

ÍNDICES DE LA CALIDAD DE LOS INSTRUMENTOS DE EVALUACIÓN

Roberto Salomón Villamarín Guevara*

*rvillamarin@unach.edu.ec

Profesor Facultad de Ciencias de la Educación
Universidad Nacional de Chimborazo.

Riobamba – Ecuador

científica. Lima, Perú: Universidad Nacional Mayor de San Marcos.

Ramírez, R. (2013). Tercera ola de transformación de la Educación Superior en Ecuador. Senescyt, 53.

UNACH, H. C. U. de la U. N. de C. (2016). Reglamento de Régimen Académico. Editorial Universitaria.

Urquiza, A. (2005). Como realizar la tesis o una investigación (Vol. 1). Gráficas Riobamba.

Resumen

La Evaluación educativa es quizás el componente más importante de todo el proceso educativo, al cual no se le ha brindado la importancia que éste amerita, pues a partir de sus resultados se derivan aspectos relevantes que determinan el éxito o fracaso de todo el proceso educativo, pues sabemos que a partir de dichos resultados se define la promoción o no del estudiante, y es a su vez considerada como un indicador de la calidad de la educación o del proceso educativo.

Esta investigación busca determinar cuáles son los valores de los indicadores de calidad de los instrumentos de evaluación (índice de confiabilidad, índice de discriminación y grado de dificultad) aplicados a los estudiantes de la Facultad de Ciencias de la Educación, Humanas y Tecnologías de la Universidad Nacional de Chimborazo, en los periodos académicos correspondientes desde abril 2015 hasta agosto 2016. La muestra seleccionada es aleatoria, el diseño de la investigación es no experimental de tipo longitudinal. La investigación concluye con cifras que llaman la atención sobre la calidad de los instrumentos de evaluación, puesto que éstos no están garantizando un aprendizaje de calidad a los estudiantes de la Facultad de Ciencias de la Educación de la Universidad Nacional de Chimborazo.

Palabras claves: calidad, evaluación, indicadores, índices, correlación.

1. Abstract

The Educational Evaluation is perhaps the most important component of the whole educational process, which has not been given the importance it deserves, because from its results, relevant aspects are derived that determine the success or failure of the whole educational process, since we know that from these results it's defined the

promotion of the student, and at the same time it's considered as an indicator of the quality of education or the educational process.

This research aims to determine the quality of the indicators of the evaluation instruments (reliability index, discrimination index and degree of difficulty) applied to the students of the Faculty of Educational Sciences, Humanities and Technologies of the Universidad Nacional de Chimborazo in the corresponding academic periods from April 2015 to August 2016. The sample selected is random, the research design is non-experimental of longitudinal type.

The research concludes with figures that call attention to the quality of the evaluation instruments, since they are not guaranteeing a quality learning to the students of the Faculty of Education Sciences of the National University of Chimborazo.

Key words: quality, evaluation, indicators, indexes, correlation.

2. Introducción

La Evaluación educativa constituye uno de los pilares del proceso de educativo en todos sus niveles, siendo particularmente importante en la educación superior, pues a partir de los resultados que de la evaluación se deriven se deciden aspectos de vital importancia tanto para el estudiante, el profesor, la institución educativa, el entorno familiar y finalmente para todo el sistema educativo (Aranda, Pastor, Oliva, & Romero, 2013).

La Universidad Nacional de Chimborazo (Ecuador) cuenta con el Modelo Pedagógico denominado: Aproximación Epistemológica-Metodológica, desde la Complejidad, para el desarrollo integral de la persona, rearticulando la Investigación, Formación y Vinculación, en el cual se establecen los lineamientos generales

de los aspectos que se deben considerar para la evaluación, sin embargo no se llega a determinar la directivas para la elaboración de los instrumentos de evaluación, de la misma forma que no se establecen políticas claras para garantizar la calidad de los mismos (Loza et al., 2014).

Este modelo manifiesta que la Evaluación debe ser considerada como parte del proceso [educativo], y que a la misma se la debe organizar a través de proyectos de investigación inter-disciplinarios, desarrollándose mediante procesos de autoevaluación, coevaluación y heteroevaluación de forma natural, en el mismo ambiente de aprendizaje y cuyo objetivo sea apuntalar las debilidades cognitivas, de cara a la toma de decisiones oportunas para tutorizar a los estudiantes de manera individual, atendiendo a sus particulares formas y estilos de aprendizaje. Así mismo, el modelo establece algunos criterios sobre: ¿cómo?, ¿cuándo? y ¿con qué evaluar?. Es en este último aspecto que esta investigación a centra sus esfuerzos, pues tiene como objetivo determinar los índices de calidad de los instrumentos aplicados en la evaluación a los estudiantes.(Alvarez & Miranda, 2010)

En este contexto resulta importante para todos los actores del sistema educativo determinar si los instrumentos de Evaluación Educativa aplicados por los docentes de la Facultad de Ciencias de la Educación, Humanas y Tecnologías cumplen con criterios de calidad, de tal manera que puedan garantizar resultados objetivos y veraces, que permitan retroalimentar todo el proceso educativo, que finalmente es el objetivo superior de la evaluación, según lo sostiene (Ramírez, 2013).

Para su desarrollo se han recopilado los resultados (datos) de las evaluaciones de todas las carreras de la Facultad de Ciencias de la Educación a partir del semestre abril 2015, hasta agosto 2016 (3 semestres académicos), las

mismas que fueron registradas en la plataforma educativa virtual institucional.

3. Metodología

La presente investigación es de carácter longitudinal, correlacional. La población está determinada por las evaluaciones registradas en la plataforma b-learning de la Universidad Nacional de Chimborazo, durante los periodos de abril 2015 hasta febrero 2016. La Facultad cuenta con 10 carreras, de las cuales con 8 niveles de formación en cada una de ellas. Cada nivel de formación tiene 6 asignaturas. Por cada asignatura se debe registrar dos notas parciales (cada 8 semanas), por cada nota se debe considerar los componentes de la Evaluación que determina el Reglamento de Régimen Académico de la (UNACH, 2016), que son: Componente de docencia (actividades donde principalmente las ejecuta el docente), componente de Experimentación (actividades que desarrolla el estudiante, bajo la guía del docente) y, finalmente las actividades del componente de Trabajo Autónomo (que son actividades de aprendizaje que desarrolla autónomamente el estudiante). Dando una población total de más de 5000 registros, por lo que se considera una población infinita. Para la muestra, al ser una investigación correlacional, y acogiendo el criterio de (Lind, Marshall, & Wathen, 2013; Urquizo, 2005, Mejía, 2005), donde se considera para un estudio correlacional es conveniente tener una muestra grande, se ha seleccionado 1567 registros elegidos de manera aleatoria de las distintas carreras que conforman la Facultad de Ciencias de la Educación. Por cada instrumento de evaluación se calculó los indicadores de calidad (índice de confiabilidad y de discriminación, y el grado de dificultad) que son criterios para medir la calidad del instrumento de evaluación, según (Mejía, 2005); con la finalidad de determinar el grado de correlación que pudiera existir entre

éstos indicadores de calidad, con el rendimiento académico (promedio de las notas) de cada uno de ellos (Guamán, 2009; Mejía, 2005).

Para calcular la variable Grado de dificultad, índice de confiabilidad e índice de discriminación se utiliza el procedimiento descrito por (Mejía, 2005), cuyos resultados se interpretan utilizando para ellos las tablas de Küder – Richardson. A continuación, se describen cada una de las fórmulas utilizadas para el cálculo de cada una de las variables:

Grado de Dificultad

Para hallar el grado de dificultad de una prueba se la puede determinar numéricamente por la relación matemática entre promedio de los puntajes obtenidos y el puntaje máximo posible de alcanzarse en la prueba.

$$G_d = \frac{\bar{X}}{P_m} * 100\%$$

Fórmula 1.- Grado de Dificultad

Donde:

G_d = Grado de dificultad

\bar{x} = Promedio de los puntajes obtenidos

P_m = Puntaje máximo posible de obtenerse en la Prueba

A continuación se muestran los valores de Küder-Richardson para establecer el grado de dificultad del instrumento

Tabla 1.- Valores de Küder Richardson para el cálculo del Grado de Dificultad del instrumento.

VALOR	CRITERIO
81% o más	Muy fácil
61% - 80%	Relativamente fácil
51% - 60%	Dificultad adecuada
31% - 50%	Relativamente difícil
11% 30%	Difícil
11% o menos	Muy difícil

En esta variable de estudio se considera una categoría adecuada, aquella que corresponde a Dificultad adecuada y Relativamente difícil. Las categorías Muy Fácil, y Muy Difícil no se consideran adecuadas, puesto que la primera permite que sean promovidos estudiantes que aún no han desarrollado las destrezas necesarias del nivel de instrucción y por el otro lado, la categoría muy difícil constituiría una barrera a la promoción de los estudiantes que ya han adquirido las destrezas mínimas necesarias para aprobar el nivel.

Índice de Confiabilidad

La Confiabilidad se refiere al proceso de establecer cuan fiable, consistente, coherente o estable es el instrumento que se ha elaborado; se realiza para determinar si el instrumento inspira confianza (Guamán, 2009a). Se determina si es confiable o no, si al ser aplicado en repetidas ocasiones arroje los mismos resultados. La confiabilidad, es la seguridad, exactitud, precisión o consistencia que debe poseer una prueba, es decir se la concibe la consistencia o estabilidad de las medidas cuando el proceso de medición se repite (López, 2011; Prieto, 2010).

Cuanto más alto sea el coeficiente de correlación positiva entre las dos versiones de la prueba, la prueba será más confiable. Además está decir que el coeficiente de correlación perfecto es 1,00 y para asignar alto grado de confiabilidad a una prueba se debe esperar obtener coeficientes de, más o menos, 0.80; (Mejía, 2011), señala las siguientes:

- Técnica de la replicación o test - retest.
- Aplicación de dos versiones paralelas de una misma prueba.
- Aplicación de pruebas en mitades.
- Técnica de Aplicación en tiempo diferido.

Para calcular la confiabilidad del instrumento de

evaluación, desde el punto de vista numérico se aplica la siguiente fórmula:

$$C_f = \frac{n}{n - 1} \left[\frac{\bar{x}(n - \bar{x})}{n * \partial^2} \right]$$

Fórmula 2.- Cálculo de la Confiabilidad

A continuación se muestran los valores de Küder-Richardson para establecer el grado de confiabilidad del instrumento.

Tabla 2.- Valores de Küder Richardson para el cálculo del índice de Confiabilidad del instrumento

VALOR	CRITERIO
0.53 o menos	Confiabilidad nula
0.54 – 0,59	Confiabilidad Baja
0,60 – 0,65	Confiable
0,66 – 0,71	Muy Confiable
0,72 – 0,99	Excelente Confiabilidad
1.0	Confiabilidad Perfecta

Las categorías adecuadas para esta variable se consideran desde confiable hasta la confiabilidad perfecta, las categorías baja y nula; no se consideran adecuadas puesto que indican que el instrumento no garantiza un mínimo de confiabilidad.

Índice de Discriminación del Instrumento

El índice de discriminación está determinado por la capacidad del instrumento de discernir entre el grupo que conoce los contenidos evaluados, del grupo que no los conoce.

Para hallar el índice de discriminación de una prueba se forman dos grupos de estudiantes en función de sus respuestas correctas. El grupo superior está constituido por los estudiantes que respondieron mejor y el grupo inferior por los que alcanzaron puntajes más bajos (Mejía, 2005).

El índice de discriminación se establece analizando las diferencias de las respuestas correctas del grupo superior y las respuestas correctas del grupo inferior. Se supone que los

estudiantes del grupo superior tendrán mejores respuestas que los del grupo inferior y el índice de discriminación será más alto cuanto mayor sea la diferencia de las puntuaciones correctas del grupo superior con respecto a las del grupo inferior.

$$I_d = \frac{pm_s - pm_i}{Pm} * 100\%$$

Fórmula 3.- Cálculo del Índice de Discriminación

Donde:

- I_d = • Índice de discriminación
- pm_m = • Puntaje máximo posible de respuestas correctas del grupo superior.
- pm_i = • Puntaje máximo posible de respuestas correctas del grupo inferior.
- Pm = • Puntaje máximo posible de respuestas correctas del grupo inferior

A continuación se muestran los valores de Küder-Richardson para establecer el Índice de Discriminación del instrumento.

Tabla 3. Valores de Küder Richardson para la variable Índice de Discriminación

VALOR	CRITERIO
40% o más	Buen índice
30% - 39%	Razonable
20% 29%	Regular
10% o meno	Deficiente

En la variable índice de discriminación, se considera categorías aceptables la que corresponde a “Buen índice y Razonable”, puesto que éstas permiten discriminar con claridad, entre quienes dominan los conocimientos y quienes no lo han hecho aún.

4. Resultados

Análisis del comportamiento de la Grado de Dificultad.

Las variables en estudio, grado de dificultad, índice de confiabilidad, e índice de discriminación se ha clasificado según las tablas de Küder-Richardson explicadas en la sección Grado de Dificultad cuyo resultado se muestra en el cuadro y su representación gráfica se puede apreciar en la figura de la misma manera en la gráfica; se puede apreciar la variación del rendimiento académico respecto de cada una de las categorías del Grado de dificultad del instrumento de Evaluación.

Tabla 4.- Distribución de los datos con respecto de la variable Grado de dificultad

Categoría	F. A	F. %
Muy Fácil	628	40,1
Relativamente Fácil	547	34,9
Dificultad Adecuada	162	10,3
Relativamente difícil	177	11,3
Difícil	48	3,1
Muy difícil	5	0,3
TOTAL	1567	100

Fuente: Villamarín R. (2017)

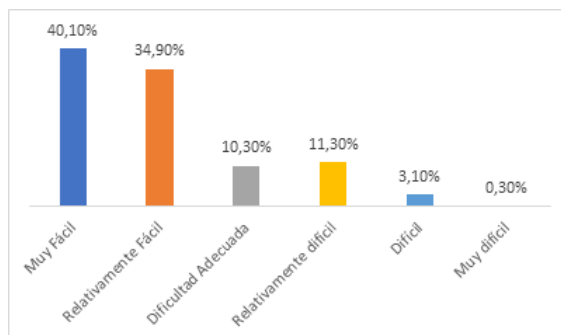


Ilustración 1.- Distribución de los datos con respecto de la variable Grado de dificultad

Análisis del comportamiento de la Grado de Dificultad.

Al analizar los datos, en la variable Índice de Confiabilidad, éstos se distribuyen de la siguiente manera:

Tabla 5.- Distribución de los datos, respecto de la variable índice de confiabilidad

Categoría	F. A	F. %
Nula	216	13,8
Baja	37	2,4
Confiable	45	2,9
Muy Confiable	54	3,4
Excelente Confiabilidad	883	56,3
Perfecta	224	14,3
No se puede determinar	108	6,9

Fuente: Villamarín R. (2017)

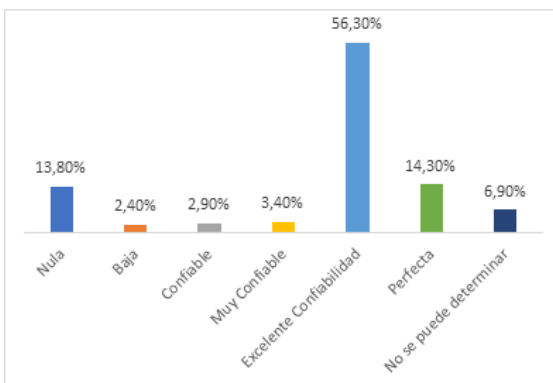


Ilustración 2.- Distribución de los datos, respecto de la variable índice de confiabilidad

Análisis del comportamiento de la variable índice de discriminación.

La variable, Índice de discriminación tiene el siguiente comportamiento:

Tabla 6.- Distribución de los datos respecto de la variable índice de discriminación

Categoría	F. A	F. %
Bueno	229	14,6
Razonable	134	8,6
Regular	242	15,4
Deficiente	962	61,4
Total	1567	100,0

Fuente: Villamarín R. (2017)

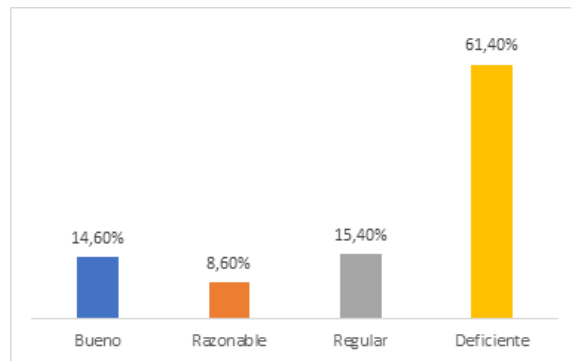


Ilustración 3.- Distribución de los datos, con respecto de las categorías del índice de discriminación

5. Discusión

La distribución de los datos de la variable Grado de dificultad muestra que un 75% de los datos se encuentran en la categoría muy fácil y relativamente difícil; y sólo un 10% de los datos se encuentran en la categoría de dificultad adecuada.

Al analizar las categorías del índice de confiabilidad se puede apreciar que un porcentaje significativo de los datos está en la categoría excelente, solo un 14.30% se encuentra en la categoría perfecta. Hay dos categorías que llaman la atención, la categoría nula con un 13,8% del total de los datos, y la categoría en la que no se puede determinar el índice de confiabilidad, esto ocurre en los datos donde todos los estudiantes tienen una misma nota, generalmente la nota máxima (10 de 10 puntos posibles), esto hacer que el valor de la desviación estándar sea igual a cero, dando como resultado una división para cero, lo cual desde el punto de vista matemático, arroja un error. Las categorías confiable y muy confiable son muy bajas en comparación con las restantes categorías.

Del análisis de la variable índice de discriminación se puede apreciar que un 61% de los datos se encuentran en la categoría deficiente, y solo un 23,2% se encuentran en la categoría bueno y

razonable, lo cual indica que los instrumentos aplicados no permiten realmente distinguir entre quienes dominan los conocimientos y quienes no lo hace.

6. Conclusiones

- 1) Del análisis de la variable Grado de dificultad se puede deducir que dichos instrumentos no están garantizando un nivel adecuado de aprendizaje, dado que el 75% de los datos se encuentra en las categorías, Muy Fácil y relativamente Fácil. Esto sugiere que hay estudiantes que están siendo promovidos al siguiente nivel de educación sin haber adquirido las destrezas necesarias, que permitan garantizar un nivel de conocimiento o de adquisición de destrezas adecuadas para dicho nivel.
- 2) En la variable Grado de confiabilidad se puede apreciar que un 68,6% tiene una confiabilidad excelente y perfecta, lo cual dice que los instrumentos cumplen con dicha norma; sin embargo, existen dos valores que sumados son preocupantes (nula = 13,8% y 6,9% que no se puede determinar su confiabilidad, 20,7% del total de datos), indica que aproximadamente en una de cada 5 instrumentos de evaluación no existe confianza en los resultados que arrojan dichos instrumentos.
- 3) Al analizar los datos cuyo índice de confiabilidad "No se puede determinar", vemos que en dichos registros todos los estudiantes tienen la misma nota, lo que sugiere que los docentes no están aplicando adecuadamente las normas y los principios de la evaluación, puesto que se está asignando a todos los estudiantes una misma nota, sin un análisis previo de la calidad de la tarea presentada.
- 4) Estudiando la variable índice de discriminación, se puede apreciar que el

61,4% de los datos tienen un índice de discriminación deficiente, lo cual indica claramente que los instrumentos de evaluación no permiten discriminar entre quienes saben y no, por lo que se puede deducir una vez más que hay estudiantes que están siendo promovidos, sin que cumplan con los requisitos necesarios para hacerlo, y; muy probablemente otros estudiantes que están siendo reprobados cuando deberían ser promovidos al siguiente nivel.

- 5) Del análisis global de los datos vemos que es necesaria y urgente emprender en el establecimiento de políticas educativas que permitan mejorar la práctica evaluativa con el fin de garantizar una calidad en el servicio de educación superior que ofrece la Universidad Nacional de Chimborazo.

7. Bibliografía

- Alvarez, R. A., & Miranda, M. O. (2010). Fiabilidad y validez en la evaluación docente universitaria. *Revista Electrónica« Actualidades Investigativas en Educación»*, 10(1), 1–38.
- Aranda, A. F., Pastor, V. M. L., Oliva, F. J. C., & Romero, R. (2013). La evaluación formativa en docencia universitaria y el rendimiento académico del alumnado. *Aula abierta*, 41(2), 23-34.
- Castejón, J. L. (2014). *Aprendizaje y Rendimiento Académico*. Editorial Club Universitaria. Alicante España.
- Guamán, G. (2009). *La Evaluación del Aprendizaje: Del Modelo Criterial, hacia una propuesta Alternativa-crítica referida a las competencias profesionales*. Primera parte (Vol. 1).
- Lind, D., Marshall, W., & Wathen, S. (2013).

Estadística aplicada a los negocios.
Mexico, Mexico DF Editorial Mac Hill.

Loza, C., Murillo, M., Tenezaca, R., Montalvo, C., García, C., & Piñas, B. (2014).
Modelo Pedagógico de la Universidad Nacional de Chimborazo: Aproximación Epsitemológico - Metodológica para el desarrollo integral de la persona rearticulando la investigación, formación, y vinculación.

Mejía, E. (2005). Metodología de la investigación científica. Lima, Perú: Universidad Nacional Mayor de San Marcos.

Ramírez, R. (2013). Tercera ola de transformación de la Educación Superior en Ecuador. Senescyt, 53.

UNACH, H. C. U. de la U. N. de C. (2016). Reglamento de Régimen Académico. Editorial Universitaria.

Urquiza, A. (2005). Como realizar la tesis o una investigación (Vol. 1). Gráficas Riobamba.