

Capítulo 1

EVALUACIÓN DEL, PARA Y COMO APRENDIZAJE

Melchor Sánchez Mendiola

“Evaluación es un intento de conocer a la persona.”

DEREK ROWNTREE

*“No todo lo que puede ser contado cuenta,
y no todo lo que cuenta puede ser contado.”*

ALBERT EINSTEIN

INTRODUCCIÓN

La palabra evaluación genera múltiples reacciones en las personas, no todas agradables. Solemos decir que las actividades organizacionales deben ser evaluadas, que para mejorar hay que evaluar, que sin evaluación es imposible saber exactamente en dónde estamos, entre otras sentencias, pero cuando se trata de que la evaluación sea personal la situación cambia, sobre todo cuando es sumativa de alto impacto. Al tener noticia de que seremos evaluados, frecuentemente nos genera una sensación de incomodidad, de angustia o hasta miedo, y si podemos diferirla o exentarla en ocasiones, lo hacemos. Evaluar se dice fácil, pero en los hechos no lo es tanto, requiere una actitud reflexiva y madura, recursos para realizarla, personal con entrenamiento y experiencia en sus matices metodológicos y técnicos, tiempo para planearla, realizarla y analizarla, así como infraestructura para documentar las diferentes etapas del proceso. Los procesos de evaluación efectivos requieren estructuras organizacionales participativas, no tan verticales o jerárquicas, que estén dispuestas a aceptar los resultados con entusiasmo y transparencia, para actuar en consecuencia y mejorar la estructura, procesos y resultados del sistema. Se requiere pensamiento sistémico y visión de largo plazo para que el proceso de evaluación se integre adecuadamente al sistema, y participación activa de las personas que conforman los diferentes elementos del mismo. En resumen, la evaluación no es una tarea fácil ni sencilla, se requiere esfuerzo a nivel individual y colectivo, así como apoyo de los diversos niveles de la estructura organizacional.

Reflexionemos sobre los siguientes escenarios educativos:

- Un profesor de matemáticas en la escuela de arquitectura hace un pequeño examen de tres preguntas al inicio de cada clase, sobre el tema de la misma. Al final del curso

promedia los resultados de estas pruebas y los incluye como 5% de la calificación final.

- El departamento de Biología Molecular y Genómica de una escuela de medicina decide, por acuerdo del profesorado, que los exámenes departamentales elaborados por su cuerpo colegiado sean el principal elemento de la calificación final, siendo el 70% de la misma.
- Una docente de la clase de filosofía de la educación en la carrera de pedagogía, rechaza el uso de los exámenes escritos en sus cursos, solamente usa exámenes orales porque son más realistas y permiten explorar pensamiento crítico.
- Un aspirante a ingresar a la universidad, que tiene niveles altos de empatía, habilidades de comunicación, creatividad y curiosidad, no alcanza la puntuación requerida en el examen de conocimientos. Como consecuencia presenta un cuadro importante de depresión.
- Un profesor de cinematografía utiliza gran parte de su tiempo en reuniones individuales con los alumnos, en esos espacios les da retroalimentación detallada y constructiva. Ni los estudiantes ni él identifican estos eventos como actos de evaluación.
- Una profesora de ciencias políticas en la universidad, acostumbrada a usar reactivos de falso y verdadero en sus exámenes, se niega a dejar de utilizarlos a pesar de la abundante evidencia publicada que documenta limitaciones de este formato de preguntas.
- Un estudiante de ingeniería con promedio de 8.5 en el curso, le exige al profesor que redondee la calificación a 9 ya que no sería justo que le disminuyan la puntuación a 8, lo que afectaría su promedio en la carrera.
- El departamento académico de una universidad compara los resultados de los exámenes finales de los cursos de los últimos cinco años, sin tomar en cuenta que los exámenes han sido de diferentes grados de dificultad en cada ciclo.

Todas y cada una de estas viñetas son realistas y tienen diversas implicaciones educativas. En cada una de ellas hay diferentes actores que tienen una serie de experiencias y sesgos hacia la evaluación, que determinan su actitud ante estas situaciones. Estos escenarios tienen implicaciones técnicas, pedagógicas, didácticas, psicológicas, afectivas y socioemocionales, que deben ser analizadas y atendidas con profesionalismo y niveles apropiados de conocimientos sobre el proceso evaluativo. La evaluación como campo de estudio ha crecido y se ha hecho más sofisticada en el último siglo, con incremento sustancial en trabajos de investigación, publicaciones, programas de posgrado, agrupaciones académicas especializadas, y congresos sobre el tema. Sin embargo, han ocurrido una serie de vaivenes en la actitud de la sociedad, las universidades, el profesorado y el estudiantado hacia este campo disciplinario, que con frecuencia generan confusión y hasta rechazo. Es necesario que el profesorado que participa en educación a todos los niveles tenga una comprensión operativa básica de los conceptos relevantes y pertinentes de evaluación en educación, para poder lograr cabalmente el cumplimiento de la labor de la enseñanza.

En este libro nos enfocamos en el uso de la evaluación y sus implicaciones para el aprendizaje, en virtud de que consideramos que los logros educativos de los educandos son de primordial importancia, y del hecho que es indispensable no solo documentar el aprendizaje, sino apoyarlo y propiciarlo. La enseñanza debe integrarse con los mecanismos e

instrumentos de evaluación disponibles para lograr sus metas. Recordemos los comentarios de Sigmund Freud sobre las *profesiones imposibles*, aquellas en las que puedes estar seguro de lograr resultados insatisfactorios, como son el psicoanálisis, el acto de gobernar y la educación (Freud, 1937). Los docentes creemos que lo que enseñamos es aprendido por los estudiantes, desafortunadamente esto no es así, siempre hay un gradiente entre lo que se enseña y lo que se aprende, y también se aprenden cosas diferentes de lo que se pretende enseñar. Por ello, la única manera de tener cierta claridad sobre el efecto de la educación universitaria y su impacto en los estudiantes, es llevar a cabo una evaluación técnicamente adecuada, alineada con el currículo y los métodos de enseñanza, que proporcione resultados interpretables y útiles para los diferentes actores del proceso educativo (Brennan, 2006; Martone y Sireci, 2009). En este capítulo describiremos algunas de las definiciones relevantes a la evaluación del, para y como aprendizaje, y sus implicaciones educativas.

¿QUÉ ES EVALUACIÓN?

Hay muchas definiciones del término *evaluación* en el área educativa, Miller la define como: “un término genérico que incluye un rango de procedimientos para adquirir información sobre el aprendizaje del estudiante y la formación de juicios de valor...” (Miller, 2012). Ello implica un proceso sistemático de acopio de información mediante la aplicación de diversos instrumentos, como son exámenes escritos u orales, para ser analizada con rigor metodológico y así fundamentar la toma de decisiones. La edición más reciente de los Estándares para Pruebas Educativas y Psicológicas de AERA-APA-NCME, define evaluación como: “método sistemático de obtención de información, usado para formular deducciones sobre las características de personas, objetos o programas; proceso sistemático para medir o evaluar las características o el desempeño de individuos, programas u otras entidades con la finalidad de hacer inferencias; en ocasiones se usa como sinónimo de prueba” (AERA, APA y NCME, 2014).

Una prueba (examen o test) es un tipo particular de evaluación que consiste en un conjunto de preguntas o retos, administrados en un periodo de tiempo fijo y bajo condiciones razonablemente comparables para el estudiantado (Miller, 2012). De acuerdo a los Estándares de AERA-APA-NCME una prueba es un “dispositivo de evaluación o procedimiento en el cual se obtiene y puntúa una muestra sistemática del comportamiento de un examinado en un dominio específico, a través de un proceso estandarizado” (AERA, APA y NCME, 2014). Las pruebas son uno de los instrumentos utilizados primordialmente para evaluación sumativa y diagnóstica.

Es pertinente enfatizar la diferencia entre “evaluación” y “medición”, ya que medición es la asignación una descripción numérica a los resultados de una prueba u otro tipo de evaluación de acuerdo a reglas específicas. En ese sentido el enfoque de una medición es más limitado y cuantitativo, mientras que evaluación se conceptualiza como un término más amplio e incluyente, que puede incluir elementos tanto cuantitativos como cualitativos, e incluye juicios de valor sobre el proceso y los resultados (Miller, 2012).

Por otra parte, es importante comentar el uso de las palabras en otros idiomas, ya que en español usamos la misma palabra para evaluación del aprendizaje y evaluación de programas,

mientras que en inglés el término más frecuentemente usado para evaluar aprendizaje es “*assessment*”, y para evaluar programas es “*evaluation*”. Esto puede prestarse a confusión al usar la literatura académica internacional, la cual predominantemente está en inglés (aunque esto no es una regla, ya que varía dependiendo del contexto y del país) (McKean y Aitken, 2016). La terminología utilizada en educación y particularmente en evaluación educativa es frecuente fuente de discusión, desacuerdos y confusión, por lo que es importante mantener una actitud abierta y reflexiva a los diversos usos de los términos en diferentes países y grupos profesionales, así como consultar los glosarios internacionales de estos términos, como el de la UNESCO <https://learningportal.iiep.unesco.org/en/glossary>.

Independientemente de las definiciones técnicas que usemos de evaluación y sus conceptos cercanos, el profesorado que tiene interacciones con estudiantes debe internalizar la evaluación desde una visión más profunda, como sugirió en 1977 Derek Rowntree, de Australia: “Cuando una persona, con algún tipo de interacción directa o indirecta con otra, obtiene e interpreta información de manera consciente sobre el conocimiento y la comprensión, habilidades y actitudes de la otra persona. Hasta cierto punto *evaluación es un intento de conocer a esa persona*” (Rowntree, 1977). Se deben tener en cuenta los siguientes principios generales en evaluación educativa (Miller, 2012):

- 1) Es prioritario especificar claramente lo que se va a evaluar.
- 2) La evaluación es un medio para un fin, no un fin en sí mismo.
- 3) Los métodos de evaluación deben elegirse con base en su relevancia, tomando en cuenta las características que se van a evaluar.
- 4) Se requiere una variedad de procedimientos e instrumentos para que la evaluación sea útil y efectiva.
- 5) Su uso adecuado requiere conocimiento sobre las bases técnicas de la evaluación y sus instrumentos, así como sobre las bondades y limitaciones de cada método.

En este libro nos enfocamos en los aspectos de la evaluación que tienen implicaciones en el aprendizaje, por lo que describiremos los conceptos de evaluación **del** aprendizaje, **para** el aprendizaje y **como** aprendizaje. Como será aparente para el lector, el uso de estas preposiciones y mediadores terminológicos son de particular importancia en la actualidad, cuando la evaluación debe convertirse en un elemento consustancial de la enseñanza y el aprendizaje. Debemos potenciar el uso de los avances logrados en este campo para que los estudiantes y docentes no sufran la evaluación, que no la vivan como una serie interminable de exámenes que generan estrés, ansiedad y desgaste, para en su lugar lograr un aprendizaje más auténtico e integrado (Bennett, 2015; Brennan, 2006; Moreno-Olivos, 2010). Al final del día, la evaluación y el aprendizaje son partes de un todo que puede visualizarse como un ciclo (Figura 1).

En el esquema del ciclo pueden apreciarse diversos elementos, todos fundamentales para que el proceso se convierta en un círculo virtuoso: la misión de la institución educativa con sus metas y objetivos de aprendizaje, la identificación de los constructos a evaluar, la

colección de evidencia de aprendizaje durante el proceso, la interpretación de dicha evidencia y su uso para retroalimentación y mejora continua, y la implementación reflexiva de cambios en el proceso educativo, para aplicarse en subsecuentes ciclos. Los conceptos descritos a continuación son analizados con mayor profundidad en otros capítulos del libro, se presentan aquí como parte del panorama general del proceso de evaluación y sus implicaciones para el aprendizaje.

Figura 1. El ciclo de la evaluación y el aprendizaje (adaptado de Maki, 2010)



CLASIFICACIÓN TRADICIONAL DE TIPOS DE EVALUACIÓN

Evaluación diagnóstica, sumativa y formativa. Una de las clasificaciones tradicionales de la evaluación educativa, desde el punto de vista de su objetivo, es: diagnóstica, sumativa y formativa.

La evaluación *diagnóstica* se realiza al principio de un curso o actividad académica, con la finalidad de determinar el nivel de conocimiento, habilidad o actitud del estudiante ([Capítulo 3](#)). Esta información puede ser de utilidad para las instituciones educativas y el profesorado, ya que permite hacer adecuaciones en el contenido y en la implementación de las actividades programadas que correspondan a las características del alumnado.

La evaluación *sumativa* es aquella compuesta por la suma de valoraciones efectuadas durante un curso o unidad didáctica, a fin de determinar el grado con que los objetivos de la instrucción se alcanzaron, otorgar calificaciones o certificar competencias ([Capítulo 5](#)). Este tipo de evaluación se inserta en el marco de “evaluación **del** aprendizaje”, que se comentará

más adelante en este capítulo. Un tipo de exámenes sumativos merece atención especial, los llamados *exámenes de altas consecuencias o de alto impacto* ([Capítulo 5](#)) (Sánchez y Delgado, 2017).

La evaluación *formativa* es la que se utiliza para monitorizar el progreso del aprendizaje con la finalidad de proporcionar retroalimentación al estudiante sobre sus logros, deficiencias y oportunidades de mejora ([Capítulo 4](#)). Esta evaluación ocurre a lo largo de todo el proceso educativo del estudiantado, puede ser *formal* o *informal*. La evaluación formativa tiene un poderoso componente educativo, ya que permite identificar aquellas actividades que se llevan a cabo correctamente para continuar realizándolas de dicha manera, y aquellas que poseen alguna deficiencia, poder detectarlas a tiempo y corregirlas (Man Sze Lau, 2016; Martínez Rizo, 2009). Este tipo de evaluación forma parte de la *evaluación para el aprendizaje*, el cual se comentará más adelante.

EVALUACIÓN REFERIDA A NORMA Y CRITERIO

Otra forma tradicional de clasificar a la evaluación es de acuerdo con la interpretación de los resultados. Puede ser con referencia a norma (relativa), o con referencia a criterio (absoluta). Cuando los resultados de la evaluación se interpretan con *referencia a norma*, el resultado se describe en términos del desempeño del grupo y de la posición relativa de cada uno de los estudiantes evaluados (Miller, 2012; Montgomery y Connolly, 1987). Este tipo de evaluación se utiliza para colocar a los alumnos en escalas de rendimiento y puntaje, con la finalidad de asignarles un lugar dentro del grupo. Ejemplos de este tipo de interpretación son: la cantidad de errores de ortografía en un manuscrito es igual o mayor que la del 60% de su grupo; la puntuación en el examen de biología del estudiante está en el percentil 75 de una muestra nacional que ingresó a la universidad.

La evaluación con *referencia a criterio* describe el resultado específico que se encontró, de acuerdo a criterios o metas preestablecidos. Este tipo de evaluación busca la comparación del estudiante con relación a un estándar establecido previamente. Ejemplos de este tipo de evaluación son: los exámenes de inglés como segunda lengua (*Test of English as a Foreign Language*, TOEFL), en el que hay niveles de desempeño previamente determinados y los resultados se interpretan de acuerdo a dichos estándares, no de acuerdo al desempeño del grupo de sustentantes; un examen de certificación de una especialidad médica, en el que se establece un nivel determinado de desempeño específico para acreditarla. En el contexto de la evaluación criterial, puede darse el caso de que ningún estudiante apruebe el examen al no alcanzar el mínimo establecido, lo que no ocurre con las evaluaciones con referencia a norma o relativas.

La evaluación normativa versus la evaluación criterial tiene diversas aristas (Lok et al, 2016). Mientras que los resultados de un examen criterial pueden utilizarse para jerarquizar a los estudiantes por la puntuación obtenida, en un examen normativo también pueden definirse niveles de desempeño. Desde un punto de vista estricto, la interpretación con referencia a norma o a criterio se refiere solo al método de interpretación de los resultados, no tanto al examen en sí, de manera que estas interpretaciones serán más significativas y útiles

en tanto que las evaluaciones se diseñen específicamente para la interpretación a realizar. Los dos tipos de evaluación descritos se pueden conceptualizar como extremos de un continuo, más que una dicotomía estricta, en el que el extremo de la evaluación con referencia a criterio se enfoca en la descripción del desempeño de la persona, y el extremo con referencia a norma enfatiza la discriminación o identificación de los individuos (Miller, 2012). En el Tabla 1 se describen las principales diferencias entre estos tipos de evaluación.

Tabla 1. Comparación de las características de la evaluación con referencia a criterio y con referencia a norma (adaptado de Lok et al, 2016; Miller, 2012; Montgomery y Connolly, 1987)

	Evaluación con referencia a criterio (absoluta)	Evaluación con referencia a norma (relativa)
Propósito	<ul style="list-style-type: none"> • Refleja el progreso del desarrollo de estudiantes individuales. • Usada para planeación de programas y certificación de individuos. 	<ul style="list-style-type: none"> • Genera un ordenamiento de los estudiantes por sus posiciones relativas en el grupo • Utilizada para diagnóstico y colocación de estudiantes.
Diseño de tareas evaluadas	Alineadas con el contenido y resultados de aprendizaje esperados.	Discrimina a los estudiantes de alto y de bajo desempeño.
Unidad de interpretación de las puntuaciones	Individual.	Grupal.
Presentación de los resultados	Calificaciones ligadas a criterios determinados.	Calificaciones derivadas de puntuaciones crudas, generalmente presentadas en una campana de distribución o en niveles gruesos de puntuaciones.
Ventajas	<ul style="list-style-type: none"> • Capacidad de reflejar el desempeño real y el progreso de cada estudiante. • Permite la incorporación explícita de habilidades cognitivas de alto nivel. • Es más defendible si se establecen adecuadamente los estándares de desempeño. 	<ul style="list-style-type: none"> • Permite comparar a los individuos dentro del grupo. • Puede ajustarse estadísticamente para tener una amplitud específica (por ejemplo: desviación estándar). • Puede evitar la “inflación de calificaciones”.
Desventajas	<ul style="list-style-type: none"> • No hay control en la distribución de las calificaciones, puede llevar a “inflación de calificaciones”. • Limitada capacidad de comparar el desempeño individual contra el del grupo. • Puede enfrentar obstáculos para su adopción, si la organización desconoce la metodología y está acostumbrada a la evaluación con referencia a norma. 	<ul style="list-style-type: none"> • El desempeño del grupo puede influir las calificaciones individuales, lo que puede ser injusto. • Puede no mostrar el rango de diferencias reales entre los estudiantes individuales • No nos dice con precisión si los individuos son competentes o tienen las habilidades requeridas para una tarea.

INSTRUMENTOS DE EVALUACIÓN DEL APRENDIZAJE

Un aspecto indispensable de la evaluación educativa son las herramientas o instrumentos que utiliza. Los instrumentos de evaluación son técnicas de medición y recolección de datos que tienen distintos formatos, atendiendo a la naturaleza de la evaluación. Existe una gran variedad de instrumentos con diversas ventajas y limitaciones para documentar el aprendizaje de los conocimientos, habilidades y destrezas de los estudiantes. Es responsabilidad del profesor y de las instituciones educativas elegir los métodos apropiados para el proceso de evaluación. Estos pueden clasificarse en las siguientes categorías:

- *Evaluaciones escritas*: ensayos, preguntas directas de respuesta corta, exámenes de opción múltiple, relación de columnas, disertaciones, reportes.
- *Evaluaciones prácticas*: exámenes orales, exámenes prácticos con casos, examen clínico objetivo estructurado (ECOFE).
- *Observación*: reporte del profesor, listas de cotejo, rúbricas.
- *Portafolios y otros registros del desempeño*: libretas de registro, portafolios, registros de procedimientos.
- *Autoevaluación y evaluación por pares*: reporte del educando, reporte de los compañeros.

Cada uno de estos instrumentos tiene ventajas y desventajas, características psicométricas, así como recomendaciones para su implementación. Es responsabilidad de los profesores y responsables de la evaluación en las instituciones educativas diseñar, seleccionar, implementar y acumular evidencia de validez de los instrumentos más apropiados para evaluar el aprendizaje de los estudiantes, de acuerdo al currículo y las características del contexto local.

CRITERIOS PARA UNA BUENA EVALUACIÓN

La evaluación educativa es tan buena como la metodología utilizada y el uso que se hace de los resultados. Varias organizaciones internacionales han propuesto criterios sobre las *buenas prácticas* en evaluación, a continuación, describimos brevemente algunos (Norcini et al, 2011):

- 1) **Validez**. Uno de los conceptos más importantes para que los resultados de los procesos de evaluación tengan sustento sólido y uso apropiado, es el de *validez*. El [capítulo 2](#) de este libro explora este tema con mayor profundidad.
- 2) **Confiabilidad o fiabilidad**. Es la capacidad del examen de arrojar un resultado consistente cuando se repite, es decir, la reproducibilidad del examen. El concepto también es analizado en el [capítulo 2](#).
- 3) **Justicia y equidad**. En las últimas décadas las principales organizaciones de evaluación educativa del mundo han enfatizado la necesidad de justicia y equidad en todo el proceso educativo, incluyendo la evaluación del aprendizaje, para ser congruentes con el principio social de la educación (AERA, APA y NCME, 2014).

- 4) **Equivalencia.** Se refiere a que los exámenes proporcionen puntuaciones o decisiones equivalentes, cuando se administran en diferentes instituciones o tiempos (AERA, APA y NCME, 2014; Norcini et al, 2011). Es un concepto poco conocido por el profesorado, a pesar de su importancia para interpretar exámenes aplicados de manera periódica que pretenden evaluar el mismo constructo, o exámenes en diferentes contextos en los que queremos asegurar que sean de la misma dificultad, sobre todo en evaluación sumativa de alto impacto (Carter, 1984). Para lograr equivalencia se requieren procedimientos estadísticos de varios grados de sofisticación, que se encuentran en el grupo de métodos de equiparación o *igualación* de exámenes. Uno de ellos es el uso de *reactivos ancla* (preguntas con un grado de dificultad similar y comportamiento estadístico bien documentado) en un porcentaje de reactivos de cada versión del examen. Para el uso óptimo de estas técnicas se requieren profesionales de evaluación educativa expertos en estos procedimientos ([capítulo 20](#)).
- 5) **Factibilidad y aceptabilidad.** Estas propiedades se refieren a que las evaluaciones sean prácticas, realistas y apropiadas a las circunstancias del contexto, incluyendo instalaciones físicas, recursos humanos y financieros. Por ejemplo, el método más utilizado en el mundo para evaluar la competencia clínica en medicina es el Examen Clínico Objetivo Estructurado, que consiste en una serie de múltiples estaciones estandarizadas, en las que el sustentante se enfrenta en cada una a un reto que requiere la aplicación de competencias específicas, como pueden ser habilidades de comunicación, realizar un diagnóstico, o interpretar estudios de laboratorio (Boursicot et al., 2011). Este tipo de examen requiere gran cantidad de recursos humanos, instalaciones apropiadas y mucha dedicación en términos de disciplina, tiempo y organización. Esta disponibilidad de recursos puede no estar al alcance de varias escuelas, de manera que, aunque el examen tenga abundante evidencia de validez y confiabilidad, si no es factible deben buscarse otras alternativas como los exámenes orales o la observación directa. Otros ejemplos podrían ser el uso de exámenes adaptativos por computadora, o simuladores de alta fidelidad, herramientas que requieren gran inversión inicial y de mantenimiento. Las evaluaciones deben ser aceptables tanto para los estudiantes como para los profesores. Si hay un rechazo de la comunidad a algún tipo de evaluación, por ejemplo, a la evaluación por pares, eso hará difícil su implementación.
- 6) **Efectos educativo y catalítico.** Los métodos de evaluación, sobre todo los de índole sumativa, tienen efectos en los métodos de estudio y las prioridades de aprendizaje de los estudiantes (Newble, 1983). Aunque el profesor le diga al estudiante que un tema es fundamental, la pregunta inevitable es: “¿eso va a venir en el examen?” La cultura prevalente en varios sitios es que si algo no cuenta para el examen no se le da mucha importancia, por lo que la manera como se aplica la evaluación tiene efectos en la motivación de los estudiantes y en sus métodos de estudio. La evaluación tiene también un efecto *catalítico*, ya que puede tener influencia en los demás docentes, los departamentos académicos y la misma institución (Norcini et al., 2011). Si se privilegian los exámenes escritos de opción múltiple, habrá un efecto en cascada en los diferentes actores del proceso educativo. Si

se fomenta la evaluación formativa, de la misma manera pueden cambiar las actitudes hacia la evaluación de los participantes, sobre todo cuando viven sus efectos positivos.

¿QUÉ ES LA EVALUACIÓN DEL, PARA Y COMO APRENDIZAJE?

En las últimas décadas ha ocurrido un movimiento pendular del énfasis en la evaluación sumativa de alto impacto, con exámenes escritos de opción múltiple, estandarizados a gran escala, hacia una evaluación más amplia y profunda, con una visión más holística de su rol en el proceso de enseñanza-aprendizaje (Bennett, 2015; Hodges, 2013). El paradigma preponderante desde mediados del siglo XX, dominado por el discurso de la psicometría y la objetividad, en el que las conductas y tareas humanas son convertidas en números, está siendo modificado por el movimiento hacia un uso más adecuado de la evaluación como parte del proceso de aprendizaje. Las ventajas tradicionales de los exámenes estandarizados objetivos, de aparente justicia y equidad con igualdad de condiciones para los participantes, han sido fuertemente cuestionadas y el concepto de evaluación **para** el aprendizaje ha adquirido fuerza.

La importancia de evaluaciones más auténticas, en ambientes de trabajo más dinámicos e interprofesionales que requieren trabajo en equipo, han impulsado un importante cambio en las prioridades de la sociedad, las universidades, el profesorado, el mercado laboral y el estudiantado. Las evaluaciones sumativas de alto impacto requieren gran inversión en recursos financieros, humanos, técnicos y de tiempo, que si bien siguen teniendo importancia en contextos específicos, deben balancearse con otras estrategias y visiones de la evaluación que ayuden a los estudiantes a aprender, que rescaten los aspectos éticos, socioemocionales y humanos de la educación, usando metodologías cualitativas, mixtas, que exploten la riqueza de la subjetividad que ocurre en los espacios de aprendizaje formales e informales. En diversos capítulos del libro se abordan temas relativos a ello ([capítulos 35, 37, 38, 40](#)). Por otra parte, la pandemia por COVID-19 ha generado un verdadero tsunami educativo a nivel global, con un fuerte impacto en los exámenes sumativos en ambientes controlados, con una migración a las evaluaciones en línea, incluso en casa de los estudiantes. Esto tiene diversas implicaciones que se comentan en los capítulos correspondientes ([capítulos 5 y 9](#)).

En los últimos años ha adquirido fuerza el concepto de “evaluación **del-para-como** aprendizaje” (“*assessment of-for-as learning*” en inglés), que pretende modificar el énfasis que ha existido en la evaluación sumativa, exámenes y calificaciones, hacia un panorama más amplio e integrado que nos lleve a anclar todo el proceso de evaluación con el aprendizaje, la meta fundamental del proceso educativo (Ashford-Rowe et al, 2014; Bennett, 2015; Harapnuik, 2021; Maki, 2010; NFETLHE, 2017a). Si bien en español se escucha un poco extraño combinar las palabras del, para y como, con evaluación y aprendizaje, sobre todo cuando se verbalizan oralmente, creemos que el argumento más importante para apoyar el uso y aplicación de este concepto en nuestro medio, es centrar los esfuerzos en el aprendizaje. Propiciar que la evaluación no sea un elemento con cierto carácter punitivo que empuja o determina el aprendizaje, sin tomar en cuenta la libertad de cátedra, el currículo formal

y oculto, el bienestar de los estudiantes, para poder migrar a un panorama más integrado y holístico. En última instancia el propósito principal de la evaluación es potenciar y mejorar el aprendizaje, no solo documentarlo, por lo que el aprendizaje idealmente debe ser el centro de todas las evaluaciones (McKean y Aitken, 2016). La enseñanza, el aprendizaje y la evaluación son conceptos y actividades que están ligados de manera inextricable, y el alineamiento entre estos elementos con la planeación, diseño e implementación curricular es indispensable y se convierte en un elemento clave para el éxito educativo. En el siguiente apartado describiremos por separado los tres elementos “del”, “para” y “como”, para después integrarlos en un modelo que toma lo mejor de los tres y lo liga con el resto del proceso educativo.

- **Evaluación DEL aprendizaje** (“*assessment OF learning*”). De acuerdo con varios autores este tipo de evaluación es equivalente a la evaluación sumativa ([capítulo 5](#)), para documentar que el aprendizaje ocurrió y el nivel del mismo. Su naturaleza es evaluar actividades que ya ocurrieron, después o al final de un periodo de aprendizaje, y enfatiza los aspectos cuantitativos y numéricos, asociándose con las calificaciones o grados. Cuando esta evaluación tiene consecuencias importantes en el estudiante, se le denomina “evaluación de alto impacto” ([capítulo 5](#)). En este tipo de evaluación el actor principal es el profesor o la organización que aplica la evaluación, quienes son los principales tomadores de decisiones, y el estudiantado es un participante pasivo que recibe o a quien se le aplica el examen o la prueba, en contraste con la evaluación para el aprendizaje ([Figura 2](#) y [Tabla 2](#)).
- **Evaluación PARA el aprendizaje** (“*assessment FOR learning*”). Como se comentó anteriormente, la principal meta de la evaluación debe ser mejorar el aprendizaje, no solo medirlo, por lo que cuando hablamos de evaluación para el aprendizaje nos referimos a la evaluación tradicionalmente llamada formativa, ligada a la retroalimentación (Maki, 2010; Man Sze Lau, 2016; Martínez Rizo, 2009; Wiliam, 2011). Esta evaluación ocurre durante todo el proceso de enseñanza y aprendizaje, es más longitudinal, representa un diálogo que ocurre entre docentes y estudiantes a lo largo de sus múltiples interacciones. Está enfocada en ayudar al estudiante, identificar sus áreas de oportunidad y logros, para orientarlo a progresar de una mejor manera en el proceso educativo, sin generarle estrés o desgaste, tratándolo como persona. Pretende moverse de una acción que se le hace **al** estudiante, a un proceso que se hace **con** el estudiante. Esta evaluación es inseparable de la enseñanza y apoya fuertemente el aprendizaje, si se lleva a cabo con profesionalismo y responsabilidad ([Figura 2](#), [Tabla 2](#)).
- **Evaluación COMO aprendizaje** (“*assessment AS learning*”). En este tipo de evaluación el estudiante es empoderado, tiene mayor responsabilidad en el proceso de aprendizaje y puede ser el tomador clave de decisiones. El estudiantado necesita adquirir habilidades para el uso de los conceptos básicos de evaluación en su desarrollo personal. El aprendizaje autodirigido a lo largo de la vida, el aprendizaje autónomo, el pensamiento crítico, entre otros, requieren evaluar datos e información sobre situaciones laborales y vivenciales, analizarlos, establecer juicios de valor, tomar decisiones sobre la problemática

personal y profesional. Todo ello requiere autoevaluación y capacidad de tomar decisiones con base en evaluación de contextos y realidades complejas. Aunque el profesorado generalmente es quien tiene el poder jerárquico en el proceso educativo formal, la evaluación como aprendizaje mueve este locus de control externo a un control más intrínseco por parte del estudiantado. Sin embargo, el estudiante requiere apoyo de docentes y sus pares para ejercer cabalmente las habilidades mencionadas. La evaluación como aprendizaje ayuda a los estudiantes a aprender a aprender, fomenta la meta-cognición y el aprendizaje autorregulado (Figura 2, Tabla 2).

Diversos autores también incluyen en el paraguas conceptual de evaluación como aprendizaje, al aprendizaje potenciado por los exámenes (*test-enhanced learning*), uno de los beneficios importantes de la evaluación para mejorar el aprendizaje a través de pruebas, descrito en el [capítulo 7](#) (Green et al, 2018; Roediger et al, 2011).

Figura 2. Elementos de la evaluación del, para y como aprendizaje
(adaptado de National Forum for the Enhancement of Teaching and Learning in Higher Education, 2017a)

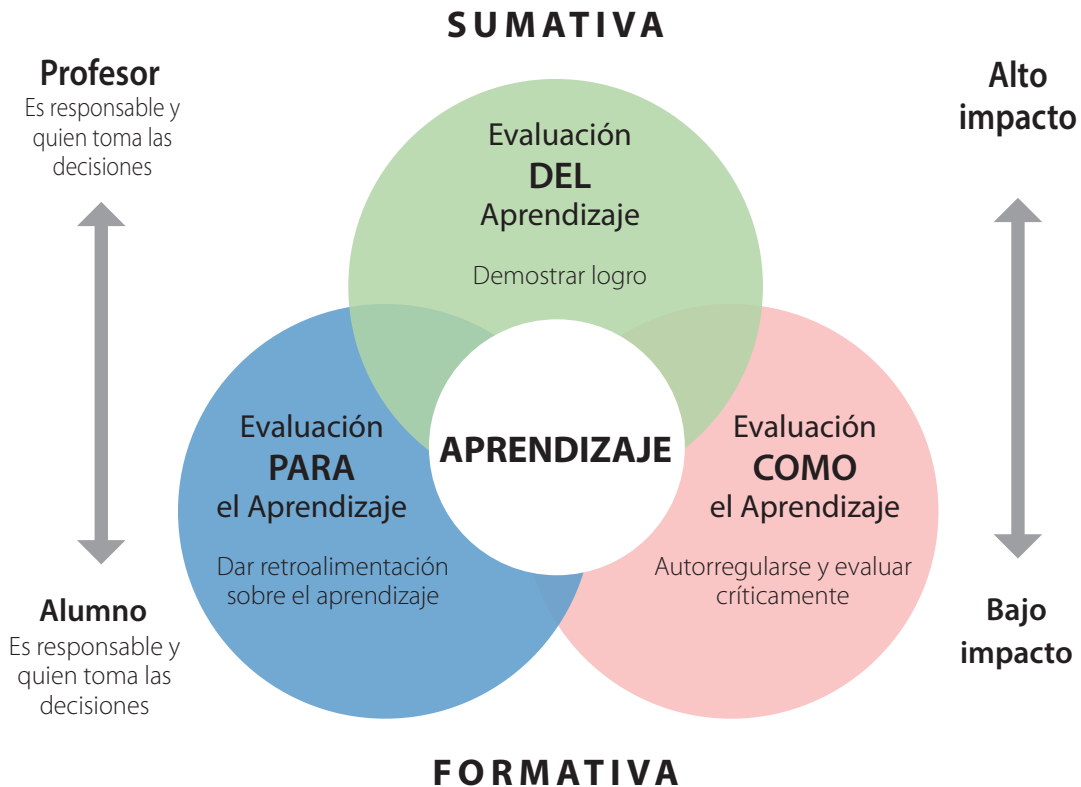


Tabla 2. Características de la evaluación del, para y como aprendizaje (adaptado de Harapnuik, 2021)

Evaluación	Del aprendizaje	Para el aprendizaje	Como aprendizaje
Tipo	Sumativa	Formativa	Formativa
Qué	El profesorado determina el progreso del estudiante o la aplicación de conocimientos y habilidades comparados con un estándar.	Docentes y pares valoran el progreso y el aprendizaje para ayudar a los estudiantes a mejorar.	El educando toma responsabilidad de su propio aprendizaje y hace preguntas sobre el proceso, explora cómo mejorar.
Quién	Profesor(a).	Profesor(a) y pares	Estudiante y pares.
Cómo	Evaluaciones formales para coleccionar evidencia del progreso del estudiante, se usa para medir logro y otorgar calificaciones.	Actividades de evaluación formales e informales como parte del aprendizaje, para mejorarlo e informar la planeación del aprendizaje futuro.	Los estudiantes usan la retroalimentación formal e informal, así como la auto-evaluación, para ayudar a su comprensión de los siguientes pasos en su aprendizaje. Usan además los exámenes como herramienta de aprendizaje.
Cuándo	Exámenes y reportes periódicos.	Retroalimentación continua	Práctica reflexiva continua
Por qué	Clasificar, documentar, verificar, promover.	Mejorar el aprendizaje.	Lograr aprendizaje más profundo, aprender a aprender.
Énfasis	Puntuaciones, calificaciones, competitividad.	Retroalimentación, apoyo, colaboración.	Colaboración, reflexión, auto-evaluación.

Después de analizar por separado estos tres elementos de la evaluación, es aparente que forman parte de un modelo conceptual entrelazado con interacciones dinámicas y cierto grado de superposición entre ellos, en el que las tres categorías (del, para y como) no son mutuamente excluyentes. La [Figura 2](#) ejemplifica sus relaciones y puntos de contacto, así como el rol cambiante de docentes, estudiantes e impacto de las evaluaciones en cada uno de ellos. Por ejemplo, las evaluaciones de bajo impacto, como la que hacemos a las participaciones en clase de los estudiantes, cuyos propósitos son diagnósticos, motivacionales y para retroalimentación, son un ejemplo de superposición entre los círculos de evaluación del y para el aprendizaje. Cuando los estudiantes reciben tareas o exámenes para auto-eva-

luación, hay superposición entre evaluación del y como aprendizaje. Existen varias combinaciones potenciales entre estas tres grandes estrategias de evaluación y sus implicaciones para el aprendizaje, pero el concepto central que las une es que todas ellas deben facilitar el aprendizaje de los estudiantes.

Cada componente del modelo tiene sus particularidades, pros y contras, énfasis en diferentes tiempos y propósitos, pero cada uno tiene un rol clave en el aprendizaje de los estudiantes y el éxito del proceso educativo. La comprensión de estos elementos y sus interacciones por el profesorado, el estudiantado y las autoridades educativas, permitirá un diseño e implementación más adecuada de las experiencias de aprendizaje. Eventualmente sería deseable que la necesidad del uso diferenciado de estos términos pase a la historia, ya que idealmente toda la evaluación debería ser para el aprendizaje. En palabras de Black y Wiliam (1998), “el usuario último de la información de evaluación que se obtiene para mejorar el aprendizaje es el estudiante”, por ello es fundamental que los estudiantes no sean ajenos a las actividades de evaluación que diseñan e implementan las universidades y el profesorado. La evaluación debe motivar al estudiantado y apoyarle en su transición a ser un aprendiz para toda la vida.

EVALUACIÓN AUTÉNTICA

Uno de los principales retos de la evaluación, sobre todo en escenarios artificiales estandarizados controlados, es la poca autenticidad de los exámenes escritos de baja fidelidad, por lo que es importante comentar el concepto de la evaluación auténtica. El término “evaluación auténtica” no es un tipo separado de evaluación, sino más bien un adjetivo que denota características de la misma que se acercan a los retos que las personas enfrentan en la vida cotidiana, en el trabajo y el ejercicio profesional, y que promueve el aprendizaje (McKean y Aitken, 2016). Este tipo de evaluación pretende replicar las tareas y el desempeño encontrados típicamente en el mundo del trabajo, y se ha documentado que tiene un impacto positivo en el aprendizaje, la autonomía, la motivación, la meta-cognición y la auto-regulación (Villarroel et al, 2018). Autenticidad se entiende como realismo, que implica enlazar el conocimiento con la vida diaria y el trabajo, en contextos relevantes, con problematización de los retos a resolver definidos en el currículo. Este tipo de evaluación intenta integrar lo que ocurre en la escuela y el aula con el empleo y el escenario del trabajo.

Se han descrito criterios para que una evaluación sea auténtica: representatividad (que evalúe adecuadamente el panorama del dominio); significado (tareas relevantes y pertinentes a la vida real); complejidad cognitiva (habilidades de orden superior); y cobertura de contenido (material incluido amplio) (Ashford-Rowe et al, 2014; McKean y Aitken, 2016). En una revisión sistemática del tema, Villarroel y colaboradores identificaron 13 características consistentes en la evaluación auténtica, que se pueden organizar en tres dimensiones conceptuales (Tabla 3).

Tabla 3. Dimensiones y características de la evaluación auténtica (adaptado de Villarroel et al, 2018)

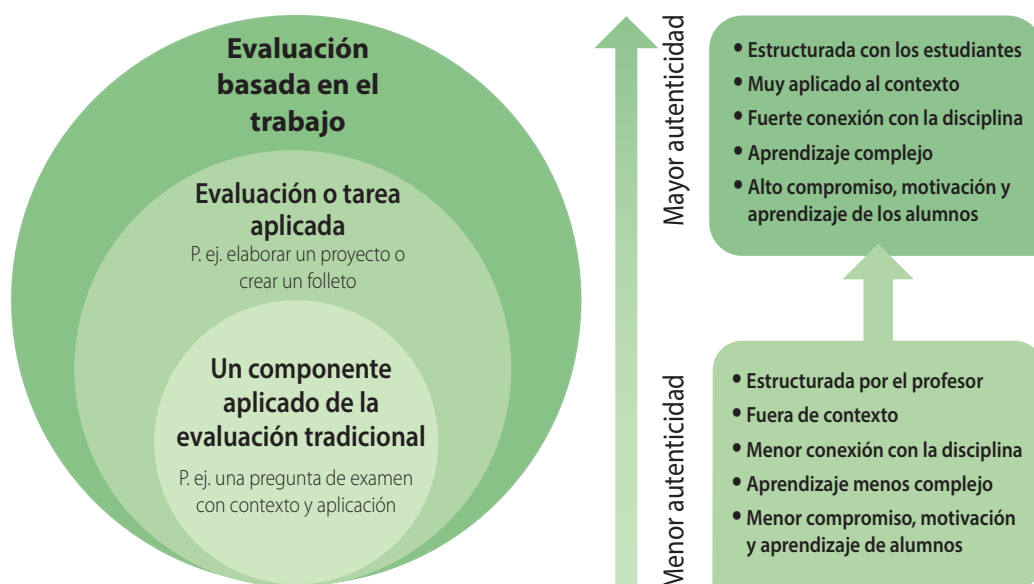
Dimensiones	Características
Realismo	<ul style="list-style-type: none"> • Problemas contextualizados de la vida diaria. • Relevancia más allá del salón de clase. • Desempeño auténtico. • Competencias para el desempeño en el trabajo. • Tareas similares al mundo real. • Valor práctico.
Reto cognitivo	<ul style="list-style-type: none"> • Pensamientos de orden superior. • Habilidad para resolver problemas. • Habilidad para tomar decisiones.
Juicio evaluativo	<ul style="list-style-type: none"> • Retroalimentación. • Sentimiento formativo. • Conocimiento a priori de los criterios de evaluación.

La evaluación auténtica ha adquirido cada vez más relevancia en el mundo moderno, en el que los empleadores identifican carencias y limitaciones en las habilidades para resolver problemas de los graduados de las universidades. Esto ha motivado que se incrementen los métodos de evaluación que utilizan esta perspectiva, combinando el uso de la tecnología, agregando escenarios y viñetas que den contexto y problematización a los reactivos de los exámenes escritos, utilizando centros de simulación con simuladores reales o virtuales, entre otras herramientas e instrumentos.

A pesar de sus beneficios, existen barreras para la adopción generalizada de la evaluación auténtica que deben tomarse en cuenta: la tradición en algunas universidades en las que los exámenes escritos de conocimiento decontextualizados son la norma; la dificultad para incorporar los principios de evaluación auténtica en el día a día de la docencia y evaluación, por cuestiones de tiempo y recursos; así como la limitada formación docente sobre el tema. De cualquier forma, es fundamental que en las evaluaciones modernas se haga el esfuerzo de incorporar la evaluación auténtica, para mejorar el aprendizaje de los estudiantes y su aceptación de las evaluaciones. Al sentir que los exámenes son más realistas y apegados a los

problemas que enfrentarán en la vida real, estudiantes y docentes se sentirán más satisfechos y motivados a participar en el proceso. Por otra parte, se debe conceptualizar a la autenticidad como un continuo, con niveles variables de fidelidad (Figura 3). Este esquema ayuda a pensar en el diseño de evaluaciones que pueden ir de un nivel menor a uno mayor de autenticidad y realismo, de cualquier forma, diseñar pruebas y exámenes requiere que estos estén alineados con las metas de aprendizaje del currículo y los métodos de enseñanza de la asignatura o curso. La evaluación auténtica es fuertemente determinada por la disciplina o área del conocimiento de la profesión que se estudia (medicina, música, leyes, enfermería, etc.), ya que el contexto de la práctica influye en la percepción de la importancia y legitimidad de lo que se debe aprender. Los estudiantes anticipan los escenarios en donde aplicarán lo aprendido en la universidad.

Figura 3. El continuo de la evaluación auténtica (adaptado de NFETLHE, 2017).



EVALUACIÓN ALTERNATIVA

El término de evaluación alternativa se ha usado con mayor frecuencia en los últimos años, y es particularmente importante en el contexto actual de la pandemia por COVID-19, en el que ha habido un movimiento amplio y profundo hacia la evaluación en línea, a distancia y mediada por tecnología (Jiao y Lissitz, 2020). El tema de la evaluación en línea y sus implicaciones en la pandemia es analizado en el [capítulo 9](#). En este apartado planteamos el concepto de evaluación alternativa y sus implicaciones en el proceso educativo. Este término se refiere a una aproximación a la evaluación en forma diferente a la que se ha hecho tradicionalmente

(generalmente exámenes escritos con papel y lápiz). Stobart y Gipps describen tres niveles de evaluación alternativa que se refieren a las diferentes facetas del concepto: en primera instancia, en el nivel más básico, puede referirse a formatos alternativos, por ejemplo, el uso de una computadora en lugar del examen con papel y lápiz, sin modificar el contenido de la prueba, o usar preguntas abiertas en lugar de reactivos de opción múltiple; el segundo nivel puede ser el uso de un enfoque diferente a la evaluación, por ejemplo evaluación del desempeño, uso de portafolios; y el tercero sería el uso de evaluación formativa para el aprendizaje, en lugar (o complementario a) la evaluación sumativa del aprendizaje con exámenes de alto impacto, como se ha comentado en este capítulo.

Más allá de utilizar el término de evaluación alternativa como algo diferente a lo que se hace tradicionalmente, es importante revisar nuestras premisas de lo que es la educación, enseñanza, aprendizaje y evaluación, y explorar evaluaciones alternativas que pueden tener diferentes premisas. Anderson (1998) propuso que la evaluación alternativa tiene los siguientes supuestos: existen múltiples significados de lo que es el conocimiento; el aprendizaje es construido socialmente y activamente; tanto el proceso como el producto son relevantes; debe enfocarse en habilidades multimodales; incluye la subjetividad; el control del aprendizaje es compartido por profesor y estudiante, como un proceso colaborativo; y primordialmente, el propósito primario es facilitar el aprendizaje.

CONCLUSIONES

El término evaluación en educación incluye un abanico amplísimo de conceptos, estrategias e instrumentos, cuya comprensión y aplicación tiene enorme potencial para mejorar el proceso educativo. Es necesario que docentes, estudiantes y autoridades se involucren en asimilar los principios básicos de esta temática, para generar un andamiaje de apoyo que facilite el aprendizaje de los estudiantes, lo que implica apropiarse de los instrumentos, conceptos y estrategias de la evaluación del, para y como aprendizaje. Citando a un académico mexicano, el Dr. Tiburcio Moreno, la evaluación tiene muchas caras, y en países como el nuestro ha estado permeada por una visión empirista que descansa en el principio: “*Todos sabemos de evaluación, porque alguna vez hemos sido evaluados*” (Moreno Olivos, 2010). Debemos mejorar nuestros conocimientos y habilidades en evaluación, es una obligación ética y moral de todo el profesorado.

REFERENCIAS

- American Educational Research Association, American Psychological Association, National Council on Measurement in Education, y Joint Committee on Standards for Educational and Psychological Testing. (2014). *Standards for educational and psychological testing*. Washington, D.C.: AERA. <https://www.testingstandards.net/open-access-files.html>
- Anderson, R. S. (1998). Why talk about different ways to grade? The shift from traditional assessment to alternative assessment. *New Directions for Teaching and Learning*, 74, 5–16.

- Ashford-Rowe, K., Herrington, J., & Brown, C. (2014). Establishing the critical elements that determine authentic assessment. *Assessment & Evaluation in Higher Education*, 39(2), 205-222. <https://doi.org/10.1080/02602938.2013.819566>
- Au, W. (2007). High-Stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher*, 36(5), 258-267.
- Bennett, R. E. (2015). The Changing Nature of Educational Assessment. *Review of Research in Education*, 39(1), 370–407. <https://doi.org/10.3102/0091732X14554179>
- Black, P., Wiliam, D. (1998) Assessment and Classroom Learning, *Assessment in Education: Principles, Policy & Practice*, 5:1, 7-74, DOI: 10.1080/0969595980050102
- Boursicot K., Etheridge L., Setna Z., Sturrock A., Ker J., Smee S., et al. (2011). Performance in assessment: consensus statement and recommendations from the Ottawa conference. *Med Teach*, 33(5), 370-383. DOI: 0.3109/0142159X.2011.565831
- Brennan, R. L. (2006). Perspective on the Evolution and Future of Educational Measurement. En R. L. Brennan (Ed.), *Educational Measurement. National Council on Measurement in Education and American Council on Education* (pp. 1-16). Westport, Connecticut: Praeger Publishers.
- Carter, K. (1984). Do teachers understand principles for writing tests? *Journal of Teacher Education*, 35(6), 57-60.
- Freud, S. (1937). Analysis Terminable and Interminable. *Int. J. Psycho-Anal.*, 18:373- 405.
- Green, M., Moeller, J. y Spak, J. (2018). Test-enhanced learning in health professions education: A systematic review: BEME Guide No. 48. *Medical Teacher*, 40(4), 337-350. <https://doi.org/10.1080/0142159X.2018.1430354>
- Harapnuik, D. K. (2021). Assessment OF/FOR/AS Learning. En blog: It's About Learning. Creating Significant Learning Environments. https://www.harapnuik.org/?page_id=8900
- Hodges B. (2013). Assessment in the post-psychometric era: learning to love the subjective and collective. *Medical teacher*, 35(7), 564–568. <https://doi.org/10.3109/0142159X.2013.789134>
- Jiao, H. and Lissitz, R.W. (2020), What Hath the Coronavirus Brought to Assessment? Unprecedented Challenges in Educational Assessment in 2020 and Years to Come. *Educational Measurement: Issues and Practice*, 39: 45-48. <https://doi.org/10.1111/emip.12363>
- Lok, B., McNaught, C., & Young, K. (2016) Criterion-referenced and norm-referenced assessments: compatibility and complementarity, *Assessment & Evaluation in Higher Education*, 41(3), 450-465, DOI: 10.1080/02602938.2015.1022136
- Maki, P. L. (2010). *Assessing for Learning: Building a Sustainable Commitment across the Institution*. Stylus Publishing, LLC.
- Man Sze Lau, A. (2016). “Formative good, summative bad?” – A review of the dichotomy in assessment literature. *Journal of Further and Higher Education*, 40(4), 509-525. DOI: 10.1080/0309877X.2014.984600
- Martínez Rizo, F. (2009). Evaluación formativa en aula y evaluación a gran escala: hacia un sistema más equilibrado. *Revista Electrónica de Investigación Educativa*, 11(2). Recuperado de <http://redie.uabc.mx/redie/article/view/231>
- Martone, A., Sireci, S. G. (2009). Evaluating Alignment Between Curriculum, Assessment, and Instruction. *Review of Educational Research*, 79(4), 1332–1361.

- McKean M., Aitken E.N. (2016) Educational Renovations: Nailing Down Terminology in Assessment. In: Scott S., Scott D., Webber C. (eds) *Leadership of Assessment, Inclusion, and Learning. The Enabling Power of Assessment*, 3. Springer, Cham. https://doi.org/10.1007/978-3-319-23347-5_2
- Miller, M. D., Linn, R. L. y Gronlund, N. E. (2012). *Measurement and Assessment in Teaching* (11^{va} ed.). Londres: Pearson.
- Montgomery, P.C., & Connolly, B.H. (1987). Norm-referenced and criterion-referenced tests. Use in pediatrics and application to task analysis of motor skill. *Physical Therapy*, 67 12, 1873-6.
- Moreno-Olivos, T. (2010). Lo bueno, lo malo y lo feo: las muchas caras de la evaluación. *Revista Iberoamericana de Educación Superior*, 1(2), 84-97. <https://www.ries.universia.unam.mx/index.php/ries/article/view/6>
- Mueller, J. (2016). Authentic Assessment Toolbox. North Central College, Naperville, IL. <http://jfmuller.faculty.noctrl.edu/toolbox/index.htm>
- National Forum for the Enhancement of Teaching and Learning in Higher Education (2017a). Assessment OF/FOR/AS Learning. <https://www.teachingandlearning.ie/our-priorities/student-success/assessment-of-for-as-learning/>
- National Forum for the Enhancement of Teaching and Learning in Higher Education (2017b). Authentic Assessment in Irish Higher Education. <https://hub.teachingandlearning.ie/wp-content/uploads/2021/06/NF-2017-Authentic-Assessment-in-Irish-Higher-Education.pdf>
- Newble, D. I., y Jaeger, K. (1983). The effect of assessments and examinations on the learning of medical students. *Med Educ.*, 17(3), 165-71.
- Norcini, J., Anderson, B., Bollela, V., Burch, V., Costa, M. J., Duvivier, R., (2011). Criteria for good assessment: consensus statement and recommendations from the Ottawa 2010 Conference. *Med. Teach.*, 33(3), 206-214.
- Roediger, H. L. III, Putnam, A. L., & Smith, M. A. (2011). Ten benefits of testing and their applications to educational practice. In J. P. Mestre & B. H. Ross (Eds.), *The psychology of learning and motivation: Cognition in education* (pp. 1–36). Elsevier Academic Press. <https://doi.org/10.1016/B978-0-12-387691-1.00001-6>
- Rowntree, D. (1977). *Assessing students: How shall we know them?* Londres: Kogan Page Ltd.
- Sánchez-Mendiola, M., Delgado-Maldonado, L. (2017). Exámenes de alto impacto: Implicaciones educativas. *Inv Ed Med.*, 6(21), 52-62. DOI: 10.1016/j.riem.2016.12.001
- Stobart, G., & Gipps, C. (2010). Alternative assessment. In B. McGraw, E. Baker, & P. Peterson (Eds.), *International Encyclopedia of Education* (3rd ed., pp. 202–208). Oxford, UK: Elsevier.
- UNESCO (2020). COVID-19. A glance of national coping strategies on high-stakes examinations and assessments. Paris: UNESCO. https://en.unesco.org/sites/default/files/unesco_review_of_high-stakes_exams_and_assessments_during_covid-19_en.pdf
- Villarroel, V., Bloxham, S., Bruna, D., Bruna, C., & Herrera-Seda, C. (2018) Authentic assessment: creating a blueprint for course design, *Assessment & Evaluation in Higher Education*, 43(5),840-854, DOI:10.1080/02602938.2017.1412396
- Wiliam, D. (2011). What is assessment for learning?, *Studies in Educational Evaluation*, 37(1), 3-14, <https://doi.org/10.1016/j.stueduc.2011.03.001>.