

Optimizando las herramientas virtuales para evaluar a estudiantes de Estadística

Caro, Norma Patricia - Ahumada, María Inés
Facultad de Ciencias Económicas, Universidad Nacional de Córdoba
pacaro@eco.unc.edu.ar - ahumada.mi@gmail.com

XII Foro Educación Superior, innovación e internacionalización

Resumen

La evaluación de los estudiantes es un aspecto importante en el proceso educativo y el contexto cambiante en el que éste último se desarrolla, motiva a los docentes a buscar estrategias innovadoras para mejorar dicha tarea.

Desde hace varios años se viene aplicando la metodología Blended-Learning en los cursos de grado de Estadística. Sin embargo recién en 2016 se comenzó a evaluar a los estudiantes desde la virtualidad, haciendo uso intensivo de los cuestionarios Moodle. La finalidad de tal innovación fue complementar las evaluaciones escritas tradicionales con otras instancias evaluativas intermedias, apelando a tecnologías amigables y con bajo impacto en la carga de trabajo para los docentes, al considerar la masividad de los cursos.

Se logra un impacto positivo de la experiencia motivando a los alumnos a dar mayor continuidad al proceso de aprendizaje y obteniendo mejoras cualitativas en sus rendimientos finales. Asimismo es necesario exceder la mera observación de resultados, abarcando también un análisis exhaustivo de la fiabilidad de los instrumentos de evaluación. Los indicadores psicométricos reportados por el módulo permiten detectar fácilmente las debilidades del instrumento, que derivan en múltiples estrategias para realizar ajustes e incrementar la calidad del mismo. Se aplicaron con éxito las propuestas de mejoras.

Palabras Clave: Cuestionarios, Fiabilidad, Evaluación formativa, Estadística.

1. Introducción

El Blended-Learning, es considerado en un contexto de enseñanza virtual, como una modalidad de enseñanza-aprendizaje que incluye formación no presencial y presencial. El equipo docente que viene participando en la experiencia de aplicar esta metodología en un curso masivo de Estadística descriptiva y Probabilidad, en la Facultad de Ciencias Económicas de la UNC, siempre consideró muy relevante la adecuada selección y aplicación de las TICs, desde la elección del software estadístico e incluyendo la elaboración y actualización permanente de los materiales (Tutoriales, Instructivos y bases de datos) para que se encuentren disponibles en formato digital.

Al inicio se avanzó sobre la conformación del Aula Virtual, en una plataforma propietaria, la cual permitió desplegar la propuesta de enseñanza-aprendizaje. Cada año se exploraban y aplicaban las herramientas disponibles en la plataforma: Noticias, Foros, Encuestas, publicación de archivos, entre otras.

En 2015 se abandonó la plataforma del inicio, migrando el Aula a una plataforma Moodle. En 2016 se incorporaron cambios importantes en las técnicas de evaluación, sostenidos en el

enfoque de una apropiada evaluación formativa, acorde al contexto institucional en el que se desarrolla la asignatura y con el objetivo de incidir para generar una mejora del aprendizaje que pudiera ser significativamente tangible.

La evaluación de los alumnos es un tema sensible siendo una de sus finalidades medir los logros en los resultados de aprendizaje que se pretendan alcanzar, existiendo una relación con los procedimientos idóneos que deberían seguirse en los procesos evaluativos (Blanco y Ginovart, 2012). La necesidad de generar dispositivos de evaluación confiables constituye una ventaja estratégica para cualquier docente que desee incorporar tecnologías de la información y comunicación (TIC) en su práctica académica.

Blanco y Ginovart (2012) plantean que varios estudios han puesto de relieve el papel cada vez más importante de las tecnologías de la información y la comunicación en el campo de la evaluación (Delgado y Oliver, 2006; Graff, 2003, Steegmann et al, 2008), hasta el punto de que la evaluación virtual ha pasado a ser un nuevo elemento del ámbito educativo (Brinck y Lautenbach, 2011; Crews y Curtis, 2011; Daly et al, 2010; Ferrao, 2010).

La evaluación como tema crítico puede ser abordado de distintas maneras, pudiendo clasificar a los cuestionarios en dos grupos: formativos y evaluativos.

- Cuestionarios Formativos: El objetivo es aprender. El alumno tiene la oportunidad de intentar varias veces resolver el mismo examen o la misma pregunta, pues el objetivo es, que si se identifica que ignora algo, lo aprenda al resolver el examen.
- Cuestionarios Evaluativos o Sumativos: El objetivo es valorar si se otorgan o no los créditos del tema/curso. El profesor usará la calificación del examen como un elemento importante para la calificación del curso.

Los cuestionarios en Moodle no sólo han demostrado ser útiles para llevar a cabo evaluaciones sino que pueden ser modificados y adaptados según las necesidades del proceso de enseñanza aprendizaje. Tales ajustes pueden ser propuestos con fundamentos basados en el análisis de los índices psicométricos que se obtienen directamente al finalizar la evaluación.

Este trabajo es motivado por los resultados favorables que se obtuvieron a partir de la aplicación de las evaluaciones intermedias en entorno virtual y por la aceptación generalizada por parte de los alumnos, quienes accedieron masivamente a participar, aun tratándose de evaluaciones no obligatorias para aprobar el curso. También se relevaron las valoraciones cualitativas de los alumnos participantes sobre la experiencia. Así se pudieron identificar los principales aspectos positivos y negativos para fortalecer la herramienta.

Con la certeza de que en mayor o menor medida las evaluaciones contribuían al proceso de enseñanza, fueron sorprendentes los hallazgos al profundizar el análisis sobre la calidad del instrumento una vez aplicado. Al contar con una serie de indicadores psicométricos, automáticamente disponibles en modulo, se cuenta con elementos muy objetivos para plantear mejoras concretas en el diseño del cuestionario para que la evaluación resulte más fiable.

2. Descripción de la experiencia (completar con las dos experiencias)

La presente experiencia de enseñanza-aprendizaje se encuentra contextualizada dentro de los cursos de Estadística previstos en el ciclo Formación Básica Común del actual plan de estudios. Se trata de la asignatura Estadística I, que abarca conceptos básicos de Estadística Descriptiva, Introducción a la Probabilidad, Modelos de probabilidad y Distribuciones en el muestreo. La Estadística es una de las disciplinas que pueden ser

aplicadas al estudio metódico de problemas de los más diversos campos de la realidad. De estos últimos interesan particularmente los relacionados con la economía, la administración y la contabilidad. Asimismo, se apoya en el uso de software estadístico, indispensables para el manejo de bases de datos.

A partir del año 2016 se planificó incorporar cuatro instancias de evaluaciones individuales, presentadas a los alumnos como "Autoevaluaciones" no obligatorias, con recompensa de puntos que se suman al puntaje de los correspondientes pruebas parciales, que sí son obligatorios para obtener la regularidad de la materia. Estas instancias se pensaron también como herramientas para motivar al alumno a apropiarse de los contenidos, de manera paulatina a lo largo del cursado y evitar que el momento dedicado al estudio se restrinja a los días previos a cada parcial.

2. 1. Diseño del Instrumento de evaluación

El equipo docente recorrió las siguientes etapas para la elaboración de un instrumento de evaluación: Preparación, Aplicación, Valoración y Devolución. El desafío fue repensar esas instancias pero en el contexto de la virtualidad.

La etapa de "Preparación" implicó investigar en las alternativas de construcción del "Banco de Preguntas", definir la creación de las categorías que agrupen las preguntas conforme a los contenidos de la asignatura. La importancia que tiene darle una estructura muy organizada al banco de preguntas es permitir la reutilización de las mismas, ya que el banco de preguntas es escalable, año a año se pueden sumar preguntas a las categorías de preguntas creadas, o crear nuevas categorías y también modificar las preguntas para mejorarlas, ampliarlas, etc. El hecho de que las preguntas sean reutilizables, permite crear cuestionarios a partir de una mezcla aleatoria de preguntas de diferentes categorías, dando como resultado exámenes diferentes sobre un mismo conjunto de categorías predefinidas¹.

La etapa de "Aplicación", requiere también prever aspectos operativos vinculados a decidir la manera en la que el alumno responderá el cuestionario. Por la cantidad de alumnos que cursan habitualmente se realiza una inscripción previa de aquellos que deseaban participar en la realización de cuestionario, para organizar la disponibilidad de aulas equipadas con computadoras. Dado el carácter de evaluación sumativa otorgado a estos cuestionarios, los alumnos realizaron la actividad en aulas informáticas de la Facultad, de manera individual, aprovechando las medidas de seguridad informática, en el apartado de restricciones extras para los intentos, en la Edición de ajustes del cuestionario:

- Se indicó una dirección IP fija pública (la de la Facultad) desde la cual sólo se podía responder el cuestionario.
- Se empleó una clave de seguridad para iniciar la evaluación que se otorga al alumno al ingresar al aula.
- Se eligió un orden aleatorizado para las preguntas de respuestas múltiples.

Las etapas de valoración y devolución son amplias y fácilmente aplicables en los cuestionarios de Moodle, ya que en su diseño es posible ponderar el peso de cada pregunta y también es de suma utilidad la posibilidad de incluir una "retroalimentación" a cada respuesta (correcta o incorrecta) y/o una "retroalimentación" global para el ítem. El esmero en completar de manera clara esta información representa un aporte sumamente valorado por el alumno, porque le permite comprender mejor cuál fue su error o sino reforzar el concepto o contenido del ítem cuando responde correctamente la pregunta.

¹ documento: https://docs.moodle.org/all/es/Gestionando_preguntas

Por último, luego de concluida la evaluación, cada alumno puede volver a consultar en cualquier momento su cuestionario con el objetivo de coadyuvar al estudio y a la preparación para la instancia del examen escrito, el cuál sí es obligatorio para aprobar el curso.

3. Resultados y discusión

Para el periodo de análisis las conclusiones obtenidas a partir de los resultados de dichas experiencias evidencian una mejora en las calificaciones obtenidas por los alumnos. Cabe aclarar que para tales conclusiones se podrían incorporar a futuro un análisis de otros factores para validar el impacto de la metodología sobre el aprendizaje.

Asimismo, es esperable y razonable pensar que los alumnos que realizan las autoevaluaciones obtengan mejores notas en promedio, que aquellos que no participan en la actividad. Sin embargo las diferencias resultan significativas.

Se realizó una comparación a través de un test t de diferencia de medias (Cuadro 1) entre quienes realizaron tres o más autoevaluaciones, lo que significa que tuvieron una mayor participación respecto a quienes realizaron 2 o menos, siendo mayor la calificación promedio en el primer grupo. Por otro lado, se comparó a quienes tuvieron una amplia participación versus quienes decidieron no realizar autoevaluación alguna, siendo el promedio de las calificaciones mayor en este primer grupo. En ambos test se obtuvieron diferencias significativas, lo que se observa con los valores de probabilidad (p-value).

Cuadro 1: Diferencia de medias

Variable	Grupo 1	Grupo 2	Dif de medias	LI (95%)	LS(95%)	p-value
Nota del examen	3 o mas autoev. (n=78)	2 o menos autoev. (n=101)	1,21	0,56	1,86	0,0003
Nota del examen	3 o mas autoev. (n=78)	Ninguna autoev (n=101)	1,08	0,29	1,86	0,0076

3.1 Revisión del cuestionario

Las sucesivas aplicaciones de los cuestionarios permiten avanzar sobre nuevas etapas de revisión de los mismos, apelando a las herramientas que proporciona el aula virtual, en este caso la plataforma Moodle dentro del Reporte de estadísticas del cuestionario.

A través del cálculo de indicadores psicométricos, se evalúa individualmente el cuestionario en dos aspectos uno referido a características del mismo en su conjunto, como instrumento de evaluación. El otro aspecto apunta a analizar individualmente cada uno de los ítems que lo conforman.

A través de la interpretación de estos indicadores que se presentan en el “Reporte de estadísticas de examen”, se pretende aprovechar el potencial de este recurso, proponiendo ajustes de diseño fundamentados en análisis cuantitativos objetivos.

En los Cuadros 2 y 3 se presentan respectivamente, los resultados para cada ítem o pregunta, para un cuestionario del año 2016 y uno mejorado del año 2017.

Cuadro 2: Comportamiento individual de cada ítem- Cuestionario 2016 (*)

VALORES DE REFERENCIA		entre 30% y 70% es deseable	Valores arriba del 40% son insatisfactorios	muy por debajo del estimado, revisar	Entre 30% y 50% (adecuado) - >50% Muy bueno	> 50% optimo			
Q#	Tipo de pregunta	Nombre de la pregunta	Índice de dificultad	Desviación estándar	Valores aleatoria estimada	Peso estimado	Peso efectivo	Índice de discriminación	Eficiencia discriminativa
1	Opción múltiple	PREGUNTA 1	0.254	0.4388	0.2	9,09%	0.0805	0.1371	0.2029
2	Opción múltiple	PREGUNTA 2	0.2857	0.4554	0.2	9,09%	0.0857	0.162	0.2196
3	Numérica	PREGUNTA 2B	0.0794	0.2725	0	9,09%	0.0496	0.0792	0.1679
4	Numérica	Pregunta 3	0.8413	0.3684	0	9,09%	0.0948	0.416	0.6095
5	Numérica	PREGUNTA 4	0.4762	0.5034	0	9,09%	0.1215	0.4983	0.603
6	Numérica	PREGUNTA 5	0.4444	0.5009	0	9,09%	0.091	0.1507	0.1836
7	Opción múltiple	PREGUNTA 6	0.5238	0.5034	0.2	9,09%	0.0918	0.1545	0.1775
8	Opción múltiple	PREGUNTA 7	0.7143	0.4554	0.2	9,09%	0.1011	0.3254	0.4091
9	Opción múltiple	PREGUNTA 8	0.2698	0.4474	0.2	9,09%	0.0678	0.0249	0.0357
10	Opción múltiple	Tablas de contingencia	0.4286	0.4988	0.25	9,09%	0.1162	0.4332	0.5463
11	Opción múltiple	Tablas y gráficos	0.3968	0.4932	0.25	9,09%	0.0999	0.2495	0.3201

(*) Se resaltan valores no adecuados

Este reporte es generado de manera automática por cada cuestionario que haya sido respondido por un grupo de alumnos en una misma instancia, en esta experiencia sólo tienen la posibilidad de resolverlo una vez (por el carácter de evaluación sumativa). Se muestra una descripción del comportamiento individual de cada ítem.

Para cada pregunta es posible identificar una posible acción a realizar con el objetivo de mejorar la calidad del cuestionario. Por ejemplo, para el ítem #4 se podría revisar y aumentar el nivel de dificultad, por medio del enunciado o por medio de las alternativas propuestas. Mientras que para casos similares al ítem #5 sería mejor eliminarlo o reemplazarlo si se desea conservar la extensión del cuestionario. Notar que el ítem es de bajo nivel de dificultad, y al mismo es baja su eficiencia discriminativa.

Realmente las alternativas para mejorar cada ítem son amplias y requieren de cierta dedicación para que efectivamente los ajustes redunden en mejoras de los indicadores.

A continuación se muestran los resultados obtenidos a partir de los arreglos efectuados a los ítems del cuestionario anterior:

Cuadro 3: Comportamiento individual de cada ítem- Cuestionario 2017 (*)

VALORES DE REFERENCIA		entre 30% y 70%		Valores arriba del 40% son insatisfactorios		muy por debajo del estimado, revisar	Entre 30% y 50% (adecuado) - >50% Muy bueno	> 50% optimo	
Q# Tipo de pregunta	Nombre de la pregunta	Índice de dificultad	Desviación estándar	Calificación aleatoria estimada	Peso estimado	Peso efectivo	Índice de discriminación	Eficiencia discriminativa	
1	Opción múltiple	Conceptos básicos	65.63%	48.26%	20.00%	10%	9.80%	63.83%	67.86%
2	Opción múltiple	Tipo de variable y Escala de medición	53.13%	50.70%	25.00%	10%	9.68%	57.34%	69.24%
3	Opción múltiple	EQUIPAMIENTO TÉCNICO Y CALIDAD	59.38%	49.90%	20.00%	10%	10.72%	77.49%	88.61%
4	Opción múltiple	Empresas exportadoras	62.50%	49.19%	20.00%	10%	10.91%	82.82%	91.06%
5	Opción múltiple	Pymes Industriales Familiares	40.63%	49.90%	20.00%	10%	9.26%	51.96%	69.84%
6	Opción múltiple	Innovación	46.88%	50.70%	20.00%	10%	9.77%	58.81%	74.83%
7	Emparejamiento	Rubro de Actividad e Inversión en Tecnología	66.67%	47.14%	12.50%	10%	11.10%	91.75%	96.99%
8	Opción múltiple	Interpretando "Frecuencias"	15.63%	36.89%	20.00%	10%	6.71%	35.01%	73.70%
9	Opción múltiple	Interpretando Frecuencias	59.38%	49.90%	25.00%	10%	10.95%	81.98%	93.16%
10	Opción múltiple	Interpretando Frecuencias	65.63%	48.26%	25.00%	10%	11.10%	88.91%	95.52%

(*) Se resaltan valores no adecuados

Al aplicar el cuestionario en 2017 que fue respondido por una cantidad similar de alumnos, que tomaron el mismo curso, es decir conservando el mismo equipo docente, bajo el mismo plan de estudios y los mismos recursos en cuanto a material de estudio, etc. Los resultados de la valoración por ítem, mejora significativamente.

3.2 Estadísticas globales del cuestionario

El reporte estadístico incluye también información general del cuestionario, en el cual se hace una valoración global en la que influyen aspectos como la cantidad de alumnos que respondieron, la extensión del instrumento en cuanto a la cantidad de ítems que lo conforman e incluso la ponderación que se le asigna a los ítems al realizar el diseño del mismo. El Cuadro 4 se muestra de manera comparativa, los resultados de ambos años:

Cuadro 4: Información general sobre el cuestionario - Años 206 y 2017

Nombre del cuestionario	Grupo 3 Año 2016	Grupo 3 Año 2017	Indicador de Referencia
Nombre del curso	Estadística I	Estadística I	
Número de primeros intentos	63	29	
Número total de intentos	63	32	
Promedio de los primeros intentos	42,86%	54%	
Promedio de todos los intentos	42,86%	54%	promedio de 50% a 75%
Calificación media de los últimos intentos	42,86%	57%	
Calificación media de los mejores intentos	42,86%	57%	
Calificación media (de todos los intentos)	45,45%	70%	
Desviación estándar (para todos los intentos)	19,55%	36%	
Asimetría de la distribución de puntuaciones (para todos los intentos)	-0,2388	-0.56	
Curtosis de la distribución de puntuaciones (para todos los intentos)	-0,9296	-1.36	
Coefficiente de consistencia interna (para todos los intentos)	56,10%	92%	>80%
Ratio de error (para todos los intentos)	66,26%	29%	<50%
Error estándar (para todos los intentos)	12,96%	11%	entre 5% y 8%

3.2.1 Promedio de los intentos

El promedio de intentos fue del 42,86% y 54% respectivamente. Este valor es la media aritmética de las calificaciones obtenidas por los alumnos que finalizaron el cuestionario, que en este caso al momento de diseñar el cuestionario se optó por una escala de 0 a 100, pero Moodle ofrece varias alternativas. Para exámenes discriminantes con retroalimentación diferida, instituciones expertas de referencia, sugieren que en la aplicación de este tipo de recurso los exámenes eficientes deben obtener una calificación promedio de 50% a 75%. Los valores fuera de estos límites, sugieren repensar la evaluación.

3.2.2 Calificación media (de todos los intentos)

Esta calificación fue del 45,45% en 2016 y 70% en 2017. Bajo este título Moodle reporta el valor de la **Mediana** de las calificaciones. Es decir, es el valor central de la distribución. Entonces su interpretación es que la mitad de los alumnos obtuvieron una calificación inferior al 70% y el 50% restante, alcanzaron una calificación mayor que 70%, por ejemplo para el valor de 2017.

3.2.3 Desviación estándar (para todos los intentos)

Es una medida de la dispersión de las calificaciones alrededor del Promedio. En este caso es del 19,55% y 36% para el segundo año, lo que se encuentra incluido en el rango que va entre el 12% y 18%, para el 2016 que es lo que se sugiere como adecuado. Pero para 2017 el desvío estándar resultó mayor a 12% se refleja que las calificaciones se dispersan más, respecto al promedio.

3.2.4 Asimetría de la distribución de puntuaciones (para todos los intentos)

Se trata de una medida de la asimetría de la distribución de calificaciones. Si se obtiene "Cero" implica que la distribución de las calificaciones es perfectamente simétrica, mientras que los valores positivos indican presencias de unos pocos valores de calificaciones altas

('cola' a la derecha) y los coeficientes negativos indican presencia de pocos valores bajos y calificaciones concentradas en valores altos de la escala. En esta oportunidad el valor es de -0,23 y -0,56.

Las recomendaciones para este indicador es intentar obtener un valor de sesgo de -1 a 1. Si fuera un valor negativo muy elevado, puede estar indicando una falta de discriminación entre los alumnos, a los que les va mejor que al promedio. Asimismo, valores positivos muy altos (mayores a 1) pueden revelar falta de discriminación cerca de la calificación umbral para pasar/reprobar.

3.2.5 Curtosis de la distribución de puntuaciones (para todos los intentos)

La Curtosis (0,92 y -1,36) es una medida de forma e indica que tan aplanada es la distribución, es decir aporta una idea acerca de la deformación vertical de la misma. Para este indicador se busca obtener como valor deseable en el rango entre 0 y 1. Ya que un coeficiente de curtosis mayor que 1 puede indicar que el cuestionario no está discriminando muy bien entre los alumnos muy buenos (o los muy malos) y aquellos que son "alumnos promedio".

3.2.6 Coeficiente de consistencia interna (para todos los intentos)

Este coeficiente (Alpha de Cronbach) aporta una medición global, o también llamada "*medición compuesta*" sobre todos los ítems (o *componentes de medición*), como una estimación (inferior) de la fiabilidad de una prueba psicométrica. Aquí es deseable alcanzar valores mayores, ya que un valor bajo indica, o bien que algunas de las preguntas no son muy buenas para discriminar entre alumnos con diferentes niveles de conocimiento del tema que se evalúa y por esto las diferencias entre las puntuaciones totales están más bien asociadas al azar; o bien puede revelar que algunas de las preguntas está funcionando con una calidad diferente al restos de los ítems, provocando que el examen en su conjunto no sea homogéneo.

En la práctica es casi imposible obtener consistencia interna mayor al 90%, por ello la mayor parte la bibliografía que trata el tema indica que un valor de 75% o más es satisfactorio. Se acepta que si el valor del coeficiente es inferior a 64%, el examen completo es insatisfactorio, y en nuestro caso se obtuvo un 56,1% que luego de considerar la revisión del instrumento para adoptar medidas correctivas llega al 92%.

3.2.7 Ratio de error (para todos los intentos)

El ratio de error arroja un valor del 66,26% y luego disminuye al 29%. Esta tasa de error se relaciona con el coeficiente de consistencia interna anterior, según la bibliografía de referencia, de acuerdo con lo presentado en el Cuadro 6.

La tasa de error estima el porcentaje de la desviación estándar que se debe a efectos aleatorios en lugar de diferencias auténticas en la habilidad entre los estudiantes. De manera que cuestionarios con valores de tasa de error superiores al 50%, no deberían considerarse satisfactorios. Así, se muestra que el ratio de error para el cuestionario bajo análisis fue igual a 76%, evidenciando la necesidad de revisarlo. Tal valor implica que un cuarto de la desviación estándar se debe a diferencias en habilidad de los alumnos y que el resto son efectos aleatorios.

Cuadro 6: Relación entre Coeficiente de Consistencia Interna y la Tasa de Error

Coeficiente de Consistencia Interna	100	99	96	91	84	75	64	51
Tasa de Error	0	10	20	30	40	50	60	70

3.2.8 Error estándar (para todos los intentos)

El Error estándar es la desviación estándar multiplicada por la Tasa de error, y dividida por 100. Es una medida de la incertidumbre en la calificación de cualquier alumno dado. Hipotéticamente, si ese mismo estudiante resolviera otro cuestionario equivalente en la misma cátedra, se esperaría que su calificación estuviera dentro de más-menos un error estándar de la calificación anterior. Las interpretaciones aceptadas para este indicador plantean que cuanto menor sea su valor, mejor es cuestionario, inferior al 5% o 6%. Mientras que si el error estándar es mayor al 8%, es probable que una proporción sustancial de los alumnos se encuentren erróneamente calificados. En esta oportunidad esta medida fue del 12, 96 y 11%%.

3.3.1 Índice de Facilidad / Nivel de Dificultad

Es simplemente la puntuación promedio de los alumnos en el ítem, es igual al porcentaje de aprobación de los ítems sobre el total de intentos del mismo.

Cuadro 4: Índice de Facilidad: Valores de referencia.

Índice de Facilidad (%)	hasta 5	6 - 10	11- 20	20 - 34	35 - 64	66 - 80	81 - 89	90 - 94	95- 100
Interpretación	Extremadamente difícil, o algo está mal en la pregunta	Muy difícil	Difícil	Moderadamente difícil	Correcta para el alumno promedio	Bastante fácil	Fácil	Muy fácil	Extremadamente fácil

Como posible regla general, los índices de facilidad entre 30% y 70% suelen aportar diferencias importantes entre el nivel de conocimiento, habilidad y preparación entre los alumnos. Es importante que el docente tenga claro cuál es el propósito de la evaluación y las características de los estudiantes, para poder plantear niveles de dificultad deseados acordes.

3.3.2 Desviación estándar

Es una medida de la dispersión de las calificaciones respecto a la media, mide la magnitud de cuánto puede discriminar la pregunta. Ante un índice de facilidad muy alto (o muy bajo), la dispersión será grande, sin embargo una buena desviación estándar no implica necesariamente una buena discriminación. Según la tabla de referencia un valor de 1/3 del máximo de la pregunta (33%) es insatisfactorio.

3.3.3 Puntaje esperado aleatoriamente

Este es el promedio de calificación que se esperaría que los estudiantes obtuvieran si respondieran al azar. Sólo disponibles para las preguntas de opción múltiple y válido para cuestionarios de retroalimentación diferida. Son deseables valores por debajo del 40%.

3.3.4 Ponderación deseada y efectiva

La ponderación deseada o peso de la pregunta, Moodle la expresa como un porcentaje del puntaje general. Y la ponderación efectiva es una estimación de la ponderación o peso que realmente tiene la pregunta para contribuir a la dispersión total de las calificaciones.

3.3.5 Índice de discriminación y Eficiencia de discriminación

Esta es la correlación entre las calificaciones ponderadas en la pregunta y las del resto del examen. Revela qué tan efectivo es el ítem para clasificar o separar a los alumnos más capaces de los menos capaces (Cuadro 5).

Cuadro 5: índice de discriminación de los ítems.

Índice (%)	valores negativos	0 - 19	20 - 29	30 – 50	50 o superior
Interpretación	Ítem probablemente inválido	Discriminación muy débil	Discriminación débil.	Discriminación Adecuada	Discriminación Muy buena

3.3.6 Eficiencia de discriminación

El indicador intenta estimar que tan bueno es el índice de discriminación en relación con la dificultad de la pregunta. Un ítem muy fácil o muy difícil no discrimina entre alumnos con habilidades diferentes, debido a que la mayoría de ellos obtendrán el mismo puntaje para esta pregunta y viceversa. En el Cuestionario aparecen ítems con valores negativos que deberían ser revisados.

4. Conclusiones finales

Se encontraron evidencias sobre el impacto que tiene la aplicación de los cuestionarios de Moodle sobre el rendimiento de los alumnos, medido a partir de la nota del examen final, hecho que motiva a continuar mejorando la propuesta de enseñanza. No basta con tomar evaluaciones y esperar los resultados que tienen los alumnos, ya que puede ser engañoso y crear un pensamiento negativo hacia el estudio realizado por los alumnos, cuando es el instrumento el que no es confiable para evaluarlos.

Las buenas prácticas para mejorar el diseño de los cuestionarios para evaluar en la virtualidad deben considerar el análisis de medidas psicométricas similares a las presentadas en este trabajo para fundamentar objetivamente tales mejoras.

Se logró aprovechar el enorme potencial latente en los cuestionarios que se presentan a los alumnos como una instancia de autoevaluación, en el sentido de que se espera que sean un reflejo de cómo se están preparando para las instancias de evaluación que en definitiva les permiten aprobar la asignatura.

Referencias

- Blanco, M., & Ginovart, M. (2012). Los cuestionarios del entorno Moodle: su contribución a la evaluación virtual formativa de los alumnos de matemáticas de primer año de las titulaciones de Ingeniería. *RUSC. Universities and Knowledge Society Journal*, 9(1).
- Brink, R., & Lautenbach, G. (2011). Electronic assessment in higher education. *Educational Studies*, 37(5), 503-512.
- Crews, T. B., & Curtis, D. F. (2011). Online course evaluations: Faculty perspective and strategies for improved response rates. *Assessment & Evaluation in Higher Education*, 36(7), 865-878.
- Daly, C., Pachler, N., Mor, Y., & Mellar, H. (2010). Exploring formative e-assessment: using case stories and design patterns. *Assessment & Evaluation in Higher Education*, 35(5), 619-636.
- Delgado García, A. M., & Oliver Cuello, R. (2006). La evaluación continua en un nuevo escenario docente. *RUSC. Universities and Knowledge Society Journal*, 3(1).
- Ferrão, M. (2010). E-assessment within the Bologna paradigm: evidence from Portugal. *Assessment & Evaluation in Higher Education*, 35(7), 819-830.
- Graff, M. (2003). Cognitive style and attitudes towards using online learning and assessment methods. *Electronic Journal of e-learning*, 1(1), 21-28.
- Iglesias Rodríguez, A., Olmos Migueláñez, S., Torrecilla Sánchez, E. M., & Mena Marcos, J. J. (2014). Evaluar para optimizar el uso de la plataforma moodle (studium) en el departamento de didáctica, organización y métodos de investigación. *Tendencias pedagógicas*.
- Llorente Cejudo, M. (2007). Moodle como entorno virtual de formación al alcance de todos. *Comunicar*, 15(28).
- Martín Galán, B., & Rodríguez Mateos, D. (2012). La evaluación de la formación universitaria semipresencial y en línea en el contexto del EEES mediante el uso de los informes de actividad de la plataforma Moodle. *RIED. Revista iberoamericana de educación a distancia*, 15(1).
- Steggmann, C., Huertas, M. A., Juan, Á. A., & Prat, M. (2008). E-learning de las asignaturas del ámbito matemático-estadístico en las universidades españolas: oportunidades, retos, estado actual y tendencias. *RUSC. Universities and Knowledge Society Journal*, 5(2), 1-14.
- San Martín, E. (2004). Elementos de Psicometría- Teoría de Medición, Teoría Clásica de Test y teoría de Respuesta al Ítem. *CLATSE VI Sexto Congreso Latinoamericano de Sociedades Estadísticas. Departamento de Estadística - Pontificia Universidad Católica de Chile*.