

REVÁLIDAS, EVALUACIÓN DE COMPETENCIAS Y CALIDAD DE LOS APRENDIZAJES*

Ángel Pérez Gómez
Universidad de Málaga

RESUMEN

El artículo se propone analizar y debatir el sentido y fundamento de las propuestas de reforma educativa avanzadas por el Ministerio español de Educación en junio del 2012, en particular la implantación de pruebas externas de evaluación, reválidas, en tres momentos de la vida escolar. Sobre la base de los informes de investigación más actuales se debate que el recurso a los tests, pruebas objetivas o reválidas externas induce generalmente aprendizajes superficiales, es decir, reproducción de datos, fórmulas y algoritmos simples. La evaluación de las competencias o cualidades humanas singulares de cada aprendiz que requiere la complejidad de la era digital contemporánea no cabe, es incompatible con las reválidas, tests y pruebas estandarizadas iguales para todos. Las cualidades más importantes del aprendizaje en la era digital —autodirección, iniciativa, creatividad, pensamiento crítico, solución de problemas y autoevaluación— son difícilmente evaluables mediante instrumentos baratos y masivos «de papel y lápiz», como los tests estandarizados o las reválidas externas, requieren, por el contrario, herramientas y procesos más complejos, plurales y flexibles, como la observación continua de los equipos docentes sobre el quehacer y los productos de los aprendices.

PALABRAS CLAVES: test, evaluación educativa, educación personalizada, calidad de los aprendizajes, evaluación de competencias.

ABSTRACT

«Revalidation, competencies assessment and quality of learning». The paper attempts to analyze and discuss the meaning and substance of education reform proposals advanced by the Spanish Ministry of Education in June 2012, specially the introduction of external evaluation tests, revalidation, in three stages of school life. On the basis of current research and reports discuss how the resources to test, or external «paper and pencil» assessment usually induces only superficial learning, ie reproduction of data, simple formulas and algorithms. The assessment of competencies or unique human qualities of each learner that requires the complex contemporary digital age, is incompatible with the bar exams, and standardized tests for everyone. The most important qualities of learning in the digital age, —self-direction, initiative, creativity, critical thinking, problem solving and self-evaluation—are difficult to evaluate using cheap and massive “paper and pencil” instruments as standardized test, instead require tools and processes more sensible, complex, plural and flexible as the continuous observation of teaching teams about the learners work and products.

KEY WORDS: test, educational assessment, personalized education, quality of learning, competencies assessment.



EVALUACIÓN Y FRACASO ESCOLAR. LAS RAZONES PROCLAMADAS DE UNA CONTRARREFORMA

No es fácil sobrevalorar la importancia de la evaluación en la configuración de toda la vida escolar. La evaluación se constituye en el verdadero y definitivo programa, ya que indica «*lo que realmente cuenta*» en la vida escolar. Pocos dudan en la actualidad, y los estudios internacionales así lo confirman (Mackinsey, 2011, PISA TIMSS¹) que los modos de evaluar los aprendizajes de los estudiantes condicionan sustancialmente los procesos de enseñanza de los docentes, la selección de los contenidos del currículo, la determinación de las prácticas de enseñanza y sobre todo la configuración de las experiencias y estilos de aprendizaje de los estudiantes, así como el clima de relaciones sociales y los ambientes de aprendizaje escolar. Esta consideración, ampliamente compartida por la experiencia y por la investigación, aconseja prestar atención especial a las reformas que ha anunciado el Ministerio de Educación del partido conservador en España en junio del 2012².

Dos parecen ser los motivos explícitamente utilizados por el Ministerio para resucitar las reválidas propias del periodo preconstitucional español, a los 12, 16 y 18 años: el elevado índice de fracaso y abandono escolar (en torno al 30%) y el deterioro del nivel del sistema educativo español que pone de manifiesto los resultados mediocres de los estudiantes españoles en las pruebas internacionales, en particular en PISA.

Con respecto al *primer argumento*, conviene recordar desde el principio que los estudiantes contemporáneos, en su mayoría, no fracasan en la escuela, ni abandonan los estudios de manera prematura, por el nivel de dificultad de una exigencia escolar dura, sino, principalmente, por aburrimiento, por ausencia de interés. La mayoría de las investigaciones internacionales (Darling-Hammond, 2010; Willingham, 2009; Bellanca, 2010; Wagner, 2012) confirman que la relevancia de lo que se trabaja en la escuela se ha convertido en el factor crucial para garantizar la permanencia de los individuos en la escuela, su rendimiento y su desarrollo satisfactorio, especialmente para los estudiantes de niveles socioculturales inferiores.

Una política educativa como la anunciada en dicho documento, que incrementa el nivel de dificultad de las pruebas y exámenes para superar los cursos y sobre todo las etapas, manteniendo el mismo nivel de relevancia e interés de los contenidos y actividades escolares, y sin modificar el resto de los factores que intervienen en

* Fecha de recepción: 01/09/2012; Fecha de aceptación: 30/10/2012.

¹ TIMSS (Trends in International Mathematics and Science Study: Estudio Internacional de Tendencias en Matemáticas y Ciencias).

² Produce cuando menos rubor y vergüenza propia y ajena constatar la extrema debilidad académica del documento que se propone para fundamentar y justificar el borrador de propuestas de reformas educativas publicado en junio de 2012. Es difícil encontrar en la historia de la política educativa española un documento tan pobre, confuso, escaso de argumentos, de apoyos y evidencias científicas y plagado de errores, medias verdades y utilización interesada y poco rigurosa de datos estadísticos nacionales e internacionales.



los procesos de aprendizaje, no puede suponer sino un incremento del absentismo, del fracaso, de la repetición de curso y del abandono temprano. Cuanto mayor sea el índice de repetición, mayor será el desajuste del sistema y la desmotivación de los estudiantes. Por otra parte, ¿realmente se pretende incrementar en España el índice de repetición de curso que se encuentra en torno al 45% al concluir la ESO? La repetición de curso no puede considerarse una estrategia pedagógica eficaz, porque no implica modificación del tratamiento pedagógico para los repetidores, sino separarlos de su grupo de edad y trasladar el problema a los docentes del próximo curso. Ya hay demasiadas evidencias de que esta estrategia por sí misma no conduce sino a incrementar, por lo general, el fracaso y a consumir de modo estéril importantes recursos económicos y humanos (Darling-Harmmond, 2011; OCDE, 2012). Finlandia, el país cuyos estudiantes obtienen las mejores puntuaciones en PISA, ha desarrollado en las últimas décadas la política contraria, evitar decididamente la repetición de curso, salvo casos muy extremos (en torno al 1-2%), interviniendo decididamente ante las primeras señales de dificultad, mediante importantes medidas de apoyo y atención personalizada dentro y fuera del aula.

Pocos dudan ya de que la implicación real y entusiasta del aprendiz en las tareas escolares es la condición fundamental para que florezca el aprendizaje satisfactorio. Sin embargo, la motivación casi exclusiva en el actual panorama escolar español es claramente extrínseca, se relaciona con la necesidad de superar pruebas, exámenes y cursos para adquirir las acreditaciones burocráticas que requiere la carrera académica y profesional habitual de la mayoría de los ciudadanos. Y ésta, precisamente, parece ser la obsesión del gobierno conservador, fortalecer la presencia de la motivación extrínseca mediante el incremento de nuevas pruebas externas o reválidas. A mi entender, una política educativa suicida si lo que nos preocupa es la calidad de los aprendizajes. Para quienes no perciben en el contexto familiar la importancia de los requisitos académicos para su vida personal y laboral futura, precisamente los más desfavorecidos social y culturalmente, la motivación extrínseca se desvanece y el esfuerzo sin sentido del aprendizaje escolar se vuelve insuperable. El absentismo, el fracaso y el abandono se convierten en sus salidas habituales y «naturales». Para quienes gozan de un contexto familiar y social cultivado, la motivación extrínseca les permite superar las pruebas y exámenes pero en poco ayuda a aprender de manera significativa, relevante y creativa.

En segundo lugar, conviene detenerse en el argumento del deterioro del nivel educativo. Al amparo de los mediocres resultados obtenidos por los estudiantes españoles a los 15 años en las pruebas de PISA³, la propuesta del actual gobierno conservador supone una «vuelta a lo básico», a la tradición academicista. Se critica el descenso del nivel educativo que ha supuesto la comprensividad establecida desde

³ Los resultados en comprensión lectora, matemáticas y ciencias, se mueven ligeramente por debajo la media de los países de la OCDE, cerca de países como EEUU, Italia...



la LOGSE⁴, y en particular la política de no favorecer las repeticiones de curso, que conduce, en su opinión, a la relajación y al deterioro del «nivel». Varias flagrantes contradicciones se encierran, a mi entender, en esta posición política, entre las que destacamos las siguientes:

- La primera contradicción supone o mala fe o ignorancia ciertamente culpable. El relativo fracaso en las pruebas de PISA no puede inducir una política de intensificación de los métodos y propuestas que lo han generado. PISA no pretende medir el grado de posesión por parte de los estudiantes de los contenidos del currículum disciplinar y enciclopédico, sino el grado de comprensión y uso de modelos y mapas mentales que utilizan los estudiantes en tres ámbitos básicos: comprensión lectora, pensamiento matemático y pensamiento científico. En cada uno de estos tres campos del saber y del hacer, PISA no solicita la reproducción de datos, informaciones y hechos, aprendidos de memoria, sino la utilización de los mismos en modelos, esquemas o mapas mentales que ayudan al sujeto a orientarse ante problemas verosímiles, más o menos reales y actuales. Utilizar los resultados de PISA para reforzar un manera tradicional de entender la enseñanza y aprendizaje como mera acumulación y reproducción de datos, hechos, informaciones, fórmulas y algoritmos, es, a mi entender, o ignorancia o mala fe, toda vez que los principios que sustentan las pruebas de PISA y sus desarrollos concretos defienden todo lo contrario.
- La segunda contradicción hace referencia a la pretensión de justificar las reválidas o pruebas externas en el prestigio o relevancia que han adquirido las pruebas PISA. Las pruebas PISA tienen un propósito bien concreto y limitado: detectar el desarrollo medio y diferencial de los ciudadanos de diferentes países de las capacidades implicadas en tres ámbitos del saber para resolver problemas y situaciones de la vida cotidiana. Son pruebas muestrales que no pretenden conocer y diagnosticar el desarrollo de individuos concretos, ni explicar los procesos de formación de sus singulares talentos o debilidades, ni acreditar o validar individualmente su rendimiento para autorizar el tránsito al curso o etapa posterior, o prescribir la repetición del mismo. Las reválidas o pruebas externas propuestas en el borrador de Ley son pruebas censales, que abarcan a todos los estudiantes, y tienen importantes efectos de acreditación y prescripción para cada individuo de sus posibilidades académicas sucesivas.

Por otra parte, desarrollar el rigor e independencia de las pruebas PISA en todos los campos del saber y en tres momentos distintos de la vida académica de todos los ciudadanos supondría un coste económico de proporciones tan elevadas

⁴ LOGSE: Ley de Ordenación General del Sistema Educativo, promulgada en 1989 por el gobierno del partido socialista en España.



como estériles serían sus efectos y utilidades, desde el punto de vista educativo, como veremos a continuación.

El problema del abultado volumen de fracaso y abandono en España radica, a mi entender, en la irrelevancia y obsolescencia de la escuela convencional que permanece en nuestro contexto. La escuela academicista actual, aunque en un estadio más elaborado y sofisticado, sigue el mismo esquema de la escuela industrial, por ello difícilmente puede responder a las exigencias de un mundo ya no mecanizado, sino abierto, flexible, cambiante, creativo e incierto. Lo que se enseña y aprende en la escuela convencional: un currículum de carácter enciclopédico y saturado de datos, fragmentado en disciplinas desconectadas entre sí, y de los problemas reales y actuales, que se aprende de manera generalmente memorística y mecánica, para reproducir fielmente en pruebas de papel y lápiz, no prepara a los ciudadanos para comprender, ni para intervenir en el mundo complejo y cambiante en el que viven y, por ello, difícilmente motiva a aprender por el placer y el sentido de aprender, descubrir, crear o resolver problemas.

CURRÍCULUM ENCICLOPÉDICO DE TALLA ÚNICA

Si la evaluación determina en gran medida lo que se aprende y cómo se aprende, cabría preguntarnos: ¿Qué merece la pena enseñar y evaluar en la era digital? La escuela convencional ha generado progresivamente, a lo largo del siglo xx, un currículum monstruoso, que pretende abarcar todos los ámbitos del saber en su desarrollo actual. Un currículum enciclopédico de extensión ilimitada y de escasa profundidad. Como lo expresa Toni WANGER (2012), un currículum con kilómetros de extensión, por milímetros de profundidad. Es a todas luces una pretensión inalcanzable que solamente se puede perseguir a costa de la profundidad, desarrollando un modo superficial de relación con el saber, confundiendo e identificando el conocimiento con la reproducción memorística de datos, fechas, informaciones, fórmulas y algoritmos.

En la era de la información global y digital, esta pretensión además de estéril me parece inaceptable. Por una parte, el crecimiento del conocimiento es exponencial, se duplica aproximadamente cada dos años (Riegle, 2007; Davidson, 2011), por tanto es inútil la pretensión de que toda la información quepa en el currículum escolar. Por otra, la acumulación ilimitada y reproducción fiel de los datos es una tarea a la medida perfecta de las herramientas tecnológicas actuales, accesibles fácilmente y en cualquier momento para cualquier aprendiz.

La escuela convencional por su insistencia en un currículum enciclopédico que abarque todo el contenido de los diferentes campos del saber prioriza aquellas estrategias de aprendizaje que mejor permiten abarcar la superficie extensa del currículum, es decir, la memorización de datos, hechos, informaciones, clasificaciones y algoritmos, aprendizaje superficial del eslabón inferior del conocimiento (Robinson, 2011; Darlin-Hammond, 2011; Bellanca, 2012; Biggs, 2007; Sunderman, 2008; McThiege y Seif, 2011; Wagner, 2012). Pero memorizar no significa necesariamen-



te comprender, y menos transferir, aplicar y recrear, es decir, no ayuda a generar aprendizajes de orden superior de los componentes más elaborados del conocimiento.

El incremento de las exigencias académicas sin sentido para cada estudiante, el recurso a la repetición de curso, la implantación de pruebas externas censales en tres momentos de la vida escolar, o la anticipación de la segregación de los estudiantes en función de su rendimiento académico, para nada mejoran el problema porque no inciden en el origen del mismo ni atacan a los factores que lo provocan. Como pone de manifiesto un volumen cada vez más importante de estudios e informes (Willingham, 2009; Richardson, 2011; Gato, 2005; McKinsey 2007; Pérez Gómez, 2012), los factores que amenazan la calidad y relevancia de la vida escolar pueden sintetizarse en los siguientes: la irrelevancia de lo que se enseña y aprende en la escuela para la gran mayoría de los fracasados; el currículum enciclopédico, fragmentado en disciplinas, de talla única; la metodología de enseñanza uniforme, a través de la transmisión, que provoca que todo el alumnado estudie los mismos contenidos, al mismo ritmo, con el mismo método y con idénticos recursos; así como las formas, instrumentos y procedimientos de evaluación mediante exámenes y pruebas idénticas para todos, de papel y lápiz, que requieren fundamentalmente la reproducción.

LA EDUCACIÓN PERSONALIZADA

Un abrumador volumen de datos de la investigación pedagógica y didáctica (Darling-Hamon, 2011; Stodart, 2010; Elmore, 2011; Davidson, 2011; Sizer, 2000; Sandler, 2012; Stenhouse, 1987; Wagner, 2012) están insistiendo desde hace varias décadas en la necesidad de proponer la enseñanza personalizada como la estrategia didáctica más satisfactoria y eficaz, capaz de responder a las exigencias del complejo y cambiante contexto contemporáneo, al mismo tiempo que a las diferencias individuales de intereses, estilos de aprendizaje, ritmos y motivaciones de cada estudiante. La diferenciación personalizada⁵ del currículum escolar desde épocas muy tempranas induce y estimula la autonomía en el desarrollo personal, de modo que cada individuo vaya elaborando y construyendo su propio y singular proyecto de vida, personal, social y profesional. La propuesta actual del Ministerio conservador en España, pretende mantener el modelo de talla única hasta los 15 años, y en ese momento provocar la segregación en dos vías con prestigio social, profesional y

⁵ Me parece imprescindible distinguir la diferenciación personalizada de la diferenciación grupal. La diferenciación cuando se utiliza para segregar y clasificar grupos humanos en función de características o estereotipos, ya sea en función del sexo, como pretende ahora el Ministerio de Educación, o del color de la piel, o del nivel de rendimiento estandarizado..., se convierte en una estrategia pedagógica que conduce fácilmente a la discriminación y desigualdad. La diferenciación educativa que permite y estimula la personalización del aprendizaje de cada uno de los individuos, promoviendo el desarrollo individual diferenciado y singular dentro del mismo grupo social heterogéneo de individuos diferentes, puede considerarse por el contrario la estrategia pedagógica más adecuada para fomentar el desarrollo de cada individuo hasta el máximo de sus posibilidades.



cultural claramente desequilibrado, manteniendo en cada una de las vías el mismo modelo de talla única y uniformidad didáctica que ahoga la personalización.

La enseñanza personalizada se propone dar cabida a las diferencias individuales sin provocar discriminación y desigualdad, es decir, ofreciendo dentro del mismo grupo natural de desarrollo oportunidades diferenciadas de valor equivalente, permitiendo y estimulando la búsqueda de la propia identidad singular, el propio talento, la propia pasión (Robinson, 2008). Las reválidas y pruebas externas de papel y lápiz, que imponen estándares comunes hasta los 18 años, responden al primer modelo pedagógico uniforme y de talla única, que excluye y expulsa a los más desfavorecidos e impide el desarrollo personalizado de cada uno de los estudiantes. La enseñanza personalizada no cabe, es incompatible con las reválidas, tests y pruebas estandarizadas iguales para todos, requiere por el contrario un sistema de evaluación sensible a las diferencias individuales, atento a la singularidad y originalidad de los desarrollos y productos de cada aprendiz. Como afirma Linn STODARD en su interesante ensayo *Human Greatness* (2010), la misión que exige la escuela actual a los docentes es una misión imposible: imponer un currículum homogéneo para uniformizar el desarrollo de los estudiantes en una era global, cambiante y diversificada. En sentido convergente se expresa ELMORE, en una reflexión sobre su vida profesional (Elmore, 2011), al afirmar con énfasis que lo que más le ha sorprendido en su dilatada experiencia profesional es la multiplicidad de significados diferentes que despierta en los estudiantes un mismo proceso, contenido o tarea de aprendizaje. La diversidad de interpretaciones reafirma la heterogeneidad del desarrollo humano y niega la creencia en la conexión simplista y casi automática entre lo que se enseña y lo que se aprende.

La educación estandarizada, convencional, no sólo aburre a los estudiantes, sino que también les prepara para trabajos que ya no van a existir en el futuro. El modelo de talla única, uniforme, es un modelo fijado en el pasado, obsoleto, propio de la ideología y de la metodología de las cadenas de montaje de la era industrial, basada en la división estricta del trabajo mecanizado y en la jerarquía de las relaciones laborales y sociales. La mayoría abrumadora de las investigaciones (Novak, 1998; Vergnaud, 1990; Pérez Gómez, 2012; Bellanca, 2010; McThiege y Seif, 2010; Wagner, 2012) confirman reiteradamente que el aprendizaje es un proceso permanente, apoyado en aprendizajes previos, no un producto acabado, que los aprendizajes requeridos en la era digital son los aprendizajes de orden superior y no los meramente reproductivos, que los aprendices son diferentes desde el punto de vista cultural, genético, social, emocional e intelectual, tienen y desarrollan recursos personales singulares de aprendizaje, modos de percibir, organizar, reaccionar y evaluar claramente diferentes unos de otros, diversificados y cambiantes, y que precisamente la riqueza del individuo y de la sociedad reside en el respeto y fomento de la diversidad creativa de todos y cada uno de sus miembros.

Así pues, supone una sorprendente paradoja, como destaca DAVIDSON (2011), que cuando los aprendices contemporáneos tienen fácil acceso a la ilimitada y libre exploración *on-line*, cuando necesitan aprender habilidades de valoración y pensamiento crítico de orden superior, para ir formando sus criterios de discriminación, valoración, selección y propuesta de alternativas, la escuela priorice y a veces



reduzca toda su tarea a formar y evaluar capacidades de pensamiento reproductivas, de orden inferior. Las calculadoras relativizaron, en su día, la importancia del aprendizaje memorístico de algunos aspectos del cálculo, y los motores de búsqueda como Google o las enciclopedias virtuales como Wikipedia han convertido en obsoletos los aprendizajes memorísticos de meros datos y las pruebas de elección múltiple, hoy día escasamente útiles, excepto para el juego del «Trivial» o para los concursos de TV.

Para tomar en consideración la dimensión personalizada del proceso educativo, habría que hablar más de oportunidades de valor equivalente y adaptadas a las peculiaridades que singularizan a cada sujeto, que de igualdad de oportunidades, tan frecuentemente interpretada como homogeneidad y uniformidad de contenidos, métodos, ritmos y evaluaciones, por la pedagogía de talla única. Imponer un traje de talla única para todos en la escuela, provoca, por una parte, el fracaso y la exclusión de los que no logran adaptarse a la talla predeterminada y, por otra, que la mayoría de los ciudadanos concluyan la escuela sin desarrollar la mayoría de sus talentos, facultades y capacidades singulares y creativas.

LA EVALUACIÓN EDUCATIVA DE LOS APRENDIZAJES. EVALUACIÓN Y CALIFICACIÓN

Un currículum enciclopédico de extensión ilimitada, cuyo aprendizaje memorístico hay que comprobar conduce por lo general a un instrumento de calificación sencillo y económico, de papel y lápiz, denominado prueba objetiva de respuestas cortas y normalmente de elección múltiple. Conviene destacar con DEDE (2007) que estas herramientas pueden medir con fiabilidad pero miden por lo general el objeto erróneo o el objeto irrelevante, aprendizajes superficiales: ¿En qué año y quién descubrió América?, títulos y fechas de las obras de García Lorca, por ejemplo. No hay manera de detectar mediante tales pruebas el desarrollo de las capacidades mentales de orden superior, las cualidades o competencias que requiere la formación del ciudadano contemporáneo, como por ejemplo: observar, indagar, analizar y valorar los factores que están condicionando la contaminación atmosférica y el deterioro del medio ambiente.

Un análisis de las implicaciones políticas y pedagógicas de los procesos externos de evaluación requiere, a mi entender, diferenciar claramente dos términos y conceptos que se confunden en el uso cotidiano, incluso por parte de los profesionales implicados en los procesos de enseñanza y aprendizaje: evaluación y calificación. Evaluación, evaluación educativa o evaluación formativa, refiere a un proceso complejo y lo más completo posible y flexible de diagnóstico, de descripción e interpretación, del desarrollo de los individuos, sus cualidades, sus fortalezas y debilidades, el grado de configuración actual de cada uno de los componentes de las competencias —conocimientos, habilidades, emociones, actitudes y valores—, así como del funcionamiento de las competencias como conjuntos, como sistemas de comprensión, toma de decisiones y actuación. La evaluación educativa concluye en un informe comprensible y accesible a todos, pero principalmente a la persona



evaluada, para que la información aportada pueda ayudar a tomar decisiones y autorregularse en el futuro, a corto, medio y largo plazo.

Calificación o evaluación sumativa es la concreción, la reducción del diagnóstico a una categoría, numérica o verbal, para facilitar la comparación, la clasificación y la selección de los individuos. Todo proceso de reducción de un análisis e informe complejo y completo a una categoría supone inevitablemente una deformación perversa de las posibilidades del diagnóstico educativo. La categoría no sólo pierde la riqueza de la información descriptiva de una situación o momento de desarrollo, también pierde la posibilidad de entender procesos, factores intervinientes, y por tanto proponer formas de mejorar o rectificar los errores, lagunas o deficiencias. La calificación solamente se justifica, a mi entender, cuando necesitamos clasificar a los diferentes componentes de un grupo humano a efectos de poder seleccionar a los más adecuados para una tarea, programa, puesto o responsabilidad. Cuando, como en la enseñanza obligatoria —que se propone la formación básica, no profesional, de los ciudadanos—, no hay ninguna necesidad de seleccionar, pues deseamos que todos se formen de manera singular y diferenciada hasta el límite de sus posibilidades, tampoco hay ninguna necesidad de clasificar y por tanto de calificar. En la educación básica y obligatoria será necesario fomentar la evaluación, la evaluación formativa y reducir e incluso erradicar la calificación.

La evaluación educativa —aquella que ayuda a conocerse para formarse de manera autónoma— requiere que los aprendices entiendan lo que están aprendiendo y el sentido, utilidad y valor de lo que aprenden, y reciban el *feedback* necesario para valorar cómo lo están haciendo así como el apoyo requerido para saber cómo hacerlo mejor en el futuro (JAMES, 2007; FENWICK y PARSONS, 2009).

¿Cómo valorar de manera suficientemente sostenible y válida el desarrollo de los componentes básicos de las competencias o cualidades humanas de los estudiantes? El problema de toda evaluación radica en establecer los criterios de valoración y elaborar los instrumentos de diagnóstico. La enorme complejidad que implica la medición del desarrollo humano ha conducido, por lo general, a estrechar el objeto de valoración a lo que es fácilmente medible, lo que cuesta menos medir y al mismo tiempo permite una medición más fiable. Son precisamente estos objetos de medición, relacionados con las respuestas exactas en las pruebas de papel y lápiz, los que inducen la perversión de los procesos de evaluación, pues desnaturalizan la comprensión del desarrollo de las cualidades humanas a la mera reproducción memorística de informaciones y datos, dejando fuera más del 90% de los componentes que constituyen las competencias básicas de los seres humanos, es decir, los recursos que utilizan en la vida cotidiana (conocimiento, habilidades, actitudes, emociones y valores), como sistemas de comprensión, toma de decisiones y actuación. Digámoslo un vez más, ni la educación personalizada ni por tanto el desarrollo de competencias o cualidades humanas cabe o es compatible con los tests, pruebas objetivas o reválidas. ¿Cómo medir las emociones, las habilidades, los procesos, los comportamientos complejos, las actuaciones creativas en objetivas y económicas pruebas de papel y lápiz? (Reeves, 2010; Dede, 2010).



LA PERVERSIÓN DE LOS TESTS Y PRUEBAS OBJETIVAS

En principio, hay acuerdo generalizado en que los tests y las denominadas pruebas objetivas miden normalmente los procesos de pensamiento de orden inferior, el escalón inferior del conocimiento, la reproducción de datos y la ejecución de algoritmos simples, precisamente las tareas que hoy pueden hacer perfectamente las máquinas digitales, pero eluden o son ciegos a los procesos mentales de orden superior: comprensión, indagación, valoración, creatividad e innovación, cualidades que los ordenadores, por ahora, no pueden replicar. Con demasiada frecuencia, los tests y pruebas de evaluación externa enfatizan los hechos y datos y no los conceptos, esquemas y modelos explicativos o interpretativos, favorecen la memorización más que el pensamiento lógico, los hechos y datos sin contexto, los detalles desconectados del análisis, la uniformidad y despersonalización sobre la idiosincrasia y diversidad⁶. Ahora bien, como ya he indicado anteriormente, el aprendizaje y reproducción de los hechos no garantiza sino un aprendizaje superficial. El aprendizaje profundo y significativo requiere que los hechos se interpreten sobre la base de modelos conceptuales, tomando en consideración el impacto de las variables sociales, históricas, políticas, lingüísticas y culturales que rodean la vida de los individuos.

Así pues, estas evaluaciones externas basadas en amplias muestras y pruebas de papel y lápiz no pueden ni entenderse ni proponerse como evaluaciones formativas. Su pretensión legítima se circunscribe a proporcionar datos que ayudan a elaborar diagnósticos parciales de situación, estimar algunas fortalezas y debilidades que demuestran los aprendices en ciertos ámbitos del saber, suponen llamadas de atención, identifican deficiencias, síntomas que hay que interpretar y en el mejor de los casos estimulan la reflexión de los implicados para indagar por debajo de tales síntomas e identificar las posibles causas y proponer estrategias de mejora. En definitiva, constituyen una evaluación sumativa que, en el mejor de los casos, puede inducir o desencadenar una evaluación formativa.

Además, las cualidades más importantes del aprendizaje en la era digital —autodirección, iniciativa, creatividad, pensamiento crítico, solución de problemas y autoevaluación— son difícilmente evaluables mediante instrumentos baratos y masivos «de papel y lápiz», como los tests estandarizados y difícilmente estimulables mediante la pedagogía transmisiva y reproductora (McThiege y Seif, 2011).

¿Cómo medir y valorar los procesos de aprender, desaprender y volver a aprender, que constituyen, en mi opinión, la esencia de los procesos educativos?

La obsesión contemporánea por la rendición de cuentas mal entendida, en términos de resultados fácilmente medibles en tests, reválidas o pruebas externas no sólo

⁶ ROBINSON (2011) considera que existen cuatro problemas esenciales relacionados con la evaluación educativa: –Un énfasis desmedido en la evaluación sumativa, en forma de test. –Un énfasis desproporcionado en resultados medibles para establecer comparaciones. –La dificultad y complejidad de todo proceso de evaluación de la creatividad. –La presión que en muchos países ejercen las evaluaciones nacionales sobre los docentes y las escuelas.



olvida la finalidad educativa de la escuela sino que margina también la enseñanza de los contenidos disciplinares y restringe los procesos de enseñanza y aprendizaje a una preparación academicista para superar las pruebas o los tests. Más que explicar y trabajar contenidos, se entrenan habilidades para superar con éxito los posibles ítems de una previsible prueba, mediante repetición de los ítems de pruebas anteriores (Engel y Meier, 2010).

El «*Salthousestudy*» (1989) hizo una crítica exhaustiva y bien fundamentada de las contradicciones y carencias de fundamentación de los tests desde Kelly y Binnet, y concluye que en la actualidad en EEUU gran parte de la enseñanza se reduce a entrenamiento para pasar contradictorias, inconsistentes e inconclusivas pruebas de pensamiento de orden inferior. Incluso Diane RAVITCH, responsable clave en la configuración de las políticas educativas basadas en los tests bajo la presidencia de Clinton y Bush, ha criticado recientemente (Ravitch, 2010) y de forma despiadada el valor de los tests, denunciando el desastre que han comportado en el sistema educativo de EEUU. La situación es tan grave y tan ridícula que, además del desconsuelo y desencanto general, está provocando que muchas escuelas innovadoras estén formando cualidades de orden superior de septiembre a marzo, y dediquen los últimos meses de cada curso a desarrollar las capacidades de orden inferior, memorización, que requieren los tests⁷.

RESNICK (2007) es de la misma opinión cuando afirma que los tests han sustituido al currículum en las escuelas de EEUU, y en consecuencia, cuanto más se implantan las pruebas estandarizadas de evaluación objetiva, en nuestro caso serían las reválidas, menor es el desarrollo en los estudiantes de las capacidades de pensamiento de orden superior que requiere la compleja y cambiante vida contemporánea en la era digital. Con demasiada frecuencia el rigor en los procesos de evaluación se ha identificado con el rigor en la definición y concreción de estándares de valoración⁸. A mi entender no es el rigor formal lo que cualifica y proporciona valor a los estándares, sino la calidad y relevancia de su contenido. Podemos, como suele ser

⁷ Como analizan acertadamente Angela ENGEL y Debora MEIER, en «*Seed of tomorrow*» (2010), los tests como absolutos científicos han recuperado en EEUU la popularidad que alcanzaron en los años 20. Actualmente se utilizan para medir el rendimiento de los estudiantes, indicar la calidad de una escuela, la eficacia de los docentes, la validez de las reformas educativas, y el nivel del salario de los docentes. Hay dos tipos de tests estandarizados: los tests de rendimiento en las diferentes disciplinas y los tests de habilidades cognitivas, tests de aptitudes, que pretenden medir los procesos de pensamiento. En las escuelas de EEUU, los primeros son los que dominan de forma absoluta el panorama de la evaluación. Conviene recordar con ENGEL y MEIER (2010), que en los ocho años de desarrollo del programa NCLB, *No Child Left Behind*, implantado por Bush en EEUU, miles de escuelas han sido reformadas o suprimidas, el currículum se ha estrechado y el número de abandonos se ha incrementado.

⁸ REEVES (2010) afirma al respecto que la naturaleza de los tests y pruebas objetivas, con sus condiciones de estandarización, secreto y resultados exclusivamente individuales, supone la antítesis del nuevo marco de evaluación educativa caracterizado por la necesidad de comprensión, exploración, creatividad y colaboración. No es posible reconciliar las demandas de capacidades del siglo XXI con las condiciones del contexto tradicional de evaluación mediante pruebas objetivas y tests.



habitual, medir muy bien y con rigor el contenido erróneo o irrelevante. Cuanto más se sube en la escalera de estándares fijos y uniformes de evaluación mediante tests, más baja el nivel de formación de los estudiantes para afrontar las exigencias de la compleja era digital⁹. ¿Es éste el modelo que se nos propone en la reforma actual?

A diferencia de la evaluación estandarizada mediante tests, la evaluación educativa, auténtica considera que la mejor estrategia de evaluación, más rigurosa y más justa, es aquella que utiliza pluralidad de instrumentos y *procedimientos congruentes con el sentido* de los procesos de aprendizaje y las finalidades deseadas: ensayos, entrevistas, trabajos y proyectos, observación y tutorización continua del docente, portafolios, exposiciones orales, diarios y cuadernos de campo, procesos de autoevaluación, seminarios de debate y reflexión... (Stobart, 2010; Soto Gómez, en prensa; Wiggins, 1996; Monereo, 2008). Conviene destacar, hasta el límite de la reiteración, que la calidad de la educación no se reduce a lo que se puede medir, los criterios y estándares públicos que realmente merecen la pena no necesariamente deben ser fáciles de medir y cuantificar¹⁰.

REVÁLIDAS Y EVALUACIÓN EDUCATIVA. RECUPERAR LA CONFIANZA EN LOS DOCENTES

La obsesión por las calificaciones deteriora las posibilidades educativas de la evaluación concebida como diagnóstico para ayudar a la mejora de los procesos de aprendizaje. Con demasiada frecuencia en nuestro panorama escolar las recomendaciones de evaluación continua se han mal entendido o convertido en exigencias de examen y calificación permanente. El aprendizaje deja de tener sentido en sí mismo y se convierte en un medio para conseguir las calificaciones deseadas. Ahora bien, la escuela y el docente que se apoyan exclusivamente en la fuerza de los exámenes y calificaciones para estimular o forzar el aprendizaje de su alumnado han tirado la toalla. El aprendizaje así forzado tiene escaso o nulo valor educativo porque impide que los estudiantes descubran el valor y la aventura del saber, la pasión por el descubrimiento, la utilidad de los saberes para mejorar la vida y la convivencia, así como la magia de las experiencias de apropiación y recreación de la cultura, el arte, la ciencia.

Evaluar competencias fundamentales requiere *evaluar sistemas de comprensión y acción* y, por tanto, evaluar actuaciones, a sabiendas de la complejidad de elementos

⁹ Me detengo en el análisis de la realidad en EEUU en este aspecto, por el enorme poder que tiene en educación para marcar tendencias, y porque ha anticipado hace 15 años lo que ahora se propone hacer el Ministerio del gobierno conservador, resucitando las reválidas.

¹⁰ Puede consultarse al respecto la página web creada por Jon MUELLER en Illinois sobre herramientas de evaluación auténtica, donde proporciona rúbricas y estándares para medir y mejorar la comprensión de los estudiantes. Es especialmente recomendable el artículo «¿Cómo construir evaluaciones auténticas?». <http://jonathan.mueller.faculty.noctrl.edu/toolbox/>. En el mismo sentido cabe recomendar el excelente libro de Tara FENWICH y James PARSONS (2009): *El arte de la evaluación*, donde puede encontrarse de forma clara y sencilla desde los argumentos teóricos más fundamentales hasta las estrategias e instrumentos que pueden ayudarnos a promover la evaluación educativa.



presentes en las actuaciones humanas. Implica, sin duda, la búsqueda de nuevos propósitos y la utilización de nuevos modelos, estrategias e instrumentos de evaluación, adecuados para *captar la complejidad*, más allá de las convencionales pruebas de papel y lápiz. La evaluación de competencias requiere cuestionar el valor y sentido de las calificaciones y enfatizar el valor de los informes detallados y completos de diagnóstico sobre los procesos, los resultados y los contextos.

Por tanto, la mejor manera de desarrollar la evaluación continua y *formativa* es la atención y tutoría continua del equipo de docentes que acompañan y orientan el aprendizaje de cada estudiante, no las reválidas ni las pruebas externas puntuales. Estas herramientas, por su propia naturaleza, no pueden utilizar el requerido y extenso abanico de múltiples y complementarios procedimientos de evaluación educativa, que, además de ayudar a aprender y a mejorar a cada estudiante, permiten elaborar un diagnóstico más sostenible y fiable del grado de desarrollo tanto de los elementos que componen las competencias o cualidades humanas fundamentales de cada estudiante como de su funcionamiento integrado como sistemas de comprensión, toma de decisiones y actuación. Si lo que merece la pena estimar es el desarrollo de las cualidades humanas o competencias básicas, el instrumento más adecuado nunca serán las pruebas puntuales de papel y lápiz, sino la observación continua de los docentes, que como equipo de profesionales detectan las fortalezas y debilidades de los distintos componentes que conforman las competencias desarrolladas por cada aprendiz. Así pues, las reválidas ni sirven para ayudar a aprender, función educativa de la evaluación, ni sirven para diagnosticar el grado de desarrollo de las competencias en un momento dado, función sumativa y calificadora de la evaluación. Miden en todo caso el nivel de desarrollo de habilidades de orden inferior, memorizar y reproducir datos y fórmulas, hoy día claramente prescindibles o sustituibles.

En España ya experimentamos en su día el nulo valor pedagógico de las reválidas (Libro Blanco, 1970) y hemos continuado comprobando el carácter estéril de otra prueba externa obsoleta, «la selectividad», abandonada ya como procedimiento de selección por parte de las universidades más prestigiosas del contexto internacional.

En definitiva, en mi opinión, la estrategia para intentar disminuir drásticamente el abandono temprano no es incrementar el nivel de dificultad, ni disminuirlo, ni incrementar las repeticiones de curso, ni establecer itinerarios grupales, ni saturar el currículum obligatorio de pruebas externas o reválidas, como plantea el borrador del Ministerio, sino priorizar la pertinencia y relevancia de contenidos y métodos, es decir, fortalecer los programas personalizados que se comparan con los proyectos personales de cada estudiante, y sustituir los estándares homogéneos y uniformes por criterios de exigencia académica sensibles a las singularidades, a las diferencias, a las trayectorias diferenciadas de valor equivalente.

Los países con mejores resultados académicos en las pruebas internacionales (PISA), como Finlandia, no tienen ninguna prueba externa; países con resultados mediocres, como EEUU, tienen pruebas externas todos los años de escolarización. Son, por tanto, otras las variables y factores que pueden explicar las diferencias en el rendimiento. Las pruebas externas, estandarizadas, de papel y lápiz, miden en todo caso el grado de «fiebre» o «temperatura corporal académica» de cada estudiante,



pero son incapaces de detectar y diagnosticar las causas de la misma y tampoco pueden por tanto proponer y desarrollar tratamientos adecuados.

REFERENCIAS

- BELLANCA, J. (2010). *21st Century Skills: Rethinking how student learn*. Bloomington. Solution Tree Press.
- (2012) *How to Teach Thinking Skills Within the Common Core*. Bloomington, Solution Tree.
- BERRY, B. (2011) *Teaching 2030*. N.Y. Teacher College Press.
- BIGGS, J. (2007). *Teaching for Quality Learning at University*, Buckingham: Open University Press/ McGraw Hill.
- CARVALHO, L. (2009). Production of OECD's «Programme for International Student Assessment» (PISA) Knowledge and Policy. www.knowandpol.eu.
- COSTA, A.L. y KALLICK, B. (2004). *Assessment Strategies for Self-directed Learning*. N.Y. Corwin Press.
- DARLING-HAMMOND, L. (2010). *The flat world and education*. Nueva York: Teachers College Press.
- DAVIDSON, C. (2011). *Now You See It: How the Brain Science of Attention Will Transform the Way We Live, Work, and Learn*. Nueva York: Penguin Books.
- DEDE, C. (2007). *Transforming Education for the 21st Century*. Cambridge: Harvard Education Press.
- DEDE, Ch. (2010). Comparing frameworks for 21st Century Skills. En James Bellanca, 2010, *op. cit.*
- ELMORE, R. (2011). *I used to think... And now I think...* Cambridge: Harvard Education Press.
- ENGEL, A. y MEIER, D. (2010). *Seed of tomorrow*. New York, Paradigm Publishers.
- GATTO, J.T. (2005). *Dumbing Us Down: The Hidden Curriculum of Compulsory Schooling*. Canadá: New Society Publisher.
- HOWE, K.R. y ASHCRAFT, C. (2005). Deliberative democratic evaluation: Successes and limitations of an evaluation of school choice. *Teachers College Record*, 107(10), 2274-2297.
- JAMES, M. (2007). *Learning how to learn*. Londres: Routledge.
- FENWICK, T y PARSONS, J. (2009). *The art of Evaluation*. EEUU: Thomson Educational Publishing.
- ROCHEX, J.Y. (2006). Social, Methodological and Theoretical Issues Regarding Assesment: Lessons From a Secondary Analysis of PISA 2000 Literacy Tests. *Review of Research in Education*, 2006.
- KAHNEMAN, D. (2011). *Thinking fast and Slow*. New York. Farrar, Straus and Giroux.
- MCKINSEY and Colbs. (2007). *How the World's Best-performing School Systems Come Out on Top*. Londres: McKinsey & Co.
- MCTIGHE, J y SEIF, E. (2010). An implementation Framework to Support 21st Century Skills. En James Bellanca, 2010, *op. cit.*
- MONEREO, C. (2009). *PISA como Excusa. Repensar la evaluación para cambiar la enseñanza*. Barcelona. GRAO.
- NEVO, D. (2002). Dialogue evaluation: Combining internal and external evaluation, en D. Nevo (ed.), *School-based evaluation: An international perspective* (pp. 3-16). Oxford, UK: Elsevier.
- NOVAK, J.D. (1998). *Learning, Creating and Using Knowledge: Concept maps as facilitative tools for schools and corporations*. Mahwah, N.J.: Lawrence Erlbaum & Assoc.



- OCDE, (2005): *Informe PISA, 2003: Aprender para el mundo del mañana*. OCDE, Madrid, Santillana.
- PÉREZ GÓMEZ A.I., SOTO GÓMEZ E., SOLA FERNÁNDEZ, M., SERVÁN NÚÑEZ, M^a.J. (2008). *La evaluación como aprendizaje*. Guías para el Espacio Europeo de Educación Superior, núm. 6. Akal. Madrid.
- (2012). *Educarse en la era digital*. Madrid, Morata.
- (2008). ¿Competencias o pensamiento práctico? La construcción de los significados de representación y de acción. En Gimeno y V.A. *Educación por competencias, ¿qué hay de nuevo?* Morata, Madrid.
- (2007). La naturaleza de las competencias básicas y sus aplicaciones pedagógicas. *Cuadernos de Educación*. 1. Gobierno de Cantabria.
- RAVITCH, D. (2010). *The Death and Life of the Great American School System: How Testing and Choice Are Undermining Education*. Nueva York: Basic Books.
- REEVES, D. (2010). A framework for Assessing 21st Century Skills, en James Bellanca, *op. cit.*
- RESNICK, L.B. y MATSUMURA, L.C. (2007). «Academic proficiency: Bright hopes, blurry vision». *Voices in education*, 14.
- RICHARDSON, W. (2011). «Investing in Teachers as Learners». *Education Week*, 30, pp. 21-24.
- RIEGLE, R. (2007). *Education in the Information Age*, en: <http://people.coe.ilstu.edu/rpriegle/www-docs/educationinfoage.thm>.
- ROBINSON, K. (2008). *El elemento*. Barcelona, Grijalbo.
- ROBINSON, K. (2011). *Out of our minds: Learning to be creative*. United Kingdom: Capstone Publishing.
- SALTHOUSE, T.A. (1989). Aging and skilled performance. En A.M. Colley & J.R. Beech (eds.), *Acquisition and performance of cognitive skills* (pp. 247-263). Chichester, UK: John Wiley & Sons.
- SANDLER, S. (2012). People vs. 'Personalization' Retaining the human element in the high-tech era of education. *Edweek*, 29/02/2012.
- SIZER, Th. (2005). *The Red Pencil: Convictions from Experience in Education*. N.Y., Donnelly and Sons.
- SOTO GÓMEZ, E. (en prensa). «La reflexión, corazón y alma del portafolios educativo». En PÉREZ GÓMEZ y varios: *La función pedagógica del portafolios educativo*. Madrid, Akal.
- STENHOUSE, L. (1987). *La investigación como base de la enseñanza*. Madrid, Morata.
- STOBART, G. (2010). *Tiempos de pruebas: Los usos y abusos de la evaluación*. Madrid, Morata.
- STODDARD, L. (2010). *Educating for human Greatness*. Florida: Peppertree Press.
- SUNDERMAN, G. (2008). *Holding NCLB accountable: Achieving accountability, equity, & school reform*. Thousand Oaks, CA: Corwin Press.
- SUSKIE, L. (2001). *Assessment to Promote Deep Learning*. Virginia: Stylus Publishing.
- VERGNAUD, G. (1990). La théorie des champs conceptuels. *Récherches en Didactique des Mathématiques*, 10(23): 133-170.
- WAGNER, T. (2012). *Creating Innovators*. New York. Scribner.
- WIGGINS, G.P. (1996). Practicing what we preach in designing authentic assessments. *Educational Leadership*, 54(4): 1-25.
- WILLINGHAM, D. (2009). *Why doesn't Student like Schools*. Nueva York: Wiley. (Traducción castellana en GRAO.)

