

LA UTILIDAD DE LAS EVALUACIONES DE SISTEMAS EDUCATIVOS: EL VALOR DE LA CONSIDERACIÓN DEL CONTEXTO¹

The usefulness of educational systems evaluations: the value of the consideration of context

ARTURO DE LA ORDEN HOZ
Universidad Complutense de Madrid

JESÚS M. JORNET MELIÁ
Universitat de València

En este artículo se hace una revisión de los procedimientos de evaluación de sistemas educativos. A partir de esta revisión, se analizan algunos factores relacionados con la baja utilidad de este tipo de evaluaciones. El análisis se dirige hacia el diseño deficiente de los cuestionarios de contexto. Se observa que los cuestionarios de contexto no llegan a explicar el rendimiento. Su baja capacidad de explicación del rendimiento es la causa de la baja utilidad de las evaluaciones de sistemas educativos; pues no permiten establecer recomendaciones de mejora. Se propone una base teórica y metodológica alternativa para diseñar los cuestionarios de contexto. Este tipo de metodología para diseñar los cuestionarios de contexto para la evaluación de sistemas educativos tiene como finalidad última poder asegurar la calidad teórica de los constructos que se miden, y aportar una base sólida para definir un modelo explicativo que permita extraer mayor utilidad de la información con que se trabajan las evaluaciones de sistemas. Esta propuesta indica la necesidad de tomar como referencia las aportaciones de la investigación educativa, un modelo sistémico que ayude a dar coherencia a los planes de evaluación, considerar las necesidades de información de las partes interesadas en la evaluación y un procedimiento más riguroso de diseño de instrumentos. Esta metodología es la base de otros trabajos que se presentan en este mismo número de la revista y que se citan en este artículo.

Palabras clave: *Calidad educativa, Evaluación de sistemas educativos, Cuestionarios de contexto, Explicación del desempeño, Diseño de instrumentos de medida.*

Introducción

Los *sistemas educativos* se identifican con la organización a través de la cual una sociedad estructura y vertebrada los recursos, medios y

contenidos de la educación. Por tanto, representan la respuesta de un Estado a las necesidades individuales y sociales de educación (*a quién, para qué, cómo, cuándo... educar*). En línea con lo señalado por la profesora García

(2000), la evaluación de los sistemas educativos, aunque no constituye un fenómeno nuevo, se ha incrementado en los últimos cuarenta años, asociándose a la apertura democrática, a la rendición de cuentas y al control del gasto público, a la mejora de la competitividad socio-económica y a la descentralización (por su internacionalización). Consideremos brevemente cuál es el sentido de la evaluación de sistemas sociales.

En otra parte (De la Orden, 1981) nos referíamos a la evaluación como una característica inherente a toda actividad humana intencional que se manifiesta tanto en la vida cotidiana, escasamente planificada, cuanto en la acción sistemática de las empresas individuales o colectivas del hombre. En el devenir de la rutina diaria, el hombre evalúa o valora constantemente personas, objetos, situaciones y acontecimientos, juzgándolos, previo su contraste con instancias o modelos más o menos explícitamente representados; pero es en la actividad colectiva sistematizada, práctica y con base racional (la política, la administración, la actividad económica, la educación, la medicina, etc.), donde la evaluación aparece como un componente esencial. El énfasis actual en la evaluación, característico del ámbito general de la actividad social planificada y, por tanto, de la educación, es una manifestación de la importancia que las sociedades contemporáneas atribuyen al control de las empresas comunitarias fundamentales. En realidad, la evaluación está orientada a determinar el valor de todas las facetas esenciales de la sociedad y facilitar su optimización.

Esta constatación nos induce a intentar una explicación del papel primordial que la evaluación desempeña en el ámbito de la actividad social organizada —intencional y sistemática—. En primer lugar, la actividad social organizada está integrada por objetos o entidades creados por el hombre con una finalidad específica, es decir, *artificiales*. Las empresas productoras o distribuidoras de bienes o servicios, los programas

sociales de toda índole (educativos, sanitarios, urbanísticos, de bienestar, etc.) y las instituciones de todo tipo son creaciones humanas que genéricamente consideramos como *sistemas artificiales* en contraposición, por ejemplo, con el sistema solar, el sistema nervioso humano o el sistema defensivo de las abejas que consideramos como *sistemas naturales*. Los sistemas naturales, incluidos los que podríamos llamar *sistemas sociales espontáneos*, disponen de mecanismos automáticos de *auto-mantenimiento* (o auto-regulación) que se activan cuando las condiciones del ambiente en que se desenvuelven cambian y amenazan su existencia o su función. Sin embargo, los sistemas artificiales carecen de tales mecanismos. En estos sistemas es justamente la evaluación la que sustituye a los mecanismos automáticos de auto-mantenimiento y, en consecuencia, se decide y aplica voluntariamente con la intención de producir los cambios deseables o deseados en la estructura y/o en el funcionamiento del sistema para asegurar su permanencia y funcionalidad. Esto explica por qué la evaluación, como podemos constatar en la actualidad, se desarrolla y adquiere relevancia máxima en los momentos históricos caracterizados por un intenso cambio social. En realidad, la evaluación se identifica con un mecanismo permanente de ajuste (*artificial*) de las decisiones sobre las actividades sociales organizadas (*sistemas artificiales*) y cuya función básica es la retroalimentación de los mismos para optimizar sus alternativas de adaptación a los cambios del entorno. Si no hubiera cambio, no habría alternativas y, en consecuencia, la evaluación perdería su sentido.

Así pues, la mera presencia de la evaluación *sensu stricto*, es decir, la información y el juicio de valor sobre el objeto evaluado, en este caso sobre los sistemas sociales, no garantiza una alta calidad en los productos y servicios de los mismos; se limita a asegurar una condición necesaria para ello. La eficacia que optimiza la evaluación exige, además, que sean tomadas las decisiones adecuadas, basadas en los resultados evaluativos, que aseguren las acciones reguladoras pertinentes (Cicchinelli y Barley, 1999).

Hoy, en un ambiente de reforma generalizada en los sistemas educativos para adaptarse a los profundos y permanentes cambios en el contexto social en el que se insertan, la evaluación educativa cobra también su máximo protagonismo y presenta características similares a las apuntadas para el campo en su conjunto.

Aunque en este tipo de procesos la vinculación real entre la aportación de la información evaluativa y la toma de decisiones está más desconectada que en otros tipos de planes de evaluación referidos a unidades de análisis más específicas², lo cierto es que su finalidad última es la toma de decisiones y, en cualquier caso, se trata de aportar información para el conocimiento y la comprensión de la realidad y favorecer el debate social. De este modo, a estos procesos se asocian funciones tales como las diagnósticas (mejora del sistema), retroalimentación de información (para los gestores —apoyo a la toma de decisiones—, o para la sociedad —rendición de cuentas—), así como otras acepciones más específicas (que constituyen matices respecto a las anteriores), como por ejemplo, el control, la acreditación y/o certificación o la prospectiva (Tiana, 1993, 1996; Tiana y Muñoz, 1997; Oroval, 1995; Landsheere, 1996; García, 2000).

Panorama de las evaluaciones de sistemas educativos

Sin pretender realizar una revisión exhaustiva de las aportaciones realizadas en la evaluación de sistemas educativos, estimamos necesario hacer un breve recorrido a través de las propuestas e instituciones que han propiciado el desarrollo de los enfoques que se dan en la actualidad.

En el plano internacional, probablemente, los dos organismos que mayor impacto han tenido en el impulso de las evaluaciones de sistemas educativos han sido la International Association for the Evaluation of Educational Achievement (IEA) y la Organización para la Cooperación y el Desarrollo Económico (OCDE).

La IEA se constituye en 1958. Se trata de una organización de cooperación internacional independiente de las instituciones nacionales de investigación y agencias gubernamentales de evaluación. Tiene como objetivo aportar información fiable y válida para poder valorar el efecto de las políticas y prácticas educativas basándose en estudios de carácter comparado. De esta institución fue director el profesor Alejandro Tiana Ferrer, actual director del Centro de Altos Estudios Universitarios de la OEI y catedrático de Educación Comparada de la UNED (España), en cuya etapa se produjeron grandes avances institucionales, y que es en la actualidad una de las referencias de mayor prestigio y calidad nacional e internacional en el ámbito de la evaluación de sistemas y en la educación comparada, sectores de conocimiento íntimamente relacionados.

La IEA, desde su creación, ha desarrollado más de treinta estudios comparados internacionales, entre los que destacan la *Encuesta internacional sobre matemáticas y ciencias* (TIMMS, 1995, 1999, 2003, 2007, 2008 —avanzadas— 2011), o el *Estudio internacional sobre competencia en lectura* (PIRLS, 2001, 2006, 2011). Asimismo, ha llevado a cabo otros estudios de gran trascendencia como, por ejemplo, *TIMSS-R Vídeo estudio de prácticas en el aula*, *Educación cívica* (CIVED y CIEC), *Tecnología de la información en la educación* (SITIOS-M1, M2-SITIOS, SITIOS 2006), sobre el uso de ordenadores en la educación (COMPED) y la educación pre-primaria (PPP), o en educación superior: *Maestros de Educación para el Desarrollo y Estudio en Matemáticas* (TEDS-M)³.

Por otra parte, la OCDE, y promovido por el CERI (Centro para la Investigación e Innovación Educativa) a partir de 1988 pone en marcha su proyecto INES para llegar a disponer de un sistema de indicadores que coadyuven a los países de la OCDE a valorar y orientar sus políticas educativas. En este marco, la OCDE entiende la educación como un instrumento económico, que ayuda al desarrollo personal

y social. Ello marca el estilo de indicadores que incluye en sus estudios. En 1992, como evolución de este proyecto, comienza a publicarse de forma anual el informe *Panorama de la Educación (Education at a Glance)*.

En este contexto, el *Informe Delors* (1996) constituyó un elemento de referencia que influyó en la concepción acerca de cómo orientar la educación (cuya finalidad sustantiva debería, según dicho informe, dirigirse esencialmente al desarrollo de competencias para la vida), y en consecuencia, supuso un impulso para transformar en parte los estudios evaluativos de carácter internacional hacia esta perspectiva.

A partir de 1997 se inicia el Proyecto Internacional de Evaluación de Estudiantes (PISA) dirigido a valorar en diferentes materias (especialmente las instrumentales) las competencias que tienen adquiridas los estudiantes de quince años en diferentes países. Se pretende con ello disponer de un indicador de producto o resultado educativo que permita mejorar la información disponible que se había ido desarrollando a partir de los proyectos anteriormente citados. En ese mismo año, se pone en marcha el Proyecto DeSeCo (Design and Selection of Competencies), que se publicó en 2003 y cuya finalidad es la identificación de las competencias clave para el desarrollo personal y social.

En este mismo marco, en la Unión Europea (UE), la Red Eurydice (2002-2003) realiza un estudio acerca de las competencias básicas que deberían atenderse desde la educación obligatoria. Este estudio es la base para identificar las competencias que se asumen como referencia de desarrollo de los diseños curriculares para educación primaria y secundaria obligatoria de diversos países.

Además de los estudios internacionales, hay que destacar las aportaciones que se han ido realizando desde diversas instancias nacionales. En este sentido, probablemente el país en el que mayor uso y reconocimiento se ha dado a la evaluación de sistemas es Estados Unidos

de América. Entre una gran cantidad de aportaciones, quizás cabe destacar:

- La aprobación de la Primary and Secondary Education Act (1965) como ley federal que regula diversos aspectos de la educación en EE.UU. y que señala expresamente la obligación de aspirar a altos estándares de calidad, identificando la rendición de cuentas —a través de la evaluación—.
- El informe de Coleman *et al.* (1966), que tuvo un gran impacto al analizar los factores asociados al rendimiento escolar. Su conclusión, aunque se plantee de forma simplista, asociaba el logro educativo al nivel socioeconómico familiar, restando valor a la aportación escolar.
- El NAEP (National Assessment of Educational Progress) comenzó a fraguarse en 1964, a partir de aportaciones de la Carnegie Corporation. Su objetivo es realizar evaluaciones periódicas de diversas materias en todos los estados de EE.UU., estableciendo así una referencia común que permitiera valorar la situación educativa del país y derivar informaciones para orientar las políticas públicas. En 1969 comenzaron a llevarse a cabo evaluaciones con carácter voluntario y se ha ido extendiendo por todo el país a través de los años como una referencia necesaria para la evaluación de la educación estadounidense. Su impacto, más allá de los resultados, es que se ha convertido en un centro de referencia internacional sobre las metodologías que se pueden utilizar y pone el énfasis en poder analizar las tendencias de evolución del logro educativo.

Del mismo modo que en EE.UU., en los últimos años se han ido poniendo en marcha instituciones regionales y nacionales para la evaluación del sistema educativo.

- En Iberoamérica se han desarrollado instituciones tanto nacionales como internacionales, que realizan aportaciones de

gran valor en este ámbito. Entre ellas, cabe destacar:

- a) El Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (LLECE), dependiente de la UNESCO, que como se señala en su página web «es la red de sistemas de evaluación de la calidad de la educación de América Latina. Lo coordina la Oficina Regional de Educación de la UNESCO para América Latina y el Caribe y tiene su sede en Santiago de Chile». El LLECE desarrolla diversas actividades, tanto de carácter evaluativo (por ejemplo, estudios como el SERCE, y otros relativos a factores asociados al rendimiento), como de refuerzo para la mejora de los sistemas de evaluación de los países de la región (actividades de encuentro entre investigadores, y formación).
 - b) El Programa de Promoción de la Reforma Educativa en América Latina y el Caribe (PREAL). Es un proyecto conjunto del Diálogo Interamericano y la Corporación de Investigaciones para el Desarrollo, CINDE. Su finalidad es mejorar la calidad y la equidad de la educación, promoviendo el debate sobre políticas educativas, basado en informaciones de calidad, así como identificar y difundir las mejores prácticas y realizar un seguimiento de la mejora educativa.
- A nivel nacional, prácticamente todos los países han ido desarrollando sus instituciones de evaluación, normalmente dependientes del poder público. Entre ellas cabe destacar: el Instituto Nacional para la Evaluación de la Educación de México (INEE), el Instituto Nacional de Estudios e Pesquisas Educacionais Aníxio Teixeira (INEP) de Brasil, el Instituto Colombiano para la Evaluación de la Educación (ICFES), la Dirección Nacional de

Información de la Calidad Educativa (DINIECE) de Argentina, el Sistema de Medición de la Calidad en Educación (SIMCE) de Chile, la Estadística de Calidad Educativa (ESCALE) de Perú, o el Instituto de Evaluación (IE) de España.

Sintetizando el desarrollo de la evaluación de sistemas educativos, podemos señalar los siguientes aspectos:

- Actualmente se puede afirmar que ya es una práctica habitual la evaluación de sistemas, que se sustenta sobre estudios periódicos nacionales e internacionales. Los esfuerzos aportados, aunque son ingentes, parece que no llegan a satisfacer las expectativas que hay acerca de su utilidad final, sea para la orientación de políticas públicas, o para usos más específicos dirigidos a mejorar los procesos organizacionales y/o los relativos a los procesos de enseñanza-aprendizaje.
- Los enfoques o modelos de estudio que se siguen, tanto en las evaluaciones nacionales como internacionales, presentan una forma de hacer muy similar: el criterio final de calidad se identifica en los resultados o logro escolar en diversas materias, primordialmente, las denominadas como instrumentales; aunque el espectro de análisis se ha ido ampliando progresivamente a todas las disciplinas. Es decir, se reducen a aproximaciones en las que a partir de un indicador, generalmente de desempeño del alumnado, se pretende inferir la calidad global del sistema.
- En los estudios nacionales suelen darse dos tipos de tendencias:
 - Referencia de la evaluación al currículo nacional (en caso de existir); en este sentido se evalúa la consecución de metas curriculares (bien sean expresadas a través de objetivos —lo más frecuente—, o a través de competencias

- supuestamente más innovador, aunque menos habitual—);
 - o bien, dirigir la evaluación hacia las llamadas competencias genéricas y/o comunes, en el caso en que no exista un currículo único nacional.
- En los estudios internacionales, las referencias necesariamente son genéricas con el fin de no incluir fuentes importantes de sesgo debidas a las diferencias curriculares entre los países participantes. Sin embargo, no suelen aportarse estudios lógicos acerca de la alineación de las pruebas respecto a los currículos de cada país, aportándose exclusivamente los marcos lógicos generales de las evaluaciones (como referencia teórica) y los resultados empíricos en las pruebas como expresión de la calidad del sistema.
 - En todos los casos, se pretende que los resultados de las evaluaciones sirvan, cuanto menos, para la orientación de las políticas públicas. No obstante, como señalamos anteriormente, las expectativas sociales que crean suelen demandar mayor utilidad de sus resultados y se espera que puedan impactar más concretamente en las organizaciones escolares, las aulas y, en definitiva, en los procesos de enseñanza-aprendizaje.

Enfoques usuales en la evaluación de sistemas educativos

Las informaciones que aportan la mayor parte de instituciones de evaluación, sean nacionales o internacionales, se estructuran en dos grandes enfoques: indicadores y estudios. Sobre ellos nos detendremos a continuación.

Los *enfoques basados en indicadores* pretenden aportar datos o resultados que permitan informar acerca de diversas características de los sistemas. Los indicadores se agrupan en *mapas* que intentan reflejar en su conjunto el comportamiento

del sistema. Las fuentes de información suelen ser las que proveen estadísticas nacionales o internacionales periódicas. Pueden integrar diversas tipologías de indicadores: económicos, estructurales, y funcionales del sistema.

Su utilidad, en cualquier caso, y dado su carácter macro-analítico, debe circunscribirse en relación a la orientación de políticas educativas. Los mapas de indicadores suelen estar diseñados para dar respuesta a preguntas específicas del sistema en cada país, por lo que de uno a otro país pueden encontrarse variaciones en cuanto a los indicadores que se incluyen. No obstante, es frecuente localizar un número importante de indicadores comunes, que permiten realizar una lectura transnacional, de carácter comparativo. Como ejemplo, en la tabla 1 se presenta el mapa de indicadores del Instituto de Evaluación de España, en su versión para el año 2011.

Como puede observarse, se incluyen indicadores que están preferentemente orientados a informar acerca de la respuesta que ofrece el sistema educativo, en cuanto a escolarización y resultados. Asimismo, permite informar de algunos aspectos acerca del valor social de la educación (Jornet, Perales y Sánchez-Delgado, 2011).

- Esfuerzo económico que realiza el Estado en ofrecer servicios educativos (por ejemplo, a través de los indicadores de gasto en educación).
- Consecuencias en el estatus socioeconómico personal asociadas a la educación, por ejemplo, a través de los indicadores de tasa de actividad y desempleo, según el nivel educativo, o con el de diferencias de ingresos laborales según el nivel de estudios.

En otros institutos o agencias se incluyen referencias más precisas relativas a elementos de entrada y contexto. Por ejemplo, en la tabla 2 se presentan algunos ejemplos acerca de otros indicadores que se incluyen en los mapas de otras instituciones.

TABLA 1. Mapa de indicadores del Instituto de Evaluación de España, 2011

Escolarización y población	Escolarización y población escolarizable en las edades de 0 a 24 años
	Escolarización según la titularidad
	Esperanza de vida escolar a los 6 años
Tasas de escolarización en las edades teóricas de los niveles no obligatorios (CINE 0, 3, 4 y 5)	Educación infantil
	Educación secundaria segunda etapa
	Educación superior
Alumnado extranjero	
Alumnos por grupo y por profesor	Alumnos por grupo educativo
	Alumnos por profesor
Participación en el aprendizaje permanente	
Gasto total en educación	Gasto total en educación con relación al PIB
	Gasto público total en educación
	Gasto público destinado a conciertos
Gasto en educación por alumno	
Competencias básicas en cuarto curso de educación primaria	Competencia básica en comunicación lingüística
	Competencia básica en matemáticas
	Competencia básica en el conocimiento y la interacción con el mundo físico
	Competencia básica social y ciudadana
Competencias básicas en segundo curso de educación secundaria obligatoria	Competencia básica en comunicación lingüística
	Competencia básica en matemáticas
	Competencia básica en el conocimiento y la interacción con el mundo físico
	Competencia básica social y ciudadana
	Conocimiento cívico y ciudadano (ICSS)
Competencias clave a los 15 años de edad	Competencias clave a los 15 años en lectura
	Competencias clave a los 15 años en matemáticas
	Competencias clave a los 15 años en ciencias
Idoneidad en la edad del alumnado	Idoneidad en la edad del alumnado de educación obligatoria
	Alumnado repetidor
Abandono temprano de la educación y la formación	
Tasas de graduación (CINE 2, 3, 5 y 6)	Tasa bruta de graduación en educación secundaria obligatoria
	Tasa bruta de graduación en estudios secundarios segunda etapa
	Tasa bruta de graduación en estudios superiores
	Tasa de graduados superiores en ciencias, matemáticas y tecnología
Nivel de estudios de la población adulta	
Tasa de actividad y de desempleo según el nivel educativo	Tasa de actividad según nivel educativo
	Tasa de desempleo según nivel educativo
Diferencias de ingresos laborales según el nivel de estudios	Ingresos laborales y nivel de estudios
	Ingresos laborales por sexos y nivel de estudios

TABLA 2. Algunos ejemplos de otros indicadores, con expresión de la institución que los utiliza

Tipo de indicador	Algunos ejemplos
Indicadores de capacidad económica	
Producto Interior Bruto (PIB)	PIBC: El PIB <i>per cápita</i> es la relación entre el valor total de mercado de todos los bienes y servicios finales generados por la economía de una nación, durante un año, y el número de habitantes de ese año (INEE, México).
Producto Interior Bruto per Cápita (PIBC)	
Indicadores de desigualdad	
Tasa de Riesgo de Pobreza (TRP)	TRP: Porcentaje de personas que viven en hogares cuya renta total equivalente es inferior al 60% de la renta mediana nacional equivalente (Eurostat) –Laeken–.
Desigualdad de la Distribución de Ingresos (DDI)	DDI: Q80/Q20: ratio entre los ingresos totales percibidos por el 20% de la población con los ingresos más altos (quintil superior Q80) y el recibido por el 20% de la población con los ingresos más bajos (quintil inferior Q20). Se entiende por ingresos a la renta disponible equivalente (Eurostat) –Laeken–.
Índice de marginación	IM: Es una medida que agrupa los valores relativos de nueve indicadores socioeconómicos, con la finalidad de diferenciar unidades territoriales según las carencias padecidas por la población como resultado de falta de acceso a la educación, residencia en viviendas inadecuadas, ingresos monetarios insuficientes y residencia en localidades pequeñas (INEE, México).
Índice de Desarrollo Humano (IDH)	IDH: Medida de calidad de vida que resulta de promediar los avances relativos logrados por un país o demarcación geográfica, en cuanto a tres aspectos básicos requeridos por la población: gozar de una vida larga, contar con educación y disfrutar de un nivel de vida decoroso (INEE, México).

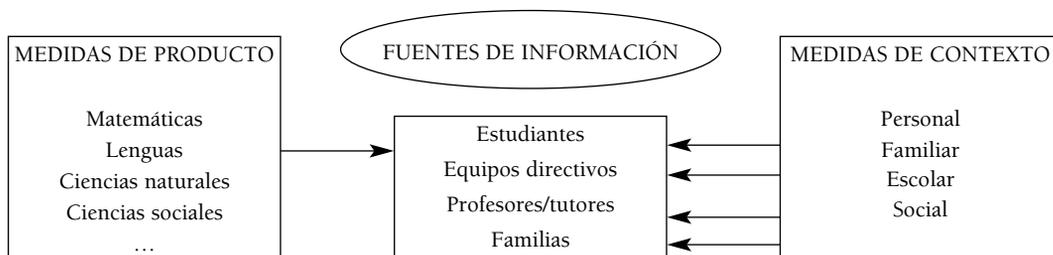
Por otra parte, los *enfoques basados en estudios* han adoptado un modelo que prácticamente se replica en todas las instituciones y agencias que realizan evaluaciones de sistemas, sean nacionales o internacionales. Normalmente se trata de *estudios muestrales*⁴. Sus características principales pueden sintetizarse en las siguientes:

- Se estructuran tomando como referencia el producto o resultado educativo que muestra el alumnado en diferentes materias a través de pruebas estandarizadas. Normalmente, las pruebas están bien construidas y están sujetas a fuertes controles métricos para asegurar su fiabilidad

y validez. El dominio educativo al que se refieren es el diseño curricular en el caso de evaluaciones nacionales (por ejemplo, ver pruebas EXCALE del INEE de México), o constructos teóricos competenciales en el caso de pruebas internacionales (ver por ejemplo proyecto PISA). Los desarrollos de este tipo de pruebas, al pretender representar la amplitud del currículo —o de un constructo amplio de referencia—, se basan en modelos de muestreo matricial, por lo que sus puntuaciones no pueden interpretarse a nivel individual ni, en muchos casos, a nivel de escuela, sino que son leídas para unidades muestrales

- más genéricas, como provincias, regiones, estados o países.
- De este modo, se parte del desempeño educativo del alumnado, mostrado en pruebas estandarizadas, como un indicador de resultado educativo del sistema y, en consecuencia, sobre el que informar acerca de la calidad de la educación en diferentes niveles de organizaciones educativas.
 - De este modo, pretenden responder a cuestiones del tipo:
 - ¿Cuánto saben de matemáticas o de español los alumnos de sexto de primaria?
 - ¿Hay diferencias por sexo?
 - ¿Hay diferencias por titularidad del centro educativo?...
 - Están normalmente estructurados como planes de evaluación de aplicación periódica, si bien no suelen tener bien resuelta la comparabilidad o análisis longitudinal de los resultados a través de los años, por lo que uno de sus déficits es la limitada posibilidad de informar acerca de la evolución de los sistemas.
 - En todos los casos, se suelen incluir los denominados *cuestionarios de contexto*. La base que justifica la inclusión de este tipo de informaciones reside en la constatación, a través de la investigación educativa, de la influencia que tiene el contexto en el rendimiento educativo (Tejedor y Caride, 1988; Murillo y Hernández-Castilla, 2011a y 2011b). Sin embargo, aunque los denominemos cuestionarios de contexto, más que un único cuestionario, se trata de *sistemas de cuestionarios* que se dirigen a diversas fuentes de información: al alumnado, a los equipos directivos, al profesorado y a las familias. El objetivo, en este caso, es recabar información que facilite una mejor comprensión acerca de los niveles de logro de forma que puedan, en última instancia, identificarse las claves de buenas prácticas que permitan mejorar los resultados educativos. Intentan, pues, aportar información acerca de los contextos personal, escolar, familiar y social.
 - Las *características* que suelen tener estos cuestionarios son, entre otras (Jornet, 2009):
 - Incluyen datos (personales, familiares, sociales), indicadores simples (representados por un ítem o reactivo), e indicadores complejos (representados por escalas referidas a constructos, como por ejemplo, el nivel socioeconómico y cultural de las familias o el clima social del aula).
 - De este modo, pretenden describir elementos del proceso y del contexto habitual en que se produce la enseñanza. Así, aunque los denominemos de forma simplificada *cuestionarios de contexto* recogen realmente información de entrada, proceso y contexto.

FIGURA 1. Esquemización del enfoque de estudios evaluativos de sistemas educativos



- Suelen elaborarse de forma bastante deficiente, sin referencias a modelos explicativos del aprendizaje, ni en relación a las posibles preguntas que deberían contestarse con la evaluación, y que podrían interesar a las partes interesadas (o audiencias en términos de Weiss). Generalmente, y hasta fechas muy recientes en que se ha incrementado el interés por su mejor desarrollo, se ha prestado una escasa atención a su diseño.
- En Pedagogía, uno de los problemas no resueltos, que probablemente afecta en gran medida a su desarrollo disciplinar, es la carencia de teorías bien fundamentadas acerca de los fenómenos que estudian. La definición de los constructos teóricos que se incluyen en los cuestionarios de este tipo suelen estar definidos de forma simplista, y la selección de las variables, indicadores y/o constructos que los representan se suele realizar sin referencia sistemática a criterios precisos, con escasa atención a una teoría o modelo de referencia o información contrastada sobre resultados de investigaciones anteriores que justifiquen su valor como elemento explicativo o predictor del desempeño.
- Adicionalmente, en la recogida de información, se pueden controlar mejor los cuestionarios de alumnos, profesores/tutores y equipos directivos que los correspondientes a las familias, lo cual se traduce en una peor representación y mayor posibilidad de incidencia de sesgos en los elementos contextuales y/o procesales que están fuera del marco escolar (González-Montesinos y Backhoff, 2010). Téngase en cuenta que lo más frecuente es que la mayor parte de las respuestas provenga de las familias que mayor atención prestan al aprendizaje de sus hijos, mientras que el colectivo de familias

con menor preocupación tienen por este tema, suele quedar escasamente representado en este tipo de estudios (Perales, Cardona y Gómez, 2009).

En síntesis, en el *enfoque de estudios*, podemos identificar como sus mayores aportaciones las siguientes:

- La buena definición de las medidas de producto o desempeño educativo, basada en diseños en que se controla exhaustivamente los instrumentos utilizados.
- El control exhaustivo del muestreo y recogida de información (en los estudios muestrales).
- La experiencia y conocimiento, que se aporta desde el enfoque investigador.
- Y, normalmente, el alto impacto social que suelen tener ante la opinión pública y los administradores educativos.

No obstante, también tienen importantes debilidades; entre ellas podemos señalar las siguientes:

- A pesar de que las medidas de resultados de aprendizaje suelen estar bien elaboradas y existen procedimientos útiles para realizar un seguimiento y valorar la evolución de los sistemas o las consecuencias de programas de mejora o cambios legales, la falta real de un análisis longitudinal implica una sensible disminución de su valor para la orientación política en la educación.
- Por otra parte, el mayor problema que podemos señalar es la deficiente definición de las medidas de contexto, efecto de la escasa atención a sus procesos de diseño, su falta de referencias teóricas y sus bajos controles métricos.
- Asimismo, es curioso que, recogida la información en las evaluaciones habituales, esta se aprovecha muy poco. Normalmente, tan solo se analizan las relaciones del desempeño con algunas variables

demográficas. Así, podemos afirmar que adolecen de la carencia de una adecuada atención al análisis explicativo del producto basado en el contexto (sea en relación a variables de entrada, proceso o contexto propiamente dicho).

- De este modo, aunque el objetivo que anima la realización de estas evaluaciones, tal como hemos señalado anteriormente, sea principalmente facilitar información para orientar las políticas públicas, lo cierto es que resultan de escasa utilidad para este propósito.
- Finalmente, la desconexión con las necesidades de información de las partes interesadas (equipos directivos, profesorado y familias) supone una gran dificultad para extraer información de utilidad, es decir, que ayude a elaborar recomendaciones de mejora para cada una de ellas.

La utilidad de las evaluaciones y el papel del contexto

¿Podríamos incrementar la utilidad de las evaluaciones de los sistemas educativos? Es una pregunta compleja para responder en el reducido espacio de un trabajo de estas características, pero entendemos que sí. Como en la mayor parte de los problemas científicos y técnicos, no existe una única respuesta. Debemos explorar posibilidades de mejora, de manera que puedan aportarse, a través de la comprobación de sus potencialidades, orientaciones precisas acerca del modo en que se pueda optimizar esta situación. Para ello, nos referiremos en este trabajo a tres posibles líneas de actuación que pueden abrir el camino para el incremento de la utilidad de las evaluaciones:

- Definir un modelo de referencia para la selección de variables que considerar y especificar las características conceptuales de la calidad que se pretende evaluar.
- La atención a las necesidades y expectativas de las partes interesadas.

- Los procedimientos o metodologías para la elaboración de los sistemas de cuestionarios de contexto.

El primer aspecto que habría que considerar afecta a la selección de las variables que medir, que habitualmente se realiza sin considerar el rol que van a tener en el plan de evaluación; es decir, sin establecer hipótesis acerca de sus posibles efectos en relación a los resultados. La consecuencia es la que se observa en buena parte de los informes de evaluación de sistemas: la información sobre aprendizajes se presenta de forma independiente de la recogida en los cuestionarios de contexto y, a lo sumo, se establece alguna relación diferencial del aprendizaje con alguna variable como sexo, tipo de titularidad del centro educativo o nivel socioeconómico de las familias.

Las dimensiones que se evalúan en los sistemas educativos recorren una amplia variedad de opciones, dado que su selección depende en gran medida del modelo de gestión política prevalente y, en consecuencia, del enfoque que se dé a la evaluación. No obstante, se podrían clasificar las dimensiones usualmente consideradas en los planes de evaluación de sistemas educativos en cuatro categorías: *variables de entrada, de proceso, de producto y de contexto* —ver tabla 3—.

Hay que señalar que en esta estrategia se deberían contemplar al menos dos grandes procesos: a) validación de constructo del modelo de explicación, es decir, la definición del modelo que permite orientar la selección de variables, y b) la estrategia para poder incluir la medición de las mismas en los cuestionarios de contexto, de forma que sea operativa para una evaluación de sistemas. Sobre ambos aspectos tratamos a continuación.

El primer aspecto que tenemos que definir para la elaboración de cuestionarios de contexto es la identificación de un modelo que oriente de forma satisfactoria la selección de variables que hay

TABLA 3. Descripción de variables de entrada, proceso, producto y contexto

Variable	Definición
De entrada	Características personales de los alumnos, profesores y tipo de institución escolar.
De proceso	Elementos que describen cómo se produce la enseñanza: estilo de enseñanza, clima social y de aprendizaje de la clase, gestión de la integración, grado de multiculturalidad de los grupos, participación social en la enseñanza, etc.
De producto	Resultados de aprendizaje por áreas curriculares (Matemáticas, Lenguaje, CC. Sociales...).
De contexto social-familiar	Elementos familiares del estudiante: escolaridad y ocupación de los padres, indicadores socioeconómicos, utilización del tiempo extra-clase del alumno, relaciones con los padres, etc.

que considerar en los mismos. Este aspecto implica posiciones teóricas y cuestiones de utilidad de información que están vinculadas a los fines establecidos para cada evaluación particular. Por ejemplo, entendemos que no deberían interesar los mismos elementos explicativos cuando se aborda una evaluación transnacional que cuando se realiza una evaluación nacional o local.

El segundo elemento que tener en cuenta es el tipo de concepto de calidad que se requiere evaluar. De la Orden *et al.* (1997, 2007) se refieren a las definiciones de calidad de Harvey y Green (1993): calidad como fenómeno excepcional, calidad como perfección o coherencia, calidad como ajuste a un propósito, calidad como relación valor-coste y calidad como transformación o perfeccionamiento de la institución. Además de estas, especialmente en los últimos años, en la definición del tipo de calidad que se pretende comprobar en los sistemas educativos ha irrumpido otro término con gran fuerza: calidad como equidad.

No obstante lo anterior, al hablar de evaluación de sistemas no podemos olvidar que existen modelos solventes, de carácter sistémico, que permiten orientar la selección de variables o indicadores y determinar su rol como, por ejemplo, el modelo sistémico de De la Orden (1997), revisado por Gaviria y Tourón (2000) —ver figura 2—, que permite considerar la relación

entre los diversos componentes y elementos del sistema para las grandes dimensiones de la calidad como *funcionalidad, eficacia y eficiencia*, esenciales para valorar el sistema educativo en términos de la coherencia entre los elementos de entrada, procesos, productos, objetivos y contexto del sistema educativo. Las ventajas del modelo radican en los siguientes aspectos:

- a) Se trata de un modelo holista, de carácter sistémico, que considera todos los elementos relevantes del sistema educativo —o de sus instituciones—, así como sus relaciones contextuales.
- b) Permite definir teóricamente las dimensiones esenciales de la calidad, considerando las variables o indicadores de los componentes sistémicos citados.
- c) Exige y facilita el análisis de todos los componentes del sistema y sus relaciones.
- d) Al definirse como un modelo, es comprobable en términos de análisis de relaciones estructurales y causales.
- e) Se puede utilizar tanto desde una perspectiva estrictamente cuantitativa como desde otra basada en la complementariedad metodológica, tal como la describe Bericat (1998).
- f) Facilita la identificación de indicadores y la tipología de análisis que son necesarios para comprobar en términos operativos las diversas facetas o dimensiones

implicadas en el concepto de calidad que explicita el modelo.

- g) Los efectos que pueden analizarse son:
- *Funcionalidad*, como grado en que el sistema educativo satisface las necesidades educativas de los alumnos, las familias, la comunidad y la sociedad en su conjunto en el marco de la cultura o perfil axiológico vigente.
 - *Eficacia*, como expresión del logro de objetivos educativos.
 - *Eficiencia*, como grado de optimización de recursos y procesos educativos en la consecución de objetivos.

- h) Tal como indicábamos anteriormente, el término de *equidad* como elemento que considerar en la calidad del sistema se ha ido poniendo de manifiesto como una referencia explícita en buena parte de las evaluaciones de sistemas educativos. No obstante, en el modelo de De la Orden la *equidad* está presente, y subsumida como una *faceta de la funcionalidad*. De este modo, a partir de esta propuesta, la *equidad* se entiende como una manifestación del grado en que el sistema educativo cumple la función de satisfacer una aspiración social derivada de uno de los valores que caracterizan a la cultura de las sociedades occidentales. Es decir, la *equidad* es una manifestación de la dimensión *funcionalidad* en el contexto del modelo.

Un segundo nivel de concreción al definir los planes de evaluación de sistemas se sitúa en la atención a las partes interesadas: *¿qué preguntas hay que responder con la evaluación para que sea útil a las diferentes partes interesadas?* La evaluación, basada en un enfoque investigador, debe adecuarse para ofrecer respuesta a las necesidades y expectativas que la sociedad en cada momento le demanda. Ello no implica comprometer ni el rigor, ni la calidad de la evaluación, sino únicamente reconocer que se trata de una actividad que está al servicio de una sociedad

en un momento histórico determinado. Este es uno de los aspectos que diferencian la investigación básica respecto de la evaluativa.

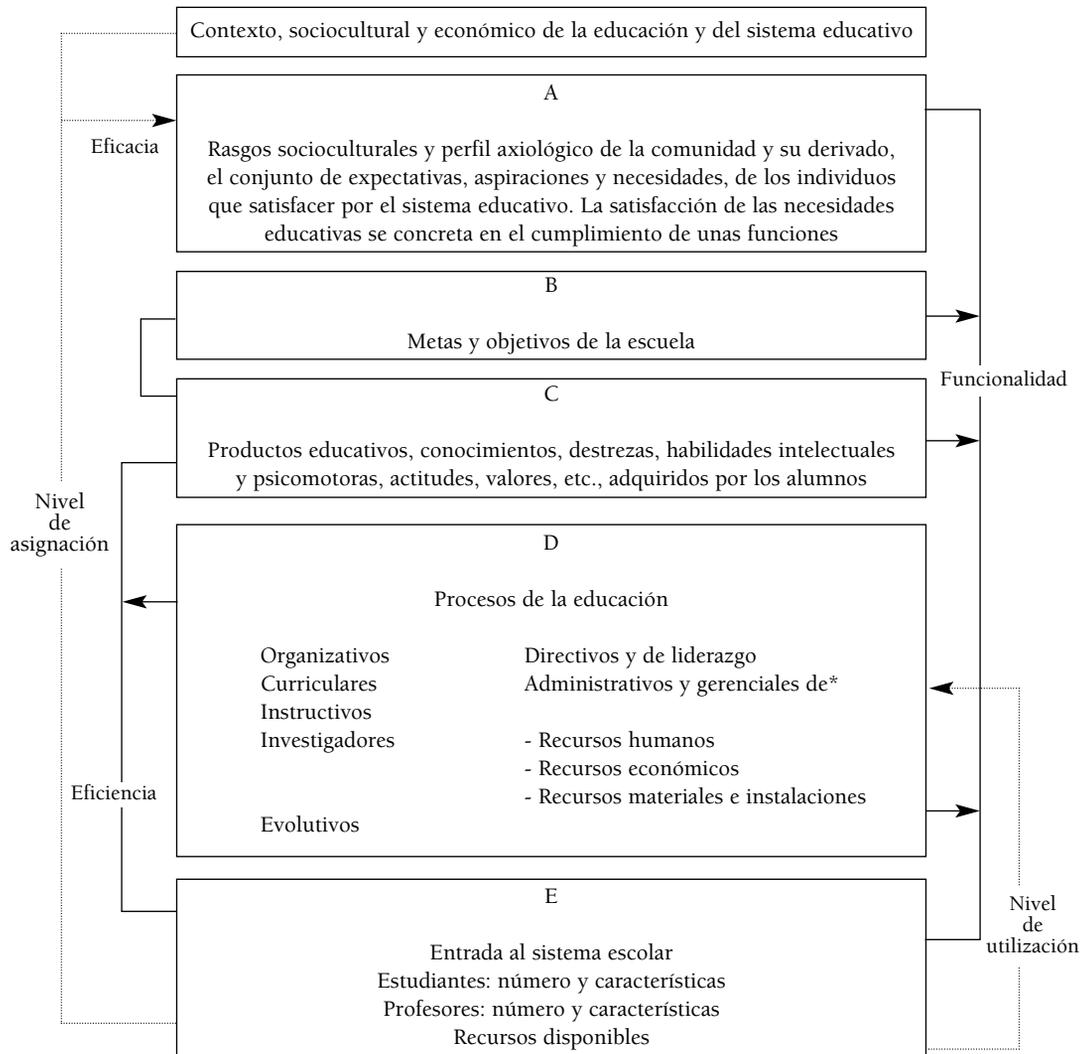
Por ello, sin descuidar los elementos que se puedan y deban considerar en cuanto a la validez interna y externa de las evaluaciones, deben priorizarse como elementos constitutivos de la validez *la credibilidad* y *la utilidad*. Dificilmente si una evaluación no resulta creíble para todas las partes interesadas resultará útil. Y, a su vez, no será creíble aquella evaluación que no aporte información para resolver los problemas que preocupan a cada una de las partes interesadas.

Además de problemas de comunicación de resultados que habitualmente están presentes en los ejercicios evaluativos que vamos comentando en este trabajo es en la orientación de la evaluación donde se puede construir la credibilidad y perseguir la utilidad.

A partir de un modelo explicativo que oriente de manera precisa el diseño global de la evaluación, es necesario derivar e identificar las necesidades de información que se requieren: a) para la orientación de la política educativa, b) para la mejora de la gestión y administración de las instituciones educativas, c) para la innovación en el aula, con el fin de aportar elementos de mejora de los procesos de enseñanza-aprendizaje y d) para la orientación de la sociedad, en especial de las familias, en su papel de acompañamiento del trabajo escolar.

La identificación de estas necesidades de información, además de sustentarse sobre el análisis lógico-teórico de un modelo de referencia como el que hemos comentado anteriormente, puede beneficiarse de procesos de investigación evaluativa dirigidos a tal cometido. Se trata, pues, de poder incluir en los planes de evaluación los elementos necesarios para que cada parte interesada en la evaluación pueda recibir información relevante y pertinente a sus necesidades, para mejorar su papel en el complejo entramado de intervenciones que constituye la educación.

FIGURA 2. Modelo sistémico de De la Orden



El tercer elemento sobre el que vamos a reflexionar acerca de las posibilidades de mejora de la utilidad de las evaluaciones de sistemas es el relativo al *diseño de los sistemas de cuestionarios de contexto*.

Como hemos comentado anteriormente, se puede afirmar que las pruebas estandarizadas de rendimiento como indicadores de sistemas educativos suelen ser instrumentos bien elaborados,

apoyados sobre modelos de medida muy evolucionados (como la teoría de respuesta al ítem) y sustentados sobre técnicas (como el muestreo matricial) que aseguran una correcta representación de los diseños curriculares. Ello permite abordar análisis de síntesis (basados en la lectura escalar de los niveles de adquisición) o bien otros de tipo criterial, orientados a identificar el desajuste en el logro de objetivos previstos para el sistema o la

detección de lagunas sistemáticas de aprendizaje.

Una adecuada detección del producto o resultado del sistema —aunque sea limitado a nivel cognoscitivo, y no considere otro tipo de posibles desarrollos (afectivo, social, emocional, valores...)— si bien es sumamente necesaria e importante, no es más que la base para que los planes de evaluación de sistemas educativos puedan aportar la utilidad que se espera de ellos: detectar la situación del sistema, así como identificar los elementos que puedan dinamizarse para la mejora del mismo. Ciertamente, esta falta de atención a los factores psico-socio-afectivos es una carencia grave, dado que en todos los sistemas educativos lo que se pretende es una formación integral de los ciudadanos. Y, tal como señala el profesor Ramón Pérez Juste en comunicación personal a los autores: «en todas las leyes que conozco de diferentes países tales objetivos aparecen claramente destacados. Y dado que lo que no se evalúa se devalúa a los ojos de gestores, directivos, profesores, padres y alumnos, esta es una carencia que, a mi juicio, es grave, aunque sea consciente de las dificultades que entraña su evaluación» (Pérez Juste, 2012).

En esta falta de utilidad destaca probablemente como aspecto peor resuelto, tal como indicamos anteriormente, el diseño de los instrumentos orientados a la recogida de información de variables de entrada, proceso o contexto son los denominados habitualmente *questionarios de contexto*. Son múltiples los estudios que ponen el énfasis en este aspecto. Tan solo a modo de ejemplo, en un estudio que realizamos comparando tres bases de datos (las procedentes de los proyectos PISA y TIMSS en México, con el de EXANI-1 —Jornet y Backhoff, 2004—), pudimos comprobar las dificultades para encontrar conclusiones comunes entre diversos estudios realizados en referencia a la misma población, e inclusive las contradicciones que se producen al pretender explicar el rendimiento observado.

La clave de la situación descrita radica claramente en la deficiente elaboración de los cuestionarios de contexto, la cual se basa habitualmente en la opinión de los responsables de los planes de evaluación; es decir, no suelen desarrollarse siguiendo un plan de diseño de instrumentos adecuado, apoyado en la información que a este respecto puede derivarse de la investigación educativa y, en especial, de las aportaciones de la pedagogía diferencial. En esta línea, hay que señalar que son precisamente las informaciones relativas al proceso (estilo de la enseñanza, estrategias y metodología didáctica, clima social en el aula, violencia escolar, interculturalidad...) las peor resueltas en este tipo de evaluaciones. En consecuencia, la utilidad final que pueden aportar estas evaluaciones al funcionamiento escolar está muy limitada. Por otra parte, es evidente que cuando establecemos los perfiles personales o de grupo de los alumnos resulta inevitable hacer alusión a estas variables contextuales, procedentes del ámbito familiar, sociocultural, del grupo de iguales...

¿Podría mejorarse esta práctica, esta forma de diseñar los cuestionarios de contexto para la evaluación de sistemas educativos? Estimamos que sí, dado que las variables que habitualmente se miden (sean de entrada, proceso o contexto) suelen ser variables sobre las que existe, en muchos casos: a) conocimiento de su funcionalidad a través de la investigación educativa, b) soluciones metodológicas adecuadas para su medición/ evaluación —aunque se hayan desarrollado instrumentos con otros propósitos— y c) experiencia en su funcionamiento en planes de evaluación de sistemas.

Las fases que tener en cuenta para su desarrollo deben subsumirse en un proceso que va más allá del diseño tradicional de instrumentos dirigidos a medir un solo constructo (tal como se describen en este mismo número monográfico en Jornet, González-Such y Perales, 2012). Así, tenemos que:

- a) Definir el concepto de calidad que evaluar, atendiendo a las necesidades de

- información de las partes interesadas. Para ello, hay que hacer operativa la definición de calidad a partir de un modelo, como por ejemplo el descrito por De la Orden (1997, 2007) y especificar las preguntas de evaluación que deberían responderse para que la información sea finalmente útil para la orientación de la política educativa, la mejora en la administración y gestión institucional y del aula y la orientación de las actuaciones familiares y sociales que inciden, de manera informal pero con gran impacto, en el hecho educativo.
- b) Establecido el marco o modelo de referencia al hacer operativo el objeto de la evaluación dirigida a un concepto de calidad específico, es necesario identificar los indicadores que deben incluirse en los cuestionarios de contexto, así como las fuentes de información más pertinentes para cada caso.
 - c) Los indicadores pueden ser simples (representados por un solo reactivo o ítem) o complejos (representados por la síntesis de una escala de valoración, o por el conjunto de diferentes reactivos categoriales).
 - d) La mayor dificultad se centra en la inclusión de los indicadores complejos. Así, por ejemplo, podemos necesitar incluir una evaluación acerca del clima social del aula o del modo en que se atiende la diversidad en los centros escolares (Biencinto-López *et al.*, 2009). Para este propósito existen diversas escalas de percepción, tanto dirigidas al alumnado como al profesorado. Generalmente, las escalas existentes presentan buenas características métricas, pero su excesiva longitud para una evaluación de estas características hace inviable su aplicación. Por ello, es necesario componer los cuestionarios de contexto a partir de indicadores simples e indicadores complejos que estén representados por micro-instrumentos derivados de escalas bien diseñadas, usuales para el diagnóstico o la evaluación de fenómenos psico-socio-educativos.
 - e) Validar los micro-instrumentos en relación a su calidad diferencial respecto a los instrumentos originales.
 - f) Validar el modelo general explicativo que ha orientado la selección de todas las variables e indicadores que se consideran en la evaluación.

A modo de conclusión

En síntesis, podemos afirmar que se pueden aportar planes de evaluación suficientemente solventes para valorar lo que consigue el sistema como producto educativo, pero no está clara la utilidad de la información que aportan estos planes ni para los gestores políticos y/o académicos, ni para los educadores, ni para las familias. De esta manera, señalar lo que se puede hacer para la mejora es probablemente la clave de utilidad de cualquier plan que no satisfacemos adecuadamente en la actualidad.

Los elementos de mejora que hemos señalado orientan hacia compromisos con enfoques innovadores para el diseño de este tipo de planes de evaluación, en los que debe incorporarse el conocimiento adquirido en la investigación educativa, los modelos explicativos del desempeño y una forma de trabajo optimizante para el desarrollo de los sistemas de cuestionarios de contexto.

Notas

¹ Este trabajo se diseña en el contexto del proyecto de investigación *Análisis de Variables de Contexto: Diseño de cuestionarios de contexto para la evaluación de sistemas educativos* (Proyecto AVACO, I+D+I, 2006-2008. Código SEJ 2005-05 923), y se desarrolla en el contexto del Proyecto *Modelos de Análisis de Variables de Contexto* (Proyecto M-AVACO, I+D+I. 2009-2012. Código EDU 2009-13485). Investigador principal: Jesús Jornet Meliá, del Departamento de Métodos de Investigación y Diagnóstico en Educación de la Universitat de València.

² Como en la evaluación de estudiantes, profesores, centros o programas, en las que la evaluación y la intervención subsiguiente están más claramente vinculadas.

³ Por la extensión de referencias, remitimos al lector a la página web de la IEA, anteriormente citada, en la que se pueden consultar todos los informes mencionados.

⁴ En este trabajo no nos referiremos a las evaluaciones censales que se dan en algunos países, estados o regiones autónomas, dado que inicialmente su propósito no es tanto la evaluación del sistema como las posibilidades de aportar información para las diferentes partes interesadas, y tienen en definitiva menor tradición y extensión de uso que las muestrales.

Referencias bibliográficas

- BERICAT, E. (1998). *La integración de los métodos cuantitativo y cualitativo en la investigación social*. Barcelona: Ariel.
- CICCHINELLI, L. F. y BARLEY, Z. (1999). *Evaluating for success. An evaluation guide for districts and schools*. Washington D.C.: McREL.
- COLEMAN, J. S., CAMPBELL, E., HOBSON, C., MCPARTLAND, J., MOOD, A., WEINFELD, F. y YORK, R. (1966). *Equality of educational opportunity*. Washington: U.S. Government Printing Office.
- DE LANDSHEERE, G. (1996). *El pilotaje de los sistemas educativos*. Madrid: La Muralla.
- DE LA ORDEN, A. (1981). Evaluación del rendimiento educativo y la calidad de la enseñanza. En GARCÍA HOZ, V. (ed.) *La calidad de la educación*, Madrid: Publicaciones del Consejo Superior de Investigaciones Científicas, 111-131.
- DE LA ORDEN, A. (2007). Evaluación de la calidad de la educación. Un modelo sistémico como base para la construcción de un sistema de indicadores. *Conceptos, metodologías y experiencias para la construcción de indicadores educativos*. México: Instituto Nacional para la Evaluación de la Educación (INEE), 6-21.
- DELORS, J. (1996). *La educación encierra un tesoro*. Madrid: Santillana - UNESCO.
- GARCÍA, M. (2000). Evaluación de sistemas educativos. En GONZÁLEZ, T. (coord.). *Evaluación y gestión de la calidad. Un enfoque metodológico*. Málaga: Aljibe.
- HARVEY, L. y GREEN, D. (1993). Defining quality. *Assessment and Evaluation in Higher Education*, 18 (1), 9-34.
- JORNET, J. M. y BACKHOFF, E. (2004). Análisis diferencial de perfiles de rendimiento y variables asociadas en los proyectos mexicanos EXANI-I, TIMSS y PISA. En TIRADO, F. (coord.) *Evaluación de la educación en México. Indicadores del EXANI-I*. México: CENEVAL, 423-482.
- JORNET, J. M. (2009). La evaluación del sistema educativo en la España democrática. En JORNET, J. M. (ed.) *La letra sin sangre entra*. Valencia: Publicaciones de la Universitat de València, 205-224.
- JORNET, J. M., GONZÁLEZ-SUCH, J. y PERALES, M. J. (2012). Diseño de cuestionarios de contexto para la evaluación de sistemas educativos: optimización de la medida de constructos complejos. *Bordón*, 64 (2), 89-110.
- MURILLO, F. J. y HERNÁNDEZ-CASTILLA, R. (2011b). Efectos escolares de factores socio-afectivos. Un estudio multinivel para Iberoamérica. *Revista de Investigación Educativa*, 29 (2), 407-427.
- OROVAL, E. (ed.) (1995). *Planificación, evaluación y financiación de sistemas educativos*. Madrid: Civitas.
- PERALES, M. J., CARDONA, L. y GÓMEZ, D. (2009). Conferencia: «Familia y transformación social. Análisis del papel de las familias en los estudios de evaluación de sistemas educativos. Introducción al estudio de validación de un cuestionario», Huelva, XIV Congreso de AIDIPE: Educación, investigación y desarrollo social.

- TEJEDOR, F. J. y CARIDE, J. A. (1988). Influencia de las variables contextuales en el rendimiento académico. *Revista de Educación*, 287, 113-146.
- TIANA, A. (1993). Evaluación de centros y evaluación del sistema educativo. *Bordón*, 45 (3), 283-293.
- TIANA, A. (1996). La evaluación de los sistemas educativos. *Revista Iberoamericana de Educación*, 10, 37-61.
- TIANA, A. y MUÑOZ, F. (1997). Indicadores educativos. *Cuadernos de Pedagogía*, 256, 49-76.

Fuentes electrónicas

- BIENCINTO-LÓPEZ, C., GONZÁLEZ-BARBERA, C., GARCÍA-GARCÍA, M., SÁNCHEZ-DELGADO, P. y MADRID-VIVAR, D. (2009). Diseño y propiedades psicométricas del AVACO-EVADIE. Cuestionario para la evaluación de la atención a la diversidad como dimensión educativa en las instituciones escolares. *Relieve*, 15, 1. <http://www.uv.es/RELIEVE/v15n1/RELIEVEv15n1_4.htm> [Fecha de consulta: 11/diciembre/2011].
- DE LA ORDEN, A. (dir.) (1997). Desarrollo y validación de un modelo de calidad universitaria como base para su evaluación. *Relieve*, 3, 1-2. <<http://www.uv.es/RELIEVE/>> [Fecha de consulta: 13/diciembre/2011].
- DINIECE. <http://www.me.gov.ar/diniece/>
- ESCALE. <http://escale.minedu.gob.pe/>
- GAVIRIA, J. L. y TOURÓN, J. (2000). La evaluación de los sistemas educativos. En: JORNET, J. M. y RAMOS, G. (coords.) *Problemas de la medición y evaluación educativas. Estándares e indicadores para analizar la realidad educativa* (publicación en CD). Valencia: GEM, 227-246. <<http://www.uv.es/gem/archivos/med.pdf>> [Fecha de consulta: 9/diciembre/2011].
- GONZÁLEZ-MONTESINOS, M. J. y BACKHOFF, E. (2010). Validación de un cuestionario de contexto para evaluar sistemas educativos con modelos de ecuaciones estructurales. *Relieve*, 16, 2. <http://www.uv.es/RELIEVE/v16n2/RELIEVEv16n2_1.htm> [Fecha de consulta: 10/enero/2012].
- ICFES. <http://www.icfes.gov.co/>
- IE. <http://www.educacion.gob.es/ievaluacion.html>
- INEP. <http://portal.inep.gov.br/>
- INEE. <http://www.inee.edu.mx/>
- IEA. <http://www.iea.nl/>
- JORNET, J. M., PERALES, M. J. y SÁNCHEZ-DELGADO, P. (2011). El valor social de la educación: Entre la subjetividad y la objetividad. Consideraciones teórico-metodológicas para su evaluación. *Revista Iberoamericana de Evaluación Educativa*, 4 (1), 52-77. <<http://www.rinace.net/riee/numeros/vol4-num1/art3.pdf>> [Fecha de consulta: 10/enero/2012].
- LLECE. <http://www.llece.org/public/>
- MINISTERIO DE EDUCACIÓN (2011). *Sistema estatal de indicadores de la educación. Edición 2010*. Madrid: Instituto de Evaluación - Ministerio de Educación. <www.institutodeevaluacion.educacion.es> [Fecha de consulta: 13/diciembre/2012].
- MURILLO, F. J. y HERNÁNDEZ-CASTILLA, R. (2011a). Factores escolares asociados al desarrollo socio-afectivo en Iberoamérica. *Relieve*, 17 (2) 2. <http://www.uv.es/RELIEVE/v17n2/RELIEVEv17n2_2.htm> [Fecha de consulta: 10/enero/2012].
- NAEP. <http://nces.ed.gov/nationsreportcard/>
- OCDE. <http://www.oecd.org>
- SIMCE. <http://www.simce.cl/>

Abstract

The usefulness of educational systems evaluations: the value of the consideration of context

This paper reviews procedures of evaluation of educational systems. Based on this review, some factors related with the low usefulness of this type of assessment are discussed. The analysis is directed toward the poor design of context questionnaires. It is observed that context questionnaires fail to explain performance. Low capacity performance explanation is the cause of the low utility of educational systems evaluations; because they do not allow for establishing recommendations for improvement. We propose an alternative theoretical and methodological basis for designing context questionnaires. This type of methodology to design the questionnaires of context for the evaluation of education systems is ultimately intended to secure the theoretical quality of the constructs that are measured, and provide a solid basis for defining an explanatory model that allows for extracting more useful information with which we work on assessments of systems. This proposal suggests the need to take as a reference the contributions of educational research, a systemic model that helps to provide coherence to the evaluation plans, consider the information needs of audiences involved in the evaluation, and a more rigorous instrument design procedure. This methodology is based on other work presented in this issue of the journal and cited in this paper.

Key words: *Educational quality, Educational systems evaluation, Context questionnaires, Performance explanation, Design of measuring instruments.*

Résumé

L'utilité des évaluations des systèmes éducatifs: l'importance de considérer le contexte.

Cet article accomplit une révision des procédures d'évaluation des systèmes éducatifs. À partir de cette révision, nous analysons certains facteurs liés à la faible utilité de ce type d'évaluations. L'analyse est dirigée vers la mauvaise conception des questionnaires contextuels. Nous observons que les questionnaires de contexte n'arrivent pas à expliquer la performance. Son faible pouvoir explicatif de la performance est cause de la pauvre utilité des évaluations des systèmes éducatifs; puisqu'elles ne permettent pas l'établissement de recommandations pour l'amélioration. Une base théorique et méthodologique alternative pour la conception de questionnaires contextuels est proposée. Ce type de méthodologie pour la conception de questionnaires contextuels utilisés dans l'évaluation des systèmes éducatifs, a comme dernière finalité assurer la qualité théorique des modèles mesurés et rapporter une base solide pour la définition d'un modèle explicatif qui permet extraire des informations plus utiles avec les évaluations des systèmes. Cette proposition indique le besoin de prendre comme référence: les contributions de la recherche éducative, un modèle systémique qui contribue à la cohérence des programmes d'évaluations, prendre en considération les besoins d'informations des parties prenantes dans l'évaluation, et une élaboration plus rigoureuse des instruments. Cette méthodologie se base sur les résultats d'autres travaux présentés dans ce numéro du journal y que nous mentionnerons dans cet article.

Mots clés: *Qualité éducative, Évaluation des systèmes éducatifs, Questionnaires de contexte, Explication de la performance, Conception d'instruments de mesure.*

Perfil profesional de los autores

Arturo de la Orden Hoz

Catedrático Emérito de la Universidad Complutense de Madrid. Actualmente dedica una parte importante de su actividad a tareas de asesoramiento en investigación y evaluación educativas en universidades y centros mexicanos. Sus líneas de investigación están centradas fundamentalmente en modelos y funciones de la evaluación educativa y en metodología de la investigación.

Correo electrónico de contacto: afordenh@gmail.com

Jesús Miguel Jornet Meliá

Catedrático del Departamento de Métodos de Investigación y Diagnóstico en Educación de la Universitat de València. Coordinador de la Unidad de Investigación GEM (www.uv.es/gem). Su trabajo se orienta en el área de medición y evaluación educativas, al diseño de instrumentos para la evaluación de competencias, cuestionarios de contexto para la evaluación de sistemas educativos y a la evaluación de la dimensión educativa de la cohesión social.

Correo electrónico de contacto: jornet@uv.es