiento diario. La evaluación de estos diferentes dominios la llevamos a cabo con el PENTRASAL. No vemos otro modo de valorar la transferencia, cuando los contextos del aprendizaje y su aplicación son los mismos. Solo queda por ver si las temáticas (educación, ocio, salud...) de las situaciones cotidianas son algún obstáculo para la aplicación de las habilidades adquiridas en cualquiera de ellas.

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, en realidad es solo una prueba de comprensión del problema que se nos plantea con el fin de discriminar entre las opciones que se nos proponen. En realidad, 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, que con lo que da, nos imposibilita para averiguar qué mecanismo se pone en marcha con cada situación-problema. Por el contrario, un formato de respuesta abierto obliga a exteriorizar 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, sencillamente, es imposible obtenerlo con un formato de respuesta cerrado. En el ejemplo 2, podemos ver estas diferencias. En el ítem 6 del test Cornell, podemos apreciar que solo se debe marcar una opción, y que es imposible saber con esta respuesta si se está «viendo» el principio de transitividad o no. En la situación 5, la respuesta debe justificarse, de modo que podemos ver fácilmente si está resolviendo el problema usando ese principio o no. Además de decir si es correcto lo que se afirma, debe explicarlo. El formato de respuesta no sólo es una cuestión de economía (de ahorro de tiempo), es mucho más, es establecer la diferencia entre saber qué se piensa frente a los problemas o, por el contrario, adivinar o especular sobre lo que se piensa. No es nada baladí esta cuestión. En este formato está algo tan esencial como poder saber si realmente estamos midiendo lo que pretendemos y no otra cosa. Los números pueden «tranquilizarnos», pero no convencernos. La estadística puede decir que existen determinadas dimensiones en una prueba, pero que esas sean de pensamiento o no, no puede, porque lo que se responde a una prueba cerrada no me permite saberlo. Son demasiado complejos nuestros mecanismos de pensamiento como para captarlos con una simple marca en una casilla.

#### Ejemplo 2



El último rasgo de nuestras pruebas consiste en plantear situaciones-problema que solo se puedan responder de una manera, esto es, diseñar ítems que sean problemas de respuesta única. El test predecesor de nuestras pruebas, el HCTAES, consta de situaciones cotidianas que se pueden responder de diferentes formas, no plantea un problema para cada proceso. Veámoslo en el ejemplo 3. En el ítem 24, se observa que las respuestas que se dan y puntúan pueden ser de lo más variado, y algunas nada tienen que ver con procesos de pensamiento, pues pedir información o consultar a un experto, que se sepa, no requiere demasiada inferencia. Esto nos plantea un problema, pues nos deja en una situación parecida a la generada por las respuestas cerradas, esto es, que no sabemos qué está haciendo mentalmente quien responde el problema. Para solucionar esta grave limitación, en nuestras pruebas hemos incorporado la metodología del análisis de tareas<sup>15</sup>. 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 «atrapar» 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, una virtud nada desdeñable.

<sup>15</sup> Idem, ibidem

#### Ejemplo 3

#### SOLUCIÓN PROBLEMAS HALPERN (CTA): Item 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 Propón dos buenas soluciones para este problema CORRECCIÓN (0-1-2) : Si hace uso de métodos creativos edir ayuda, consultar experto, mezclar pastillas, etc PENCRISAL: 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. ¿Cuál sería la mejor forma de actuar por parte de Rocio para resolver su situación? Justificalo CORRECCIÓN (0-1-2) : Si aplica bien un método Identificar problema: madre e hija no atienden el bar y necesitan alguien que trabaje en el bar Subproblemas o metas: Hacer una selección de candidatos para elegir trabajador adecuado, le llevará tiempo y tendrá que cerrar el bar. Contratar a la hija de su amiga provisionalmente mientras busca un candidato adecuado

Una vez expuestos los principios en los que se fundamentan nuestras pruebas, vamos a describirlas brevemente. En cuanto a la *validez de contenido*, nuestras pruebas constan de un conjunto de ítems que son representativos de las operaciones o procesos de pensamiento crítico, y están redactados de un modo sencello y claro, y en un lenguaje coloquial. Nuestras pruebas están formadas por 35 situaciones-problemas, con formato de respuesta abierta, que no exige respuestas técnicas, que supongan alguna formación previa en lógica o estadística, por ejemplo (ver ejemplo 4).

Los procesos de pensamiento representados, lo están en los cinco factores o dimensiones que constituyen las pruebas, a saber: razonamiento deductivo, inductivo, práctico, toma de decisiones y solución de problemas (ver tabla 1). El orden de presentación de los ítems o situaciones cotidianas es aleatorio.

#### Ejemplo 4

#### TAREA DEDUCTIVA

Envisione Proposicional

#### Situación 20: PENCRISAL

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.

#### FORMULACIÓN DEL PROBLEMA

H: hay huelga; T: llegar tarde al trabajo

FALACIA NEGACIÓN ANTECEDENTE (NA

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

#### RESPUESTA CORRECTA

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.

Tabla 1

### FACTORES PENCRISAL/PENTRASAL

DEDUCCIÓN	INDUCCIÓN	R.PRÁCTICO	TD	SP
Proposicional =4 Categórico =3 TOT=7	Causal=3 Comprobación hipótesis =1 Analógico =2 Generalizac =1 TOT=7	Argum. =4 Falacias=3 TOT=7	General=2 Probabilidad=2 Coste Inversión=1 Representación=1 Disponibilidad=1 TOT=7	General=4 Regularidades=2 Medio Fin=1 TOT=7

El proceso de valoración de los ítems o situaciones-problema de respuesta única es sencillo. Se puntúa cada ítem de cero a dos. Cero, para las respuestas incorrectas, uno, para las respuestas correctas con una justificación/ explicación incompleta, y dos, cuando es completa (ver ejemplo 5).

Ejemplo 5

# 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?

#### CORRECCIÓN

- 0 puntos: "En ambos por igual, porque las probabilidades para ambos son las mismas independientemente del tamaño de las muestras"
- Incorrecto
- 1 punto: "Es más fácil en el instituto pequeño, porque habrá menos chicas"
- Explicación imprecisa
- 2 puntos: "En el pequeño, el de 200, porque la probabilidad será más real en Rp y explicación las muestras grandes que en las pequeñas'

correctas

La puntuación global es de 70 (2 X 35), y la de las subescalas de 14 (7 X 2). Las características psicométricas de las pruebas son el ser un test de ejecución máxima y de potencia. El tiempo medio para terminar la prueba es de noventa minutos. La forma de aplicación de las pruebas puede ser individual y colectiva, mediante papel y lápiz o a través de internet. La administración colectiva permite un control de la influencia de factores externos, evita la suplantación de identidad, y estandariza la aplicación. En la aplicación individual se pierde el control de la situación, no se aplica de modo estandarizado, aunque se gana en flexibilidad y disponibilidad de quien la realiza. La aplicación informatizada ofrece la ventaja de que impide dejar ítems en blanco y se puede realizar en varias sesiones.

La forma de puntuar las pruebas busca la cuantificación más precisa; la aplicación informatizada y de manera colectiva reune el mayor número de ventajas.

Lo que nos queda por tratar ahora son los resultados que hemos obtenido con este sistema de evaluación.

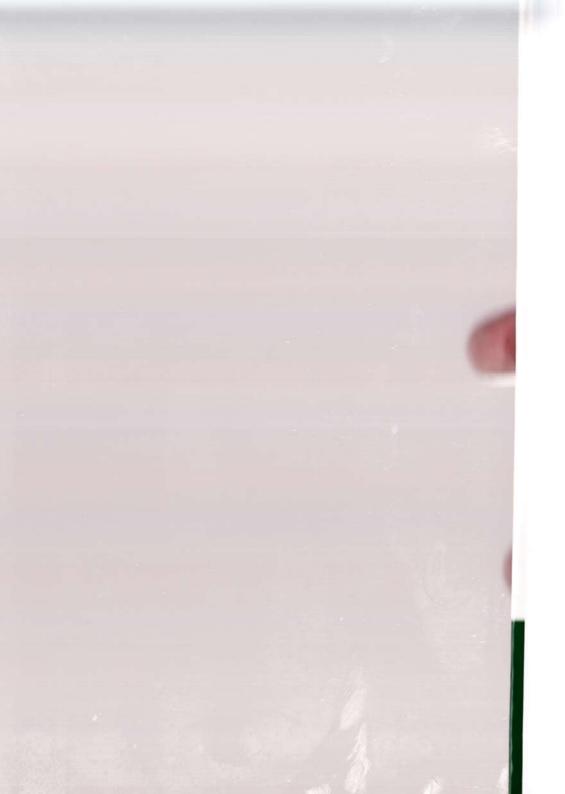
## VI. RESULTADOS Y ESTADO ACTUAL DE LA INVESTIGACIÓN

En este momento se está aplicándose la versión 3 del PENCRISAL. Esta versión ha sido modificada teniendo en cuenta los primeros resultados de la aplicación de la versión anterior. Los resultados de los que disponemos tienen que ver con la medida de la eficacia de nuestra intervención y con las propiedades psicométricas de una de las pruebas, el PENCRISAL. Al utilizar este test en un diseño antes-después, hemos encontrado que la dimensión de inducción y toma de decisiones sí mejoran después de la intervención, no así en deducción y solución de problemas, aunque en esta última sí hay una tendencia en la dirección esperada. La falta de eficacia en razonamiento práctico y en deducción se ha debido a un problema de control. Las propiedades psicométricas de la prueba son razonablemente buenas. Los factores del test explican el sesenta por ciento de la varianza y la fiabilidad entra dentro de los indices aceptables.

El estado actual de la prueba es realmente interesante. La versión 3 se está aplicando en varios países con el fin de validarla en Hispanoamérica. Estamos realizando una fiabilidad interjueces que garantice el sistema de corrección, y obteniendo la validez de criterio con la tarea de rendimiento acumulativa. Quedan aún algunas cosas por terminar antes de poder disponer del instrumento con todos los parabienes metodológicos. Esto lo lograremos a lo largo del presente año.

#### VII. CONSIDERACIONES FINALES

A lo largo de esta comunicación, nos hemos ocupado de justificar nuestro sistema de evaluación en pensamiento crítico, en contextos cotidianos. Las dificultades que acompañan a una empresa tan compleja como esta son



enormes, por lo que cualquier propuesta debe recorrer un periplo de mucho tiempo hasta que se perfila como estable. Nosotros llevamos trabajando ya tiempo en este sistema de evaluación, pero nos queda aún un buen trecho por recorrer. De momento, hemos conseguido disponer de un instrumento fiable para un programa de aprendizaje razonablemente eficaz. A pesar de todo, es mucho el trabajo que aún nos queda por delante, para disponer de un programa de instrucción satisfactoriamente eficiente y un sistema de medida válido para las competencias más universales de pensamiento. 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 de las bifurcaciones buenas, en nuestra senda por mejorar y constatar el desarrollo del pensamiento crítico.