

La evaluación de los aprendizajes en entornos virtuales.

Ignacio Arrillaga, Julio C. Siciliano

Departamento de Educación Médica
Facultad de Medicina, UDELAR
Junio 2020



La evaluación de los aprendizajes en entornos virtuales

Prof. Adjunto Lic. Ignacio Arrillaga
Prof. Agregado Dr. Julio C. Siciliano
Departamento de Educación Médica, Facultad de Medicina

Contenidos

Introducción

La evaluación en línea no debe reflejar las lógicas propias del formato presencial

Evaluación, medición, calificación

Principales tipos de evaluación

Determinantes de la calidad de la evaluación

Validez

Confiabilidad

Practicidad

Utilidad

Evaluación adaptada a los cursos en línea

Principales riesgos

Algunas herramientas para la evaluación en línea

Comparación de las pruebas en línea y presenciales

Referencias

Figuras y Tablas

Figura 1. Principales tipos de evaluación

Figura 2. Determinantes de la calidad de la evaluación.

Tabla I. Principales características de los distintos tipos de evaluación.

Tabla II. Lineamientos generales y recomendaciones para la evaluación en línea

Tabla III. Herramientas de evaluación en plataforma Moodle

Tabla IV. Principales diferencias de las pruebas objetivas desarrolladas en modalidad presencial y en línea

Introducción

El propósito de esta breve guía es apoyar a los equipos docentes en sus decisiones pedagógicas y en la adopción de las acciones adecuadas para evaluar los aprendizajes en las circunstancias actuales.

- La evaluación es un **recurso indispensable** para el perfeccionamiento de los procesos de enseñanza y de aprendizaje, y debe ser desarrollada sobre la base de un programa constituido por un conjunto de instrumentos de evaluación. El diseño de estos programas y la elección de los instrumentos a utilizar deben ser analizados y ponderados por los equipos docentes en tanto son parte de la programación didáctica de sus cursos.
- Las prácticas de evaluación de los aprendizajes por lo general tienen un efecto de tal magnitud que suelen operar en reversa, como determinantes de todas las prácticas pedagógicas, sobre todo de los aprendizajes de los estudiantes, a través de una propiedad denominada **validez retroactiva**, que se describirá más adelante (Camilloni, 1998; Muldoon, 2012; York et al, 2015; McLean, 2018).
- El concepto de evaluación evolucionó desde un proceso consistente en la comprobación del logro de ciertos objetivos o conocimientos (Tyler, 1950) hacia una concepción más global centrada en mejorar el proceso de enseñanza vinculada al asesoramiento, regulación, reorientación y ordenamiento de los aprendizajes (Castillo Arredondo y Cabrerizo Diago, 2009). En este enfoque vertebrado en el aprendizaje, adquiere especial protagonismo la retroalimentación (Ibarra Sáiz y Rodríguez-Gómez, 2010; Luo, Murray y Cropton, 2017; Nicol y Macfarlane-Dick, 2006; Olmos-Migueláñez, 2008). El concepto de **evaluación orientada al aprendizaje** incide en los tipos de tareas y en los actores participantes (Boud y Falchikov, 2006; Carless, 2007) por lo que la evaluación constituye un proceso de una enorme complejidad distribuido de forma continua a lo largo del aprendizaje. Por lo tanto definir el enfoque de evaluación en un curso y el programa de instrumentos con que se llevará a cabo es un aspecto que estructura todo el proceso de enseñanza y las formas de aprender de nuestros estudiantes, y constituye un aspecto central en las decisiones pedagógicas que deben adoptarse en la planificación de un curso.

[Volver al inicio](#)

La evaluación en línea no debe reflejar las lógicas del formato presencial

La suspensión de la actividad presencial en la Universidad (marzo 2020) determinó en forma casi inmediata la migración a la virtualidad de todas las actividades que pudieron adaptarse al formato en línea. Si bien se realizó un gran esfuerzo en la digitalización de materiales, producción de videos y se llevaron adelante clases similares a las presenciales utilizando salas de videoconferencia, lo recomendable es que la enseñanza en línea se desarrolle con propuestas específicas (Hodges et al. 2020) mediante un diseño previo cuidadoso y detallado. Es probable entonces que en el futuro, debamos rediseñar las experiencias de aprendizaje a distancia (Llorens-Largo, 2020a) poniendo en el centro de la formación la interacción (García-Peñalvo y Seoane-Pardo, 2015; Seoane-Pardo y García-Peñalvo, 2014), tanto entre profesor y estudiantes, como entre los propios estudiantes.

El otro gran desafío impuesto por la ausencia de presencialidad es llevar adelante la evaluación de los aprendizajes. Si bien la preocupación de los docentes parece excesivamente centrada en la necesidad de asegurar la identidad de la persona que se somete a la prueba y controlar el contexto físico en el que la persona está realizando dicha prueba (Carstairs y Myors, 2009; Friedman et al. 2016), sería un error considerar que los desafíos de la evaluación en línea están condicionados por este aspecto.

En principio porque los sistemas de vigilancia remota o e-proctoring (Adkins, Kenkel y Lim, 2005) no son infalibles y no pueden asegurar que no se cometa algún tipo de falta ética, dado que solamente logran reducir o minimizar la posibilidad de llevar adelante este tipo de conductas. Pero fundamentalmente porque, en forma análoga a lo que ocurre con las clases, **la evaluación en línea no puede desarrollarse mediante la reproducción de las lógicas propias de los programas de evaluación desarrollados en forma presencial.** La evaluación es un proceso mucho más rico y también, como las clases, debe adecuarse a los entornos educativos en línea.

[Volver al inicio](#)

Evaluación, medición, calificación.

Conviene precisar los términos evaluación, medición y calificación.

Medir supone las acciones orientadas a la obtención y registro de información cuantitativa sobre cualquier hecho o comportamiento (Cabrera y Espín, 1986). Es el proceso de asignar una cantidad al atributo que se considera, al compararlo con un patrón. La medición no adjudica juicios de valor, no es una evaluación. No todas las formas de evaluación requieren de una medición previa (evaluación de una obra artística o de la producción de un artefacto, por ejemplo), por lo que puede haber evaluación sin que exista medición, pero la medición sin evaluación carece de significado.

La calificación se refiere a la expresión que se hace sobre la valoración del rendimiento de los estudiantes, es decir el juicio de valor sobre el atributo en consideración.

La evaluación es la expresión de una valoración que pretende expresar el grado de suficiencia o de insuficiencia de los conocimientos, destrezas o habilidades de un estudiante como resultado de la aplicación de algún tipo de prueba, actividad, examen o proceso (Castillo et al. 2009).

[Volver al inicio](#)

Principales tipos de evaluación

Los principales tipos de evaluación pueden clasificarse de acuerdo a las dimensiones de cronología o momento, propósito, objeto evaluado, procedencia de los agentes evaluadores, evaluadores involucrados y normotipo (Casanova, 1995; Castillo Arredondo y Cabrerizo Diago, 2009; Stalljohann, 2012) (Fig.1).

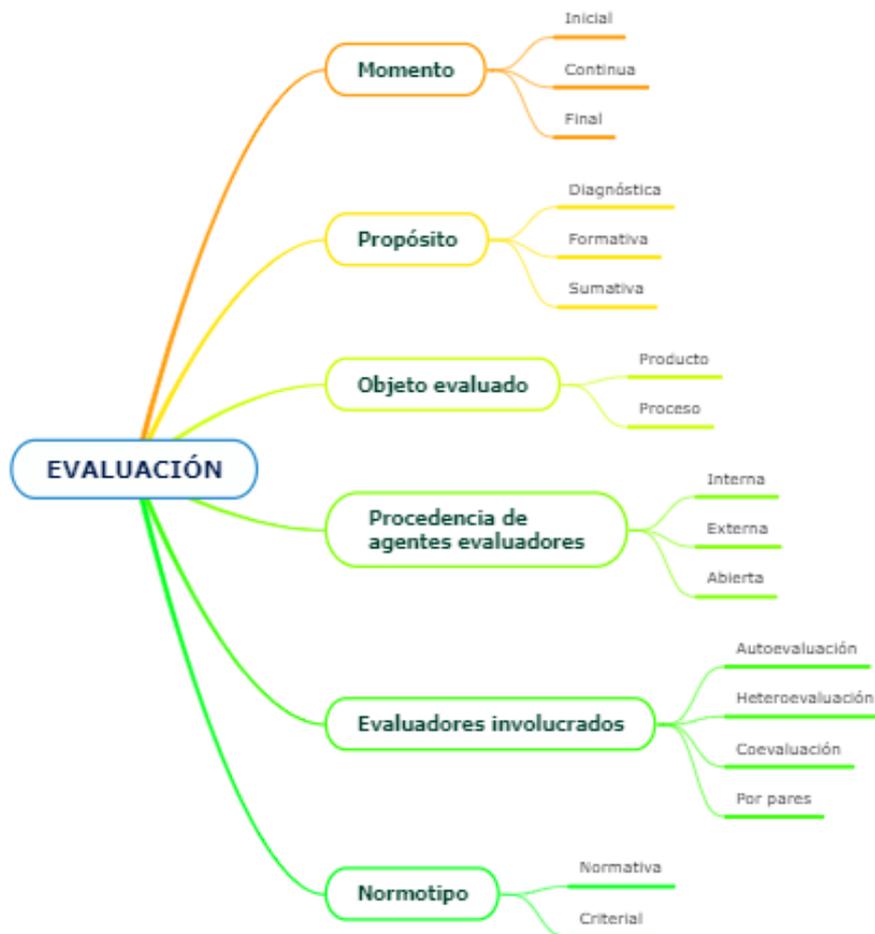


Figura 1. Principales tipos de evaluación.

Las principales características de cada tipo de evaluación se resumen en la Tabla I.

Tabla I. Principales características de los distintos tipos de evaluación.

Principales tipos de evaluación		
Dimensión	Tipo	Características
Momento	Inicial	Tiene lugar antes de comenzar el proceso de aprendizaje (Rosales, 1990).
	Continua	Se organiza a lo largo de un curso (McAlpine, 2002).
	Final	Obtención y valoración de datos al finalizar un periodo de tiempo previsto para la realización de un aprendizaje (Castillo Arredondo y Cabrerizo Diago, 2009). Además de las diferencias relativas al enfoque didáctico de evaluación final y continua, otra diferencia destacable es la carga de trabajo para el estudiante y el docente, que suele ser mucho mayor para la evaluación continua.
Propósito	Diagnóstica	Trata de predecir el futuro desempeño de los estudiantes con el objetivo de determinar el grado de preparación y pronosticar dificultades y aciertos previsibles (Rosales, 1990).
	Formativa	Diseñada para ayudar al proceso de aprendizaje proporcionando retroalimentación (McAlpine, 2002).
	Sumativa	Evaluación del aprendizaje en un período de aprendizaje con un juicio final sobre el rendimiento general del estudiantado (Earl, 2013)
Objeto evaluado	Producto	Evalúa la transmisión del conocimiento a los estudiantes
	Proceso	Intenta evaluar cómo se llevado a cabo el aprendizaje
Procedencia de agentes	Interna	Se promueve y se lleva a cabo “desde dentro” y por los propios integrantes del equipo docente.
	Externa	Evaluado y evaluador pertenecen a instancias diferentes.
	Abierta	Es la práctica de obtener y presentar credenciales demostrando lo que se ha aprendido. En este tipo la evaluación del desempeño individual se independiza del contenido del curso y de la enseñanza (Downes, 2012).
Evaluadores involucrados	Autoevaluación	Los estudiantes evalúan su desempeño.
	Heteroevaluación	Evaluadores y evaluados no son las mismas personas
	Coevaluación	Determinadas personas o grupos se evalúan mutuamente, es decir, evaluadores y evaluados intercambian su papel alternativamente (Castillo Arredondo y Cabrerizo Diago, 2009). Puede aplicarse también a situaciones en las que se permite que los estudiantes se evalúen a sí mismos, mientras que permite al equipo docente mantener el control necesario sobre las evaluaciones finales (Hall,

		1995).
	Por pares	Proceso a través del cual los estudiantes califican a sus pares (Falchikov, 1995).
Normotipo	Normativa	El referente de comparación es el nivel general de un grupo normativo determinado, es decir, establece la comparación entre el rendimiento de cada estudiante con el rendimiento medio de su grupo.
	Criterial	Se realiza con referencia a un criterio previo (criterio de evaluación), es decir, mediante la determinación precisa y concreta de los rendimientos que se pretenden alcanzar (Castillo Arredondo y Cabrerizo Diago, 2009; Popham, 1980).

[Volver al inicio](#)

Determinantes de la calidad de la evaluación

Existe una variedad de instrumentos de evaluación, y en general se acepta que cada uno de ellos utilizado aisladamente resulta insuficiente para obtener información completa sobre el aprendizaje de los alumnos. Por lo tanto en el diseño del programa de evaluación de un curso es importante seleccionar y combinar instrumentos diversos, así como decidir las oportunidades en que se administrará cada uno. En la selección de los instrumentos y construcción de las pruebas deben considerarse distintos criterios determinantes de su calidad, que son la validez, confiabilidad, practicidad y utilidad (Fig. 2). Existe una extensa bibliografía que cubre este tema en forma completa, por lo que solamente se presentará una visión sumaria. Para una revisión del significado e importancia de cada una de estas dimensiones y sus referencias bibliográficas recomendamos ver Camilloni (1998).

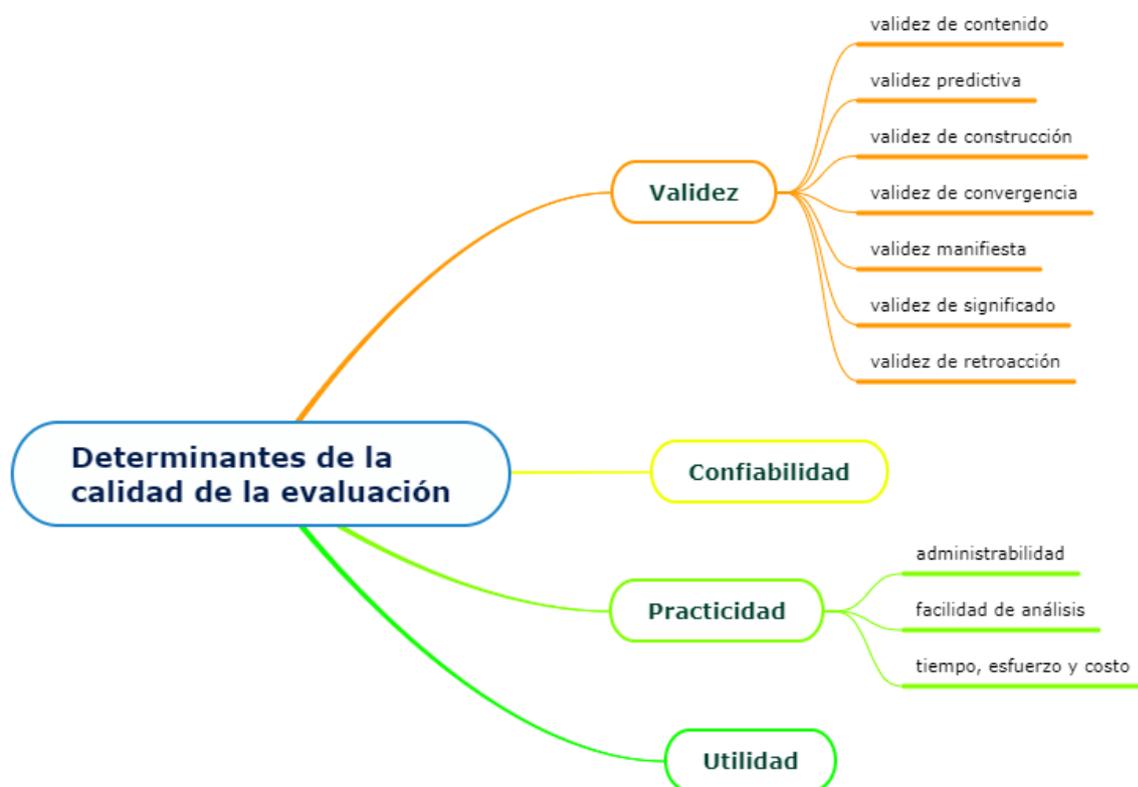


Figura 2. Determinantes de la calidad de la evaluación.

VALIDEZ

Se considera que un instrumento de evaluación es válido cuando evalúa lo que se pretende evaluar con él. La validez de un instrumento no puede ser determinada de manera absoluta, sino siempre en relación a sus propósitos específicos. Depende además de la construcción apropiada de los instrumentos. La validez nunca es absoluta dado que siempre debe ser referida a un criterio externo, solo permite hacer inferencias acerca de los saberes de los alumnos, y además porque está condicionada al ajuste posible entre el aprendizaje que se desea evaluar y los medios técnicos o instrumentos disponibles para la evaluación. Con

frecuencia se obtiene una mayor validez mediante la combinación apropiada de distintos tipos de instrumentos. Se diferencian diversas clases de validez.

Validez de contenido. Un instrumento tiene validez de contenido cuando representa una muestra significativa de los contenidos que se desean evaluar (y es equivalente a cualquier otro conjunto de contenidos que se hubieran podido seleccionar). En general cuanto mayor es la diferencia entre tiempos destinados a la enseñanza y al estudio y tiempos destinados a la evaluación, más difícil es lograr una validez de contenido aceptable. La validez de contenido es también llamada "validez curricular" porque el criterio externo que se emplea para garantizar esta clase de validez es el diseño curricular.

Validez predictiva. Corresponde a la correlación entre los resultados obtenidos en una prueba (o varias pruebas relacionadas) y el desempeño posterior del alumno en los aspectos evaluados por estas pruebas.

Validez de construcción. El programa de evaluación y cada uno de los instrumentos que lo integran deben estar contruidos de acuerdo con los principios que sostienen el proyecto pedagógico. Es decir, debe haber coherencia entre las teorías y las estrategias de enseñanza que se han utilizado y las de las instancias de evaluación que conforman el programa de evaluación.

Validez de convergencia. Es la relación que existe entre un programa de evaluación o un instrumento y otros programas o instrumentos de validez ya conocida. Por ejemplo cuando un programa o una prueba de evaluación son reemplazados por otros, debe demostrarse que son, por lo menos, tan buenos o mejores que los empleados anteriormente.

Validez manifiesta. No se refiere a la validez intrínseca del programa de evaluación o a cada prueba, sino al modo en que los instrumentos se presentan y son percibidas por los alumnos.

Validez de significado. También es un criterio que se establece desde la perspectiva de los estudiantes. Para que la prueba tenga significado y constituya un desafío que aumente su motivación y mejore su aprendizaje, las pruebas deben ubicarse en un espacio (a menudo de difícil resolución) entre los problemas demasiado difíciles para ser resueltos y los problemas demasiado fáciles en los que no se puede fallar.

Validez de retroacción. La evaluación tiene un efecto de retroacción sobre la enseñanza. Ello determina que con frecuencia, los aprendizajes encarados por los alumnos se realizan conforme a las pautas establecidas por los docentes en las evaluaciones que administran, lo cual hace que la evaluación se convierte en la verdadera reguladora de los procesos de aprendizaje. Esto no es recomendable pues el dominio de los aprendizajes en la educación es mucho más amplio que el que se puede cubrir mediante la evaluación (además como vimos la evaluación suele cubrir sólo una pequeña parte de los aprendizajes).

CONFIABILIDAD

La confiabilidad de un instrumento depende de la exactitud en la medición que proporciona y la sensibilidad en la apreciación de los rasgos que mide. Los resultados obtenidos se consideran confiables cuando permanecen semejantes y estables en todas las ocasiones que se administra ese instrumento o uno similar, y cuando pueden considerarse independientes de otros factores que hayan podido incidir (azar, fatiga, nervios).

La confiabilidad de un instrumento depende, pues, de su estabilidad, su exactitud y su sensibilidad. A estas condiciones se agrega la objetividad, que supone que la evaluación es independiente de la persona que evalúa. La confiabilidad puede ser afectada por la longitud y duración de la prueba: pruebas breves pueden arrojar resultados poco confiables, pruebas largas en exceso pueden ser afectadas por la fatiga.

También pueden incidir en la confiabilidad las condiciones físicas y materiales de administración del instrumento, el tiempo asignado, entre otros.

La confiabilidad debe permitir cubrir las necesidades planteadas por el programa de evaluación (no todos los instrumentos tienen la misma exactitud y sensibilidad y no siempre es necesario el aumento de la exactitud y sensibilidad). Por otra parte, si bien numerosos autores (Ebel, 1963; Sachs Adams, 1970; Gronlund, 1973; Pidgeon, 1979; Popham, 1983; Karmel, 1986) consideran a la confiabilidad como un requisito sin el cual no puede existir validez, ello ha sido objeto de revisión dado que sería como afirmar que los instrumentos “subjetivos” o tradicionales (exámenes o cuestionarios orales, pruebas escritas de ensayo o disertación o de respuesta extensa; pruebas de respuesta restringida o de cuestionario; monografías, tesinas y tesis; evaluación de portafolio; estudios de casos, entre otras) carecen de validez.

PRACTICIDAD

La practicidad no es, seguramente, la característica más importante de un programa o de un instrumento de evaluación, aunque es relevante ya que en general condiciona **la frecuencia** de la evaluación. Por lo general, cuando el programa es práctico o cuando el instrumento es de fácil administración e interpretación, se evalúa más frecuentemente y ello contribuye a estructurar mejor los procesos de enseñanza y los procesos de aprendizaje de los alumnos por parte del docente y de los mismos alumnos.

La practicidad de un programa o un instrumento depende de tres variables: administrabilidad, facilidad de análisis, y de la evaluación del tiempo, esfuerzo y costo necesarios para su puesta en práctica.

La administrabilidad de un programa o un instrumento es un producto de varias características:

- el tiempo de trabajo de los docentes que insume su diseño y construcción;
- el tiempo que exige su puesta en práctica;
- la claridad de la/las consigna/s, es decir, de las directivas que se dan a los estudiantes;
- los materiales, equipos y lugares especiales necesarios para su administración;

- la cantidad y preparación de las personas indispensables para la administración, cómputo, análisis e interpretación de los resultados y elaboración de las evaluaciones.

En general hay una correlación inversa entre el tiempo que demanda la construcción de un instrumento y el que exige la interpretación de sus resultados.

Es importante, como hemos dicho antes, que el programa de evaluación sea económico en tiempo, trabajo y costo. Es conveniente examinar previamente estas características para determinar la conveniencia de su empleo.

Los tres aspectos relativos a la practicidad del programa o de cada instrumento deben considerarse conjuntamente para lograr una evaluación integral de este rasgo. Cada aspecto presenta ventajas y desventajas que se compensan o equilibran cuando se analizan los otros dos. Un costo relativamente elevado, por ejemplo, podría ser compensado por un alto grado de administrabilidad, o un tiempo prolongado dedicado a la construcción de un instrumento puede ser equilibrado por un tiempo menor destinado a la corrección e interpretación de los resultados. Del mismo modo, un tiempo extenso de construcción puede compensarse por la facilidad de administración simultánea a un grupo muy numeroso de alumnos.

UTILIDAD

La utilidad de una prueba resulta de su capacidad para satisfacer las necesidades específicas relacionadas con los procesos de enseñanza y aprendizaje.

[Volver al inicio](#)

Evaluación adaptada a los cursos en línea

Al instrumentar cursos a distancia debemos evitar desarrollar las mismas actividades de enseñanza que se desarrollan en los escenarios presenciales. Un curso desarrollado en modalidad virtual debe adaptar sus prácticas y contenidos a escenarios educativos diferentes a los presenciales mediante una cuidadosa planificación que contemple las características de esta modalidad educativa (Branch & Dousay 2015).

De manera análoga, **la evaluación desarrollada en línea debe adecuarse a las especificidades de esta modalidad** sin intentar reproducir las lógicas educativas del formato presencial. Los programas e instrumentos de evaluación de los aprendizajes deben adecuarse a los procesos de enseñanza y de aprendizaje que los han generado, de modo de expresar un alineamiento curricular consistente. Por eso es recomendable diseñar los programas de evaluación y seleccionar los instrumentos que mejor se adapten a la modalidad, sin esforzarnos por trasladar en forma más o menos automática las prácticas desarrolladas en la presencialidad al formato virtual.

A continuación se mencionan algunos **lineamientos generales y recomendaciones** para la evaluación de los aprendizajes en esta modalidad (Tabla II). Conviene destacar que todas o la mayoría de estas recomendaciones deberían aplicarse también en contextos presenciales pues su implementación tendría un impacto significativo en la mejora de la evaluación de los aprendizajes desarrolladas en nuestros cursos.

Tabla II. Lineamientos generales y recomendaciones para la evaluación en línea

Evaluación continua	Permite disminuir el peso relativo de cada una de las pruebas en la decisión de promoción
Mayor número de instancias	Aumentar el número de pruebas que representen menos del 40 % del puntaje final total
Tipos de preguntas.	Banco amplio de preguntas. Preguntas dirigidas a resolver problemas, ejercicios, análisis de situaciones o casos (preguntas de “libro abierto”)
Usar tecnologías ya disponibles	Usar herramientas que brinden seguridad a docentes y estudiantes.
Diversificación	Usar distintos tipos de instrumentos además de pruebas objetivas para aumentar la cobertura (entrega de tareas, ejercicios, presentaciones, videos, etc)
Flexibilidad	Planes que aporten flexibilidad y respuesta a contingencias
Comunicación	Comunicación activa con los estudiantes

Aumentar el número de pruebas. La aplicación de un sistema de evaluación sumativa a lo largo del curso puede permitir disminuir el peso de cada una de las pruebas en la decisión de la promoción, aminorando el efecto de un reducido número de pruebas o una prueba final con los problemas de garantías de autenticidad, gestión de reclamos y posibilidades de

recuperación. Conviene realizar un mayor número de pruebas, consistente con la adquisición de conocimientos y con el desarrollo del curso, constituyendo un sistema coordinado. Cada una de estas pruebas debería representar menos de un 40 % del puntaje total final del curso, de modo que ninguna de las pruebas individuales pueda decidir la promoción. El aumento del número de estas pruebas, lo cual mejora la evaluación en el curso, en realidad sólo es posible si se utilizan pruebas en línea lo cual permite disminuir los costos y tiempos dedicados a la evaluación.

Número y Tipos de preguntas o ítems. Si se evalúan contenidos teóricos mediante pruebas objetivas de múltiple opción, las pruebas deben contar con un banco de preguntas o ítems muy superior al de la prueba que la plataforma seleccione al azar y programarse en un tiempo máximo de respuesta corto, aunque suficiente para su realización. Con estas condiciones se minimizan los riesgos de que se puedan usar libros/páginas web para buscar las respuestas y de que el colectivo de estudiantes comparta entre ellos la colección de preguntas de la prueba.

Además es altamente recomendable que las preguntas supongan la aplicación de conocimientos y conceptos (ejercicios, problemas) de modo que la búsqueda de información no suponga un problema. Puede adoptarse incluso el formato de preguntas “a libro abierto”, de modo de eliminar este inconveniente.

Usar tecnologías ya disponibles, conocidas, y lo más similares posibles a las empleadas habitualmente para el soporte de la enseñanza. Es importante que tanto los docentes como los estudiantes se sientan lo más seguros posible en el nuevo contexto de evaluación en línea y estén familiarizados con la herramienta que se utilice.

Diversificación. Es altamente recomendable diversificar los métodos de evaluación y utilizar distintos tipos de instrumentos, de modo de aumentar la cobertura y aumentar la validez del programa.

Por ejemplo se puede recurrir a solicitar a los estudiantes la **entrega de trabajos, ejercicios,** en diversos formatos incluyendo documentos y videos, individuales o grupales. En el caso de los documentos pueden utilizarse herramientas de detección de plagio si se considerara necesario. En ese caso se debe informar previamente a los estudiantes del uso de estos programas de control de plagio, así como del impacto en la calificación cuando se detecten prácticas contrarias a la ética académica.

Se puede solicitar que de forma individual o grupal se hagan videos de **presentaciones o defensas de los trabajos o resolución de problemas usando las herramientas de videoconferencia** con una duración limitada.

Para la calificación de estas entregas es recomendable que el equipo docente defina rúbricas de evaluación o escalas de valoración y que estas sean conocidas por los estudiantes previamente (Cabero-Almenara y Rodríguez-Gallego, 2013).

Eventualmente en grupos grandes se puede solicitar la entrega de tareas/trabajos y habilitar **sistemas de evaluación por pares.** Son algo más complejas de configurar en la plataforma y en este caso, el profesor debe hacer llegar a todos los estudiantes una rúbrica de evaluación o

una escala de valoración lo más precisa y objetiva posible. Su uso puede ser una forma muy interesante de promover la autoevaluación y el aprendizaje.

En el caso de requerir la realización de **pruebas orales**, por ejemplo para la recuperación de instancias de evaluación, se pueden realizar usando las herramientas de videoconferencia lo cual permite además grabar las instancias en video.

Otra herramienta de evaluación de habilidades útil, además de las entregas, es la participación en los **debates definido en los foros** de la asignatura. Se pueden abrir temas específicos de debate y pedir opiniones documentadas sobre cuestiones relevantes de la materia; se les puede solicitar que hagan aportaciones novedosas (lo que les obliga a leer las opiniones previas) y que aporten las fuentes en las que se han basado para realizar su contribución al foro. Deben ser animados y supervisados por docentes, lo cual lleva una considerable cantidad de tiempo en los grupos numerosos.

Se debe conocer y tener en cuenta las características de los estudiantes matriculados.

Deberían asumirse posiciones y métodos de evaluación que aporten la mayor flexibilidad. Por lo tanto se deben tener definidos de antemano planes de contingencia adaptados a cubrir las necesidades de los estudiantes con situaciones problemáticas o dificultades de acceso a la tecnología que requieran medidas alternativas. Estos planes, que afectarían especialmente a las pruebas y actividades sincrónicas, deberían definirse por las coordinaciones y responder a las distintas necesidades con la mayor flexibilidad posible dentro de las instrucciones de las autoridades sanitarias.

Comunicar

Una vez diseñado el método de evaluación en línea de una asignatura los docentes deben comunicarlo a sus estudiantes con la suficiente antelación, con especial atención al tipo de preguntas que se utilizarán, forma de responderlas, tiempo para realizar la prueba, herramientas a utilizar, elementos auxiliares para realizar la prueba, material de consulta permitido, forma de resolver problemas durante la prueba, normativa relevante a tener en cuenta, etc.

También es importante explicar en forma clara las indicaciones generales para la realización de las pruebas, destacando aspectos como la necesidad de tener los dispositivos bien configurados y probados para poder realizar las pruebas, si la comunicación con otros está o no permitida, el potencial uso de *software* antiplagio y control de direcciones IP desde donde se conectan los estudiantes, recomendaciones sobre el envío en el tiempo previsto, que ocurre si se sale de la prueba, recomendaciones sobre los navegadores, documentación técnica de soporte disponible, etc.

Como comentario general sobre este punto, es recomendable crear un foro para plantear y resolver las dudas que puedan surgir antes de la realización de las pruebas y dejar en la plataforma todos los puntos convenientemente aclarados.

[Volver al inicio](#)

Principales riesgos

Complementando esta serie de recomendaciones generales se plantea un conjunto de riesgos que deben contemplarse en la realización de pruebas sincrónicas en línea.

Dificultades en la realización de la prueba.

Problemas experimentados con los dispositivos electrónicos, dificultades en el acceso o conectividad, problemas derivados del ambiente en que se realiza la prueba (ruidos, distracciones, ambiente inadecuado, etc) son factores que pueden afectar el desempeño de los estudiantes. Deben tenerse previstas instancias de recuperación para remediar este tipo de situaciones.

Conductas indebidas.

Se refiere a la copia, plagio, o pasaje de información en el colectivo de estudiantes hasta el falseamiento de la identidad de la persona que realiza la prueba.

Es importante considerar que la copia y el pasaje de información no son actividades exclusivas de la enseñanza virtual. La identificación de la identidad del estudiante no debería ser la característica determinante para diseñar un programa de evaluación. El uso de las estrategias de aumento del número de instancias, evaluación continua y diversificación del tipo de instrumentos que ya se mencionaron permiten mitigar estos problemas. Es importante poner atención a los saltos cualitativos en la calidad de los trabajos. Si el contexto lo permite, centrar la estrategia en la evaluación continua incorporando elementos de reflexión y promover las instancias de interacción y participación. No obstante se sugiere diseñar las pruebas objetivas en línea de manera de disuadir estas prácticas o volverlas mucho más difíciles e improbables:

- explicar las normas en relación a la copia y otras conductas indebidas, y que se realizará un control de las direcciones IP;
- recordar a los estudiantes que las claves de acceso a la plataforma son personales e intransferibles y que un uso inadecuado de las mismas puede acarrear consecuencias serias;
- revisar, modificar y/o actualizar las preguntas y actividades de evaluación de una edición de un curso a otra;
- es preferible proponer varias pruebas cortas que una larga;
- configurar cuestionarios con preguntas seleccionadas al azar a partir de un banco de preguntas extenso;
- utilizar tiempos acotados para cada pregunta y dividir la prueba en páginas separadas impidiendo el volver hacia atrás;
- realizar la prueba en forma sincrónica para toda la cohorte.

La suplantación de la identidad podría enfrentarse al menos en teoría con los sistemas de vigilancia electrónica remota (e-proctoring), aunque estos sistemas tampoco aseguran la eliminación del problema y son hasta ahora de muy difícil e improbable utilización. Además es interesante que el uso de sistemas de vigilancia remota se concibe como un intento de igualar la incidencia de las deshonestidad académica entre las pruebas de evaluación en línea y

presenciales (Harmon y Lambrinos, 2008), entendiendo que al igual que para la supervisión de exámenes presenciales no hay un método de supervisión virtual perfecto. Si se considera estrictamente necesaria la identificación se sugiere utilizar otros medios o instancias complementarias (videoconferencias, videos explicativos, etc).

[Volver al inicio](#)

Algunas herramientas de evaluación online

A continuación se presentan algunas de las herramientas más habituales que se pueden utilizar para desarrollar los sistemas de evaluación en la plataforma Moodle (Tabla III).

Tabla III. Herramientas de evaluación en plataforma Moodle (adaptado y modificado de Abella García et al. 2020).

Herramienta de evaluación	Utilidad	Ventajas	Inconvenientes	Consejos
Tareas	Permite evaluar un trabajo/problema	Fácil configuración	La corrección puede demandar mucho tiempo en grupos numerosos	Conviene que se conozcan los criterios de corrección (rúbricas por ejemplo) Si se realizan trabajos grupales conviene realizar alguna defensa mediante videoconferencia registrada en video
	permite subir archivos de video y de audio	el video permite seguir el proceso de resolución	requiere que el estudiante tenga un ancho de banda adecuado	se puede combinar con la entrega de una tarea escrita conviene que los videos no sean subidos a la plataforma sino a la nube (por ejemplo en canal YT) y dejar en la plataforma el enlace
Cuestionarios de moodle	Escenarios de evaluación sumativa (conceptos, hechos, etc)	permite evaluar en forma rápida un gran número de estudiantes	Debe prestarse especial atención a la configuración y buscar asesoramiento si es necesario. Aunque el sistema en uso actualmente ha permitido la realización sincrónica de pruebas en cohortes muy numerosas (> 2000) su realización podría	Conviene tener un amplio banco de preguntas Es recomendable que las preguntas se seleccionen y aparezcan en forma aleatoria y que se cambie al azar el orden de las opciones en cada pregunta. Se recomienda limitar el tiempo de respuesta, pero teniendo en cuenta el tiempo necesario para

			afectar el sistema por lo que debe coordinarse su uso con la UDI	resolver con solvencia el examen. Con preguntas que requieren poco tiempo de respuesta y cuestionarios de pocas preguntas por página el riesgo de sobrecargar el sistema crece.
	Realizar autoevaluaciones de conceptos, hechos, procedimientos, etc. (escenario de evaluación formativa)	Permite evaluar un gran número de estudiantes	Debe cuidarse la configuración	Se pueden utilizar para autoevaluaciones, orientadas a una evaluación formativa Pueden estar abiertos sin límite de tiempo o con una ventana temporal muy amplia, por tanto, la sobrecarga del sistema es menor Se pueden dejar varios intentos de realización si se busca que el estudiante fije conceptos, utilizando las diferentes opciones de retroalimentación que ofrece la plataforma Es útil también trabajando con grupos pequeños
Foro de Moodle	Realizar evaluaciones, principalmente formativas, tanto de conocimientos teóricos como de competencias adquiridas o de resolución de problemas	Configuración muy sencilla Todas las contribuciones quedan registradas en el campus virtual	Pueden ser aportaciones largas; la principal limitación será el número de estudiantes	Elaborar muy bien la cabecera del foro, indicando claramente el tipo de tarea que se espera Plantear un tema que requiera un estudio previo, reflexión y opinión de los estudiantes Las aportaciones deben ser originales y se puede solicitar

				que aporten alguna fuente de información de referencia Seleccionar bien el tipo de foro, para que se ajuste al objetivo buscado, y el formato de suscripción de quien participa en la asignatura Según su uso puede ser recomendable activar la herramienta de análisis del plagio
Taller de Moodle	Evaluación/ autoevaluación entre pares	Distribuye de forma automática los trabajos a evaluar entre los estudiantes	Complicado de configurar	Se deben elaborar las rúbricas con criterios claros de evaluación Proporcionar ejemplos de respuestas buenas/pobres que sirvan de guía a los estudiantes en la evaluación Supervisar las evaluaciones que realicen los/las estudiantes Hacer saber a los/las estudiantes que el equipo docente es el último responsable de la calificación (anular las evaluaciones inadecuadas de los/las estudiantes) Necesario activar la
Lección de Moodle	Crea experiencias de aprendizaje adaptativo y personalizado	Permite incluir preguntas para evaluar o para guiar el itinerario de aprendizaje	Diffcil de configurar Conlleva mucho tiempo hacer un diseño adecuado	Recomendable hacer un esquema en papel con los distintos itinerarios

[Volver al inicio](#)

Comparación de las pruebas en línea y presenciales

Tabla IV. Principales diferencias de las pruebas objetivas desarrolladas en modalidad presencial y en línea (EVA).

Dimensión	Características	Presenciales	En línea
1. Construcción de los ítem	1.1. Número	Igual al número de ítem que tenga la prueba	Requiere la construcción de un banco de preguntas amplio).
	1.2 Momento en que deben redactarse.	Tiempo necesario para impresión, compaginación, distribución de las pruebas.	Tiempo necesario para configurar el cuestionario
2. Características del cuestionario	2.1. Configuración	Pueden armarse distintos prototipos alterando el orden de las preguntas.	Más compleja. Es recomendable que las preguntas aparezcan en orden aleatorio. Es recomendable seleccionar la distribución al azar de las opciones en cada pregunta. Configurar hora de apertura y cierre del cuestionario y calcula tiempos breves (pero suficientes).
3. Administración de la prueba	3.1. Recursos necesarios	Coordinación de utilización de locales. Ejemplares impresos de las pruebas Planillas de cómputo	Estudiantes deben contar con: - dispositivos - conectividad - espacio adecuado para realizar la prueba
	3.2. Tareas requeridas	Impresión y compaginación. Distribución de las pruebas. Locales (ordenamiento e higiene) Supervisión y vigilancia Ordenamiento y conteo de planillas previo a la corrección	Configuración de la prueba Monitorización en tiempo real del desarrollo de la prueba Soporte técnico UDI.
	3.3. Tiempo asignado	Puede asignarse tiempo extra para control del llenado de planillas, necesidades especiales, etc.	Es recomendable limitar el tiempo de respuesta, pero teniendo en cuenta el tiempo necesario para resolver con solvencia el examen y los posibles contratiempos técnicos que puedan

			ocurrir (evitar tiempo "sobrante").
	3.4. Número de instancias previstas	2 o más instancias si el número de estudiantes excede la capacidad locativa. Eventualmente instancias posteriores para estudiantes con impedimentos para realizar la prueba (en pruebas parciales).	Es recomendable que exista una instancia prevista para los estudiantes que experimenten dificultades de conectividad durante la prueba.
	3.5. Límite del número de estudiantes	Depende de la capacidad locativa.	Depende de la capacidad del servidor y el ancho de banda de la institución.
4. Corrección y retroalimentación	4.1. Corrección	Lectura y corrección de planillas asistida por ordenador.	Automática en la plataforma al cerrar el cuestionario.
	4.2. Retroalimentación	Realizada por los docentes, genérica o específica, y/o en atención a reclamos específicos.	Excelente capacidad de retroalimentación automática: Puede configurarse retroalimentación global para la prueba, general para cada pregunta, y específica para cada opción.
5. Determinantes de la calidad	5.1. Valoración de tiempo, esfuerzo y costos de realización	Requiere coordinación de la utilización de locales y adecuación a las necesidades institucionales, movilización de numerosos estudiantes y docentes, utilización plena de locales, cantidad importante de personal.	No requiere la movilización de recursos físicos y personal, lo cual permite una aplicación más frecuente del instrumento (esto es relevante pues se considera que evaluar frecuentemente contribuye a estructurar mejor los procesos de enseñanza y mejora la validez predictiva de este tipo de pruebas).
	5.2. Administrabilidad	Requiere pruebas impresas, planillas de cómputo, movilización de docentes para supervisión y administración, personal para lectura y corrección.	Requiere configurar un banco grande de preguntas, claridad de las directivas y consignas dadas a los estudiantes para la realización de la prueba.
	5.3. Identificación de los estudiantes	A cargo del docente supervisor en el salón	Baja (acceso a la plataforma con ID) Vigilancia remota (no

			se realiza en la actualidad).
	5.4. Prácticas para disminuir conductas inapropiadas	Utilización de prototipos con distinto orden de las preguntas. Vigilancia presencial	Cuestionarios configurados al azar con preguntas extraídas de un amplio banco. Orden aleatorio de la opción correcta y los distractores en cada pregunta. División del cuestionario en páginas independientes con imposibilidad de volver atrás una vez cerrada la página. Administración síncrona para toda la cohorte Tiempos acotados.
	5.5. Validez predictiva	Variable, depende de las características específicas de cada prueba.	Podría incrementarse ampliando el número y variedad de los instrumentos de control del aprendizaje (por lo cual la Practicidad resulta un aspecto relevante como se señala más arriba)
	5.6. Validez manifiesta o aparente.	La tradicional para las pruebas objetivas.	Baja. En nuestro medio las experiencias anteriores se realizaron en contextos muy diferentes y no resultan comparables. En la literatura especializada las experiencias son heterogéneas y no existen conclusiones claras al respecto.
	5.7. Limitaciones	Capacidad y disponibilidad locativa. Número de docentes disponibles para la administración de la prueba. Riesgo de suspensión en atención a activación de protocolos de emergencia.	Riesgo de sobrecarga del sistema informático. Pueden existir estudiantes con dificultades de acceso.

[Volver al inicio](#)

Referencias

- Abella García, V., Grande de Prado, M., García-Peñalvo, F. J. y Corell, A. (2020). *Guía de recomendaciones para la evaluación online en las Universidades Públicas de Castilla y León. Versión 1.1*. Castilla y León, España: Universidad de Burgos, Universidad de León, Universidad de Salamanca y Universidad de Valladolid.
doi:<https://doi.org/10.5281/zenodo.3780661>.
- Adkins, J., Kenkel, C. y Lim, C. L. (2005). Deterrents to online academic dishonesty. *The Journal of Learning in Higher Education*, 1(1), 17-22.
- Boud, D. y Falchikov, N. (2006). Aligning assessment with long-term learning. *Assessment & Evaluation in Higher Education*, 31(4), 399-413.
doi:<https://doi.org/10.1080/02602930600679050>
- Branch R & Dousay T (2015) Survey of instructional design models. Association for Educational Communications and Technology, Indiana USA.
- Cabero-Almenara, J. y Rodríguez-Gallego, M. (2013). La utilización de la rúbrica en el diseño de materiales para la e-formación. *EduTec. Revista Electrónica De Tecnología Educativa*(43).
doi:10.21556/edutec.2013.43.3
- Cabrera, F. A. y Espín, J. V. (1986). *Medición y evaluación educativa. Fundamentos teórico-prácticos*. Barcelona, España: Promociones y Publicaciones Universitarias.
- Camilloni, Alicia R. W (1998) la calidad de los programas de evaluación y de los instrumentos que los integran. En: *La evaluación de los aprendizajes en el debate didáctico contemporáneo* Alicia R. W. de Camilloni, Susana Celman, Edith Litwin y M. del Carmen Palou de Maté. Paidós (Buenos Aires).
- Carless, D. (2007). Learning-oriented assessment: conceptual bases and practical implications. *Innovations in Education and Teaching International*, 44(1), 57-66.
doi:<https://doi.org/10.1080/14703290601081332>
- Carstairs, J. y Myors, B. (2009). Internet testing: A natural experiment reveals test score inflation on a high-stakes, unproctored cognitive test. *Computers in Human Behavior*, 25(3), 738-742. doi:<https://doi.org/10.1016/j.chb.2009.01.011>
- Casanova, M. A. (1995). *Manual de evaluación educativa*. Madrid, España: La Muralla.
- Castillo Arredondo, S. y Cabrerizo Diago, J. (2009). *Evaluación educativa de aprendizajes y competencias*. Madrid, España: Pearson Educación.
- Downes, S. (2012). *Connectivism and Connective Knowledge. Essays on meaning and learning networks* Canada: Stephen Downes. Recuperado de <https://bit.ly/2V3zg5k>
- Earl, L. M. (2013). *Assessment as learning: Using classroom assessment to maximise student learning* (2nd ed.). Thousand Oaks, CA, USA: Corwin Press.
- Ebel, Robert L.: "Must all Tests be Valid?", en *Human Learning in School*, ed. John De Cecco, Nueva York, Holt, Rinehart & Winston, 1963.
- Fernández Enguita, M. (2020). Una pandemia imprevisible ha traído la brecha previsible. Recuperado de <https://bit.ly/2VT3kzU>
- Friedman, A., Blau, I. y Eshet-Alkalai, Y. (2016). Cheating and feeling honest: Committing and punishing analog versus digital academic dishonesty behaviors in higher education. *Interdisciplinary Journal of e-Skills and Life Long Learning*, 12, 193-205.

- doi:<https://doi.org/10.28945/3629>
- García-Peñalvo, F. J. y Seoane-Pardo, A. M. (2015). Una revisión actualizada del concepto de eLearning. Décimo Aniversario. *Education in the Knowledge Society*, 16(1), 119-144. doi:<https://doi.org/10.14201/eks2015161119144>
- Gronlund, Norman E.: *Medición y evaluación en la enseñanza*, México, Pax, 1973.
- Hall, K. (1995). *Co-assessment: participation of students with staff in the assessment process: A report of work in progress*. Trabajo presentado en Second European Electronic Conference on Assessment and Evaluation.
- Harmon, O. R. y Lambrinos, J. (2008). Are Online Exams an Invitation to Cheat? *The Journal of Economic Education*, 39(2), 116-125. doi:<https://doi.org/10.3200/JECE.39.2.116-125>
- Hodges, C., Moore, S., Lockee, B., Trust, T. y Bond, A. (2020). The difference between emergency remote teaching and online learning. *Educause Review*. Recuperado de <https://bit.ly/3b0Nzx7>
- Ibarra Sáiz, M. S. y Rodríguez-Gómez, G. (2010). Los procedimientos de evaluación como elementos de desarrollo de la función orientadora en la universidad. *Revista Española de Orientación y Psicopedagogía*, 21(2), 443-461. doi:<https://doi.org/10.5944/reop.vol.21.num.2.2010.11558>
- Karmel, Louis J.: *Medición y evaluación escolar*, México, Trillas, 1986.
- Llorens-Largo, F. (2020a, 16 de marzo). Recomendaciones para una docencia no presencial y apoyada con tecnología. Recuperado de <https://bit.ly/34kYXS5>
- Luo, T., Murray, A. y Cropton, H. (2017). Designing Authentic Learning Activities to Train Pre-Service Teachers About Teaching Online. *International Review of Research in Open and Distributed Learning*, 18(7), 141-156. doi:<https://doi.org/10.19173/irrodl.v18i7.3037>
- McAlpine, M. (2002). *Principles of assessment*. UK: CAA Centre, University of Luton.
- McLean H (2018): This is the way to teach: insights from academics and students about assessment that supports learning, *Assessment & Evaluation in Higher Education*, DOI: 10.1080/02602938.2018.1446508
- Muldoon R (2012) Is it time to ditch the traditional university exam? *Higher Education Research & Development* Vol. 31, No. 2, 263–265
- Nicol, D. y Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199-218. doi:<https://doi.org/10.1080/03075070600572090>
- Olmos-Migueláñez, S. (2008). *Evaluación formativa y sumativa de estudiantes universitarios: Aplicación de las tecnologías a la evaluación educativa*. Salamanca, España: Ediciones Universidad de Salamanca.
- Pidgeon, Douglas y Yates, Alfred: *Evaluación y medida del rendimiento escolar*, Madrid, Anaya, 1979.
- Popham, James W.: *Evaluación basada en criterios*, Madrid, Magisterio Español, 1983.
- Rosales, C. (1990). *Evaluar es reflexionar sobre la enseñanza* (3ª ed.). Madrid, España: Narcea.
- Sachs Adams, Georgia: *Medición y evaluación*, Barcelona, Herder, 1970.
- Seoane-Pardo, A. M. y García-Peñalvo, F. J. (2014). Pedagogical Patterns and Online Teaching. En F. J. García-Peñalvo y A. M. Seoane Pardo (Eds.), *Online Tutor 2.0: Methodologies and Case Studies for Successful Learning* (pp. 298-316). Hershey, PA: IGI Global. doi:<https://doi.org/10.4018/978-1-4666-5832-5.ch015>
- Stalljohann, P. W. (2012). *A Generic Platform for Open Assessment Management in Higher*

Education. A modular architecture for assessment management within a centrally hosted learning and teaching portal. (PhD), Aachen University, Aachen, Germany.

Tyler, R. (1950). *Basic principle of curriculum and instruction*. Chicago, USA: Chicago University.

York, T, et al. (2015). Defining and Measuring Academic Success. *Practical Assessment, Research, and Evaluation* Vol20, Article 5.

[Volver al inicio](#)