
Evaluación de un instrumento para examinar a estudiantes universitarios en conocimientos de estadística

Evaluation of an instrument to examine university students in knowledge of statistics

11

Recibido: 10/01/2022

Aprobado: 10/03/2022

Norma Patricia Caro¹
Maria Inés Ahumada²
Verónica Arias³

Resumen

En los procesos de enseñanza y aprendizaje un aspecto relevante es la evaluación del estudiantado, con estrategias innovadoras que tienen como objetivo obtener instrumentos de calidad. En el presente trabajo se pretende verificar el grado de dificultad y capacidad de discriminación de las preguntas que examinan al estudiantado en los conocimientos adquiridos en estadística, a través de una evaluación virtual con medios tecnológicos. Para ello, se construyó una prueba con 21 preguntas sobre Inferencia estadística y se la implementó en 124 estudiantes de un curso de Estadística en la Facultad de Ciencias Económicas de la Universidad Nacional de Córdoba, Argentina. A través de un modelo de respuesta al ítem (IRT), con variables de respuesta binaria, para medir los conocimientos adquiridos en el curso de estadística se analizó la relación de cada pregunta del cuestionario con el rasgo latente (conocimientos de estadística) y se determinó el grado de dificultad y discriminación de cada una de las preguntas. Los resultados indican que aquellas preguntas que aportan al constructo, que poseen dificultad media y que son las que más discriminan entre el estudiantado pueden ser retenidas en el cuestionario, ya que permiten además desarrollar otras capacidades que también deben aprenderse, mientras que aquellas preguntas no significativas o muy fáciles y que no discriminan pueden ser reemplazadas con la finalidad de obtener mejores instrumentos de evaluación para lograr los objetivos de aprendizaje. Esta experiencia fue de suma utilidad en el contexto de pandemia donde la evaluación solo fue de carácter virtual.

¹ Docente e investigadora de la Facultad de Ciencias Económicas de la Universidad Nacional de Córdoba, Argentina y miembro del Centro de investigaciones en Ciencias Económicas, CIES UNC – CONICET.

² Docente e investigadora de la Facultad de Ciencias Económicas de la Universidad Nacional de Córdoba, Argentina Córdoba, Argentina.

³ Docente e investigadora de la Facultad de Ciencias Económicas de la Universidad Nacional de Córdoba, Argentina.

Palabras clave: evaluación, cuestionario, estadística, modelos de ecuaciones estructurales.

Abstract

12

In the teaching and learning processes, a relevant aspect is the evaluation of the student, with innovative strategies that aim to obtain quality instruments. In the present work it is intended to verify the degree of difficulty and discrimination capacity of the questions that examine the students in the knowledge acquired in statistics, through a virtual evaluation with technological means. For this, a test with 21 questions on Statistical Inference was constructed and it was implemented in 124 students of a Statistics course at the Faculty of Economic Sciences of the National University of Córdoba, Argentina. Through an item response model (IRT), with binary response variables, to measure the knowledge acquired in the statistics course, the relationship of each question in the questionnaire with the latent trait (knowledge of statistics) was analyzed and determined the degree of difficulty and discrimination of each of the questions. The results indicate that those questions that contribute to the construct, that have medium difficulty and that are the ones that most discriminate between the student body can be retained in the questionnaire, since they also allow the development of other capacities that must also be learned, while those non-significant questions or very easy and that do not discriminate can be replaced in order to obtain better assessment instruments to achieve the learning objectives. This experience was extremely useful in the context of a pandemic where the assessment was only virtual.

Keywords: evaluation, questionnaire, statistics, structural equation models.

Introducción

Todo proceso de enseñanza que promueve el desarrollo de aprendizajes aplicados a la actuación profesional va acompañado de un proceso de evaluación, que refleja si el estudiantado ha adquirido saberes y habilidades que les permita acreditar sus estudios y aplicar dichos conocimientos en su ámbito profesional.

La evaluación del estudiantado es un tema complejo, contar con instrumentos y procedimientos idóneos ayuda a mejorar los procesos educativos y, por ende, influye en el rendimiento académico (Blanco y Ginovart, 2012). La necesidad de generar dispositivos de evaluación confiables es una deuda aún pendiente en las instituciones de educación superior. En esta oportunidad se presenta una experiencia que incorpora recursos tecnológicos, cuya sistematización ayuda a su comprensión y validación.

En el marco del *Blended-Learning*, como una modalidad de enseñanza y de aprendizaje que incluye formación virtual y presencial, hace bastante tiempo que se comenzaron a realizar instrumentos de evaluación con tecnologías que permiten

evaluar cursos masivos. Si bien, desde hace varios años se viene implementando esta modalidad, con motivo de la pandemia de Covid- 19, desde 2020, la modalidad virtual fue la alternativa para continuar con estos procesos.

En los formatos virtuales (plataformas) se ofrecen una serie de recursos, entre los que se encuentran las evaluaciones que pueden gestionarse en tiempo real, como son los cuestionarios en la plataforma Moodle, que pueden ser evaluados de manera continua, logrando su fiabilidad y validez (Iglesias Rodríguez, et al, 2014, Larena y Paparo, 2006, Llorente Cejudo, 2007, Martín Galán y Rodríguez Mateos, 2012, Casal, 2010; Conde, 2005).

En 2019 se evaluó a los estudiantes de un curso de Estadística en la Facultad de Ciencias Económicas de la Universidad Nacional de Córdoba, Argentina con un cuestionario, en el que se midieron los conocimientos adquiridos en estadística y las habilidades para identificar problemas (constructo latente). Para cuantificar dicho constructo se aplicó un modelo de ecuaciones estructurales generalizado, como es el modelo de respuesta al ítem (IRT). Asimismo, se analizó la relación de cada ítem con el concepto estadístico y cómo el cuestionario en su conjunto se relaciona con dicho rasgo latente. A través del análisis del grado de dificultad y de discriminación que tuvo cada pregunta para reflejar el constructo latente de interés, es posible lograr mejores instrumentos de evaluación.

La motivación de este trabajo consiste en avanzar hacia propuestas de cuestionarios fiables y válidos que permitan evaluar lo que deben evaluar, es decir que, además de realizar instrumentos de evaluación con tecnologías, evaluarlos tanto con los indicadores que puede proporcionar la herramienta cuestionario en las plataformas educativas, como así también con índices de dificultad y de discriminación de cada pregunta o ítem en torno al concepto latente.

Lo trabajado en 2019 tuvo su aplicación directa en el contexto de pandemia donde la evaluación virtual fue la única alternativa que permitió a la universidad la continuidad pedagógica del dictado de sus carreras, con énfasis en un modelo virtual, por lo que el cuestionario elaborado en ese momento constituyó la base para la elaboración de nuevos cuestionarios que se utilizaron desde 2020 en adelante.

Siguiendo a Medina Rebollo, *et. al.* (2019) se hace necesario el análisis y la reflexión de esos instrumentos de evaluación, a fin de conocer los puntos fuertes para reforzarlos y mantenerlos; y determinar los puntos débiles que nos permitan plantear propuestas y alternativas de mejora, con el fin último de mejorar el proceso de aprendizaje.

A fin de verificar el grado de dificultad y capacidad de discriminación de las preguntas que examinan al estudiantado en los conocimientos adquiridos en estadística, se plantean los siguientes objetivos:

- Validar un cuestionario para una mejor evaluación del estudiantado.
- Identificar fortalezas y debilidades en los instrumentos de evaluación.

Esta experiencia es de suma utilidad en el contexto de pandemia donde la evaluación, totalmente virtual, utiliza estos instrumentos, los que pueden ser mejorados de manera continua. También pueden ser utilizados en la modalidad presencial cuando se cuenta con los medios tecnológicos adecuados: computadoras, conexión a Internet, soporte técnico, entre otros, lo que se hace difícil cuando se trata de cursos masivos.

En este artículo se presenta una introducción, luego un marco teórico con los principales antecedentes. En una tercera sección se presenta las características de la muestra, las variables, el instrumento y el método, que consiste en aplicar un modelo de respuesta al ítem, dentro de los modelos de ecuaciones estructurales generalizados que permiten medir un constructo no observable (latente), el que se refleja en indicadores de naturaleza cualitativa. A continuación, los resultados obtenidos y por último las conclusiones.

Marco teórico

En la actualidad, la educación superior está en búsqueda constante de la calidad, lo que es respaldado y exigido por los organismos de evaluación y acreditación de carreras de grado. Dentro de una serie de acciones de mejora, el uso de las plataformas de aprendizaje como un sistema integrado único y seguro para crear comunidades de aprendizaje (Tirado-Morueta y Hernando-Gómez, 2011) ha sido utilizado ampliamente, lo que conjuga no solo el uso de las tecnologías de la información y la comunicación, sino también las competencias digitales que traen los y las estudiantes para este contexto (Carrera, et al, 2011).

En lo referente al papel cada vez más importante de las tecnologías en el campo de la evaluación, Blanco y Ginovart (2012) plantean la importancia que posee la evaluación virtual en el ámbito educativo lo que es valorizado por Brink y Lautenbach (2011); Crews y Curtis (2011); Daly et al (2010) y Ferrão (2010).

Además, según Cassany y Ayala (2008), la mayoría de nuestros jóvenes están aprendiendo de manera informal un conjunto de habilidades y técnicas de manejo de información que muchos padres y docentes ignoran, por lo que es necesario orientar sobre la riqueza pedagógica de los servicios de información y comunicación a los que están expuestos.

En base a esto y dado que la evaluación es un aspecto indisoluble de la enseñanza y el aprendizaje (Anijovich y Cappeletti, 2017), constituyendo en sí mismo un proceso en el que los y las estudiantes ponen en juego sus propias competencias y progreso (Ryan, et al, 2002). Tomatis, et al (2016) destacan este aspecto y la preocupación que genera en el cuerpo docente dada sus consecuencias en el aprendizaje.

En este sentido, la cultura de evaluación, demanda de información continua, de comprensión, de conocimiento y entendimiento de hechos o situaciones (Medina et al., 2014). Uno de los aspectos más importantes en la evaluación es la validez de sus usos en lo que se refiere al grado en que la evidencia apoya las inferencias que se hacen a partir de las puntuaciones obtenidas en dicha instancia de evaluación, y el proceso de acumulación de evidencias que sustentan estas inferencias (Rojas, 2013 como se citó en Tomatis, et al, 2009, p. 360).

En 2020, en el contexto de pandemia del Covid-19 todos los sistemas educativos en Latinoamérica y en el mundo han recibido un fuerte impacto que ha afectado a estudiantes, hogares, docentes, directivos, centros educativos, (Álvarez, et al 2020) lo que ha hecho que las herramientas digitales tomaran mayor

preponderancia y la evaluación en entornos virtuales fuese adoptada con el objetivo de dar continuidad pedagógica en las trayectorias del estudiantado (Dussel, et al, 2020).

La evaluación es una oportunidad para que los estudiantes pongan en juego sus saberes, visibilicen sus logros, reconozcan sus debilidades y fortalezas como aprendices, además de cumplir con su función clásica de aprobar. En este contexto sanitario se torna un desafío pensar en la evaluación (Balmaceda en Beltramino, 2020).

Gerbaldo y Granato, en Beltranimio (2020), consideran que la pandemia se presentó como una posibilidad para rediscutir el rol de la evaluación en los procesos de acompañamiento a las trayectorias escolares y lograr un cambio de paradigma, ya que se trata de comprender los modos de aprender, valorar procesos y resultados, conocer en qué momento de su proceso de aprendizaje los y las estudiantes se encuentran, autorregular su aprendizaje y captar cómo se sintieron ante nuevos desafíos propuestos por el /la docente. Así, esta forma de evaluación permite rever y comprender su práctica educativa, reconstruir los criterios de evaluación, indicadores y/o competencias y ofrecer ejemplos y contraejemplos que permita conocer y reconocer la singularidad de cada estudiante, facilitando pensar hacia dónde orientar su propia práctica (Rossetti en Beltramino, 2020).

A su vez, este proceso tiene lugar, particularmente en estos últimos años, en espacios constituidos por plataformas educativas, como lo es Moodle, por ejemplo, entre muchos otros desarrollos realizados para incorporar el proceso que conlleva la evaluación.

Los cuestionarios en Moodle se clasifican en dos grupos: formativos y evaluativos. Los formativos tienen como objetivo, que los estudiantes aprendan a través de la reiteración y autoevaluación, mientras que los cuestionarios evaluativos o sumativos tienen como objetivo, valorar si se otorgan o no los créditos del curso (Caro y Ahumada, 2018).

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 de los procesos de enseñanza y de aprendizaje. Tales ajustes pueden ser propuestos con fundamentos basados en el análisis de los índices psicométricos que se obtienen a partir de la misma plataforma Moodle (Caro y Ahumada, 2018, p. 2).

Estas pruebas de validez (Rodríguez Castro y Gómez López, 2010), estriban en la posibilidad de contar con otros medios que permitan evaluar la efectividad de la misma, sus alcances y limitaciones. En tal sentido, Organista-Sandoval, et al (2016) y Medina Rebollo et al (2019) desarrollan cuestionarios para estimar las habilidades digitales de estudiantes universitarios.

En función del objetivo de esta investigación, las pruebas son procedimientos de recolección de información sobre un individuo o un grupo. La teoría de las pruebas (tests) constituye el marco de referencia teórico y metodológico que reúne los modelos que subyacen a la elaboración y uso de esas pruebas (Muñiz, 2010). Los modelos que componen esta teoría formalizan las interrelaciones de:

- la puntuación obtenida (el puntaje total en una prueba o la respuesta de un individuo a un ítem),

- un rasgo latente (no observado) que se pretende medir
- el error de medida que conlleva todo el proceso de medición.

Skrondal y Rabe-Hesketh (2004) mencionan a Spearman (1904) como el pionero en la formulación de la Teoría Clásica de los Tests (TCT) que ha servido como modelo para dar una interpretación a los puntajes de las personas en esas pruebas. En los años siguientes se ha dado lugar a un sinnúmero de técnicas de medición psicológica superando a la medición clásica, siendo la Teoría de Respuesta al Ítem (TRI), la más reconocida (Hambleton y Swaminathan, 1985; Lord, 1980; Martínez-Arias, 1995; Muñiz, 1995 mencionados por Skrondal y Rabe-Hesketh, 2004), con el objetivo de estimar el error que se comete al intentar medir un fenómeno psicológico específico (Embretson y Reise, 2000).

Atorresi, et al (2009) mencionan que la TRI establece una relación entre el comportamiento de un sujeto frente a un ítem y el rasgo responsable de esta conducta (rasgo latente) mediante funciones matemáticas que describen la probabilidad de dar una determinada respuesta al ítem para cada nivel del rasgo medido por éste.

El objetivo sustancial de la TRI es la construcción de instrumentos de medición con propiedades invariantes entre poblaciones. Si dos individuos presentan idéntico nivel de rasgo medido ambos tendrán igual probabilidad de dar la misma respuesta, independientemente de la población de pertenencia (Atorresi, et al, 2009, p. 180).

La TRI (de Kohan, 2010) brinda fundamentación probabilística al problema de medir constructos latentes (no observables) y considera al ítem como una unidad básica de medición y utiliza un patrón de respuesta y no la sumatoria de repuestas individuales que es lo que hace la teoría clásica. De Kohan (2010) menciona que Lord (1968,1980) fue el primero en elaborar este modelo, basándose en una distribución de probabilidad normal.

Muñiz, 2010 plantea que

El supuesto clave de los modelos TRI es que existe una relación funcional entre los valores de la variable que miden los ítems y la probabilidad de acertar estos, denominando a dicha función Curva Característica del Ítem (...)

La forma concreta de la curva viene determinada por el valor que tomen tres parámetros a, b, y c. Siendo a: el índice de discriminación; b: la dificultad del ítem y c: la probabilidad que hay de acertar el ítem al azar. Según los parámetros tomen unos valores u otros se generan distintas formas de curvas” (Muñiz, 2010, p. 64).

La mayoría de los modelos TRI asumen, que los ítems constituyen una sola dimensión (unidimensionales) y que poseen independencia local, es decir que los ítems son independientes unos de otros, la respuesta de uno de ellos no está condicionada a la respuesta dada a otros ítems.

Materiales y métodos

Muestra, variables e instrumento de recolección de datos

En esta oportunidad se realizó una investigación aplicada, cuantitativa, de campo y explicativa.

Se trabajó con una muestra de 124 estudiantes de un total de 207 pertenecientes a un curso de Estadística del primer semestre de segundo año de las carreras de ciencias económicas de la Facultad de ciencias Económicas de la Universidad Nacional de Córdoba, en 2019, quienes realizaron una evaluación sobre Inferencia Estadística. La evaluación estaba dirigida al total del estudiantado, pero con carácter optativo con el objetivo de autoevaluarse en una instancia previa al examen parcial (evaluación de proceso) que deberían realizar de manera obligatoria, con posterioridad. El 60% que realiza la evaluación constituye una muestra no probabilística, ya que los estudiantes eligen hacerla o no, pero representativa de la población, ya que mantiene sus características.

Se realizaron 21 preguntas, cuyos temas se incluyen en la Tabla 1. Se trata de ejercicios a resolver y preguntas a contestar que fueron evaluados como correctas o incorrectas (respuesta binaria). Para poder realizar la evaluación, el estudiantado se ha preparado con el material de estudio, con las clases, entre otras actividades.

El instrumento es elaborado por el equipo de cátedra, es decir con juicio de expertos en lo que se quiere evaluar. Cada pregunta tiene dos resultados posibles.

1: aprobado

0: no aprobado, ya sea por error o por ausencia de respuesta.

Tabla 1

Contenidos incluidos en el cuestionario de evaluación

Preg.	Tema principal	Subtema
P1		Planteo de hipótesis
P2	Prueba de hipótesis para la media con varianza poblacional conocida	Estadístico y distribución de probabilidad
P3		Regla de decisión: zona de rechazo y no rechazo de la hipótesis nula
P4		Decisión y conclusiones en términos del problema
P5		Interpretación del intervalo de confianza
P6	Estimación por intervalos de la media poblacional	Cálculo e interpretación de la precisión de la estimación
P7	Prueba de hipótesis para la media con varianza poblacional desconocida	Concluir a partir del valor p reportado por el software
P8	Determinación del tamaño de muestra para estimar la proporción	Cálculo del tamaño muestral
P9	Muestreo y métodos de muestreo	Similitudes y diferencias entre métodos de muestreo
P10		Planteo de hipótesis
P11	Prueba de hipótesis para la proporción poblacional	Estadístico y distribución de probabilidad
P12		Regla de decisión: zona de rechazo y no rechazo de la hipótesis nula
P13		Calcular el valor observado del estadístico

P14		Decisión y conclusiones en términos del problema
P15	Potencia de una prueba estadística	Calcular la probabilidad de cometer el error tipo II
P16		Identificación de la prueba de hipótesis a utilizar
P17		Planteo de hipótesis
P18	Prueba de hipótesis para la diferencia de proporciones	Supuestos requeridos para realizar la prueba
P19		Estadístico y distribución de probabilidad. Regla de decisión: zona de rechazo y no rechazo de la hipótesis nula
P20		Calcular el valor observado del estadístico
P21		Decisión y conclusiones en términos del problema

Fuente: elaboración propia

Se define como constructo latente aquel que representa los conocimientos adquiridos en estadística.

Para la elaboración del instrumento se cumplieron una serie de etapas:

Diseño: en esta etapa se plasmaron una serie de decisiones que iban desde evaluar desde la virtualidad, que recursos eran requeridos, cómo evaluar, entre otras cuestiones. Estas decisiones permitieron la evaluación en la plataforma Moodle para ser realizada en computadora, tanto de forma presencial como virtual.

Preparación: implicó construir un banco de preguntas, sus categorías conforme a los contenidos de la asignatura. Un banco de preguntas bien estructurado permite la reutilización de las preguntas y elaborar nuevos cuestionarios a partir de una combinación adecuada de preguntas de diferentes categorías (Caro y Ahumada, 2018).

Aplicación: El estudiantado realizó el cuestionario de manera on-line en una hora, todo/as al mismo momento por lo que se les comunicó el día y el horario. Se utilizó la versión 3.11 de Moodle. Se tuvieron en cuenta aspectos operativos vinculados a decidir la forma en que el estudiantado responderá el cuestionario, ya sea en la Facultad o bien desde sus hogares, para que fuera individual se utilizaron medidas de seguridad para su implementación⁴.

Valoración: incluye la ponderación o peso de cada pregunta

Devolución: incluir una retroalimentación a cada ítem y/o al instrumento en su conjunto. El hecho de realizar una devolución de manera clara representa un aporte al estudiantado para que comprenda sus modos de aprender, valore sus procesos y resultados, autorregule su aprendizaje y se comuniquen con su docente sobre su evaluación (Balmaceda en Beltramino, 2020). La retroalimentación permite pensar la evaluación como oportunidad (Anijovich, y Cappelletti, 2020). Además, el cuestionario

⁴Se 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 y se eligió un orden aleatorizado para las preguntas de respuestas múltiples. En 2020 se utilizaron navegadores seguros para la realización de las evaluaciones por parte del estudiantado desde sus hogares.

puede ser consultado en otras oportunidades con el objetivo de contribuir al estudio y a la preparación de instancias futuras de evaluación.

Método

En este trabajo se aplica la TRI en el campo de la educación, sobre los conocimientos adquiridos en estadística por estudiantes que cursan en una institución de educación superior. Para ello, se utiliza un modelo de ecuaciones estructurales, un modelo logístico de dos parámetros para datos dicotómicos, que tiene como variables respuesta, variables binarias que siguen una distribución logística. Nazario, et al (2021) presentan el análisis TRI utilizando el modelo logístico de dos parámetros, para variables dicotómicas, propuesto por Birnbaum (1968), que es el que se aplica en este trabajo, según la ecuación:

$$P_i(\theta) = \frac{e^{Da(\theta-b)}}{1+e^{Da(\theta-b)}} \quad \text{Para } i=1,2,3,\dots,n$$

Donde:

$P_i(\theta)$: es la probabilidad de que un o una examinado/a elegido/a al azar con aptitud θ conteste correctamente el ítem i .

n : número de ítems del test.

b : parámetro de posición o de la dificultad del ítem.

D : factor igual a 1,7, es un valor arbitrario introducido para que la función logística sea ajustada a la ojiva normal con una exactitud de 0,01.

a : parámetro de discriminación, es la pendiente de la curva característica de cada ítem en el punto b .

Los ítems con pendiente mayor son más útiles para separar al estudiantado en distintos niveles de aptitud, que aquellos que tienen menor pendiente.

González-Montesinos, M. y Backhoff, E. (2010) utilizan un modelo de ecuaciones estructurales para obtener validez de constructo. El modelo de ecuaciones estructurales utilizado, posee una notación grafica que se presenta en la figura 1. El rasgo latente lo constituyen los conocimientos en estadística y cada uno de los resultados de las preguntas son las variables observadas.

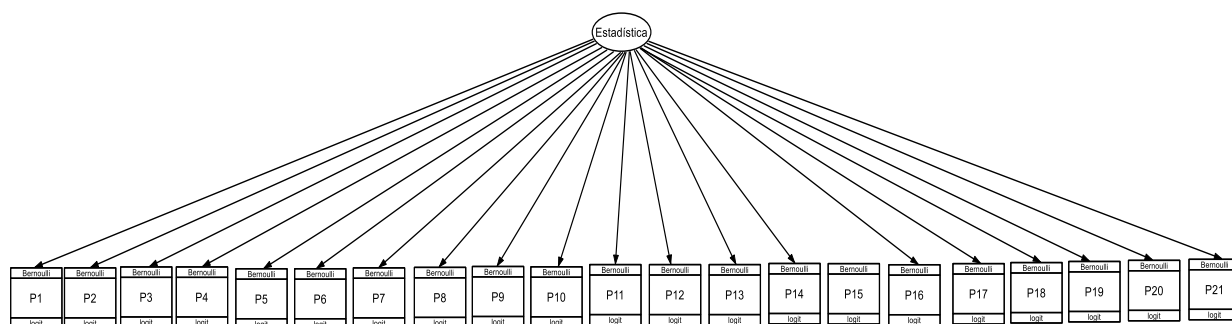


Figura 1 – Modelo de ecuaciones estructurales generalizado (teoría de respuesta al ítem)

Fuente: elaboración propia

Resultados

El estudiantado que realizó la evaluación estaba distribuido de manera uniforme entre varones y mujeres, sus edades eran en promedio de 20 años, el 80% son cursantes por primera vez y están distribuidos proporcionalmente en las tres carreras de la Facultad de Ciencias Económicas. Luego de la implementación del cuestionario, los porcentajes de aprobación se presentan en la Tabla 2. Se observa que la pregunta P15 fue aprobada sólo por el 14,52% de los estudiantes. Las preguntas P8, P17, P3 y P19 fueron aprobadas entre el 40% y menos del 50%.

Tabla 2

Porcentaje de aprobación de cada pregunta

Pregunta	P1	P2	P3	P4	P5	P6	P7
% aprobados	80,65	81,45	48,39	79,84	59,68	74,19	61,29
% reprobados	19,35	15,55	51,61	20,16	40,32	25,81	38,71
total	100,00	100,00	100,00	100,00	100,00	100,00	100,00
Pregunta	P8	P9	P10	P11	P12	P13	P14
% aprobados	41,94	57,26	72,58	83,06	68,55	74,19	61,29
% reprobados	58,06	42,74	27,42	16,94	31,45	25,81	38,71
total	100,00	100,00	100,00	100,00	100,00	100,00	100,00
Pregunta	P15	P16	P17	P18	P19	P20	P21
% aprobados	14,52	83,06	44,35	68,55	48,39	74,19	61,29
% reprobados	85,48	16,94	55,65	31,45	51,61	25,81	38,71
total	100,00	100,00	100,00	100,00	100,00	100,00	100,00

Fuente: elaboración propia

Se midió el aporte de cada pregunta (ítem) sobre el constructo (conocimientos en estadística) lo que se refleja en la tabla 3. La mayoría de las preguntas aportan significativamente al constructo lo que significa que reflejan los conocimientos y habilidades adquiridos en estadística ya que presentan una relación significativa entre cada ítem y la variable latente, excepto las preguntas P9 y P15 que no son significativas, a un nivel del 5%, ya que sus valores de probabilidad (p-valor) fueron de 0,058 y 0,123, respectivamente. Estas preguntas versan sobre:

P9: Similitudes y diferencias entre los modelos de muestreo estadístico (conceptual), lo que suele ser un tema no complejo es una pregunta que puede ser catalogada como fácil.

P15: Calcular la probabilidad de error tipo II para un valor particular de la hipótesis alternativa, lo que constituye un tema muy complejo, que de hecho fue respondido sólo por el 15% de los estudiantes en forma correcta.

Tabla 3
Coeficientes de discriminación sobre el constructo latente

Pregunta	p - valor
P1	0,029
P2	0,033
P3	0,023
P4	0,009
P5	0,000
P6	0,002
P7	0,011
P8	0,018
P9	0,058
P10	0,002
P11	0,000
P12	0,000
P13	0,002
P14	0,001
P15	0,123
P16	0,001
P17	0,002
P18	0,005
P19	0,001
P20	0,005
P21	0,003

Fuente: elaboración propia

Tabla 4
Ranking de discriminación

Pregunta	Ranking	Discriminación (coeficiente)
P9	1	0.3859
P15	2	0.4691
P8	3	0.5355
P7	4	0.6210
P1	5	0.6337
P20	6	0.6775
P17	7	0.8128
P18	8	0.8641
P2	9	0.8644
P3	10	0.9201
P21	11	0.9603
P19	12	0.9773
P13	13	1.0144
P5	14	1.0365
P6	15	1.0967
P16	16	1.2635
P10	17	1.3312
P14	18	1.4014
P12	19	1.4899
P4	20	2.1243
P11	21	13.7242

Fuente: elaboración propia

El índice de discriminación de un ítem indica en qué medida el ítem diferencia a los y las estudiantes, ya que cada uno aporta información en diferentes niveles de la escala de conocimientos de estadística. Por tanto, se pueden elegir del banco de preguntas los ítems que presenten una discriminación máxima en el nivel del rasgo asociado. Si bien se puede determinar un punto de corte, que depende del área de interés en la que se aplican estos modelos, indicados por expertos, también pueden, estos índices de discriminación ser considerados en términos comparativos, seleccionando aquellos que poseen mayores valores.

En cuanto al análisis de los coeficientes que miden discriminación, la Tabla 4 presenta de manera ordenada que preguntas discriminan menos y que preguntas discriminan más. Dejando de lado P9 y P15 que resultaron ser no significativas, se observa que P11 es la que más discrimina, es decir que es la que más aporta al

constructo (sobre el tema de selección del estadístico y su distribución de probabilidad para una prueba de hipótesis de la proporción poblacional), mientras que P8 es la que menos aporta (sobre el tema determinación del tamaño de muestra para estimar proporción).

De esta manera, en función de los objetivos planteados, se validó el cuestionario en cuanto al aporte que cada pregunta tiene sobre esos conocimientos, donde se detectaron preguntas que no aportan significativamente al constructo, por un lado y preguntas que sí aportan al conocimiento estadístico con mayores o menores grados de dificultad y de discriminación.

Esto permitió identificar fortalezas y debilidades en el instrumento de evaluación, con la finalidad de mejorar el cuestionario, o elaborar uno diferente. Aquellas preguntas que aportan al constructo, que poseen dificultad media y que son las que más discriminan entre el estudiantado se retienen en el cuestionario, ya que permiten no solo obtener los conocimientos sino desarrollar otras capacidades que también deben aprenderse, mientras que aquellas preguntas no significativas o muy fáciles y que no discriminan, pueden ser reemplazadas con la finalidad de obtener mejores instrumentos de evaluación.

En este caso particular, se elaboraron otras evaluaciones reemplazando las preguntas no significativas, las que a juicio de expertos resultaron muy fáciles de responder sin aportar al constructo y las que tenían menores puntajes en sus índices de discriminación, generando así nuevos instrumentos.

Consideraciones finales

Es de destacar la importancia de la evaluación en entornos virtuales con instrumentos idóneos. Del análisis exhaustivo del instrumento aplicado se observan algunas falencias importantes en el diseño de los mismos, como son preguntas no significativas para medir el aprendizaje y el nivel (ranking) en que cada una discrimina en el conocimiento.

De esta manera se presenta una herramienta, entre otras que permite mejorar los cuestionarios con los que se examina al estudiantado y que refleje el aprendizaje adquirido.

El hecho que los cuestionarios no evalúen lo que se pretende evaluar puede tener múltiples causas, que pueden ser detectadas a través de la evaluación de estos instrumentos y así mejorarlos, por lo que es necesario dar a conocer estos métodos y transmitir este tipo de análisis al cuerpo docente de la institución para revisar las prácticas de enseñanza y traducirlas en una evaluación de calidad.

En una segunda etapa se realizaron los ajustes al cuestionario con la finalidad de lograr una mejora continua en los instrumentos que se traducen en el proceso de evaluación, logrando evaluar los aprendizajes, y optimizar los exámenes.

Luego de sucesivos ajustes y modificaciones, se pueden hacer pruebas personalizadas, adaptadas y a medida, a fin de inferir en cada estudiante el verdadero valor del rasgo latente de manera más exacta.

La experiencia realizada en el curso de Estadística de la Facultad de Ciencias Económicas de la Universidad Nacional de Córdoba ha sido positiva, ya que ha permitido a través de la validación de cuestionarios, generar instrumentos de evaluación que contribuyen a la mejora del proceso evaluativo.

Referencias

- Álvarez, H., Arias, E., Bergamaschi, A., López, A., Noli, A., Ortiz, M., Viteri, A. (2020). La educación en tiempos de coronavirus: Los sistemas educativos de América Latina y el Caribe ante COVID-19. Banco Interamericano de Desarrollo.
- Anijovich, R. y Capelletti, G. (2017). *La evaluación como oportunidad*. Paidós.
- Anijovich, R., y Cappelletti, G. (2020). La retroalimentación formativa: Una oportunidad para mejorar los aprendizajes y la enseñanza. *Revista Docencia Universitaria*, 21(1), 81-96.
- Attorresi, H. F., Lozzia, G. S., Abal, F. J. P., Galibert, M. S., y Aguerri, M. E. (2009). Teoría de Respuesta al Ítem. Conceptos básicos y aplicaciones para la medición de constructos psicológicos. *Revista Argentina de Clínica Psicológica*, 18(2), 179-188.
- Beltramino, L. (2020). *Aprendizajes y prácticas educativas en las actuales condiciones de época: COVID-19*. 1º ed. Córdoba – Universidad Nacional de Córdoba. Recuperado de: <https://rdu.unc.edu.ar/handle/11086/19283>.
- Blanco, M., y 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., y Lautenbach, G. (2011). Electronic assessment in higher education. *Educational Studies*, 37(5), 503-512.
- Caro, N. y Ahumada, M. (2018a). Fiabilidad de los cuestionarios como instrumentos de evaluación de estudiantes de Estadística en un entorno virtual de enseñanza. [Sesión de Congreso]. XX Encuentro Internacional Virtual Educa, Salvador, Bahía, Brasil. <https://encuentros.virtualeduca.red/storage/ponencias/bahia2018/jQSGPO8KBV5EW3JMFUyp27ac2mDgygqxP4bsPTgy.pdf>
- Carrera, F., Vaquero, E., y Balsells, M. (2011). Instrumento de Evaluación de competencias digitales para adolescentes en riesgo social. *EDUTEC-E. Revista Electrónica de Tecnología Educativa* (35), 1-25.
- Cassany, D. y Ayala, G. (2008). Nativos e inmigrantes digitales en la escuela. *Participación educativa: revista del Consejo Escolar del Estado*. 2008; 9 (4): 57–75.
- Casal, S. M. S. (2010). Cuestionario de evaluación de la calidad de los cursos virtuales de la UNED. *Revista de Educación a Distancia*, (25).

- Conde, M. J. R. (2005). Aplicación de las TIC a la evaluación de alumnos universitarios. *Teoría de la Educación. Educación y Cultura en la Sociedad de la Información*, 6(2). Recuperado de: <https://www.redalyc.org/pdf/2010/201021055002.pdf>
- Flores, A. C., y Lever, J. P. (2008). El proceso de admisión como predictor del rendimiento académico en la educación superior. *Universitas psychologica*, 199-215.
- Crews, T. B., y Curtis, D. F. (2011). Online course evaluations: Faculty perspective and strategies for improved response rates. *Assessment y Evaluation in Higher Education*, 36(7), 865-878.
- Daly, C., Pachler, N., Mor, Y., y Mellar, H. (2010). Exploring formative e-assessment: using case stories and design patterns. *Assessment y Evaluation in Higher Education*, 35(5), 619- 636.
- de Kohan, N. C. (2004). Teoría de respuesta al ítem: supuestos básicos. *Revista Evaluar*, 4(1).
- Dussel, I., Ferrante, P. y Pulfer, D. (2020). La educación de pasado mañana. Notas sobre la marcha. *Análisis Carolina* (41), 1.
- Embretson, S y. Reise, S. (2000). *Item Response Theory for Psychologists*. Mahwah, NJ: Lawrence Erlbaum.
- Ferrão, M. (2010). E-assessment within the Bologna paradigm: evidence from Portugal. *Assessment y Evaluation in Higher Education*, 35(7), 819-830.
- González-Montesinos, M. y Backhoff, E. (2010). Validación de un cuestionario de contexto para evaluar sistemas educativos con Modelos de Ecuaciones Estructurales. *RELIEVE*, 16(2), 1-17.
- Graff, M. (2003). Cognitive style and attitudes towards using online learning and assessment methods. *Electronic Journal of e-learning*, 1(1), 21-28.
- Hamodi, C., López Pastor, V. M., y López Pastor, A. T. (2015). Medios, técnicas e instrumentos de evaluación formativa y compartida del aprendizaje en educación superior. *Perfiles educativos*, 37(147), 146-161.
- Iglesias Rodríguez, A., Olmos Migueláñez, S., Torrecilla Sánchez, E. M., y 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*.
- Llarena, M., y Paparo, M. (2006). Propuesta de una metodología de seguimiento y evaluación de cursos a distancia. *Revista Iberoamericana de Educación*, 37(4), 1-11.
- Llorente Cejudo, M. (2007). Moodle como entorno virtual de formación al alcance de todos. *Comunicar*, 15(28).
- Martín Galán, B., y 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 adistancia*, 15(1).

- Medina Moreno, P. y Aguirre, M. (2014). Predictibilidad de las notas de enseñanza media según establecimiento de origen sobre el rendimiento académico en estudiantes Odontología. *Revista Educación Médica Superior*. 28:65-73
- Medina-Rebollo, D., Grimaldi-Puyana, M., Del Río-Rama, M., y Sánchez-Oliver, A. (2019). Elaboración y validación de un cuestionario para evaluar el uso didáctico de la plataforma de teleformación Moodle en el contexto universitario. *Revista ESPACIOS*, 40(09).
- Ministerio de Educación de Chile (2013). *Matriz de Habilidades TIC para el aprendizaje*. Centro de Educación y Tecnología, Chile. Recuperado de: http://www.eduteka.org/pdfdir/CHILE_Matriz_Habilidades_TIC_para_el_Aprendizaje.pdf.
- Ministerio de Educación de Argentina (2020). Evaluación Nacional del Proceso de Continuidad Pedagógica. Informes preliminares. Evaluación e información educativa. <https://www.argentina.gob.ar/educacion/evaluacioninformacion-educativa/evaluacion-nacional-del-proceso-de-continuidad-pedagogica>.
- Muñiz Fernández, J. (2010). Las teorías de los tests: teoría clásica y teoría de respuesta a los ítems. *Papeles del Psicólogo: Revista del Colegio Oficial de Psicólogos*.
- Nazario, P. F., Ferreira, L., Both, J., y Vieira, J. L. L. (2021). Qualidade avaliativa do Test of Gross Motor Development 2: Uma investigação com base na Teoria de Resposta ao Item (Calidad evaluativa de la prueba de Desarrollo Motor Grueso 2: Una investigación basada en la Teoría de Respuesta al Ítem). *Retos*, 42, 696-704.
- Organista-Sandoval, J., Lavigne, G., Serrano-Santoyo, A., y Sandoval-Silva, M. (2017). Desarrollo de un cuestionario para estimar las habilidades digitales de estudiantes universitarios/Development of a questionnaire to estimate undergraduate digital skills. *Revista Complutense de Educación*, 28(1), 325.
- Rodríguez Castro, M y Gómez López, V. (2010). Indicadores al ingreso en la carrera de medicina y su relación con el rendimiento. *Revista Educación Superior*. 39,43-50.
- Rojas, M. (2013). Validez predictiva de los componentes del promedio de admisión a la universidad de costa rica utilizando el género y el tipo de colegio como variables control. *Revista Electrónica Actualidades Investigativas en Educación* 13(1).
- Ryan, S., Scott, B., Freeman, H., y Patel, D. (2013). *The virtual university: The internet and resource-based learning*. Routledge.
- Stegmann, C., Huertas, M. A., Juan, Á. A., y 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.
- Skrondal, A., and S. Rabe-Hesketh. 2004. *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. Boca Raton, FL: Chapman y Hall/CRC.

- Tirado-Morueta, R., y Hernando-Gómez, Á. (2011). Comunidades de aprendizaje a través de plataformas de teleformación. *RIED. Revista iberoamericana de educación a distancia*, 14(2), 99-120.
- Tomatis, M. C., Burrone, M. S., Romero, D., Novella, M. L., Olivero, M., Antuña, A. M., y Fernández, A. (2016). Validez predictiva del examen de ingreso a la carrera de medicina de la Facultad de Ciencias Médicas (UNC). *Revista de Educación*, (9), 357-367.