Universidad Nacional Mayor de San Marcos
Universidad del Perú. Decana de América

Dirección General de Estudios de Posgrado
Facultad de Odontología
Unidad de Posgrado

Comparación de la calidad de los instrumentos de evaluación de las asignaturas del área de Ciencias Básicas de la Facultad de Odontología de la Universidad Nacional Mayor de San Marcos

TESIS

Para optar el Grado Académico de Magíster en Estomatología

AUTOR

William CÁRDENAS SILVA

ASESOR

Dr. Carlos Alberto SUÁREZ CANLLA

Lima, Perú

2021



Reconocimiento - No Comercial - Compartir Igual - Sin restricciones adicionales

https://creativecommons.org/licenses/by-nc-sa/4.0/

Usted puede distribuir, remezclar, retocar, y crear a partir del documento original de modo no comercial, siempre y cuando se dé crédito al autor del documento y se licencien las nuevas creaciones bajo las mismas condiciones. No se permite aplicar términos legales o medidas tecnológicas que restrinjan legalmente a otros a hacer cualquier cosa que permita esta licencia.

Referencia bibliográfica

Cárdenas W. Comparación de la calidad de los instrumentos de evaluación de las asignaturas del área de Ciencias Básicas de la Facultad de Odontología de la Universidad Nacional Mayor de San Marcos [Tesis de maestría]. Lima: Universidad Nacional Mayor de San Marcos, Facultad de Odontología, Unidad de Posgrado; 2021.

Metadatos complementarios

Datos de autor	
Nombres y apellidos	William Cárdenas Silva.
Tipo de documento de identidad	DNI
Número de documento de identidad	10191752
URL de ORCID	https://orcid.org/0000-0002-0177-007X
Datos de asesor	
Nombres y apellidos	Carlos Alberto Suárez Canlla
Tipo de documento de identidad	DNI
Número de documento de identidad	21552481
URL de ORCID	https://orcid.org/0000-0003-4719-1409
Datos del jurado	
Presidente del jurado	
Nombres y apellidos	Antonia Florencia Castro Rodríguez
Tipo de documento	DNI
Número de documento de identidad	07215663.
Miembro del jurado 1	
Nombres y apellidos	Margot Margarita Gutiérrez Ilave
Tipo de documento	DNI
Número de documento de identidad	09306432
Miembro del jurado 2	
Nombres y apellidos	Marieta Petkova Gueorguieva De Rodríguez
Tipo de documento	DNI
Número de documento de identidad	09677151

Datos de investigación	
Línea de investigación	No aplica
Grupo de investigación	No aplica
Agencia de financiamiento	Sin financiamiento
Ubicación geográfica de la investigación	Universidad Nacional Mayor de San Marcos Latitud: - 11.85 Longitud: - 76.45
Año o rango de años en que se realizó la investigación	Marzo 2018 – Mayo 2021
URL de disciplinas OCDE	1. Educación general https://purl.org/pe-repo/ocde/ford#5.03.01 2. Odontología, Cirugía oral, Medicina oral https://purl.org/pe-repo/ocde/ford#3.02.14

Facultad de Odontología
Universidad Nacional Mayor de San Marcos
UNIDAD DE POSGRADO

"Año del Bicentenario del Perú: 200 años de Independencia""

ACTA N° 009-FO-UPG-2021

ACTA DE SUSTENTACIÓN DE TESIS MODALIDAD VIRTUAL
PARA OPTAR EL GRADO DE MAGÍSTER EN ESTOMATOLOGÍA
Autorizado con RR N°01357-R-20

En la ciudad de Lima, a los 04 días del mes de junio del año dos mil veintiuno, siendo las 10:00 horas, se reunieron los miembros del Jurado para llevar a cabo la sustentación de la Tesis titulada: "**COMPARACIÓN DE LA CALIDAD DE LOS INSTRUMENTOS DE EVALUACIÓN DE LAS ASIGNATURAS DEL ÁREA DE CIENCIAS BÁSICAS DE LA FACULTAD DE ODONTOLOGÍA DE LA UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS",** presentada por el Bach. don **WILLIAM CÁRDENAS SILVA** para optar el Grado de Magíster en Estomatología.

Concluida la exposición, se procedió a la evaluación correspondiente, después de la cual obtuvo la siguiente calificación:

EXCELENTE	**19**	**DIECINUEVE**
Escala	Número	Letras

A continuación, la Presidenta del Jurado, en virtud de los resultados favorables, recomienda que la Facultad de Odontología proponga que la Universidad le otorgue al Bach. don **WILLIAM CÁRDENAS SILVA** el Grado de Magíster en Estomatología.

Se expide la presente acta en cuatro originales y siendo las 11:10 horas, se da por concluido el acto académico de sustentación.

UNMSM Firmado digitalmente por CASTRO RODRIGUEZ Antonia Florencia FAU 20148092282 soft
Motivo: Soy el autor del documento
Fecha: 04.06.2021 11:27:32 -05:00

Dra. Antonia Florencia Castro Rodríguez
Presidenta

UNMSM Firmado digitalmente por GUTIERREZ ILAVE Margot Margarita FAU 20148092282 soft
Motivo: Soy el autor del documento
Fecha: 04.06.2021 11:42:38 -05:00

Mg. Margot Margarita Gutiérrez Ilave
Miembro

UNMSM Firmado digitalmente por PETKOVA GUEORGUIEVA DE RODRIGUEZ Marieta FAU 20148092282 soft
Motivo: Soy el autor del documento
Fecha: 04.06.2021 11:48:12 -05:00

Mg. Marieta Petkova Gueorguieva De Rodríguez
Miembro

UNMSM Firmado digitalmente por SUAREZ CANLLA Carlos Alberto FAU 20148092282 soft
Motivo: Soy el autor del documento
Fecha: 04.06.2021 12:09:23 -05:00

Dr. Carlos Alberto Suárez Canlla
Miembro Asesor

Escala de calificación

- Excelente 20, 19
- Muy bueno 18, 17
- Bueno 16, 15
- Aprobado
- Desaprobado 13 o menos

❖**RECOMENDACIONES**

_____

Datos de la plataforma virtual institucional del acto de sustentación:

https:

ID:

Grabación archivada en:

- **PÚBLICO ASISTENTE:** (Nombre, apellido y DNI)

A mis padres, Juan y Gudelia, por haberme dado la vida, por cuidarme durante mi niñez, por sus consejos en mi juventud y por su amistad al llegar la adultez.

A mis hermanos, Esther, Juan y Alina, por el cariño y apoyo que siempre encuentro en ellos.

A Katia, mi querida esposa, por la aventura de andar juntos; a Tiago y Vasco, mis queridos hijos.

AGRADECIMIENTOS

Al Doctor Justiniano Sotomayor Camayo, por haberme motivado a seguir aprendiendo a través de la docencia.

Al Doctor Hugo Caballero Cornejo, egregio representante de la Odontología Forense en el Perú, por su desinteresado apoyo en la plasmación de esta tesis.

Al Doctor Daniel Blanco Victorio, por su apoyo y orientación en el análisis estadístico de la presente investigación.

A mi asesor, el Doctor Carlos Suárez Canlla, por su apoyo incondicional.

ÍNDICE GENERAL

	Pág.
Carátula externa	i
Carátula	ii
Veredicto de Jurado Examinador	iii
Dedicatoria	iv
Agradecimientos	v
Índice General	vi
Lista de Tablas	viii
Lista de Figuras	x
Resumen	xii
Abstract	xiii

I : INTRODUCCIÓN

1.1 Situación Problemática	1
1.2 Formulación del Problema	3
1.3 Justificación del estudio	3
1.4 Objetivos	5
1.4.1 Objetivo general	5
1.4.2 Objetivos específicos	5

II : MARCO TEÓRICO

2.1 Antecedentes	6
2.2 Bases teóricas	10
2.2.1 Educación	10
2.2.2 Educación Universitaria	12
2.2.3 Evaluación	13
2.2.4 Evaluación del Aprendizaje	14
2.2.5 Evaluación Universitaria	16
2.2.6 Teorías de los Test	19
2.2.7 Teoría Clásica de los Test	20

2.2.8 Teoría de Respuesta al Ítem 22
2.2.9 Calidad de los Instrumentos de Evaluación del Aprendizaje 24
2. 3 Definición de términos 29

III : HIPÓTESIS Y VARIABLES

3.1 Hipótesis 31
3.1.1 Hipótesis General 31
3.2.2 Hipótesis Específicas 31
3.2 Identificación de Variables 32
3.3. Operacionalización de Variables 33

IV : METODOLOGÍA

4.1 Tipo y Diseño de Investigación 35
4.2 Unidad de Análisis 35
4.3 Población de Estudio 36
4.4 Tamaño y Selección de muestra 36
4.5 Técnicas de Recolección de Datos 37
4.6 Análisis e Interpretación de la Información 39

V : RESULTADOS 40

VI : DISCUSIÓN 97

VII : CONCLUSIONES 100

VIII : RECOMENDACIONES 103

BIBLIOGRAFÍA 105

ANEXOS 109

LISTA DE TABLAS

		Pág
N° 01	Índice de dificultad de las preguntas de Anatomía General.	41
N° 02	Dificultad por categorías de Anatomía General.	42
N° 03	Global Índice de dificultad de Anatomía General.	43
N° 04	Índice de discriminación de Anatomía General.	45
N° 05	Calidad de las preguntas de Anatomía General.	46
N° 06	Global Índice de discriminación de Anatomía General.	47
N° 07	Correlación índice dificultad x índice discriminación de Anatomía General.	49
N° 08	Índice de dificultad de las preguntas de Bioquímica General.	50
N° 09	Dificultad por categorías de Bioquímica General.	51
N° 10	Global Índice de dificultad de Bioquímica General.	52
N° 11	Índice de discriminación de Bioquímica General.	54
N° 12	Calidad de las preguntas de Bioquímica General.	55
N° 13	Global Índice de discriminación de Bioquímica General.	56
N° 14	Correlación índice dificultad x índice discriminación de Bioquímica General.	57
N° 15	Índice de dificultad de las preguntas de Farmacología II.	59
N° 16	Dificultad por categorías de Farmacología II.	60
N° 17	Global Índice de dificultad de Farmacología II.	61
N° 18	Índice de discriminación de Farmacología II.	62
N° 19	Calidad de las preguntas de Farmacología II.	63
N° 20	Global Índice de discriminación de Farmacología.	64
N° 21	Correlación índice dificultad x índice discriminación de Farmacología.	65
N° 22	Índice de dificultad de las preguntas de Fisiología General.	67
N° 23	Dificultad por categorías de Fisiología General.	68

N° 24	Global Índice de dificultad de Fisiología.	69
N° 25	Índice de discriminación de Fisiología General.	70
N° 26	Calidad de las preguntas de Fisiología General.	71
N° 27	Global Índice de discriminación de Fisiología General.	72
N° 28	Correlación índice dificultad x índice discriminación de Fisiología.	73
N° 29	Índice de dificultad de las preguntas Microbiología General.	75
N° 30	Dificultad por categorías de Microbiología General.	76
N° 31	Global Índice de dificultad de Microbiología General.	77
N° 32	Índice de discriminación de Microbiología General.	79
N° 33	Calidad de las preguntas de Microbiología.	80
N° 34	Global Índice de discriminación de Microbiología.	81
N° 35	Correlación índice dificultad x índice discriminación de Microbiología.	82
N° 36	Índice de dificultad por asignatura.	84
N° 37	Global índice de discriminación por asignatura.	86
N° 38	Global confiabilidad por asignatura.	88

LISTA DE FIGURAS

		Pág
N° 01	Índice de dificultad de las preguntas de Anatomía General.	42
N° 02	Dificultad por categorías de Anatomía General.	43
N° 03	Global Índice de dificultad de Anatomía General.	44
N° 04	Índice de discriminación de Anatomía General.	46
N° 05	Calidad de las preguntas de Anatomía General.	47
N° 06	Global Índice de discriminación de Anatomía General.	48
N° 07	Correlación índice dificultad x índice discriminación de Anatomía General.	49
N° 08	Índice de dificultad de las preguntas de Bioquímica General.	51
N° 09	Dificultad por categorías de Bioquímica General.	52
N° 10	Global Índice de dificultad de Bioquímica General.	53
N° 11	Índice de discriminación de Bioquímica General.	55
N° 12	Calidad de las preguntas de Bioquímica General.	56
N° 13	Global Índice de discriminación de Bioquímica General.	57
N° 14	Correlación índice dificultad x índice discriminación de Bioquímica General.	58
N° 15	Índice de dificultad de las preguntas de Farmacología II.	60
N° 16	Dificultad por categorías de Farmacología II.	61
N° 17	Global Índice de dificultad de Farmacología II.	62
N° 18	Índice de discriminación de Farmacología II.	63
N° 19	Calidad de las preguntas de Farmacología II.	64
N° 20	Global Índice de discriminación de Farmacología.	65
N° 21	Correlación índice dificultad x índice discriminación de Farmacología.	66
N° 22	Índice de dificultad de las preguntas de Fisiología General.	68
N° 23	Dificultad por categorías de Fisiología General.	69
N° 24	Global Índice de dificultad de Fisiología.	70
N° 25	Índice de discriminación de Fisiología General.	71
N° 26	Calidad de las preguntas de Fisiología General.	72
N° 27	Global Índice de discriminación de Fisiología General.	73

N° 28	Correlación índice dificultad x índice discriminación de Fisiología.	75
N° 29	Índice de dificultad de las preguntas Microbiología General.	76
N° 30	Dificultad por categorías de Microbiología General.	77
N° 31	Global Índice de dificultad de Microbiología General.	78
N° 32	Índice de discriminación de Microbiología General.	80
N° 33	Calidad de las preguntas de Microbiología.	81
N° 34	Global Índice de discriminación de Microbiología.	82
N° 35	Correlación índice dificultad x índice discriminación de Microbiología.	83
N° 36	Índice de dificultad por asignatura.	85
N° 37	Global índice de discriminación por asignatura.	87
N° 38	Global confiabilidad por asignatura.	88

RESUMEN

Objetivo: el objetivo de la presente investigación fue comparar la calidad de los instrumentos de evaluación de las asignaturas de Anatomía General, Bioquímica General, Microbiología General, Fisiología General y Farmacología II del área de Ciencias Básicas de la Facultad de Odontología de la Universidad Nacional Mayor de San Marcos durante el año lectivo 2016. **Metodología:** el nivel de la investigación fue descriptivo comparativo; retrospectivo, transversal y observacional. La muestra estuvo conformada por 448 exámenes y 8420 preguntas. Los datos fueron sometidos a pruebas estadísticas como Kolmogorov – Smirnov, Shapiro – Wilk , Kruskall-Wallis y alfa de Cronbach, mediante los programas estadísticos SPSS y STAT. **Resultado:** al analizar los resultados se aprecia que los exámenes teóricos de las asignaturas, en general, tienen en promedio un índice de dificultad de 0.44, que los califican como exámenes medianamente fáciles; un índice de discriminación de 0.36, calificando a los exámenes de buena calidad y un índice de fiabilidad de 0.593 catalogando a los exámenes de moderada confiabilidad. **Conclusión:** se concluye que el examen de Anatomía General fue medianamente fácil, de buena discriminación y alta confiabilidad; el examen de Bioquímica General fue de dificultad media, de buena discriminación y moderada confiabilidad; el examen de Farmacología II fue medianamente fácil, de discriminación excelente y alta confiabilidad, el examen de Fisiología General fue altamente fácil, de buena discriminación y confiabilidad moderada; el examen de Microbiología General fue medianamente fácil, de regular discriminación y moderada confiabilidad. Las pruebas estadísticas señalan diferencias significativas en los índices de dificultad y discriminación entre los exámenes de los distintos cursos.

Palabras clave: Calidad. Instrumentos. Evaluación. Índice de dificultad. Índice de discriminación. Índice de confiabilidad.

ABSTRACT

Objective: The objective of the study was to compare the quality of the evaluation instruments of the subjects of General Anatomy, General Biochemistry, General Microbiology, General Physiology and Pharmacology II of the Basic Sciences area of the Faculty of Dentistry of the National University Mayor of San Marcos during the 2016 school year. **Methodology**: The research design was descriptive; the type of research was retrospective, transversal, observational and comparative. The sample consisted of 448 exams and 8420 questions analyzed. The data were subjected to statistical tests such as Kolmogorov - Smirnov, Shapiro - Wilk and Cronbach's Alpha, using the SPSS and STAT statistical programs. **Result**: The results show that all subjects in general, the difficulty index is 0.44, which qualifies as moderately easy, the discrimination index 0.36 qualifying for good quality exams and the reliability index 0.593 cataloging the moderate reliability exams. **Conclusion:** that the subjects General Anatomy, moderately easy difficulty, good quality discrimination and reliability 0.699; General Biochemistry, medium difficulty, good quality discrimination and reliability 0.548; Pharmacology, moderately easy difficulty, excellent discrimination and 0.627 reliability; Physiology, highly easy difficulty, good quality discrimination and reliability 0.587; Microbiology, moderately easy difficulty, regular quality discrimination and reliability 0.506.

Keywords: Quality. Instruments. Evaluation. Difficulty index. Difficulty discrimination. Reliability index.

CAPÍTULO 1: INTRODUCCIÓN

1.1 Situación problemática

El presente estudio está enmarcado dentro del sistema de evaluación que todo proceso educativo debe tener.

Delgado Santa María (2004) en su texto sobre la "Evaluación en la Educación Superior" menciona que, en el año de 1937, Henry Pieron sostenía que la Docimología era la ciencia que estudiaba los exámenes con el fin de corroborar lo conseguido durante el proceso del aprendizaje. Por otro lado, señalaba también, que la Docimología estudiaba, además, la manera de calificar, así como el comportamiento de todos los involucrados en la evaluación.

Páez y Di Carlo (2012) destacando el trasfondo psicológico, didáctico y tecnológico del proceso educativo, reiteran que la Docimología estudia a los exámenes y su relación con el nivel educativo, y agregan que la Docimología tiene también una función investigativa "dirigida a la elaboración, administración y valoración de estrategias e instrumentos evaluativos".

En el proceso educativo intervienen una serie de elementos interrelacionados entre sí como los docentes, los estudiantes, las instituciones y la sociedad; así como otros elementos más específicos como la infraestructura adecuada, las estrategias y las metodologías de enseñanza aprendizaje, y una cultura evaluativa. La interacción de estos elementos durante el proceso educativo

constituye una situación compleja, donde resaltan las contradicciones entre lo que se da y lo que debería ser haciendo que la evaluación, en cualquiera de sus niveles, sea una ardua tarea, llena de tensiones y muchas veces de frustraciones (Delgado, 2004).

En la actualidad, la mayoría de las universidades, siguiendo una tendencia mundial, plantean sus currículos basados en competencias, siendo la evaluación de los aprendizajes un aspecto bastante delicado para lo cual muchas veces el docente no está adecuadamente preparado.

La evaluación del aprendizaje es un mecanismo que tiene una doble finalidad: por un lado, ayuda a regular el proceso de enseñanza aprendizaje del estudiantey por otro lado, permite la acreditación y certificación del mismo como nuevo profesional, capacitado para buscar la solución a problemas que se presenten en su entorno. Para lograr este propósito, se debe buscar obtenerevidencias de la calidad, es decir, de la validez y la confiabilidad de los instrumentos de evaluación que se aplican (exámenes) en el desarrollo de las distintas asignaturas que conforman un plan de estudios de una determinada profesión.

En los primeros años de su formación, los estudiantes de la Carrera Profesional de Odontología de la Universidad Nacional Mayor de San Marcos (UNMSM) reciben clases de asignaturas biomédicas, consideradas dentro del área de Ciencias Básicas, las cuales tienen como finalidad brindar conocimientos y competencias que luego les permitan comprender el proceso Salud – Enfermedad, en especial las afecciones que aquejan a las estructuras que conforman el sistema estomatognático, en sus aspectos anatómicos, fisiológicos y patológicos.

La observación del rendimiento académico de los alumnos en estos cursos, expresados en notas de exámenes y en promedios finales, demuestra una cantidad considerable de calificaciones bajas y desaprobatorias. Al indagar el parecer de los estudiantes, manifiestan dificultades en relación con estas asignaturas como la excesiva cantidad de contenidos, la dificultad en la

comprensión de los mismos, en no saber para qué les van a servir estos conocimientos en su futuro ejercicio profesional, en que los exámenes son muy difíciles, etc. Toda ello se refleja en un rendimiento académico deficitario y desalentador, así como en un motivo de frustración, tanto para los estudiantes como para los docentes.

El propósito principal de la presente investigación es evaluar y comparar las diferencias en la calidad de los instrumentos de evaluación de los aprendizajes, a través del análisis de tres indicadores: índice de dificultad, índice de discriminación y confiabilidad, obtenidos a partir de las respuestas de los exámenes teóricos escritos que fueron aplicados a los estudiantes que cursaron las asignaturas del área de Ciencias Básicas, durante el periodo lectivo 2016.

1.2 Formulación del Problema

¿Existen diferencias en la calidad de los instrumentos de evaluación de las asignaturas del área de Ciencias Básicas de la Facultad de Odontología de la Universidad Nacional Mayor de San Marcos durante el periodo lectivo 2016?

1.3 Justificación de la Investigación

Mediante la presente investigación se pretende evaluar y comparar la calidad de los instrumentos de evaluación, a través de la determinación y comparación de los índices de dificultad, de discriminación y la confiabilidad de los exámenes teóricos escritos que se aplicaron a los estudiantes que cursaron las asignaturas de Anatomía General, Bioquímica General, Microbiología General, Fisiología General y Farmacología II del área de Ciencias Básicas de la Facultad de Odontología de la UNMSM, durante el año lectivo 2016.

En los primeros semestres de la enseñanza de las carreras de Ciencias de la Salud, entre ellas Odontología, los cursos del área de Ciencias Básicas como Anatomía, Embriología, Histología, Biología, Bioquímica, Fisiología, Patología, Farmacología, etc., constituyen el fundamento científico de estas profesiones. Estas asignaturas se caracterizan por tener amplios y a veces complejos contenidos que deben ser

asimilados por los estudiantes, lo cual requiere un esfuerzo considerable para poder comprenderlos y tener un adecuado rendimiento académico.

Una característica relacionada con el desarrollo de los mencionados cursos es observar una gran cantidad de estudiantes con notas bajas o desaprobatorias. Este hecho constituye una situación problemática, tanto para los docentes como para los estudiantes, y también para la institución. Dentro de los múltiples factores que pueden incidir en el rendimiento académico se puede considerar al sistema de evaluación y, dentro de este sistema, los instrumentos utilizados para evaluar la enseñanza – aprendizaje, como factores determinantes.

Se desea que la presente investigación contribuya a mejorar la forma de evaluar los aprendizajes referidos a conocimientos, así como mejorar la calidad de los instrumentos empleados en recoger información acerca de los aprendizajes. Es por ello que esta investigación constituye un primer esfuerzo por valorar algunos indicadores que reflejan la idoneidad, calidad y pertinencia de los instrumentos de evaluación, a través de tres indicadores como son la dificultad, la discriminación y la confiabilidad con la idea que sirva para una mejorapermanente en la calidad de la enseñanza de los docentes y de aprendizaje de los estudiantes, haciendo que la evaluación sea un proceso justo y relacionado con los objetivos o competencias propuestas, todo lo cual debe verse reflejado en un mejor aprovechamiento de los estudiantes.

Por otro lado, se debe destacar la existencia de pocas investigaciones relacionadas con el tema en nuestro medio, y en el caso de la Facultad de Odontología de la UNMSM esel primer estudio que se realiza con respecto a esta temática. Se desea que los resultados que se obtengan contribuyan en beneficio de la comunidad académica de la facultad, que redunde en un mejor proceso de enseñanza aprendizaje de la Odontología, principalmente en el Área de Ciencias Básicas.

1.4 Objetivos

1.4.1 Objetivo General

Evaluar y comparar la calidad de los instrumentos de evaluación de las asignaturas del área de Ciencias Básicas de la Facultad de Odontología de la Universidad Nacional Mayor de San Marcos durante el periodo lectivo 2016.

1.4.2 Objetivos Específicos

1. Determinar los índices de dificultad, de discriminación y de confiabilidad de la asignatura de Anatomía General, durante el periodo lectivo 2016.

2. Determinar los índices de dificultad, de discriminación y de confiabilidad de la asignatura de Bioquímica General, durante el periodo lectivo 2016.

3. Determinar los índices de dificultad, de discriminación y de confiabilidad de la asignatura de Farmacología II, durante el periodo lectivo 2016.

4. Determinar los índices de dificultad, de discriminación y de confiabilidad de la asignatura de Fisiología General, durante el periodo lectivo 2016.

5. Determinar los índices de dificultad, de discriminación y de confiabilidad de la asignatura de Microbiología General, durante el periodo lectivo 2016.

6. Comparar los resultados.

CAPÍTULO 2: MARCO TEÓRICO

2.1 Antecedentes de la Investigación

Baladrón y Cols. (2017) realizaron un trabajo donde analizaron las respuestas al examen Médico Interno Residente (MIR) 2015 de 3712 aspirantes. Estimaron la dificultad y la discriminación de las preguntas del examen, las que se analizaron según los valores obtenidos y agrupados por curso y tipos de pregunta. Las asignaturas que presentaron mayor cantidad de preguntas con dificultad media fueron las de Fisiología, Farmacología, Geriatría, Traumatología, Neurología y Cuidados Paliativos, mientras que las asignaturas con menor cantidad de preguntas de dificultad media fueron Anatomía Patológica, Anestesiología, Cirugía Plástica, Habilidades Comunicativas, Genética yEnfermedades Infecciosas. Concluyeron que en líneas generales, la dificultad y la discriminación de las preguntas de la prueba MIRresultaron adecuados. Finalmente,propusieron el uso de métodos de la Teoría de Respuesta al Ítem como un modo de estudiar las preguntas del examen que deberíanser eliminadas.

Quiñones (2016) realizó un estudio cuyo objetivo fue destacar las características del examen estatal ordinario de Licenciatura en Enfermería, aplicado durante el curso 2012-2013 a los estudiantes del quinto año de la carrera. Para ello utilizó métodos teóricos, empíricos y estadísticos para triangular los resultados. Al analizarel examen señaló el tiempo dedicado al desarrollo de cada tema, agrupando las preguntas para analizar sus indicadores de calidad. Los resultados muestran la faltade relación entre el tiempo de rotación hospitalaria con el número de preguntas; tampoco hubo relación entre el nivel de dificultad esperado y el nivel real;

y que la mayor cantidad de preguntas correspondió a la rotación médico-quirúrgica. El examen fue considerado Medianamente Fácil, con predominio de preguntas de discriminación Pobre. En conclusión, se caracterizó el examen de Licenciatura de Enfermería del año 2012-2013, determinándose, además, su coherencia interna y su confiabilidad aceptable**.**

Baladrón y Cols. (2016) desarrollaron un trabajo cuyo objetivo fue hacer la validación estructural del examen Médico Interno Residente (MIR) del año 2015, remarcando sus aspectos cuantificables. La data empleadaen la investigación fueronlas respuestas del examen MIR 2015 en la que participaron 3712 alumnos. El promedio de dificultad fue 0,6882 y el del índice de dificultad corregido fue 0,5422; la discriminación fue de 0,2492 y la correlación biserial puntual fue de 0,2954. La confiabilidad fue 0,9459 con la fórmula 21 de Kuder- Richarson y 0,9579 con el alfa de Cronbach. Los resultados se compararoncon los exámenes de los años 1989 a 1993; concluyendo que el examen MIR es objetivo, con adecuada dificultad, discriminación y validez.

Ortiz y Cols. (2015) en su trabajo de investigación realizado en la Universidad de Ciencias Médicas de Holguín durante el 2010-2011, señalaron las variaciones en la dificultad y en la discriminación de las preguntas de la prueba finalde Metodología de la Investigación y Estadística. El estudio descriptivo fue realizadoen una muestra de 276 exámenes en los que analizaron el tiempo y su relación conlos contenidos desarrollados y el número de preguntas en la prueba; también calcularon la dificultad, así como el coeficiente de correlación del punto biserial de las preguntas. Los resultados no muestran relación entre el tiempo empleado en desarrollar cada tema con el número de preguntas planteadas; de igual modo, las preguntas fáciles fueron mayor que el esperado y que la mayoría delos ítems discriminaron de forma excelente.

Sánchez y Cols. (2015) realizaron un estudio para evaluar la calidad de los exámenes parciales tomados a 30 médicos residentes de la especialidad de Medicina General Integral, durante el 2014. Los resultados de analizar los instrumentos de evaluación mostraron diferencias y congruencias mediante

indicadores como el índice de dificultad por preguntas y temas, el nivel de discriminación y la consistencia interna; los exámenes también mostraron validezy confiabilidad.

Rivera y Díaz (2013) realizaron un trabajo cuyo objeto de estudio fue el examen teórico final escrito de la asignatura Medicina Interna en el área clínica, con el propósito de determinar la calidad del examen del 2013 al 2014. De todos los exámenes se escogió una muestra de 371. Los métodos empleados para analizar los exámenes permitieron triangular la información obtenida. Se conversó con 10 profesores que caracterizaron el examen; se determinó la relación entre el tiempo y el número de preguntas, la dificultad, la discriminación, la norma discriminativa, el coeficiente de discriminación del punto biserial y el coeficientede correlación Alfa deCronbach. Los resultados mostraron carencias en los conocimientos para planificar la prueba por parte de los profesores; también un mayor número de preguntas tipo ensayo; falta de relación entre el tiempo empleado en desarrollar cada tema y el número de preguntas presentes en el examen, así como en la distribución del nivel de dificultad esperado y real porpreguntas. La prueba fue catalogada como difícil con baja discriminación y coherencia interna aceptable en un rango bajo.

Rodríguez y Cols. (2012) realizaron un estudio cuyo propósito fue identificar elementos de calidad en los exámenes finales de las asignaturas de Metodología de la Investigación y de Estadística en los cursos 2009 – 2010 y 2010 – 2011 en la Facultad de Estomatología de la Habana. Estudiaron los exámenes finales de ambas asignaturas tomadas a los alumnos del segundo año de Estomatología; el universo de la investigación fue la totalidad de los exámenes realizados en ordinario: 201 en la primera asignatura y 173 en la segunda. Estimaron la dificultad, la discriminación y la confiabilidad con el Alfa de Cronbach. Encontraron que las preguntas más difíciles tuvieron un índice de 0,68 y una confiabilidad de 0,50 en el segundo y de 0,86 en el

primer curso que se evalúa. Concluyeron que los exámenes finales de Metodología de la Investigación y de Estadística fueron fáciles, con buena discriminación y notoria diferencia en la confiabilidad.

Carrazana y Cols. (2011) tuvieron como objetivo en su estudio, determinar el nivel de dificultad y el poder de discriminación del examen diagnóstico de Morfofisiología Humana I; lo hicieron calculando la dificultad, la discriminación y el coeficiente de discriminación en 421 pruebas. Los resultados obtenidos señalan que el examen diagnóstico de Morfofisiología Humana I del Programa de Formaciónde Médicos en la Facultad-Policlínico "Félix Edén Aguada", de Cienfuegos, Cuba fue un examen cuya dificultad varió entre Dificultad Media y Medianamente Fácil con un poder de discriminación adecuado.

Argudín y Cols. (2011) realizaron un análisis del examen ordinario de Morfofisiología Humana I del año 2005- 2006, tomado como parte del curso de Medicina Integral Comunitaria. En el análisis, un grupo de expertos revisó los contenidos temáticos y clasificaron las preguntas según sus tipos. Los resultados mostraron que las dificultades de las preguntas se relacionaban con varios aspectos, tales como: la complejidad de los temas dictados, el estilo de redacción, el tipo de pregunta y el número de preguntas por área. Igualmente, se observó queel tipo de pregunta más usada y de mayor dificultad fueron las preguntas de selección múltiple.

Muñiz y Fonseca - Pedrero (2008) analizaron la problemática de los métodos empleados en la evaluación de la educación superior. Estudiaron los aspectos técnicos y procedimentales de la evaluación universitaria en la investigación, en la docencia y en la gestión. También se revisaron las preguntas que todo sistema evaluativo se hace: qué, quién, cómo y cuándo hay que evaluar. Y, por último, resaltaron la importancia que tienen los aspectos metodológicos en la construcción de los instrumentos de evaluación.

Pérez y Cols. (2008) determinaron la dificultad y la discriminación del tercer y quinto examen parcial del curso de Cito-histología. La muestra estuvo conformada

por 388 exámenes del tercer parcial y por 350 exámenes del quinto parcial. Los resultados obtenidos muestran una media de 0,34 de dificultad para el tercer parcial (Medianamente Difícil); y, una media de 0,54 de dificultad para el quinto examen (Dificultad Media). En relación al índice de discriminación, la media obtenida por el tercer examen fue de 0,57 y la media del quinto examen fue de 0,45, ambos fueron catalogados como de Excelente Calidad. Estos índices de discriminación fueron corroborados con el Coeficiente de Discriminación que fue 0,44 para el tercer examen y de 0,54 para el quinto examen.

Backhoff y Cols. (2000) señalaron que en México el Examen de Habilidades y Conocimientos Básicos (EXHCOBA) es un examen de gran escala que ha ido mostrando sus resultados, así como sus parámetros de calidad. En el presente estudio, analizaron el nivel de dificultad y el poder de discriminación de los ítems o reactivos que conforman el examen. Los resultados señalan que la mayoría de las preguntas fueron de Dificultad Media y de buena discriminación. Igualmente, se menciona que las preguntas que mejor discriminan son del área de matemáticas y las de menor discriminación del área de ciencias sociales.

2.2 Bases Teóricas

2.2.1 Educación

La educación es un concepto amplio de múltiples dimensiones que puede definirse como “el proceso de socialización de los individuos”(Pérez, Acuña y Arratia, 2008) Otra idea del concepto educación lo declara como “un proceso humano, intencional, intercomunicativo y espiritual, en virtud del cual se realizan la instrucción, la personalización y la socialización del hombre”(Fermoso,2013).

De acuerdo con estos dos conceptos, la educación le permite al hombre desarrollarse en sentido personal y social para convertirse en un Ser Humano. En elplano personal, irá adquiriendo los rasgos distintivos que lo diferenciará de los demás; y, en el plano social lo capacitará para interactuar con los otros en forma

productiva, sean o no de su entorno sociocultural. En este sentido, la personalización y la socialización implicada en la educación, determinan su antropogénesis. En la educación formal o escolar, el sistema educativo de una sociedad presenta ordenadamente una serie de ideas, conceptos, hechos y técnicas. En este proceso, un individuo denominado profesor, docente o maestro busca influir de forma ordenada sobre otros, denominados alumnos, discentes o estudiantes con el fin de formarlos y socializarlos en concordancia con los patronessocioculturales vigentes y predominantes en una determinada época. De esta manera, el sistema educativo de una sociedad transmite y conserva su identidad colectiva a las nuevas generaciones. En la actualidad, en la era del conocimiento, la sociedad establece que el proceso educativo no termine con las etapas formales,si no que éste sea un proceso que acompañe al individuo en todas las etapas de suvida, es decir, que el ser humano debe seguir adquiriendo conocimientos y habilidades a lo largo de toda su existencia (Fermoso, 2013).

Al considerar a la educación como una actividad de múltiples dimensiones, y siguiendo lo propuesto por Fermoso (2013), las teorías que tratan de explicar el proceso educativo serían de tres clases:

- Las teorías empíricas basadas en la psicología, las cuales se fundamentan en los descubrimientos experimentales verificados por la psicología.

- Las teorías axiológicas, fundamentadas en lo antropológico, en lo ético y en lo metafísico. El proceso educativo se basa en los juicios de valor.
- Las teorías metafísicas o filosóficas que sostienen que las teorías educacionales no son saberes empíricos, sino filosóficos o una categoría intermedia entre la ciencia y la filosofía.

2.2.2 Educación Universitaria

La educación busca formar capacidades y actitudes para que las personas puedan integrarse a la sociedad en forma armoniosa, puedan regular el statu quo y puedan transformar el entorno social para mantener ciertos valores en un determinado momento histórico. En este sentido, la educación superior asume el reto de formar profesionales competentes, que resuelvan con eficiencia y eficacia los problemas sociales (Guerrero y Faro, 2012).

En estos tiempos, el proceso educativo ha sufrido un importante impacto por la transformación en la naturaleza, función, generación y transmisión del conocimiento, haciendo que el rol de éste modifique la idea acerca de la formación y la docencia, y dentro de éstos, el rol de los modelos que evalúan el aprendizaje. Así, al desarrollarse la educación, se logra relacionar al hombre con el universo a través del conocimiento, para que éste pueda conocerse, interpretar y comprender la realidad. En este sentido, educar es convertir el conocimiento en experiencia y la información en vida (Mateo y Vlachopoulus, 2013).

En las universidades, la era del conocimiento y de las tecnologías de la información determinan un cambio en el paradigma educativo, que permitan al sistema universitario adaptarse a la nueva realidad. Por ejemplo, en el campo del aprendizaje los descubrimientos de la neurociencia llevan a modificar los métodos de enseñanza; del mismo modo, conducen a repensar lo concerniente a los planes curriculares, de formación, de docencia, de instrucción y de forma especial, en las metodologías de evaluación. En este panorama, surge el concepto de educación por competencias, como una forma de preparar profesionales que tengan las capacidades requeridas para solucionar la problemática de la sociedad actual. Por ello, las universidades europeas propugnan en sus planteamientos curriculares el desarrollo de competencias en sus estudiantes; corriente que también se ha extendido a universidades latinoamericanas, entre ellas a universidades del Perú (Fermoso, 2013).

En el paradigma conductista, el cumplimiento de los objetivos es un fin fundamental del modelo educativo, que se sustenta en la transmisión y recepción de conocimientos; en una educación basada en competencias, éstas se establecen y se relacionan con el contexto del sujeto que aprende (Delgado- Maldonado y Sánchez, 2012). Por ello, al ser la educación un proceso, no resulta adecuado currículos y modelos evaluativos basados en la consecución sólo de objetivos, sino que se debe buscar desarrollar capacidades que permitan comprender, interpretar e intervenir en realidades distintas, variables y multidimensionales. En este sentido, la función de las universidades, debería maximizar el intercambio de expresiones y compartir recursos de toda índole, y en relación con la función del docente, éste debería focalizar su interés en el aprendizaje más que en la enseñanza, favoreciendo un proceso autónomo del estudiante en la adquisición de su propio conocimiento y en el desarrollo de sus esquemas mentales. Esto presupone replantear los contenidos de las asignaturas y acompañarlos en su proceso (Fermoso, 2013).

2.2.3 Evaluación

La evaluación es un aspecto consustancial de toda actividad humana que se manifiesta como un proceso. Mediante este proceso se juzga el valor de un objeto o fenómeno de la realidad, en relación con unos criterios de referencia, que deben ser adecuados con la naturaleza del objeto en evaluación y a los objetivos planteados. El proceso de la evaluación tiene una estructura y una función. Se desarrolla en relación con los fines previstos y posee un objeto que se somete a evaluación; este objeto interactúa con el evaluador haciendo que el proceso evaluativo se efectúe en determinadas condiciones. Los resultados deben analizarse y permitirnos tomar decisiones asertivas. Por otro lado, el desarrollo del proceso evaluativo, supone fases tales como: planificación, orientación, ejecución, control y regulación (Mateo y Vlachopuolos, 2013).

En relación a los procesos básicos involucrados en la evaluación destacan la recogida de información, la emisión de un juicio valorativo y la toma de una decisión (González, 2002).

Como se mencionó, la Docimología es la disciplina que, utilizando los conocimientos de la psicología, la didáctica y la tecnología educativa, estudia en forma sistemática las estrategias evaluativas, las técnicas y los instrumentos de evaluación, relacionándolo con la calidad educativa. También investiga, elabora, administra y valora los instrumentos que miden el aprendizaje. En un currículo basado en competencias, estos aspectos evaluativos adquieren una gran complejidad (Páez y Di Carlo, 2012).

2.2.4 Evaluación del Aprendizaje

El antecedente más antiguo donde se desarrolla el concepto de evaluación se encuentra en el Tetrabiblos, tratado que se le atribuye a Ptolomeo (Páez y Di Carlo, 2012).

En una breve línea del tiempo y dentro de la cultura occidental, se puede mencionar que durante la edad media la Universidad introduce los exámenes orales para evaluar a los aprendices; en el siglo XVIII, se establecen normas para evaluar con exámenes escritos; en el siglo XIX, las evaluaciones se hacen a nivel nacional y se entregan certificaciones al aprobar los exámenes escritos. En 1845, en Estados Unidos, empiezan a utilizarse las evaluaciones tipo test; publicándose en 1897 la primera investigación evaluativa en educación que consistió en un estudio comparativo entre escuelas norteamericanas sobre el valor de la instrucción en ortografía, empleando los resultados de los test. En el siglo XX, se utilizaron teorías y métodos de las ciencias sociales en los procesos evaluativos, vinculando la evaluación con la planificación y el control social, destacando el orden como un puntal. Se considera que el proceso de la evaluación termina con unjuicio de valor sobre algo. En la década de los 30, Ralph Tyler desarrolla el modelode evaluación basado en objetivos de aprendizajes; su propósito era verificar la efectividad de las escuelas, señalando aspectos susceptibles de ser mejorados. Tyler, propone el término de “Evaluación Educativa” como una alternativa a los términos de medición, prueba o examen. A partir de 1970, en base a trabajos del Instituto de Investigación de Stanford, se reconoce que cada comunidad educativa

tiene sus propias características y por lo tanto, se hace necesario buscar y encontrar la mejor manera de evaluar. De esta forma, surge una evaluación pragmática, metodológica y particular, en la que se enfatiza sobre los problemas y necesidades de los involucrados en un sistema evaluativo. El aprendizaje se relaciona con la forma de evaluar, permitiendo establecer una relación entre la forma de enseñar y aprender con la forma de evaluar en una determinada institución educativa, creándose de esta manera, una especie de cultura de la evaluación (Gonzáles, 2001).

Por ello, la evaluación del aprendizaje o evaluación educativa se convierte en un caso particular de la evaluación, que se puede definir como la actividad que tiene como propósito valorar el proceso y el resultado del aprendizaje de los estudiantes para orientar y regular la enseñanza, contribuyendo a lograr las finalidades de la formación. Sin embargo, este concepto inicial de la evaluación del aprendizaje se ha distorsionado a lo largo del tiempo, se ha convertido en sinónimo de calificar, esdecir asignar notas; se ha convertido en sinónimo de examinar, es decir tomar exámenes; la evaluación del aprendizaje es más que las dos acciones. La calificación es parte de la evaluación que expresa, de modo resumido, un juicio evaluativo, el cual toma varias formas: puede ser un número, un símbolo, un adjetivo calificativo, un informe descriptivo valorativo, etc. La nota expresa con frialdad algo que es un proceso continuo, sin resaltar las fortalezas, lasdebilidades, ni el esfuerzo puesto por el estudiante; no considera el potencial de aprendizaje que tiene el estudiante. Por esta razón, la información que aporta una calificación es una información limitada. En relación al examen, es sólo una forma de obtener información sobre el aprendizaje de los estudiantes. En los hechos, la evaluación del aprendizaje da demasiada importancia a la nota y al examen, haciendo que se aprenda sólo para aprobar el examen, valorándose poco o nada el potencial de aprendizaje que posee el estudiante, minimizándose la función formativa de la evaluación (Mateo y Vlachopoulos, 2013).

El sistema de evaluación es un elemento importante en la planificación y ejecución del proceso de enseñanza y aprendizaje, al tiempo que es uno de los elementos curriculares que más dificultades plantea en el desarrollo docente**.** En un currículo

basado en competencias, como se propugna en la actualidad, la evaluación depende del instrumento que sea más apropiado y congruente con las estrategias psicodidácticas utilizadas y con lo que se desea valorar (Páez y Di Carlo, 2012).

En la actualidad, la tendencia es mirar a la evaluación desde una perspectiva comprehensiva, resaltando su rol social, y su íntima relación con el proceso educativo; del mismo modo se debe destacar el impacto que tiene el modode evaluary la forma en que el estudiante percibe el aprendizaje. Para lograr esto laacción evaluativa supone el uso integrado y estratégico de diferentes técnicas de recogida de información, aunque la comprobación final de un nivel concreto de desarrollo de una competencia, exige el uso de pruebas basadas en ejecuciones (Gonzáles, 2002).

2.2.5 Evaluación Universitaria

Los enfoques teóricos resaltan que la evaluación debe ser integral, con técnicas diversas que permitan recolectar información sobre el desempeño del estudiante, valorando diversos aspectos de su progreso en su proceso educativo. Uno de estos enfoques, la evaluación humanista, destaca que el aprendizaje está por encima de otros aspectos como el control, la calificación, la examinación, etc., que lo que se debe buscar es un clima de confianza y comunicación con los estudiantes, desterrando el autoritarismo. En consonancia con este enfoque, la evaluación universitaria debe entenderse como un proceso de mejora continua, basada enel diálogo, y no solamente como una función fiscalizadora, calificadora, certificadoray controladora. Sin embargo, el contraste con la realidad muestra que el examen tradicional sigue siendo la forma de evaluación más empleada, de tal manera que la evaluación se convierte en sinónimo de examen, diseñada no para cumplir funciones pedagógicas sino para satisfacer necesidades burocráticas y administrativas. En este contexto, el profesor se transforma en el actor principal del proceso evaluativo y en el aula, el estudiante pasa a ser una especie de objeto y no un sujeto de evaluación. El docente en vez de evaluar, examina, califica y controla; así, el examenpasa de ser un medio, a convertirse en un fin perdiendo sucarácter formativo. Por lo tanto, es necesario que la evaluación recupere su sentido

natural, que es ser un elemento integrado de la forma de enseñar y aprender (San Martín, Jiménez y Jerónimo, 2016).

Un problema importante en la evaluación del aprendizaje es la calificación, proceso por el cual se atribuye una ponderación numérica a una realidad cualitativa y compleja, lo cual puede llevar a una valoración tergiversada del aprendizaje de un estudiante (González, 2000).

Es así que el proceso de la calificación se convierte en uno de los problemasálgidos de la evaluación del aprendizaje, relacionándolo con la problemática de la medición de cualidades o propiedades psíquicas de la persona que está en proceso de formación. Esto hace que sea abordado desde distintas dimensiones y perspectivas teóricas. El procedimiento de calificación guarda relación con el sistemade criterios que se considere y con los instrumentos que se usen para recoger información, destacando que, en una concepción formativa de la evaluación del aprendizaje, la calificación debe ser educativa y no punitiva. Por ello, el proceso calificativo para ser adecuado debe combinar procedimientos cualitativos y cuantitativos, considerando lo singular del hecho educativo y no estandarizando a los estudiantes; debe utilizar diversos sistemas de referencias; debe evaluar de acuerdo con la naturaleza del objeto de evaluación; así también, la calificación, como todo el proceso de evaluación, tiene que cumplir una función formativa en el contexto de la enseñanza y del aprendizaje (Moreno, 2009).

La evaluación del aprendizaje de los estudiantes universitarios debe ser realizado por el profesorado, contando con la opinión del estudiantado, para promover estrategias que mejoren tanto el aprendizaje de los estudiantes como la formación de los docentes; en este proceso se deben considerar diversos aspectos, tales como: el tipo de actividades que se realizaron, los materiales empleados, la metodología utilizada, los intereses del estudiantado, las relaciones y el ambiente del aula, etc. En este contexto, debe haber concordancia entre el modelo de enseñanza y aprendizaje con los objetivos o las competencias, igualmente entre los métodos didácticos con los sistemas de evaluación. Esta línea de coherencia supone que la evaluación sirva para confirmar lo aprendido, ayudando al estudiante a optimizar su aprendizaje y al docente a mejorar su enseñanza. Las principales características de un sistema evaluativo, dentro de este contexto, serían: 1) que la evaluación sea un proceso planeado, integrado y oportuno, relacionado con los objetivos o competencias a alcanzar, que busque satisfacer las demandas del desempeño profesional; 2) que la evaluación debe establecer indicadores apropiados; 3) que la evaluación, al focalizarel proceso educativo en el aprendizaje, permite que los alumnos tengan un rol activo y responsable en su aprendizaje; y, 4)que la evaluación debe ser continua y formativa, que se complemente con las evaluaciones sumativasy finales (San Martín, Jiménez y Jerónimo, 2016).

Para los estudiantes universitarios, la evaluación debería ser continua y las mejores formas de hacerlo serían: 1) la presentación y la valoración de trabajos (86%), 2) la valoración del esfuerzo (68%), 3) la realización de exámenes parciales y prácticos (54%), 4) la consideración de la asistencia a clases (48%), 5) el interés del alumno (46%) y 6) las actividades cotidianas de clase (40%). Por último, hay que destacar el rechazo a evaluar solamente con un examen final (Ortiz, 2006)

2.2.6 Teorías de los Test

Siguiendo las pautas sostenidas por Muñiz (2010), resumiremos aspectos relacionados con la Teoría Clásica de los Test y con la Teoría de Respuesta al Item:

La principal finalidad de las pruebas o test es evaluar a todos los individuos con objetividad, utilizando los mismos criterios por igual y evitando sesgos o subjetividades. Las pruebas o test se consideran que son muestras de conductas o comportamientos que permiten hacer inferencias sobre el comportamiento o habilidades de las personas.

Haciendo un poco de historia se mencionará que los hallazgos más remotos de las pruebas o test datan de hace 3 mil años antes de Cristo y se remontan a la China imperial, donde los oficiales al servicio del emperador eran sometidos a exigentes exámenes. Las pruebas o test actuales se derivan de las pruebas sensomotoras usadas por Galton (1822-1911) en su laboratorio antropométrico. El término test mental fue empleado por James McKeen Cattel (1860 - 1944), en 1890. Eran pruebas sensomotoras que no destacaban las capacidades cognoscitivas de las personas. En 1905, Binet y Simon, modifican esta prueba incluyendo tareas cognoscitivas para indagar el juicio, la comprensión y el razonamiento de los individuos. En 1916, Terman, al revisar esta prueba en la universidad de Stanford, emplea el término de Cociente Intelectual (CI); a esta revisión se le conoce como la escala Stanford-Binet. En 1917, aparecen los test colectivos alfa y beta, y posteriormente, acabada la primera guerra mundial, los test se extienden a todos los ámbitos del quehacer humano, como una forma de seleccionar a las personas con las mejores aptitudes para el desempeño de alguna actividad específica.

Para construir y analizar las pruebas o test es necesario un sustento teórico y estadístico, que permita inferir y decidir diversos aspectos de la gente; así, se asegura que las inferencias sean las correctas; por otro lado, las teorías estadísticaspermiten el cálculo de las propiedades psicométricas que garanticen que las decisiones que se tomen sean las más adecuadas; sin esas teorías, la

validez y confiabilidad de los test serían difícil de estimar. En consonancia con Muñiz (2010) se mencionará que existen dos enfoques o teorías que se utilizan para construir y analizar los test: 1) la Teoría Clásica de los Test (TCT) y 2) la Teoría de la Respuesta a los Ítems (TRI).

2.2.7 Teoría clásica de los test (TCT)

La TCT es el enfoque predominante en la construcción y análisis de los test, se basa en los estudios de Spearman, realizados a principios del siglo pasado. En 1968, Lord y Novick reformularon la teoría clásica, dando origen a la Teoría de la Respuesta al Item (TRI).

Según la TCT, al aplicar un test, lo que se obtiene son puntuaciones que no se saben si se corresponden con las puntuaciones verdaderas que obtendría una o varias personas en una determinada prueba. Por ello, se hace necesario plantear modelos para estimar el grado de precisión de las pruebas o test. Uno de los modelosmás eficaces es el ***Modelo Lineal Clásico*** que fue sustentado por Spearman; éste modelo supone que la puntuación obtenida en un test (X), está formada por la puntuación verdadera (V) y por el error (e). Formalmente, sería:

$$X= V + e$$

En la ecuación sólo se conoce la puntuación empírica (X), pero no se conocen ni la puntuación verdadera (V) ni el error (e); ante esta situación Spearman propuso tres supuestos y una definición para el modelo.

- **1° supuesto:** la puntuación verdadera se considera como la posibilidad matemática de la puntuación empírica: V= E(X); es decir, que la puntuación verdadera en un test correspondería a la media obtenida al tomar infinitas veces la misma prueba.

- **2° supuesto:** no existe relación entre el valor de las puntuaciones verdaderas y el tamaño de los errores que afectan a dichas puntuaciones:

$$r (v, e) = 0$$

- **3° supuesto:** el error de medida en un test no tiene relación con el error de medida en otro test:

$$r\ (e\ j,\ e\ k) = 0.$$

Esto supuestos no son comprobables empíricamente, serán las deducciones que se originen las que permitirán confirmarlas o rechazarlas.

El concepto que complementa el modelo lineal de la teoría clásica es el concepto de los ***Test Paralelos***, que se refiere a aquellos test que miden lo mismo con diferentes ítems. Las puntuaciones verdaderas obtenidas en los test paralelos serían las mismas, así como las varianzas de los errores.

El modelo lineal es la esencia de la teoría clásica de los test, sus variantes derivan de cómo se trate el error de medida. La **Teoría de la Generalizabilidad (TG)**, sustentada por Cronbach en 1972, es la variante más difundida. Otro enfoque surgido en el marco de la teoría clásica es el ***Test Referido al Criterio***, usado básicamente en la esfera educativa y en los ámbitos laborales; este enfoque tiene como propósito determinar si las personas tienen dominio sobre un criterio específico o sobre un campo de conocimientos, denominado criterio. Estos test se desarrollarona partir de la propuesta de Glasser en 1963.

La TCT presenta dos limitaciones que no permiten que las mediciones psicológicas sean homologables con las mediciones de otras ciencias empíricas:

- **1° las mediciones del instrumento empleado no son constantes:** esto quiere decir que, si se evalúa, por ejemplo, la inteligencia de tres personas con tres instrumentos distintos, las mediciones no son comparables porque los resultados de los tres test tienen distintas escalas; lo ideal sería que los puntajesestuviesen en la misma escala; esto es lo que logra conseguir la teoría de respuesta al ítem.

- **2° la falta de constancia de las propiedades del test respecto de las personas:** es decir, que las propiedades psicométricas de los test, como el

índice de dificultad de las preguntas o la fiabilidad de la prueba, estén en relación con las características de los individuos testeados.

Cuando en la teoría clásica se obtiene un coeficiente de fiabilidad, se presupone que éste mide a todas las personas con la misma fiabilidad; sin embargo, la evidencia empírica muestra que los test no miden por igual a todas las personas, porque esto depende del nivel de un individuo en una determinada variable. La teoría de respuesta al ítem (TRI) soluciona el problema en base a la ***Función de Información***, a través del cual se estima la fiabilidad de una prueba en función del nivel de un individuo en una variable que se esté midiendo. Los modelos de la teoríade respuesta al ítem (TRI) permiten complementar los aportes de la teoría clásica de los test (TCT), con lo cual se mejora la construcción y análisis de los test.

2.2.8 Teoría de Respuesta al Item

Continuando con los criterios expuestos por Muñiz (2010) se menciona que la Teoría de Respuesta al Ítem (TRI) o Teoría del Rasgo Latente pretende medir rasgos y constructos no observables, para ello esta teoría se focaliza en las características de las preguntas o ítems, y no en los atributos generales de las pruebas, o sea, da más importancia al ítem como unidad de análisis de la prueba que a las puntuaciones totales obtenidas, como lo hace la teoría clásica de los test (TCT).

Los distintos modelos de la TRI, asumen que existe un rasgo o aptitud potencial en el sujeto; igualmente, destaca la relación que existe entre este rasgo o aptitud potencial con el rendimiento del sujeto, describiéndolo a partir de la Curva Característica del Ítem (CCI); en esta curva se encuentra la probabilidadde responder al ítem en función del rasgo o de la aptitud potencial que posea la persona.

Los supuestos de los modelos de la Teoría de Respuesta al Ítem son:

- **Unidimensionalidad:** debe existir un rasgo o aptitud potencial o latente responsable de la respuesta que da la persona frente a un estímulo provocado por un reactivo. Es suficiente un rasgo o aptitud potencial que explique los puntajes de los individuos, así como las relaciones entre los reactivos. Es decir, el rendimiento de un individuo en una pregunta, está en relación con el nivel que posea en un solo rasgo o dimensión. Este principio vale para el test en su totalidad, es decir, que se debe esperar que los ítems que constituyen una prueba midan sólo un rasgo o dimensión.

- **Independencia local:** proviene de la unidimensionalidad. Destaca que la respuesta a un ítem es independiente a la que se da a otros, es decir, que la respuesta a un reactivo depende de sus parámetros y de la habilidad del individuo. Matemáticamente, es la probabilidad de responder correctamente un determinado número de reactivos, y que esto es igual al producto de las probabilidades de responder correctamente cada reactivo de manera separada.

- **Constancia:** ocurre en dos sentidos, primero, en el grupo de preguntas ante diferentes niveles de habilidad o potencialidades de los sujetos, y segundo, por el nivel de medición de la habilidad de una persona a partir de un conjunto diferente de reactivos. Esto significa que se puede estimar los parámetros de los ítems sin que dependan de la muestra o población que respondieron, obteniéndose la misma curva para el ítem, indistintamente de los sujetos que respondieron.

En la Teoría de Respuesta al Ítem, el error de medición (error típico de estimación) está en función del rasgo (Q) y cada nivel de rasgo o valor de Q tiene su propio error. También se puede conocer la función de información, que permite averiguar los niveles de habilidad con mayor precisión, haciendo que el error de medición sea menor.

- **Parámetros del TRI**

  El modelo de la TRI considera la capacidad de los individuos y tres parámetros logísticos: la discriminación del reactivo (a); la dificultad del reactivo (b), y el azar en la respuesta al reactivo (c). Estos parámetros caracterizan la ***función de respuesta al reactivo***. Esta función, llamada ***Curva Característica del Ítem (CCI)***, señala la probabilidad del individuo de responder correctamente, de acuerdo con su capacidad.

  La discriminación del ítem es la pendiente de la curva y se representa por la letra "a". Es el punto de inflexión de la curva característica del ítem, en donde el individuo tiene 50% de probabilidades de responder acertadamente al reactivo; esto quiere decir, que la habilidad o rasgo del sujeto (Q) es igual al índice de dificultad del reactivo (b).

  El parámetro "b" es el índice de dificultad del reactivo, representa el valor de la habilidad o rasgo del individuo en el que la probabilidad de responder correctamente el reactivo es 0.5, es decir, cuando no hay aciertos al azar. En estas circunstancias, la habilidad o rasgo del sujeto coincide con el grado de dificultad del ítem. Cuanto mayor es el valor de b, el reactivo será más difícil, por lo tanto, la probabilidad de acertar el reactivo disminuirá.

  El azar está representado por el parámetro "c". Representa la probabilidad que tiene un individuo con baja habilidad de responder bien a la pregunta. Todos los individuos, independientemente de su capacidad, rasgo o habilidad tienen la misma probabilidad de acertar por azar al responder una pregunta. Se estima que los individuos con menores capacidades recurrirán más al azar para lograr responder correctamente un ítem.

2.2.9 Calidad de los instrumentos de evaluación

La evaluación de una competencia profesional toma en consideración conocimientos, habilidades, destrezas y actitudes. Por ello, se deben confeccionar instrumentos evaluativos que tengan un alto grado de validez y confiabilidad; esto con la finalidad que la información obtenida permita incrementar la calidad de los procesos educativos, así como identificar lo que los estudiantes desean aprender.

Al aplicar un instrumento evaluativo, es necesario conocer en qué medida éste, permite conocer las competencias de los alumnos. Esto permite preguntarse: 1° ¿en qué medida el instrumento de evaluación mide lo que dice medir?, es decir, ¿qué nivel de correspondencia existe entre las preguntas y la temática desarrollada?; la respuesta determina la validez de un instrumento de evaluación; 2° ¿con qué precisión y estabilidad mide lo que está midiendo?, la respuesta a esta segunda pregunta determina la confiabilidad de un instrumento evaluativo (Soler y Salas, 1997).

En este mismo sentido Pérez, Acuña y Arratia (2008) sostienen que una vez aplicado el instrumento de evaluación surge la necesidad de conocer su calidad. En la evaluación de esta calidad se puede realizar un análisis de los reactivos o preguntas, los cuales pueden basarse en la estimación de la dificultad, del poder de discriminación y del funcionamiento de los distractores; de los tres indicadores psicométricos, se utilizan más los dos primeros al evaluar la calidad de los instrumentos evaluativos.

Un sistema de evaluación implica la elaboración, aplicación y análisis de los instrumentos a ser utilizados. Existen indicadores que permiten evaluar la calidad de los instrumentos de evaluación como el valor de discriminación, que permite descubrir preguntas mal formuladas, inútiles, que no discriminan entre los estudiantes más competentes y los menos competentes; otro es la dificultad, que esla proporción de alumnos que responden correctamente un reactivo del examen. Estos indicadores contribuyen a determinar la importancia que tienen la validez y la confiabilidad de los resultados de un instrumento de recogida de información en la evaluación del aprendizaje (Rodríguez, Fernández y Vigo,2013)

Para Alonso, Ramos y Quiñones (2014) la evaluación del aprendizaje tiene un carácter cualitativo e integrador, basado en condiciones que se relacionan entre sí; dentro de estas condiciones destacan conceptos como la objetividad, la sistematización y la continuidad, así como exigencias, como la validez y la

confiabilidad; todo lo cual debe redundar en un instrumento de evaluación de calidad.

En relación con la validez de un instrumento evaluativo Sánchez, Medina, Rodríguez, Vega y De la Torre (2014) sostienen que puede determinarse en base al índice de dificultad, al índice y al coeficiente de discriminación y a la confiabilidad, lo cual nos lleva a determinar la calidad de un instrumento evaluativo, en forma global,así como de sus respectivas preguntas.

- **Índice de dificultad**

  Lo más frecuente es entender este índice como la proporción de personas que responden correctamente un reactivo de una prueba, es decir, se estaría midiendo la facilidad de una pregunta. En el presente trabajo, se entenderá como índice de dificultad a la comparación del número de errores al responder una pregunta con el número total de examinados:

  $$Idif. = 1 - C/N$$

  Donde C representa el número de respuestas correctas y N corresponde al número total de examinados. De esta manera, el índice de dificultad toma valores que fluctúan entre 0 y 1; cuanto más cercano a 1 sea el índice, la pregunta será más difícil, y cuanto más se acerca a 0, menor será la dificultad de la pregunta o ítem (Hurtado,2018).

Valoración del índice de dificultad	
Índice de dificultad del ítem	**Categoría de los ítems**
Mayor a 0.87	Altamente difícil
Entre 0,74 a 0,86	Medianamente difícil
Entre 0,53 a 0,73	Dificultad media
Entre 0,33 a 0,52	Medianamente fácil
Menor a 0,32	Altamente fácil

- **Discriminación**

  En una prueba o ítem que mida una determinada competencia, se espera que los estudiantes con más altas puntuaciones tengan altas probabilidades de contestar correctamente un ítem, y lo contrario también, que los que obtuvieron bajas puntuaciones tengan menor probabilidad de responder correctamente la pregunta. Esto permite determinar que una buena pregunta debe distinguir a estudiantes con alto rendimiento de aquellos con menor rendimiento (Backoff, Larrazolo y Rosas, 2000)

$$Di = \frac{GA\ aciertos - GB\ aciertos}{N\ grupo\ mayor}$$

  Donde:

  - Di = Índice de discriminación del reactivo.
  - GA aciertos = Número de aciertos en el reactivo del 27% de personas con las puntuaciones más altas en el test.
  - GB aciertos = Número de aciertos en el reactivo del 27% de personas con las puntuaciones más bajas en el test.
  - **N** grupo mayor =Número de personas en el grupo más numeroso (GA o GB).

Cuando todos los individuos del GA responden correctamente una pregunta y todos los individuos del GB lo hacen mal, entonces D= 1, valor máximo que puede tomar el indicador de discriminación. Cuando sucede lo contrario, es decir, todos los individuos del GB responden correctamente la pregunta y todos los del GA responden mal, entonces D= -1, valor máximo negativo; y cuando todos los miembros del GA y del GB responden bien a la pregunta, entonces D= 0, valor mínimo de discriminación. Por lo tanto, cuanto más alto sea valor del índice de discriminación mejor diferenciará el rendimiento de los estudiantes.

D (índice de discriminación)	**Calidad del ítem**	**Recomendación**
Mayor a 0,39	Excelente	Conservar
0,30 a 0,38	Buena	Mejorar
0,20 a 0,29	Regular	Revisar
0,0 a 0,19	Pobre	Revisar a profundidad
Menor a 0,0 (negativo)	Pésima	Descartar

- **Confiabilidad**

La confiabilidad o fiabilidad es la consistencia con la que un instrumento es capaz de medir una variable dada. Esto puede conseguirse de varias maneras:

- Por la técnica de la replicación o test-retest.
- Por formas paralelas de una misma prueba.
- Por la aplicación de pruebas en mitades.
- Por la aplicación de una misma prueba en tiempos diferentes.

Desde el punto de vista numérico, se puede obtener la confiabilidad o fiabilidad de una prueba, aplicando la fórmula de Kuder - Richardson o la

fórmula Alfa de Cronbach; estas pruebas consignan el grado en que covarían los ítems que forman una prueba o test. Cuanto más se acerque a la unidad, más confiable se podrá considerar la prueba o el test (Backhoff, Larrazolo y Rosas, 2000; Villamarín, 2017).

Rangos de confiabilidad	Valoración
0,8 – 1,00	Muy alta
0,61 -0,79	Alta
0,41 – 0,60	Moderada
0,21 – 0,40	Baja
0,01 – 0,20	Muy baja

2.3 Definición de Términos

- **Evaluación**

  Proceso mediante el cual se observan y analizan las características y condiciones de un objeto, en función de unos criterios o puntos de referencia para emitir un juicio de valor que permita tomar una decisión adecuada.

- **Evaluación del aprendizaje**

  Es la evaluación del rendimiento académico del estudiante en función de las competencias, objetivos y contenidos, previamente determinados, que están enmarcados en el proceso educativo.

- **Instrumento de medición**

  En la evaluación educativa es sinónimo de prueba (lo que se usa para demostrar algo o para dar testimonio) o de examen (lo que se usa para observar o conocer algo); puede ser también un cuestionario (se emplea

para sondear y conocer opiniones) o, también, puede ser una escala (por ejemplo, una escala de actitudes o motivación). Un instrumento se puede aplicar a una población o a una muestra de ella para medircaracterísticas, atributos o rasgos de interés que se desprenden de un constructo.

- **Índice de dificultad**
  Proporción de personas que responden erróneamente una pregunta (reactivo).

- **Índice de discriminación**
  Proporción de la diferencia entre las personas de los grupos alto y bajo que responden correctamente una pregunta (reactivo).

- **Índice de confiabilidad**
  Precisión con la que un instrumento mide una variable.

CAPÍTULO 3: HIPÓTESIS Y VARIABLES

3.1 Hipótesis

3.1.1 Hipótesis General

Los instrumentos de evaluación de las asignaturas del área de Ciencias Básicas de la Facultad de Odontología de la Universidad Nacional Mayor de San Marcos aplicados durante el año lectivo 2016 tienen diferencias en cuanto a su calidad.

3.1.2 Hipótesis Específicos

- **Hipótesis Específica N°01: Según el índice de dificultad**
  - **Ho:** Según el índice de dificultad, no existen diferencias en la calidad de los instrumentos de evaluación de las asignaturas del área de Ciencias Básicas de la Facultad de Odontología de la Universidad Nacional Mayor de San Marcos durante el periodo lectivo 2016.

  - **H1:** Según el índice de dificultad, existen diferencias en la calidad de los instrumentos de evaluación de las asignaturas del área de Ciencias Básicas de la Facultad de Odontología de la Universidad Nacional Mayor de San Marcos durante el periodo lectivo 2016.

Ho: Me1 = Me2 = Me3 = Me4 = Me5

H1: Me1 ≠ Me2 ≠ Me3 ≠ Me4 ≠ Me5; al menos una mediana diferente.

Nivel de significancia: α = 0.05; nivel de confianza: 0.95

Regla de decisión: si p menor de 0.05 se rechaza la hipótesis nula.

Cálculos: H de kruskall-Wallis

Decisión:

Conclusión:

- **Hipótesis Específica N°02: Según el índice de discriminación**
  - **Ho:** Según el índice de discriminación, no existen diferencias en la calidad de los instrumentos de evaluación de las asignaturas del área de Ciencias Básicas de la Facultad de Odontología de la Universidad Nacional Mayor de San Marcos durante el periodo lectivo 2016.

  - **H1:** Según el índice de discriminación, existen diferencias en la calidad de los instrumentos de evaluación de las asignaturas del área de Ciencias Básicas de la Facultad de Odontología de la Universidad Nacional Mayor de San Marcos durante el periodo lectivo 2016.

Ho: Me1 = Me2 = Me3 = Me4 = Me5

H1: Me1 ≠ Me2 ≠ Me3 ≠ Me4 ≠ Me5; al menos una mediana diferente

Nivel de significancia: α = 0.05; nivel de confianza: 0.95

Regla de decisión: si p menor de 0.05 se rechaza la hipótesis nula.

Cálculos: H de kruskall-Wallis

Decisión:

Conclusión:

3.2 Identificación de variables

- Calidad de los instrumentos de evaluación
- Asignaturas del área de ciencias básicas

3.3 Operacionalización de Variables

Variable	Definición	Dimensiones	Indicadores	Escala	Valores
Calidad de los instrumentos de evaluación.	Característica de un examen que posee reactivos con adecuada dificultad, discriminación y confiabilidad.	Grado de dificultad	Índice de dificultad	Ordinal	• Mayor a 0,87 altamente difícil. • 0,74 a 0,86 medianamente difícil. • 0,53 a 0,73 dificultad media. • 0,33 a 0,52 medianamente fácil. • Menor a 0,32 altamente fácil.
		Grado de discriminación	Índice de discriminación	Ordinal	• Mayor a 0,39 excelente. • 0,30 a 0,38 buena. • 0,20 a 0,29 regular. • 0,00 a 0,19 pobre. • Negativo descartar.
		Grado de confiabilidad	Índice de confiabilidad por alfa de Cronbach	Ordinal	• 0,8 a 1,0 muy alta • 0,61 a 0,79 alta • 0,41 a 0,60 moderada • 0,21 a 0,40 baja • 0,01 a 0,20 muy baja

Variable	Definición	Dimensiones	Indicador	Escala
Asignaturas del área de ciencias básicas	Asignaturas que estudian las estructuras normales del cuerpo humano, así como las modificaciones fisiológicas y patológicas.	• Asignaturas	• Anatomía General. • Bioquímica General. • Microbiología General. • Fisiología General. • Farmacología II.	Nominal

CAPÍTULO 4: METODOLOGÍA

4.1 Tipo y Diseño de Investigación

La presente investigación es de tipo observacional; por el alcance de sus resultados es descriptivo comparativo; por la ocurrencia de los hechos es retrospectivo; y, transversal por la toma de la muestra, que fue en un solo momento.

4.2 Unidad de análisis

La unidad de análisis del presente estudio estuvo representada por las preguntas o ítems que conformaron cada uno de los exámenes teóricos escritos que se tomaron a los estudiantes que cursaron las asignaturas del área de ciencias básicas de la Carrera Profesional de Odontología de la UNMSM. Tales asignaturas fueron: Anatomía General, Bioquímica General, Microbiología General, Fisiología General y Farmacología II, durante el año lectivo 2016.

La cantidad total de preguntas que se analizaron fueron de 8440, distribuidas de la siguiente manera:

- Anatomía General : 2930 preguntas
- Bioquímica General : 2140 preguntas
- Microbiología General : 1860 preguntas
- Fisiología General : 890 preguntas
- Farmacología II : 620 preguntas

4.3 Población de estudio

La población de estudio estuvo constituida por los exámenes teóricos escritos que rindieron los alumnos matriculados en los cursos del área de ciencias básicas durante el año lectivo 2016:

- Anatomía General : 97 exámenes de 30 preguntas
- Bioquímica General : 107 exámenes de 20 preguntas
- Microbiología General : 93 exámenes de 20 preguntas
- Fisiología General : 89 exámenes de 10 preguntas
- Farmacología II : 62 exámenes de 10 preguntas

4.4 Tamaño y Selección de la Muestra

El tamaño de la muestra de la presente investigación se seleccionó en forma no probabilística; estuvo conformada por los exámenes teóricos escritos de las asignaturas del área de ciencias básicas y sus respectivas preguntas, que rindieron los alumnos matriculados en los cursos del área de ciencias básicas durante el año 2016 y que cumplieron con los criterios de inclusión y exclusión.

- **Criterios de Inclusión**
  - Exámenes de estudiantes matriculados durante el año lectivo 2016.
  - Exámenes de estudiantes que cursaron por primera vez los cursos de ciencias básicas incluidas en esta investigación.

- **Criterios de Exclusión:**
  - Exámenes de estudiantes con segunda o más matrículas en las asignaturas de ciencias básicas incluidas en esta investigación.

El número de exámenes que formaron parte de esta investigación fue de 448, y el total de preguntas que se analizaron fueron de 8420.

4.5 Técnica de recolección de datos

En la presente investigación la recolección de datos se realizó de las pruebas teóricas aplicados a los alumnos matriculados dentro de las asignaturas de Anatomía General, Bioquímica General, Microbiología General, Fisiología General y Farmacología II, durante el año lectivo 2016. Los datos fueron registrados en una ficha de recolección en la que se analizaron las respuestas dadas a las preguntas (reactivos) de los exámenes.

Para cada pregunta (reactivo) que formó parte de un examen se encontró su respectivo índice de dificultad que nos permitió catalogarlo como altamente fácil, medianamente fácil, dificultad media, medianamente difícil o altamente difícil; para determinar el índice de dificultad, se utilizó la siguiente fórmula, basada en Hurtado (2018)

$$IDif = 1 - C/N$$

Dónde:		
C	=	número de aciertos en el reactivo

Luego, se obtuvo el índice de dificultad de toda la prueba aplicando la siguiente fórmula, basada en Morales.

$$\text{Índice de dificultad de la prueba} = \frac{\text{media}}{\text{\# de preguntas}}$$

Del mismo modo se obtuvo de cada pregunta (reactivo) que conformó un examen el índice de discriminación, lo cual permitió valorarlos como excelentes, buenas, regulares, pobres o descartables; para obtenerlos se aplicó la siguiente fórmula: (Backhoff, 2000)

$$\mathbf{DI} = \frac{\mathbf{GA\ aciertos - GB\ aciertos}}{\mathbf{N\ grupo\ mayor}}$$

Dónde:

D_i = Índice de discriminación del reactivo

$GA_{aciertos}$ = Número de aciertos en el reactivo del 25% de personas con las puntuaciones más altas en el test.

$GB_{aciertos}$ = Número de aciertos en el reactivo del 25% de personas con las puntuaciones más bajas en el test.

$N_{grupo\ mayor}$ = Número de personas en el grupo más numeroso (GA o GB).

El índice de discriminación de toda la prueba se obtuvo de la siguiente manera

$$\text{Índice de discriminación} = \frac{\text{puntuación más alta} - \text{puntuación más baja}}{\text{Número de preguntas}}$$

Con los resultados de los índices de dificultad y de discriminación obtenidos se identificaron que preguntas (reactivos) de un examen deben ser revisados, corregidos o eliminados por ser muy difíciles, muy fáciles o que no discriminan adecuadamente.

En relación a la confiabilidad, ésta se obtuvo aplicando la fórmula matemática Alfa de Cronbach; valores mayores a 0.6 permitieron catalogar a los exámenes como confiables.

.

Obtenidos los resultados y determinados los valores de dificultad, de discriminación y de confiabilidad de las preguntas (reactivos), se procedió a compararlos para determinar las diferencias que existieron entre los exámenes tomados a los estudiantes que cursaron las asignaturas del área de ciencias básicas durante el año lectivo 2016.

4.6 Análisis e interpretación de la información

Los datos recogidos en las fichas de recolección fueron procesados, organizados, sistematizados, analizados e interpretados. Para ello se confeccionaron tablas uni y bidimensionales, indicadores y pruebas estadísticas, utilizando los programas estadísticos SPSS y STATA.

CAPÍTULO 5: RESULTADOS

El propósito del estudio fue evaluar y comparar la calidad de los instrumentos de evaluación de las asignaturas del área de Ciencias Básicas de la Facultad de Odontología de la Universidad Nacional Mayor de San Marcos durante el periodo lectivo 2016. Obtenidos los resultados que estuvieron en relación a las dimensiones que se planteó en la investigación, se realizó el análisis de los datos que se obtuvieron en la ejecución, con el apoyo del asesor y un especialista en estadística, se elaboraron tablas y gráficos que se muestran a continuación. Así como la prueba de hipótesis.

5.1 Resultados

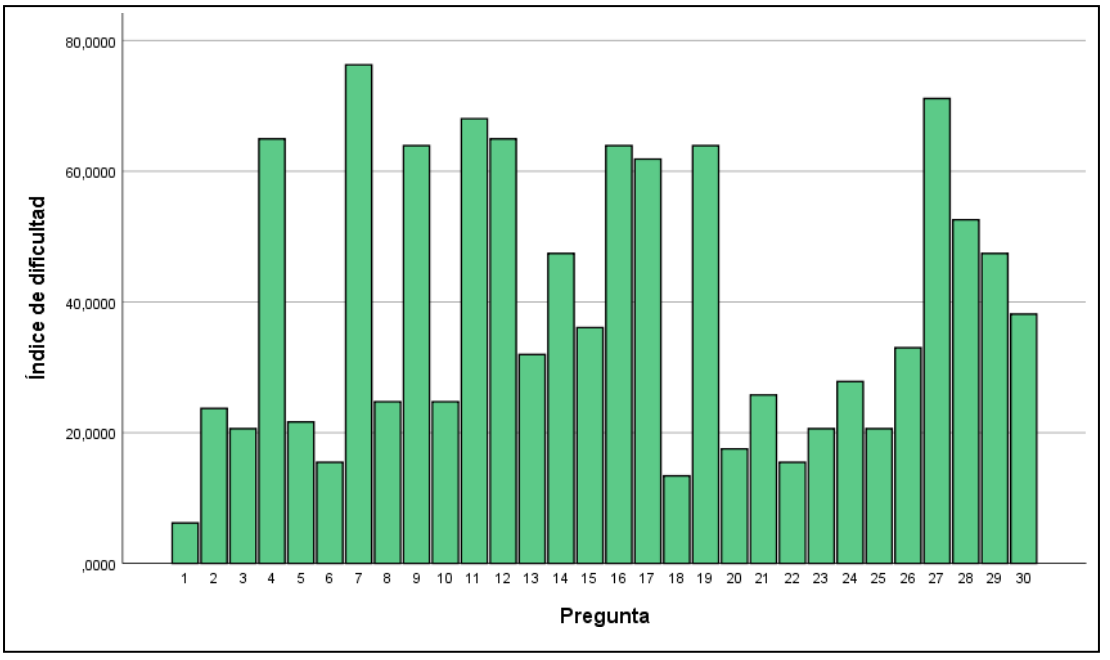
5.1.1 Anatomía General

Índice de dificultad

Tabla 1. Índice de dificultad de las preguntas de Anatomía General

Pregunta	C/97	IDif	Categoría de los ítems
1	91	.062	Altamente fácil
2	74	.237	Altamente fácil
3	77	.206	Altamente fácil
4	34	.649	Dificultad media
5	76	.216	Altamente fácil
6	82	.155	Altamente fácil
7	23	.763	Difícil
8	73	.247	Altamente fácil
9	35	.639	Dificultad media
10	73	.247	Altamente fácil
11	31	.680	Dificultad media
12	34	.649	Dificultad media
13	66	.320	Altamente fácil
14	51	.474	Medianamente fácil
15	62	.361	Medianamente fácil
16	35	.639	Dificultad media
17	37	.619	Dificultad media
18	84	.134	Altamente fácil
19	35	.639	Dificultad media
20	80	.175	Altamente fácil
21	72	.258	Altamente fácil
22	82	.155	Altamente fácil
23	77	.206	Altamente fácil
24	70	.278	Altamente fácil
25	77	.206	Altamente fácil
26	65	.330	Medianamente fácil
27	28	.711	Dificultad media
28	46	.526	Dificultad media
29	51	.474	Medianamente fácil
30	60	.381	Medianamente fácil

Figura 1. Índice de dificultad de las preguntas de Anatomía General

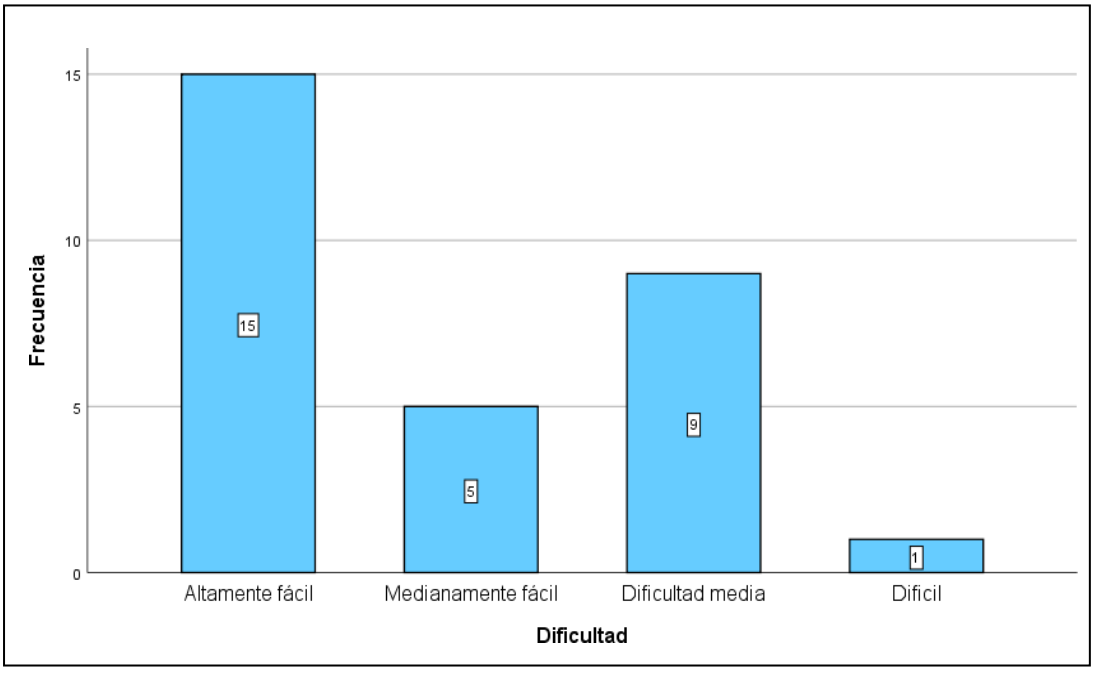


Las preguntas que tuvieron mayor índice de dificultad fueron la pregunta P7 (0,763), P27 (0,711) y P11 (0,680). Las preguntas que tuvieron menor índice de dificultad fueron la pregunta P1 (0,062), P18 (0,134) y P6 (0,155).

Tabla 2. Dificultad por categorías de Anatomía General

	Frecuencia	Porcentaje válido	Porcentaje acumulado
Altamente fácil	15	50.0	50.0
Medianamente fácil	5	16.7	66.7
Dificultad media	9	30.0	96.7
Difícil	1	3.3	100.0
Total	30	100.0	

Figura 2. Dificultad por categorías de Anatomía General

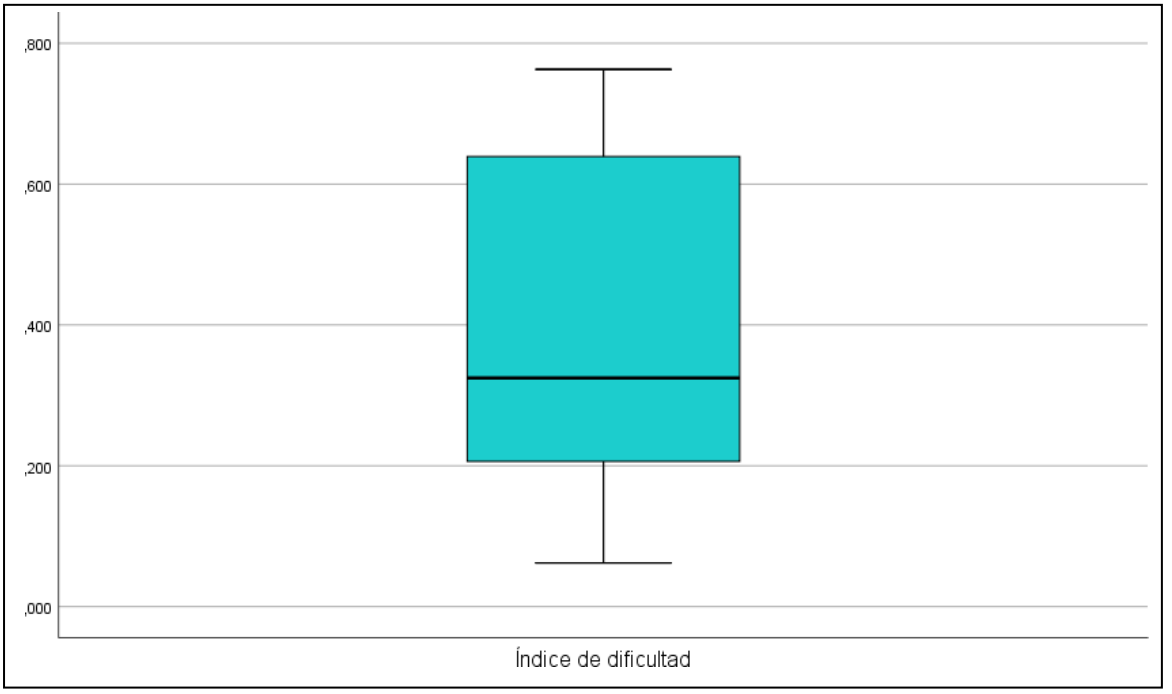


El 50% de las preguntas (15) puede ser catalogado como Altamente fácil y el 30% como Dificultad media (9). Solo el 3,3% fue catalogado como Difícil (1).

Tabla 3. Global Índice de dificultad de Anatomía General

	N	Media	Desv. Error	Mínimo - Máximo	Mediana	Dificultad
Índice de dificultad	30	.38	.03	.06 -.76	.32	Medianamente fácil

Figura 3. Global Índice de dificultad de Anatomía General



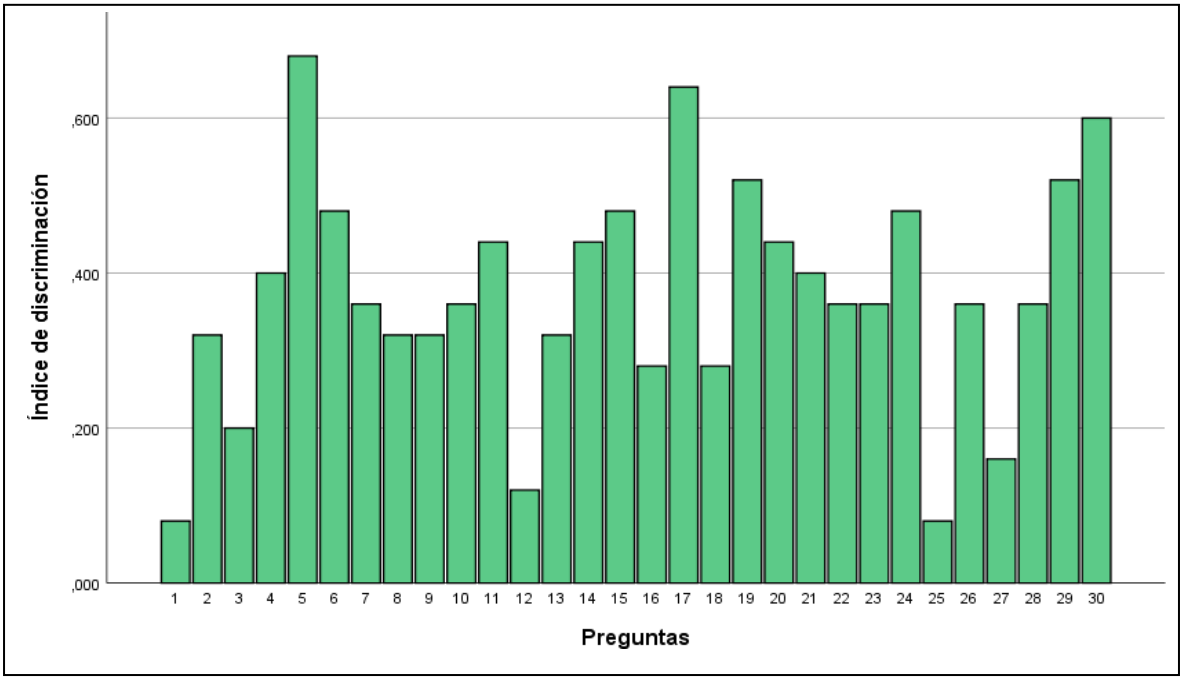
El índice global de dificultad tuvo una media de 0,38±03 un valor que puede ser catalogado como Medianamente fácil. La mediana global fue 0,32.

Índice de discriminación

Tabla 4. Índice de discriminación de Anatomía General

Pregunta	G_S	G_i	IDisc	Calidad	Recomendación
1	24	22	.080	Pobre	Revisar a profundidad
2	23	15	.320	Buena	Mejorar
3	22	17	.200	Regular	Revisar
4	13	3	.400	Excelente	Conservar
5	24	7	.680	Excelente	Conservar
6	25	13	.480	Excelente	Conservar
7	10	1	.360	Buena	Mejorar
8	23	15	.320	Buena	Mejorar
9	15	7	.320	Buena	Mejorar
10	22	13	.360	Buena	Mejorar
11	13	2	.440	Excelente	Conservar
12	10	7	.120	Pobre	Revisar a profundidad
13	22	14	.320	Buena	Mejorar
14	18	7	.440	Excelente	Conservar
15	21	9	.480	Excelente	Conservar
16	12	5	.280	Regular	Revisar
17	18	2	.640	Excelente	Conservar
18	24	17	.280	Regular	Revisar
19	17	4	.520	Excelente	Conservar
20	25	14	.440	Excelente	Conservar
21	23	13	.400	Excelente	Conservar
22	24	15	.360	Buena	Mejorar
23	23	14	.360	Buena	Mejorar
24	22	10	.480	Excelente	Conservar
25	19	17	.080	Pobre	Revisar a profundidad
26	20	11	.360	Buena	Mejorar
27	10	6	.160	Pobre	Revisar a profundidad
28	17	8	.360	Buena	Mejorar
29	20	7	.520	Excelente	Conservar
30	22	7	.600	Excelente	Conservar

Figura 4. Índice de discriminación de Anatomía General

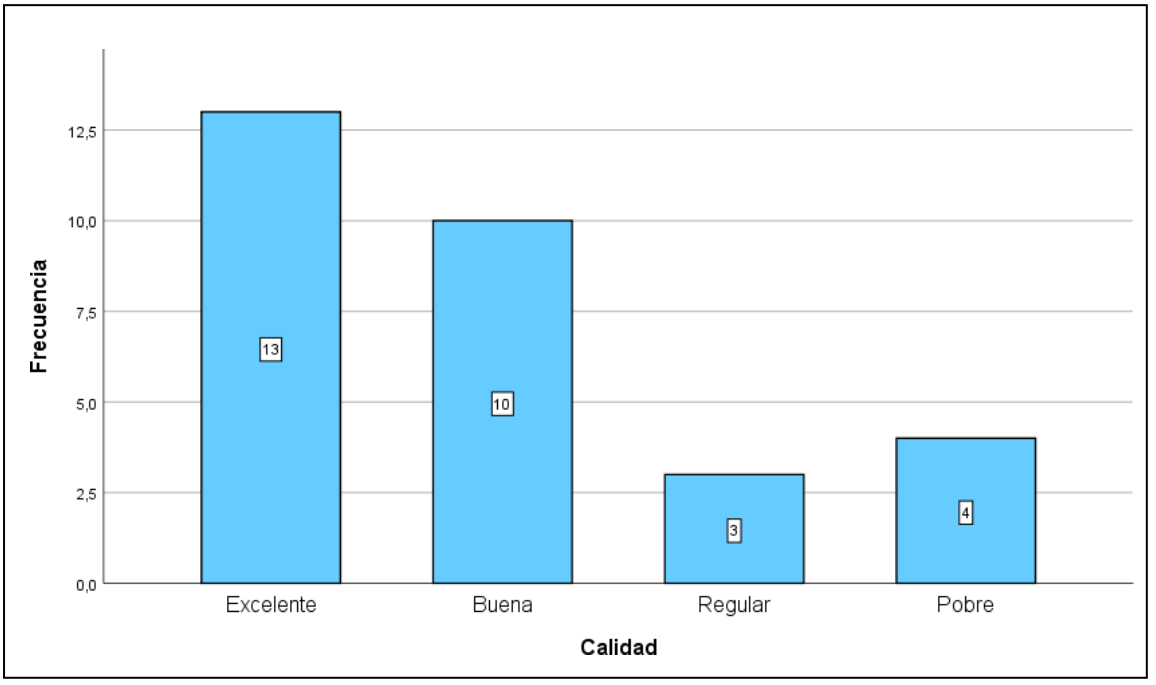


Las preguntas que tuvieron mayor índice de discriminación fueron la pregunta P5 (0,680), P17 (0,640) y P30 (0,600). Las preguntas que tuvieron menor índice de discriminación fueron la pregunta P1 (0,080), P25 (0,080) y P12 (0,120).

Tabla 5. Calidad de las preguntas de Anatomía General

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado	Recomendación
Excelente	13	43.3	43.3	43.3	Conservar
Buena	10	33.3	33.3	76.7	Mejorar
Regular	3	10.0	10.0	86.7	Revisar
Pobre	4	13.3	13.3	100.0	Revisar a profundidad
Total	30	100.0	100.0		

Figura 5. Calidad de las preguntas de Anatomía General

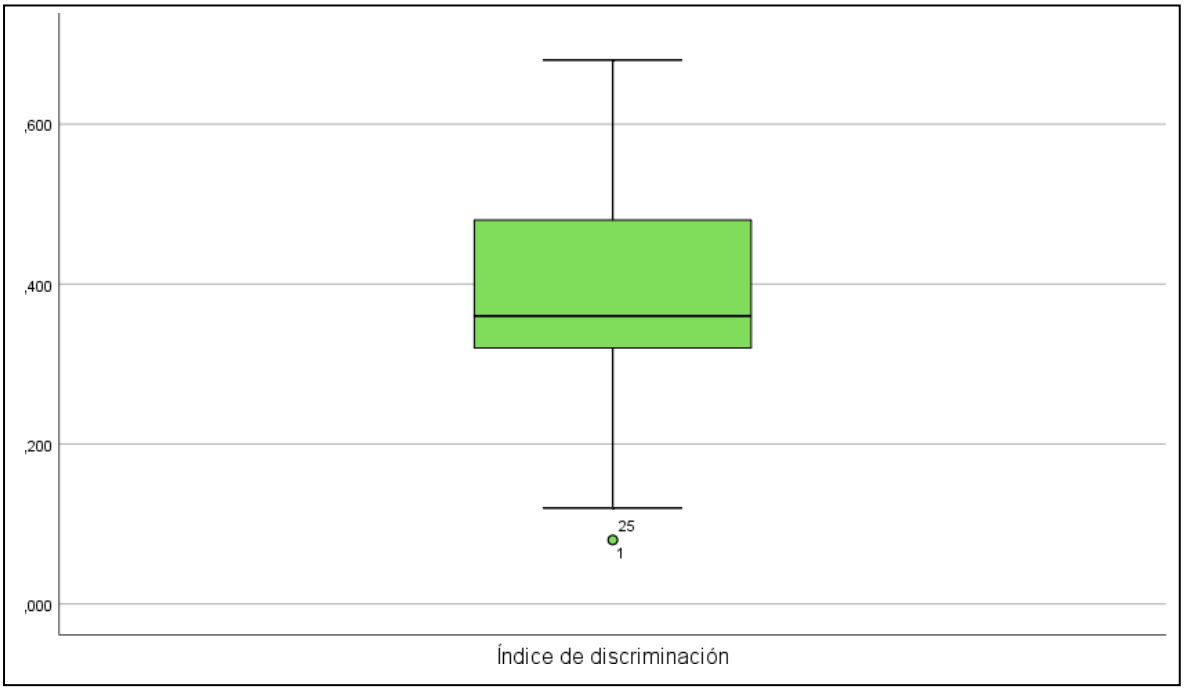


El 43,3% de las preguntas (13) puede ser catalogado con una calidad Excelente y el 33,3% con una calidad Buena (10). Solo el 13,3% fue catalogado como Pobre (4).

Tabla 6. Global Índice de discriminación de Anatomía General

	N	Media	Desv. Error	Mínimo - Máximo	Mediana	Calidad
Índice de discriminación	30	.37	.02	.08 -.68	.36	Buena

Figura 6. Global Índice de discriminación de Anatomía General

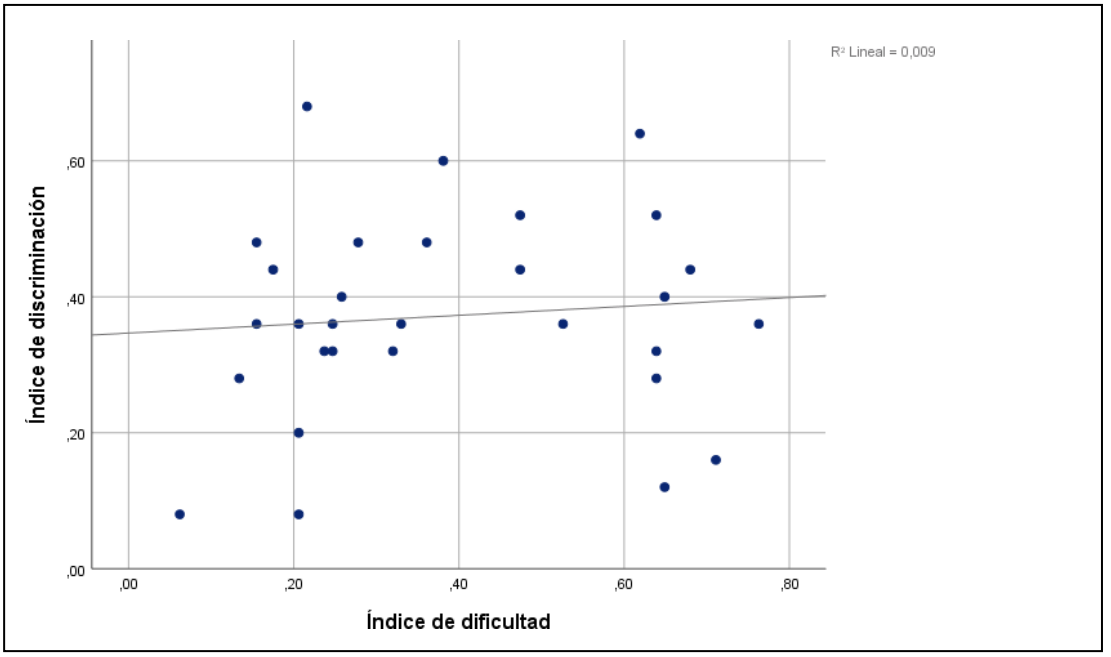


El índice global de discriminación tuvo una media de 0,37±02 un valor que puede ser catalogado con una calidad de Buena. La mediana global fue 0,36.

Tabla 7. Correlación índice dificultad x índice discriminación de Anatomía General

			Índice de dificultad	Índice de discriminación
Rho de Spearman	Índice de dificultad	Coeficiente de correlación	1.000	.132
		Sig. (bilateral)	.	.486
		N	30	30
	Índice de discriminación	Coeficiente de correlación	.132	1.000
		Sig. (bilateral)	.486	.
		N	30	30

Figura 7. Correlación índice dificultad x índice discriminación de Anatomía General



Existe una baja correlación no significativa entre el índice de dificultad y el índice de discriminación (r=0,132; p - value =0,486. Existen pocos valores con discriminación baja y dificultad baja y pocos valores con discriminación alta y dificultad alta. La mayoría de las preguntas son fáciles, pero con una buena discriminación.

Confiabilidad de Anatomía General

Estadísticas de fiabilidad	
Alfa de Cronbach	N de elementos
0,699	30

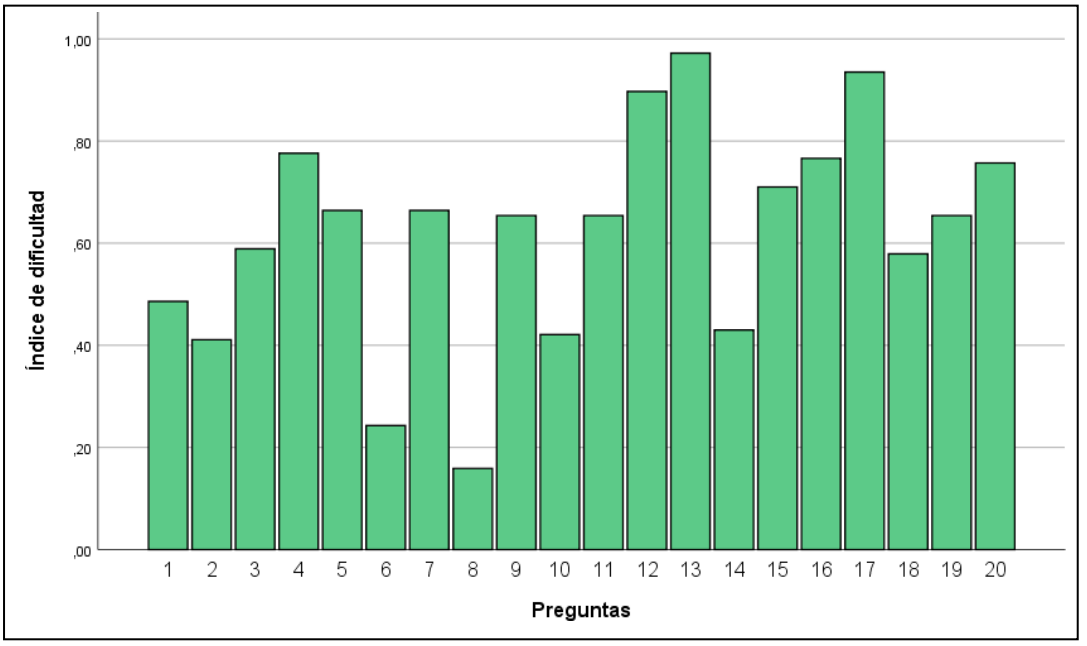
5.1.2 Bioquímica General

Índice de dificultad

Tabla 8. Índice de dificultad de las preguntas de Bioquímica General

Pregunta	C/107	IDif	Categoría de los ítems
1	55	0.486	Medianamente fácil
2	63	0.411	Medianamente fácil
3	44	0.589	Dificultad media
4	24	0.776	Difícil
5	36	0.664	Dificultad media
6	81	0.243	Altamente fácil
7	36	0.664	Dificultad media
8	90	0.159	Altamente fácil
9	37	0.654	Dificultad media
10	62	0.421	Medianamente fácil
11	37	0.654	Dificultad media
12	11	0.897	Altamente difícil
13	3	0.972	Altamente difícil
14	61	0.430	Medianamente fácil
15	31	0.710	Dificultad media
16	25	0.766	Difícil
17	7	0.935	Altamente difícil
18	45	0.579	Dificultad media
19	37	0.654	Dificultad media
20	26	0.757	Difícil

Figura 8. Índice de dificultad de las preguntas de Bioquímica General

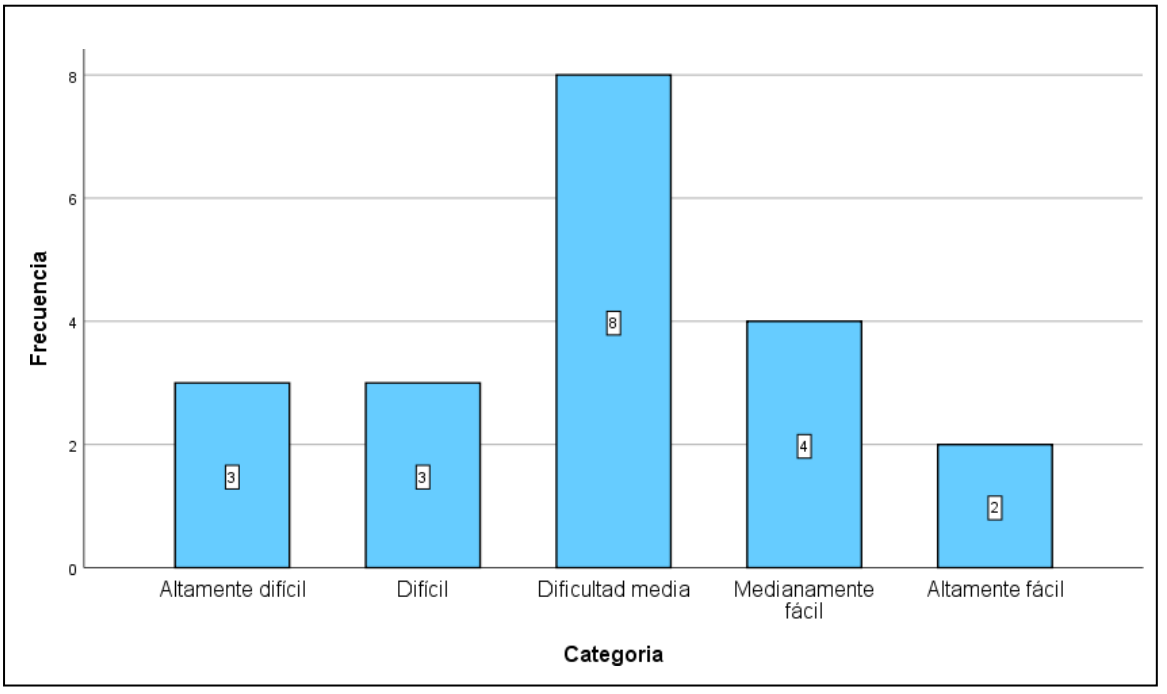


Las preguntas que tuvieron mayor índice de dificultad fueron la pregunta P13 (0,972), P17 (0,935) y P12 (0,897). Las preguntas que tuvieron menor índice de dificultad fueron la pregunta P8 (0,159) y P6 (0,243).

Tabla 9. Dificultad por categorías de Bioquímica General

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Altamente difícil	3	15.0	15.0	15.0
Difícil	3	15.0	15.0	30.0
Dificultad media	8	40.0	40.0	70.0
Medianamente fácil	4	20.0	20.0	90.0
Altamente fácil	2	10.0	10.0	100.0
Total	20	100.0	100.0	

Figura 9. Dificultad por categorías de Bioquímica General

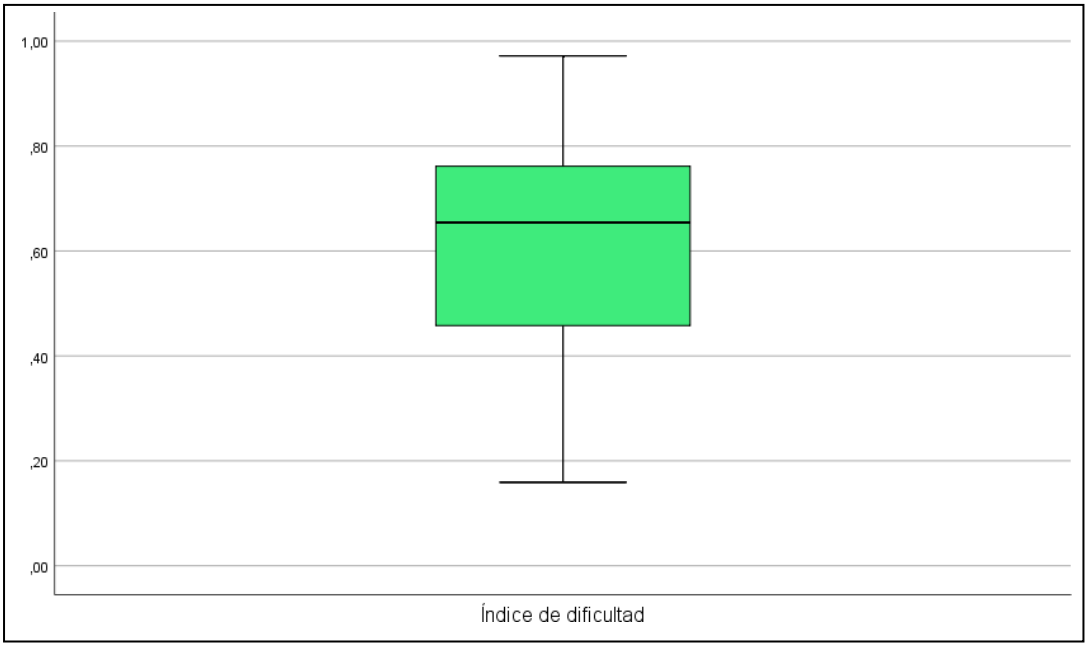


El 40% de las preguntas (8) puede ser catalogado como de Dificultad media y el 20% como Medianamente fácil (4). Solo el 10% fue catalogado como Altamente fácil (2).

Tabla 10. Global Índice de dificultad de Bioquímica General

	N	Media	Desv. Error	Mínimo - Máximo	Mediana	Dificultad
Índice de dificultad	20	.62	.04	.16 - .97	.65	Dificultad media

Figura 10. Global Índice de dificultad de Bioquímica General



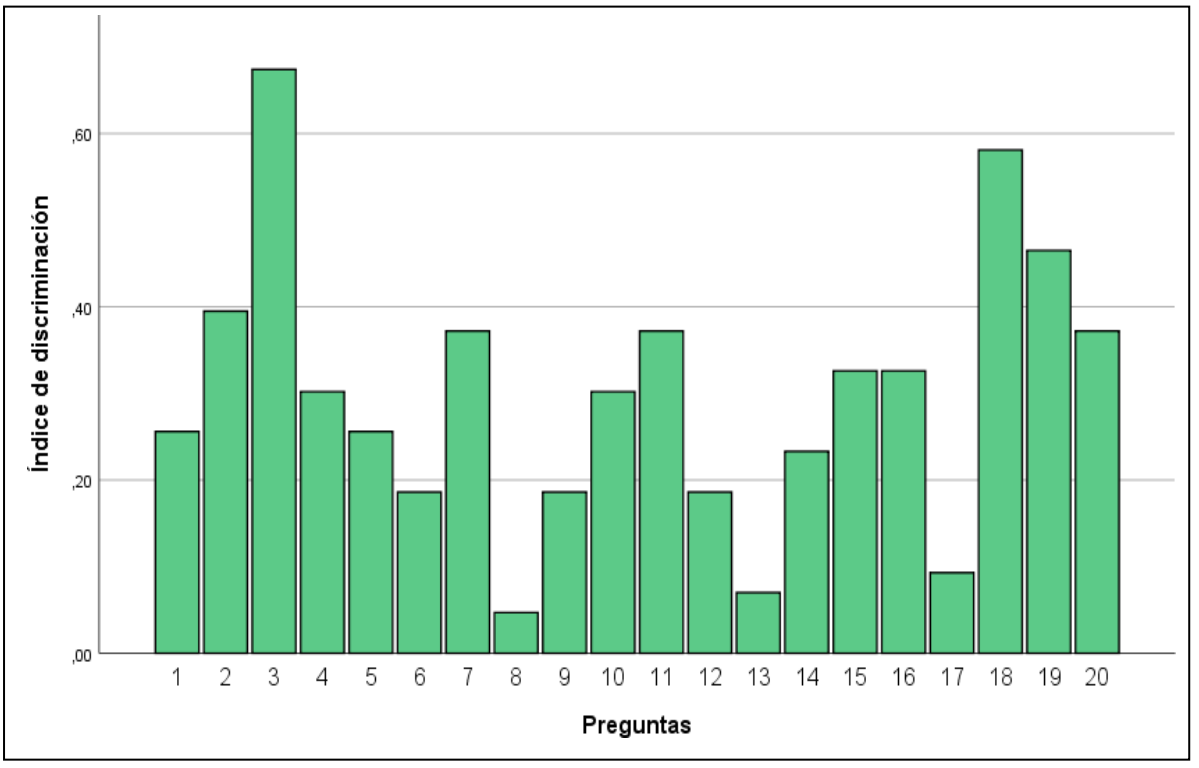
El índice global de dificultad tuvo una media de 0,62±04 un valor que puede ser catalogado como de Dificultad media. La mediana global fue 0,56.

Índice de discriminación

Tabla 11. Índice de discriminación de Bioquímica General

Pregunta	G_S	G_i	IDisc	Calidad	Recomendación
1	28	17	0.256	Regular	Revisar
2	34	17	0.395	Excelente	Conservar
3	32	3	0.674	Excelente	Conservar
4	13	0	0.302	Buena	Mejorar
5	20	9	0.256	Regular	Revisar
6	34	26	0.186	Pobre	Revisar a profundidad
7	23	7	0.372	Buena	Mejorar
8	37	35	0.047	Pobre	Revisar a profundidad
9	18	10	0.186	Pobre	Revisar a profundidad
10	30	17	0.302	Buena	Mejorar
11	21	5	0.372	Buena	Mejorar
12	9	1	0.186	Pobre	Revisar a profundidad
13	3	0	0.070	Pobre	Revisar a profundidad
14	29	19	0.233	Regular	Revisar
15	18	4	0.326	Buena	Mejorar
16	18	4	0.326	Buena	Mejorar
17	5	1	0.093	Pobre	Revisar a profundidad
18	30	5	0.581	Excelente	Conservar
19	24	4	0.465	Excelente	Conservar
20	19	3	0.372	Buena	Mejorar

Figura 11. Índice de discriminación de Bioquímica General

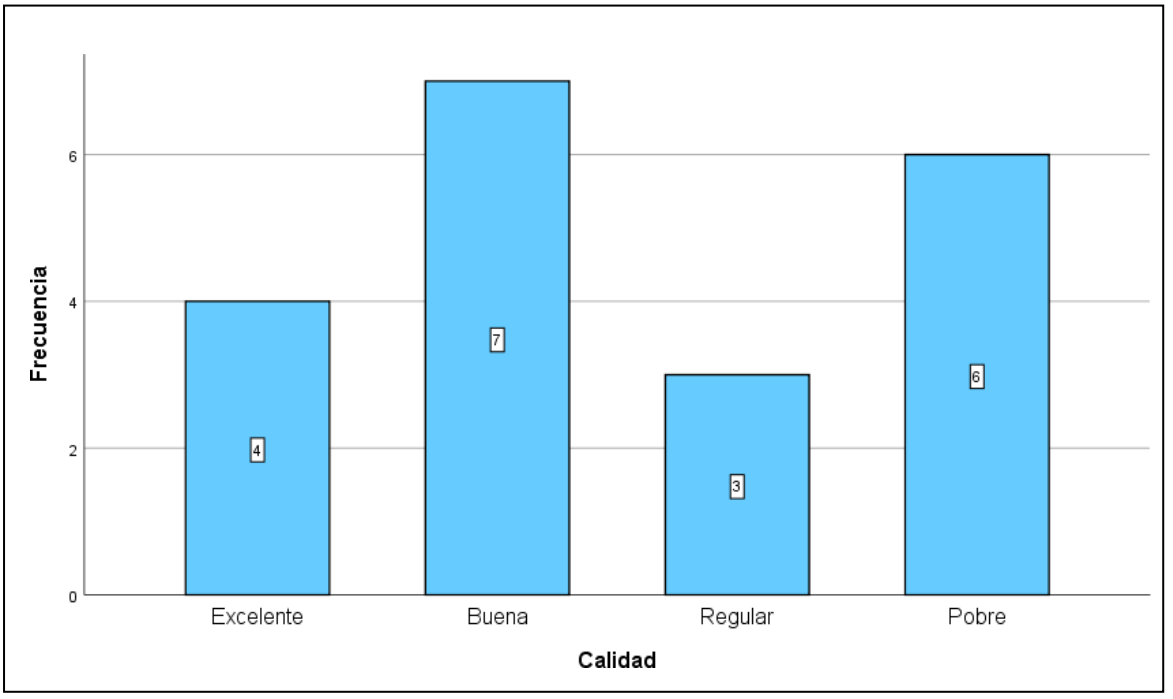


Las preguntas que tuvieron mayor índice de discriminación fueron la pregunta P3 (0,674), P18 (0,581) y P19 (0,465). Las preguntas que tuvieron menor índice de discriminación fueron la pregunta P8 (0,047), P13 (0,070) y P17 (0,093).

Tabla 12. Calidad de las preguntas de Bioquímica General

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado	Recomendación
Excelente	4	20.0	20.0	20.0	Conservar
Buena	7	35.0	35.0	55.0	Mejorar
Regular	3	15.0	15.0	70.0	Revisar
Pobre	6	30.0	30.0	100.0	Revisar a profundidad
Total	20	100.0	100.0		

Figura 12. Calidad de las preguntas de Bioquímica General

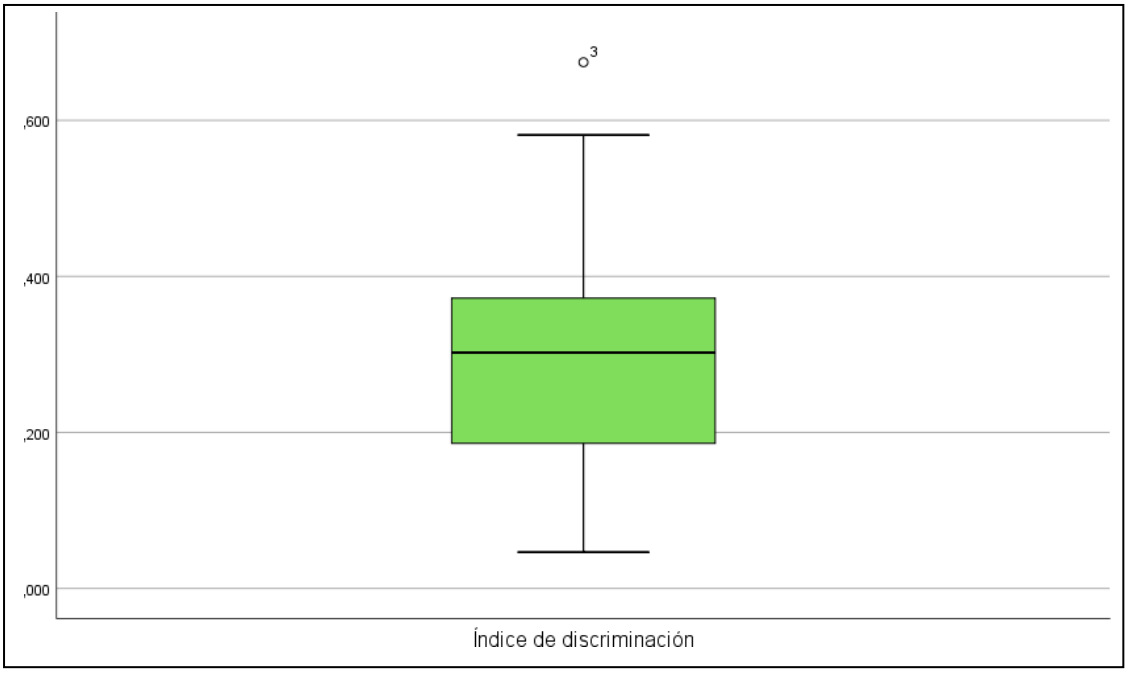


El 35% de las preguntas (7) puede ser catalogado con una calidad Buena y el 30% con una calidad Pobre (6). El 15% fue catalogado como Regular (3).

Tabla 13. Global Índice de discriminación de Bioquímica General

	N	Media	Desv. Error	Mínimo - Máximo	Mediana	Calidad
Índice de discriminación	20	.300	.035	.047 - .674	.302	Buena

Figura 13. Global Índice de discriminación de Bioquímica General

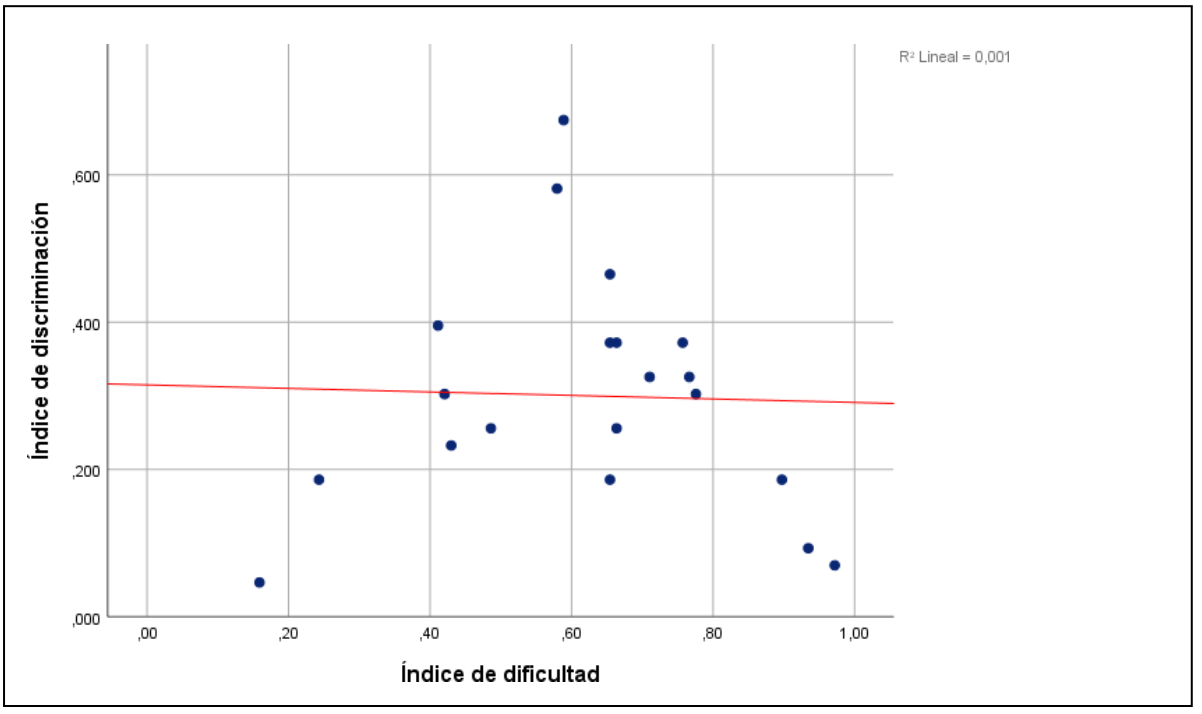


El índice global de discriminación tuvo una media de 0,30±03 un valor que puede ser catalogado con una calidad Buena. La mediana global fue 0,302.

Tabla 14. Correlación índice dificultad x índice discriminación de Bioquímica General

			Índice de dificultad	Índice de discriminación
Rho de Spearman	Índice de dificultad	Coeficiente de correlación	1.000	-.125
		Sig. (bilateral)	.	.599
		N	20	20
	Índice de discriminación	Coeficiente de correlación	-.125	1.000
		Sig. (bilateral)	.599	.
		N	20	20

Figura 14. Correlación índice dificultad por índice discriminación de Bioquímica General



Existe una correlación negativa no significativa entre el índice de dificultad y el índice de discriminación (r=-0,125; p-value=0,599). La mayoría de las preguntas son de dificultad media y discriminación media.

Confiabilidad de Bioquímica General

Estadísticas de fiabilidad	
Alfa de Cronbach	N de elementos
0,548	20

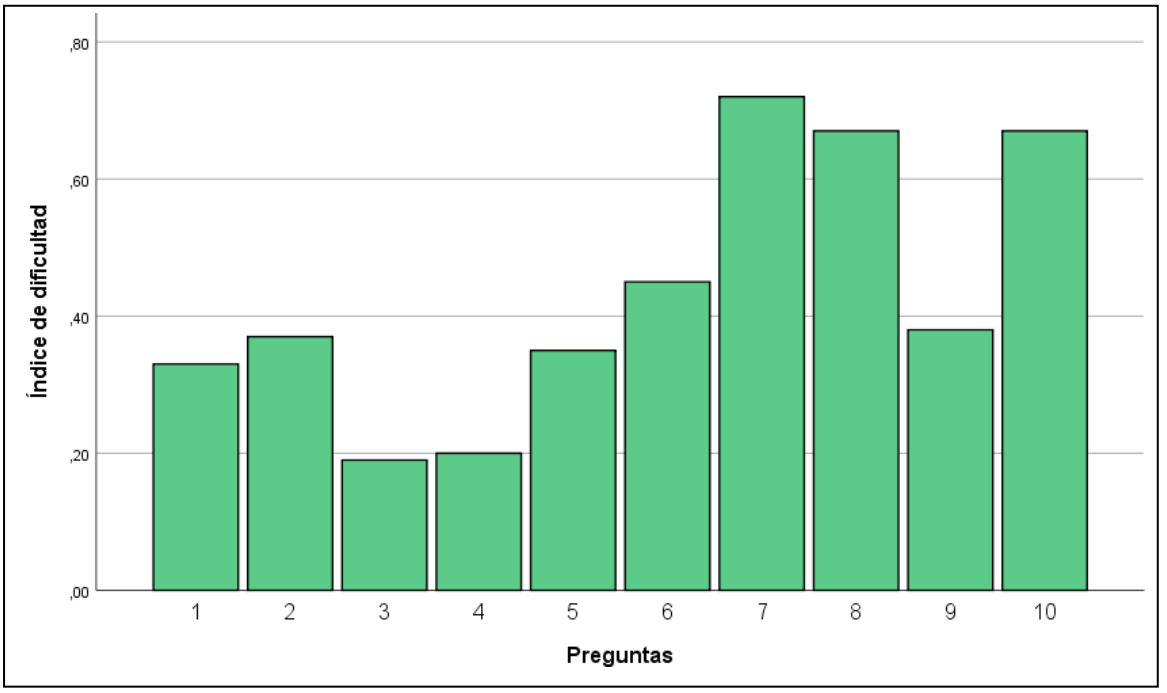
5.1.3 Farmacología II

Índice de dificultad

Tabla 15. Índice de dificultad de las preguntas de Farmacología II

Pregunta	C/62	IDif	Categoría de los items
1	41	0,33	Medianamente fácil
2	39	0,37	Medianamente fácil
3	50	0,19	Altamente fácil
4	49	0,20	Altamente fácil
5	40	0,35	Medianamente fácil
6	34	0,45	Medianamente fácil
7	17	0,72	Dificultad media
8	20	0,67	Dificultad media
9	38	0,38	Medianamente fácil
10	20	0,67	Dificultad media

Figura 15. Índice de dificultad de las preguntas de Farmacología II

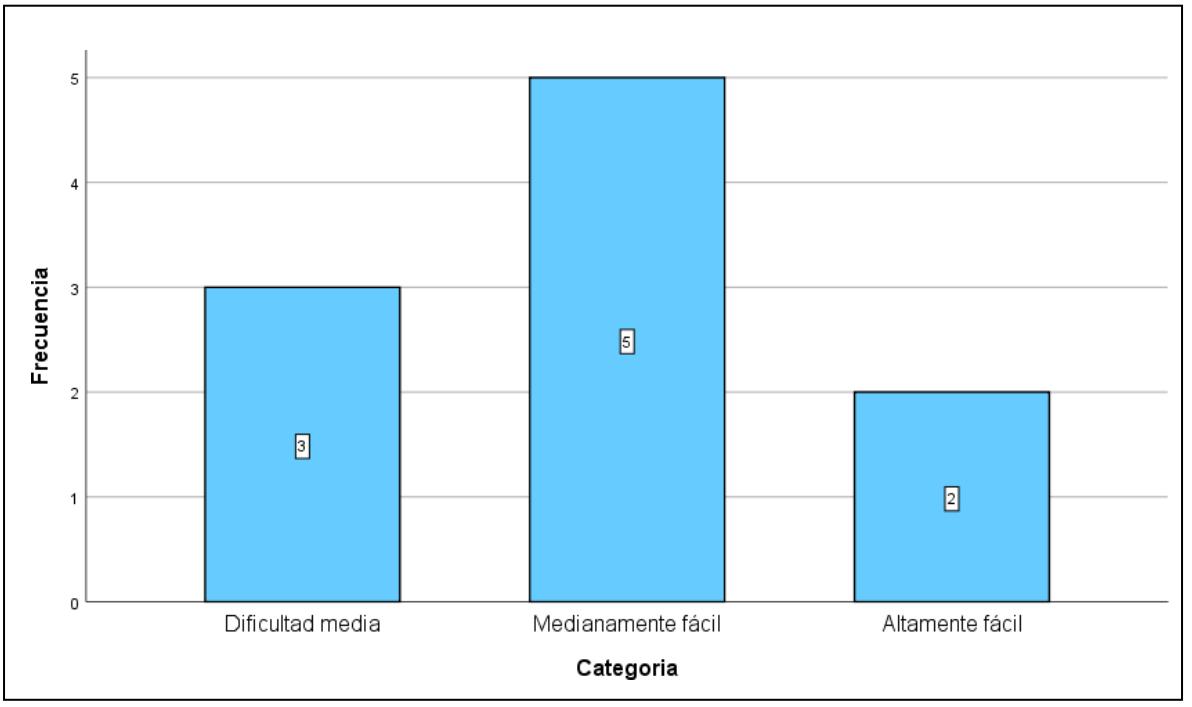


Las preguntas que tuvieron mayor índice de dificultad fueron la pregunta P7 (0,72), P8 (0,67) y P10 (0,67). Las preguntas que tuvieron menor índice de dificultad fueron la pregunta P3 (0,19), P4 (0,20) y P5 (0,35).

Tabla 16. Dificultad por categorías de Farmacología II

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Dificultad media	3	30.0	30.0	30.0
	Medianamente fácil	5	50.0	50.0	80.0
	Altamente fácil	2	20.0	20.0	100.0
	Total	10	100.0	100.0	

Figura 16. Dificultad por categorías de Farmacología II

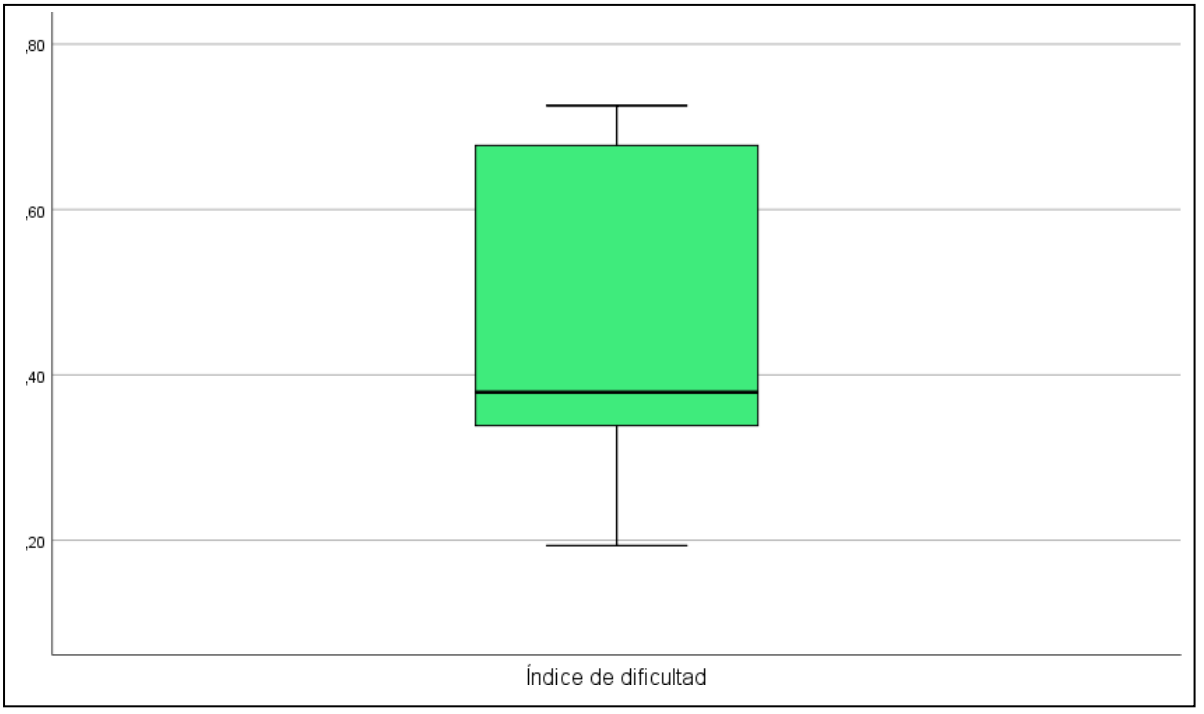


El 50% de las preguntas (5) puede ser catalogado como Medianamente fácil y el 30% como Dificultad media (3) y el 20% fue catalogado como Altamente fácil (2).

Tabla 17. Global Índice de dificultad de Farmacología II

	N	Media	Desv. Error	Mínimo - Máximo	Mediana	Dificultad
Índice de dificultad	10	.43	.06	.19 - .73	.37	Medianamente fácil

Figura 17. Global Índice de dificultad de Farmacología II



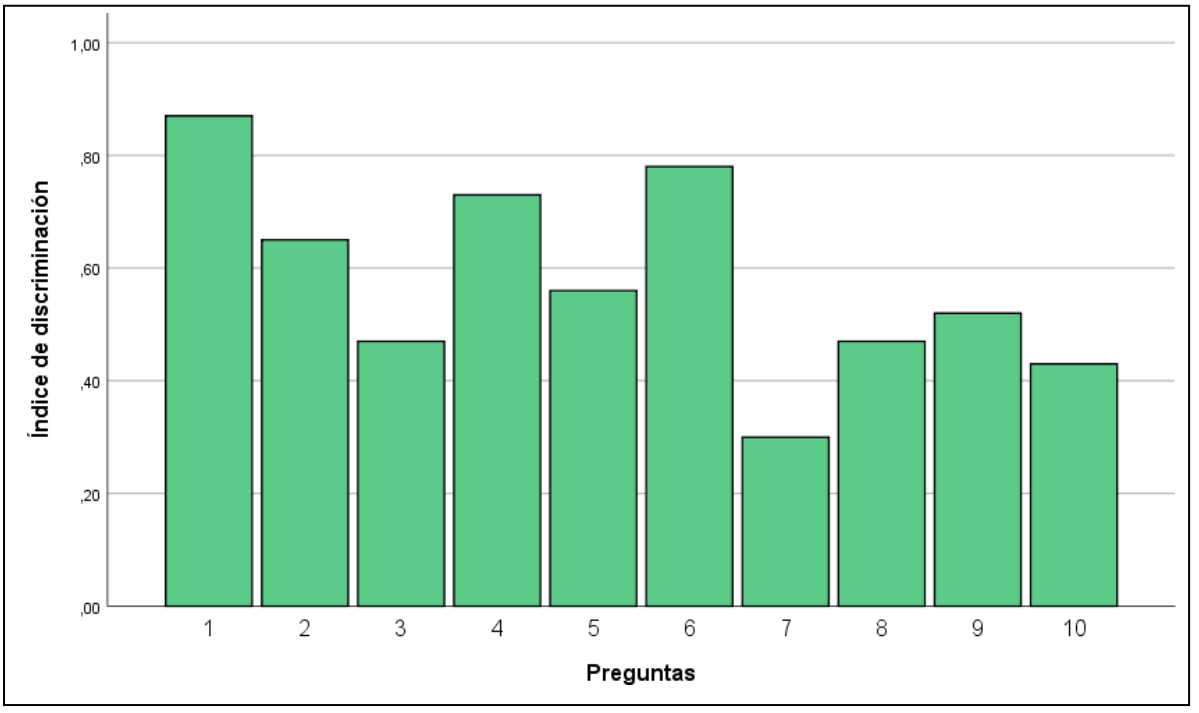
El índice global de dificultad tuvo una media de 0,43±.06 un valor que puede ser catalogado como Medianamente fácil. La mediana global fue 0,37.

Índice de discriminación

Tabla 18. Índice de discriminación de Farmacología II

Pregunta	G_S	G_i	IDisc	Calidad	Recomendación
1	23	3	.87	Excelente	Conservar
2	20	5	.65	Excelente	Conservar
3	21	10	.47	Excelente	Conservar
4	23	6	.73	Excelente	Conservar
5	20	7	.56	Excelente	Conservar
6	20	2	.78	Excelente	Conservar
7	9	2	.30	Buena	Mejorar
8	14	3	.47	Excelente	Conservar
9	18	6	.52	Excelente	Conservar
10	12	2	.43	Excelente	Conservar

Figura 18. Índice de discriminación de Farmacología II

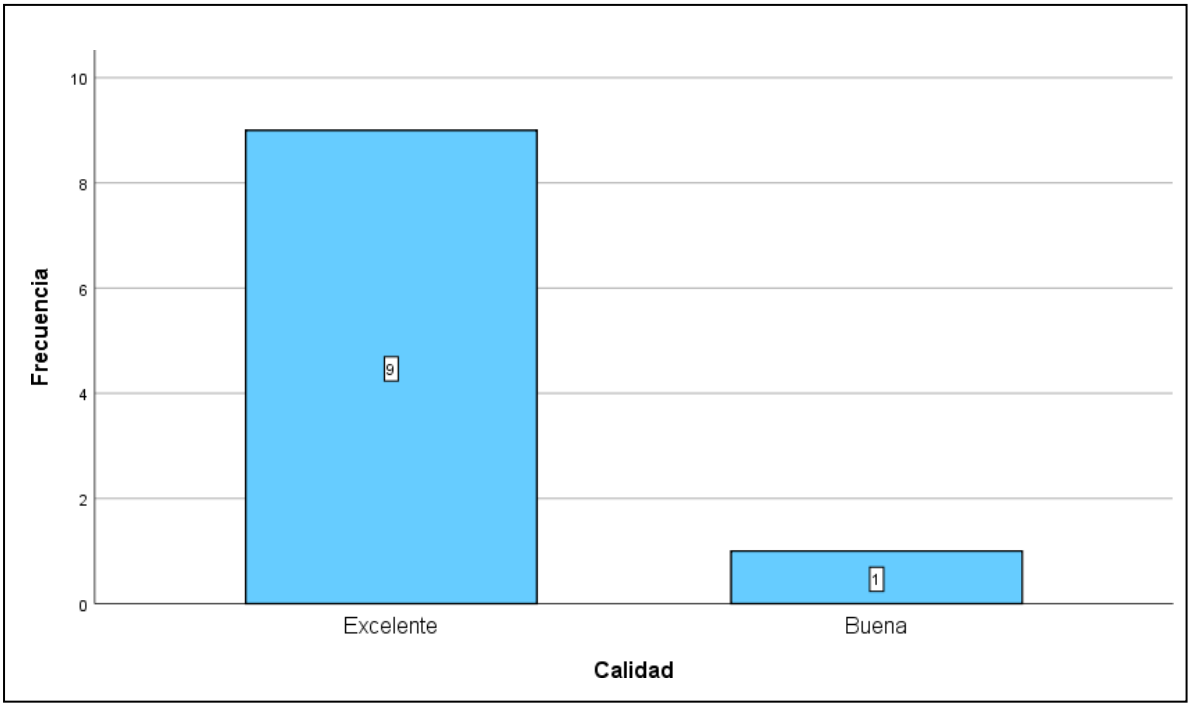


Las preguntas que tuvieron mayor índice de discriminación fueron la pregunta P1(0,87), P6(0,78) y P4(0,73). Las preguntas que tuvieron menor índice de discriminación fueron la pregunta P7 (0,30), P10 (0,43) y P8 (0,47).

Tabla 19. Calidad de las preguntas de Farmacología II

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Excelente	9	90.0	90.0	90.0
	Buena	1	10.0	10.0	100.0
	Total	10	100.0	100.0	

Figura 19. Calidad de las preguntas de Farmacología II

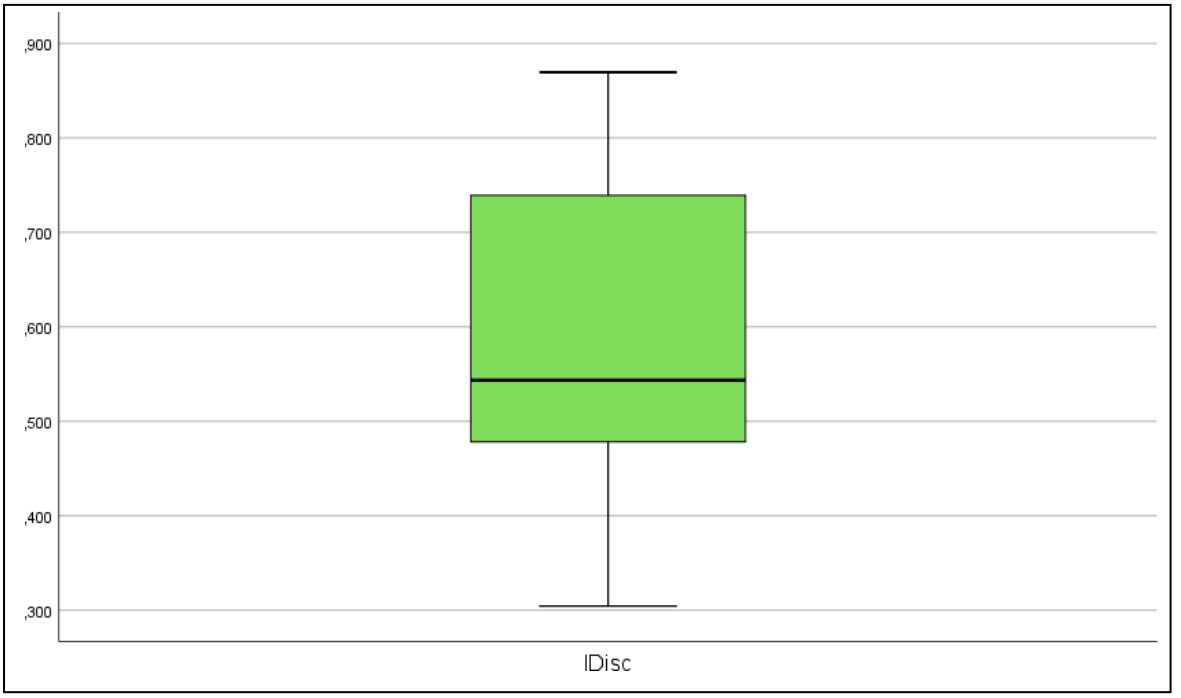


El 90% de las preguntas (9) puede ser catalogado con una calidad Excelente y el 10% con una calidad Buena (1).

Tabla 20. Global Índice de discriminación de Farmacología

	N	Media	Desv. Error	Mínimo - Máximo	Mediana	Calidad
Índice de discriminación	10	.58	0.05	.304 - .870	.54	Excelente

Figura 20. Global Índice de discriminación de Farmacología

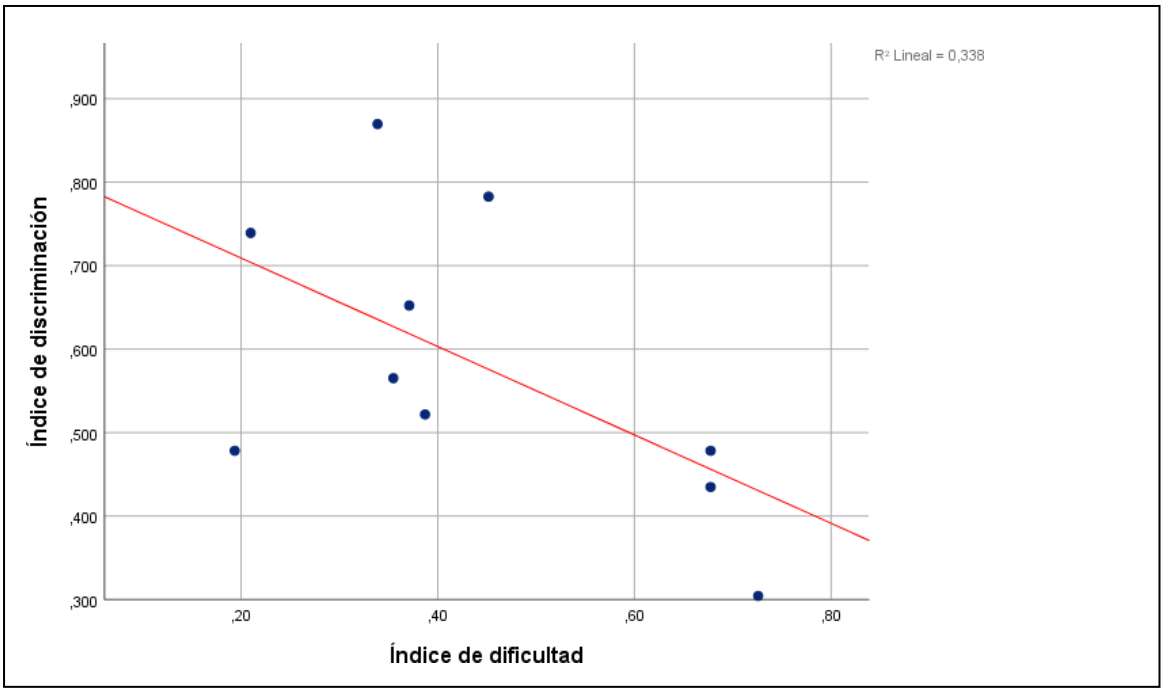


El índice global de discriminación tuvo una media de 0,58±005 un valor que puede ser catalogado con una calidad de Excelente. La mediana global fue 0,54.

Tabla 21. Correlación índice dificultad x índice discriminación de Farmacología

			Índice de dificultad	Índice de discriminación
Rho de Spearman	Índice de dificultad	Coeficiente de correlación	1.000	-.540
		Sig. (bilateral)	.	.107
		N	10	10
	Índice de discriminación	Coeficiente de correlación	-.540	1.000
		Sig. (bilateral)	.107	.
		N	10	10

Figura 21. Correlación índice dificultad x índice discriminación de Farmacología



Existe una correlación negativa no significativa entre el índice de dificultad y el índice de discriminación (r=-0,540; p-value=0,107). La mayoría de las preguntas son Medianamente Fácil y de excelente discriminación.

Fiabilidad de Farmacología

Estadísticas de fiabilidad	
Alfa de Cronbach	N de elementos
0,627	10

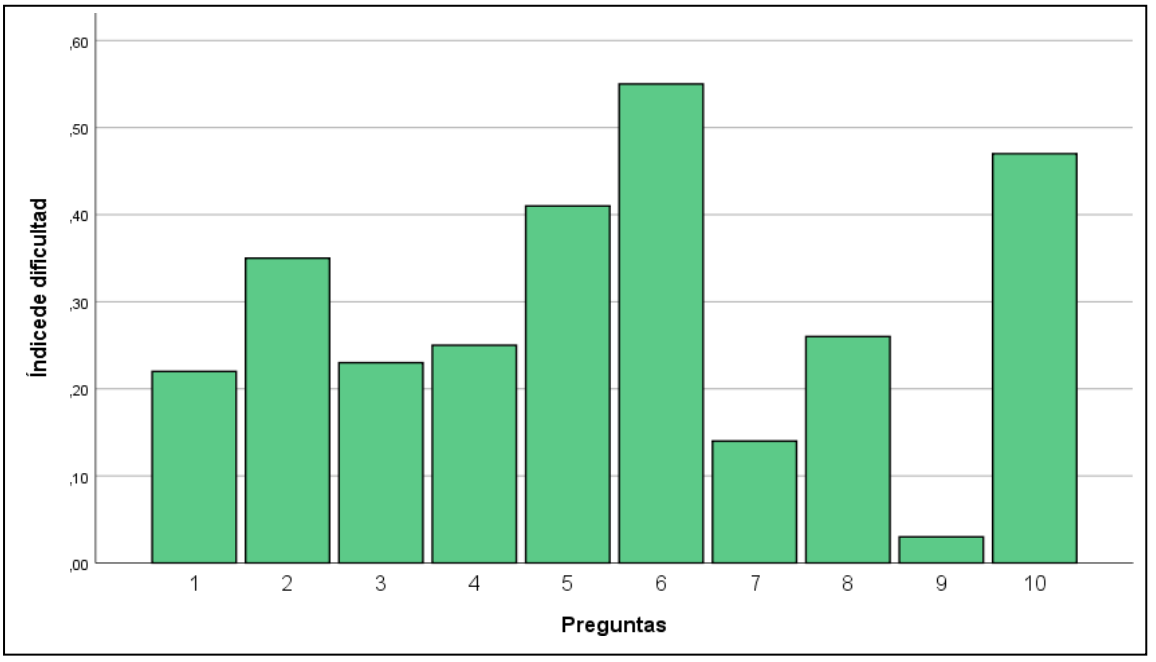
5.1.4 Fisiología General

Índice de dificultad

Tabla 22. Índice de dificultad de las preguntas de Fisiología General

Pregunta	C/89	IDif	Categoría de los ítems
1	69	0.22	Altamente fácil
2	57	0.35	Medianamente fácil
3	68	0.23	Altamente fácil
4	66	0.25	Altamente fácil
5	52	0.41	Medianamente fácil
6	40	0.55	Dificultad media
7	76	0.14	Altamente fácil
8	65	0.26	Altamente fácil
9	86	0.03	Altamente fácil
10	47	0.47	Medianamente fácil

Figura 22. Índice de dificultad de las preguntas de Fisiología General

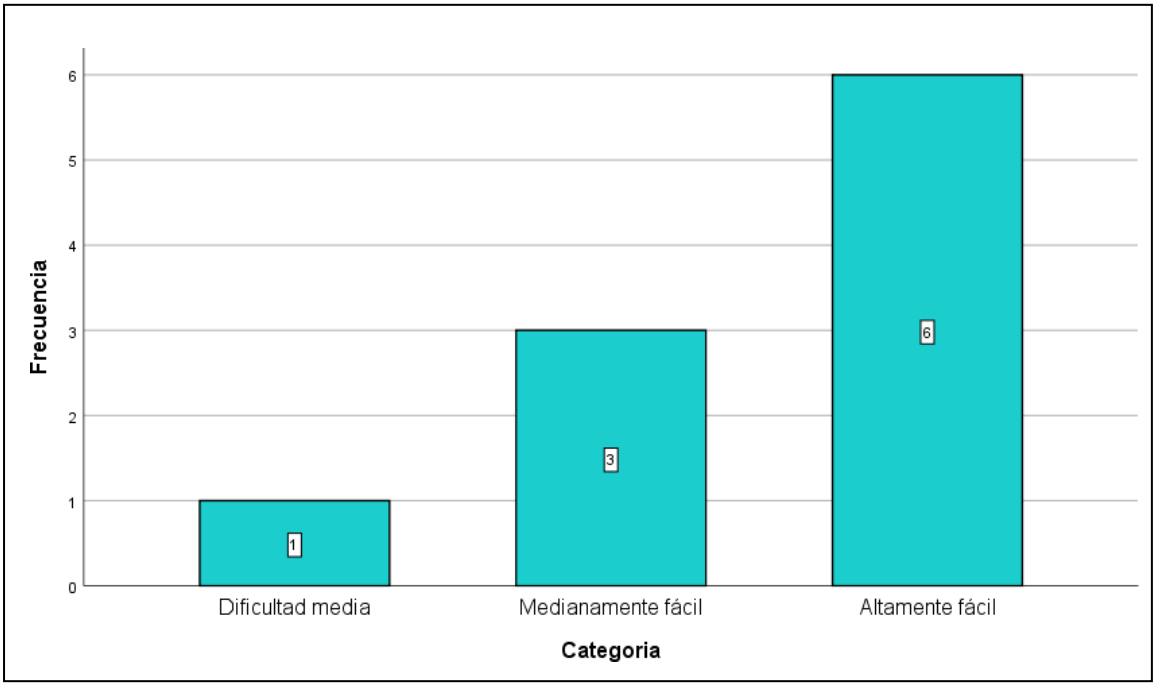


Las preguntas que tuvieron mayor índice de dificultad fueron la pregunta P6 (0,55), P10 (0,47) y P5 (0,41). Las preguntas que tuvieron menor índice de dificultad fueron la pregunta P9 (0,03), P7 (0,14) y P1 (0,22).

Tabla 23. Dificultad por categorías de Fisiología General

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Dificultad media	1	10.0	10.0	10.0
Medianamente fácil	3	30.0	30.0	40.0
Altamente fácil	6	60.0	60.0	100.0
Total	10	100.0	100.0	

Figura 23. Dificultad por categorías de Fisiología

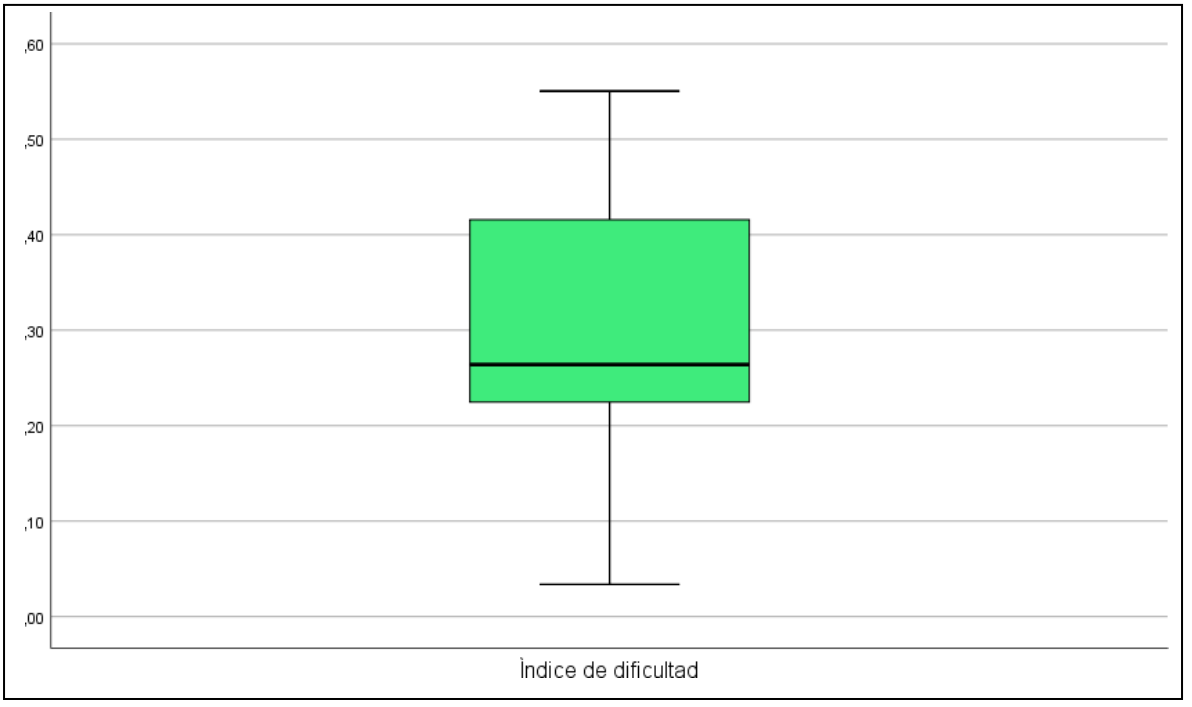


El 60% de las preguntas (6) puede ser catalogado como Altamente fácil y el 30% como Medianamente fácil (3). Solo el 10% fue catalogado como Dificultad media (1).

Tabla 24. Global Índice de dificultad de Fisiología

	N	Media	Desv. Error	Mínimo - Máximo	Mediana	Dificultad
Índice de dificultad	10	.2966	.04901	.03 - .55	.2640	Altamente fácil

Figura 24. Global Índice de dificultad de Fisiología



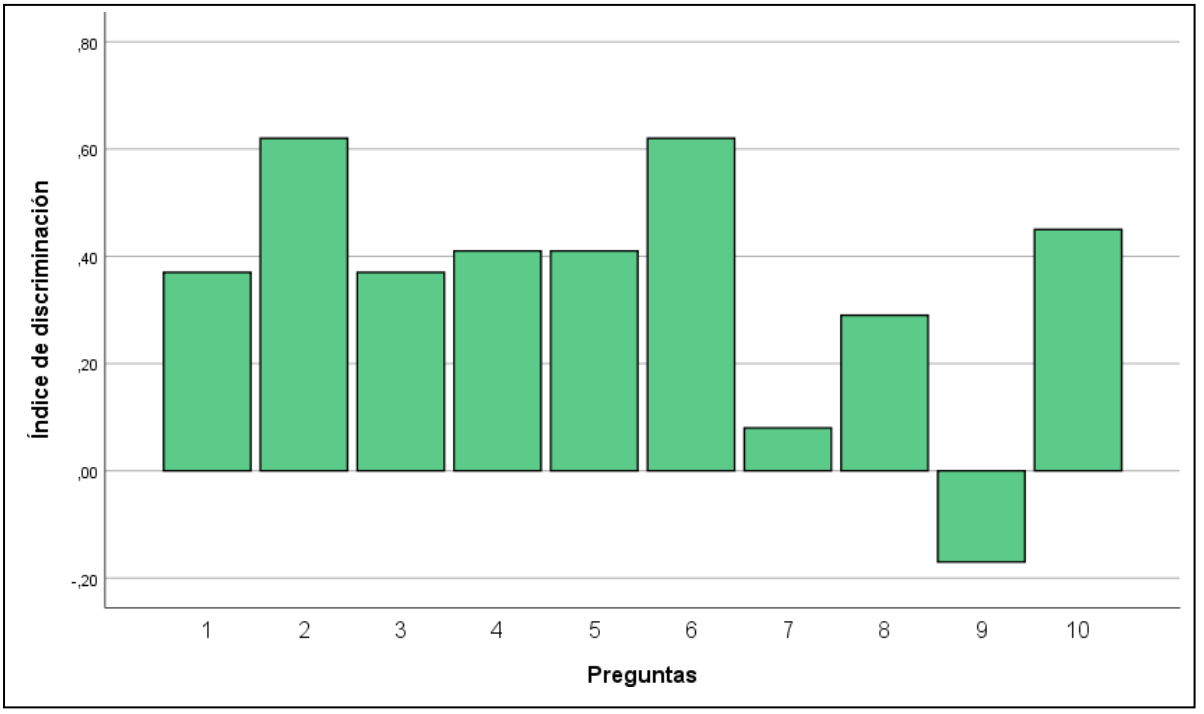
El índice global de dificultad tuvo una media de 0,29±04 un valor que puede ser catalogado como Altamente fácil. La mediana global fue 0,26.

Índice de discriminación

Tabla 25. Índice de discriminación de Fisiología General

Pregunta	G_S	G_i	IDisc	Calidad	Recomendación
1	24	15	.37	Buena	Mejorar
2	24	9	.62	Excelente	Conservar
3	23	14	.37	Buena	Mejorar
4	22	12	.41	Excelente	Conservar
5	20	10	.41	Excelente	Conservar
6	21	6	.62	Excelente	Conservar
7	23	21	.08	Pobre	Revisar a profundidad
8	24	17	.29	Regular	Revisar
9	24	28	-.17	Pésima	Descartar
10	21	10	.45	Excelente	Conservar

Figura 25. Índice de discriminación de Fisiología General

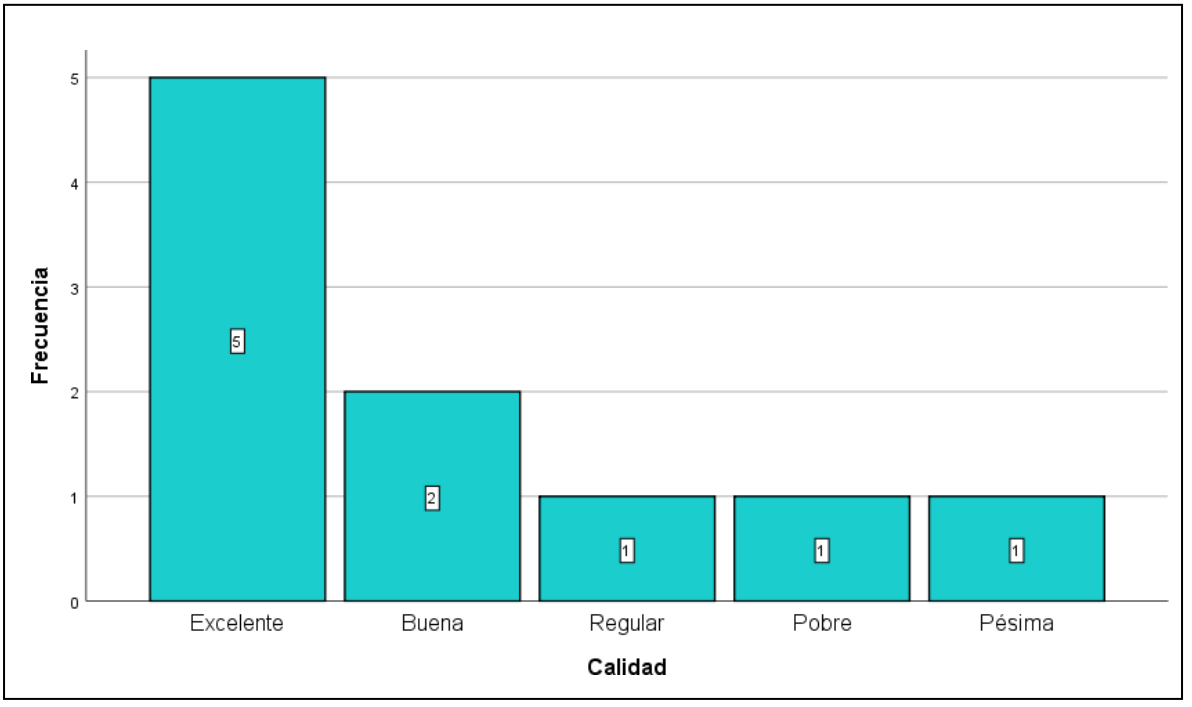


Las preguntas que tuvieron mayor índice de discriminación fueron la pregunta P2 (0,62), P6 (0,62) y P10 (0,45). Las preguntas que tuvieron menor índice de discriminación fueron la pregunta P9 (- 0,7), P7 (0,08) y P8 (0,29).

Tabla 26. Calidad de las preguntas de Fisiología General

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Excelente	5	50.0	50.0	50.0
Buena	2	20.0	20.0	70.0
Regular	1	10.0	10.0	80.0
Pobre	1	10.0	10.0	90.0
Pésima	1	10.0	10.0	100.0
Total	10	100.0	100.0	

Figura 26. Calidad de las preguntas de Fisiología General

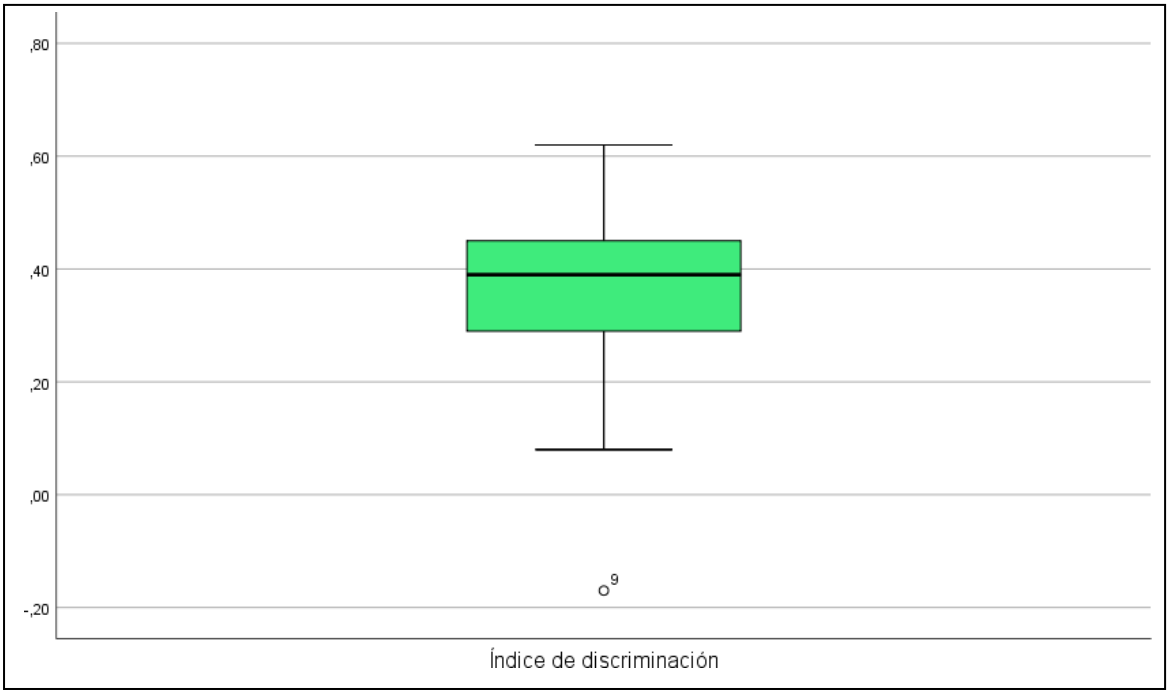


El 50% de las preguntas (5) puede ser catalogado con una calidad Excelente y el 20% con una calidad Buena (2). El 10% fue catalogado como Pobre, Regular o Pésima (1).

Tabla 27. Global Índice de discriminación de Fisiología General

	N	Media	Desv. Error	Mínimo - Máximo	Mediana	Calidad
Índice de discriminación	10	.3450	.07531	-.17 – .62	.39	Buena

Figura 27. Global Índice de discriminación de Fisiología General

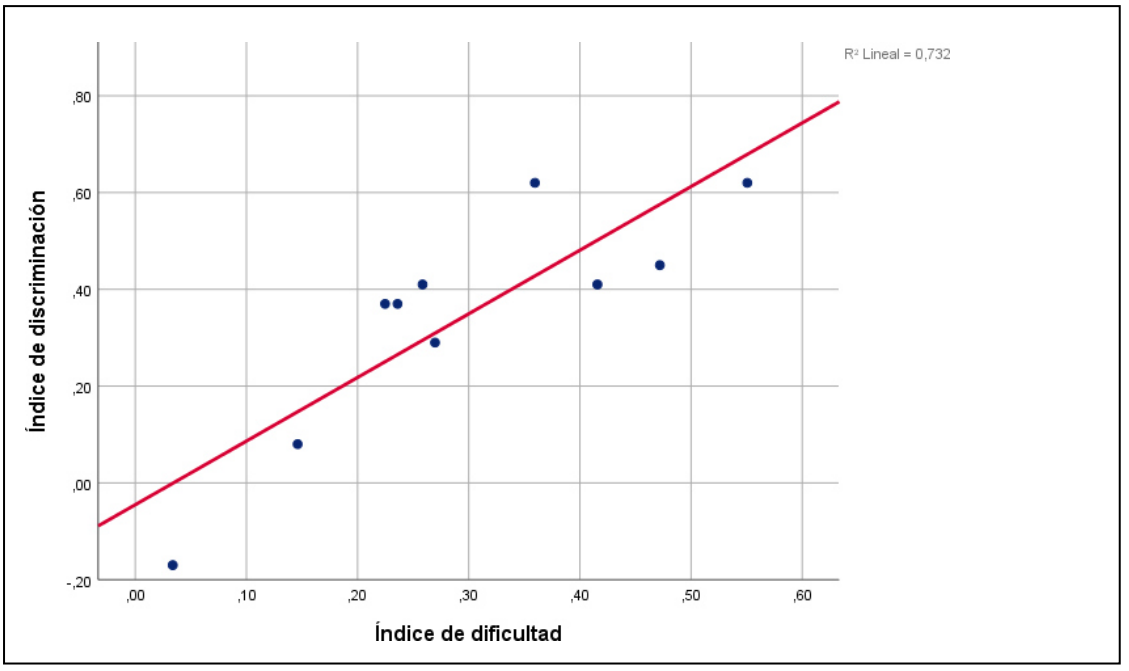


El índice global de discriminación tuvo una media de 0,34±07 un valor que puede ser catalogado con una calidad de Buena. La mediana global fue 0,39.

Tabla 28. Correlación índice dificultad x índice discriminación de Fisiología

			Índice de dificultad	Índice de discriminación
Rho de Spearman	Índice de dificultad	Coeficiente de correlación	1.000	.856**
		Sig. (bilateral)	.	.002
		N	10	10
	Índice de discriminación	Coeficiente de correlación	.856**	1.000
		Sig. (bilateral)	.002	.
		N	10	10
**. La correlación es significativa en el nivel 0.01 (bilateral).				

Figura 28. Correlación índice dificultad x índice discriminación de Fisiología General



Existe una correlación significativa entre el índice de dificultad y el índice de discriminación (r=0,856; p-value= 0,002. A medida que el índice de dificultad aumenta también aumenta el índice de discriminación.

Confiabilidad de Fisiología General

Estadísticas de fiabilidad	
Alfa de Cronbach	N de elementos
0,587	10

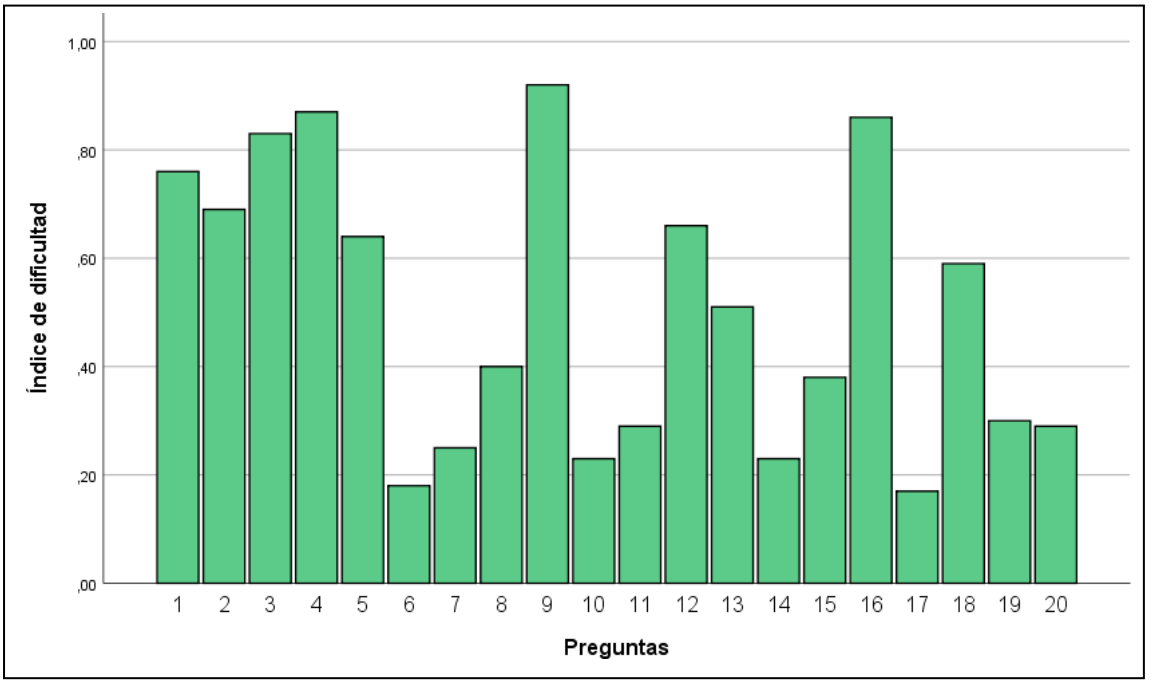
5.1.5 Microbiología General

Índice de dificultad

Tabla 29. Índice de dificultad de las preguntas Microbiología General

Pregunta	C/93	IDif	Categoría de los ítems
1	22	0.76	Difícil
2	28	0.69	Dificultad media
3	15	0.83	Difícil
4	12	0.87	Altamente difícil
5	33	0.64	Dificultad media
6	76	0.18	Altamente fácil
7	69	0.25	Altamente fácil
8	55	0.40	Medianamente fácil
9	7	0.92	Altamente difícil
10	71	0.23	Altamente fácil
11	66	0.29	Altamente fácil
12	31	0.66	Dificultad media
13	45	0.51	Medianamente fácil
14	71	0.23	Altamente fácil
15	57	0.38	Medianamente fácil
16	13	0.86	Altamente difícil
17	77	0.17	Altamente fácil
18	38	0.59	Dificultad media
19	65	0.30	Altamente fácil
20	66	0.29	Altamente fácil

Figura 29. Índice de dificultad de las preguntas Microbiología General

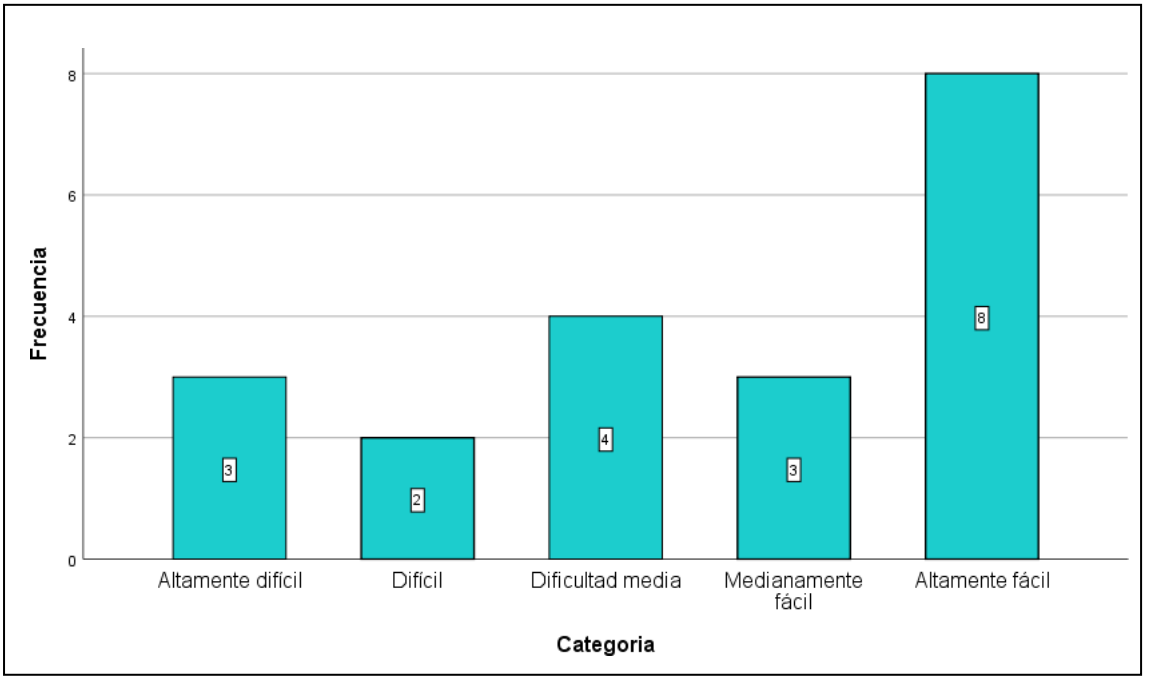


Las preguntas que tuvieron mayor índice de dificultad fueron la pregunta P9 (0,92), P4 (0,87) y P16 (0,86). Las preguntas que tuvieron menor índice de dificultad fueron la pregunta P17(0,17), P6 (0,18) y P10 (0,23).

Tabla 30. Dificultad por categorías de Microbiología General

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Altamente difícil	3	15.0	15.0	15.0
Difícil	2	10.0	10.0	25.0
Dificultad media	4	20.0	20.0	45.0
Medianamente fácil	3	15.0	15.0	60.0
Altamente fácil	8	40.0	40.0	100.0
Total	20	100.0	100.0	

Figura 30. Dificultad por categorías de Microbiología General

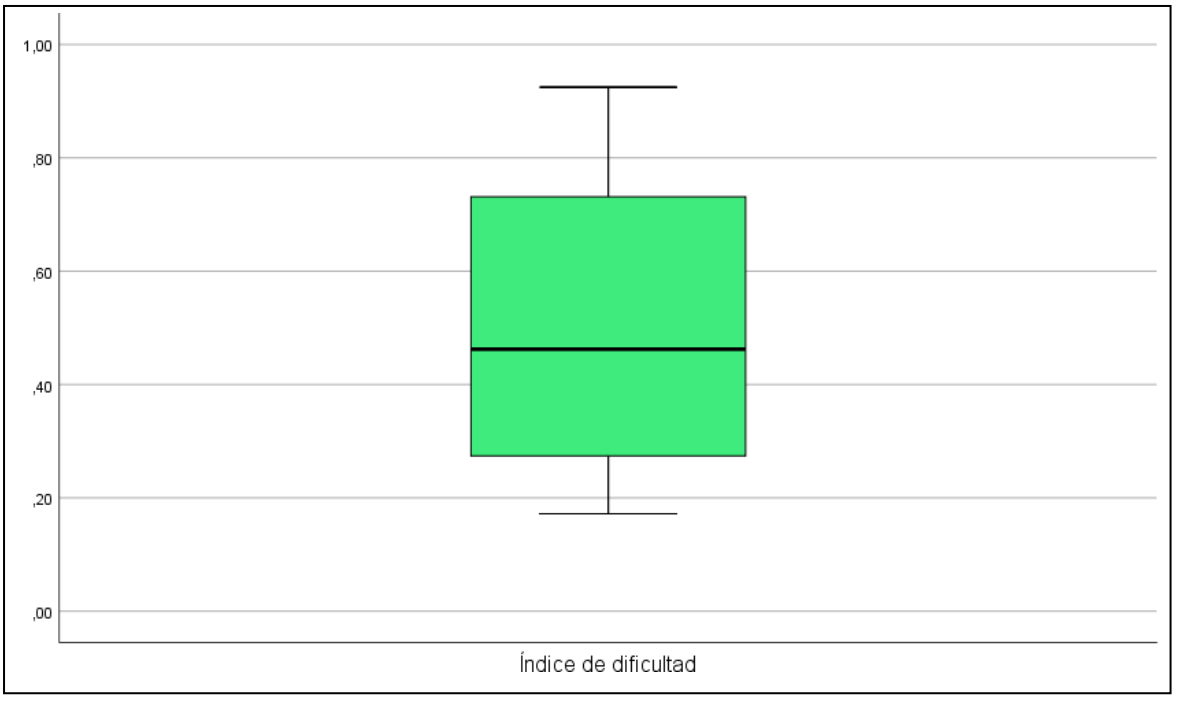


El 40% de las preguntas (8) puede ser catalogado como Altamente fácil y el 20% como Dificultad media (4). Solo el 10% fue catalogado como Difícil (2).

Tabla 31. Global Índice de dificultad de Microbiología General

	N	Media	Desv. Error	Mínimo - Máximo	Mediana	Dificultad
Índice de dificultad	20	.507	.058	.17 - .92	.46	Medianamente fácil

Figura 31. Global Índice de dificultad de Microbiología General



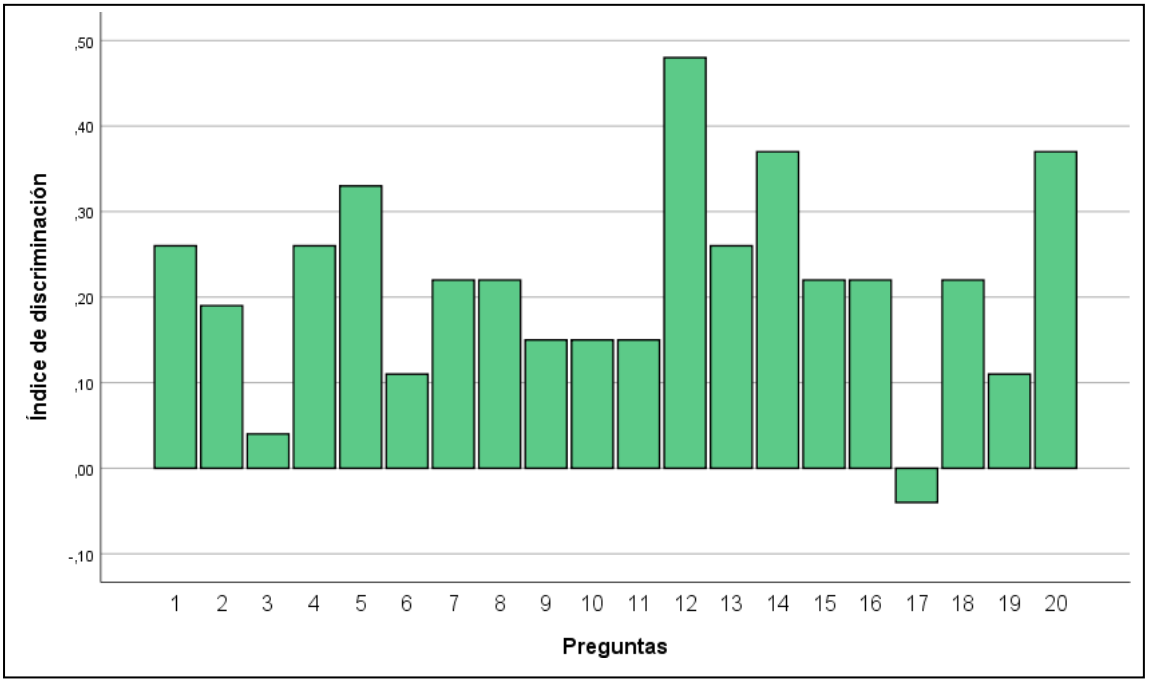
El índice global de dificultad tuvo una media de 0,50 ±05 un valor que puede ser catalogado como Medianamente fácil. La mediana global fue 0,46.

Índice de discriminación

Tabla 32. Índice de discriminación de Microbiología General

Pregunta	G_S	G_i	IDisc	Calidad	Recomendación
1	12	5	.26	Regular	Revisar
2	13	8	.19	Pobre	Revisar a profundidad
3	5	4	.04	Pobre	Revisar a profundidad
4	7	0	.26	Regular	Revisar
5	17	8	.33	Buena	Mejorar
6	27	24	.11	Pobre	Revisar a profundidad
7	23	17	.22	Regular	Revisar
8	20	14	.22	Regular	Revisar
9	5	1	.15	Pobre	Revisar a profundidad
10	24	20	.15	Pobre	Revisar a profundidad
11	24	20	.15	Pobre	Revisar a profundidad
12	18	5	.48	Excelente	Conservar
13	18	11	.26	Regular	Revisar
14	26	16	.37	Buena	Mejorar
15	22	16	.22	Regular	Revisar
16	8	2	.22	Regular	Revisar
17	25	26	-.04	Pésima	Descartar
18	14	8	.22	Regular	Revisar
19	20	17	.11	Pobre	Revisar a profundidad
20	26	16	.37	Buena	Mejorar

Figura 32. Índice de discriminación de Microbiología General

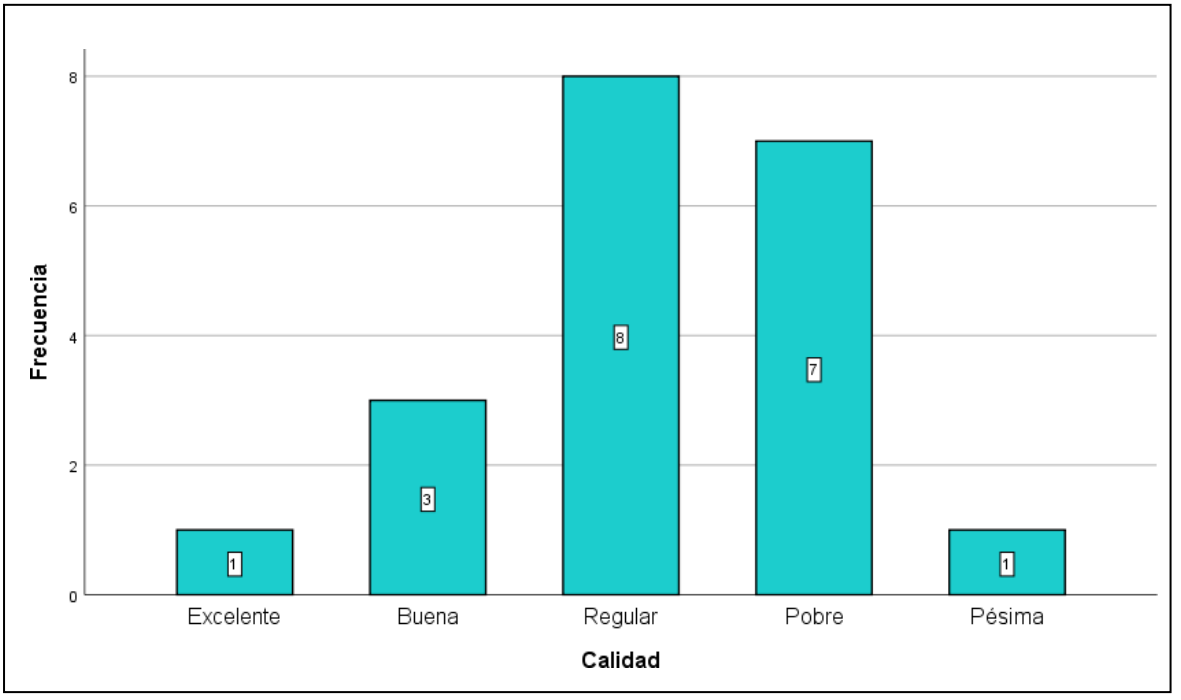


Las preguntas que tuvieron mayor índice de discriminación fueron la pregunta P 12(0,48), P14 (0,37) y P20 (0,37). Las preguntas que tuvieron menor índice de discriminación fueron la pregunta P17 (- 0,04), P3 (0,04) y P6 (0,11).

Tabla 33. Calidad de las preguntas de Microbiología

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado	Recomendación
Excelente	1	5.0	5.0	5.0	Conservar
Buena	3	15.0	15.0	20.0	Mejorar
Regular	8	40.0	40.0	60.0	Revisar
Pobre	7	35.0	35.0	95.0	Revisar a profundidad
Pésima	1	5.0	5.0	100.0	Descartar
Total	20	100.0	100.0		

Figura 33. Calidad de las preguntas de Microbiología

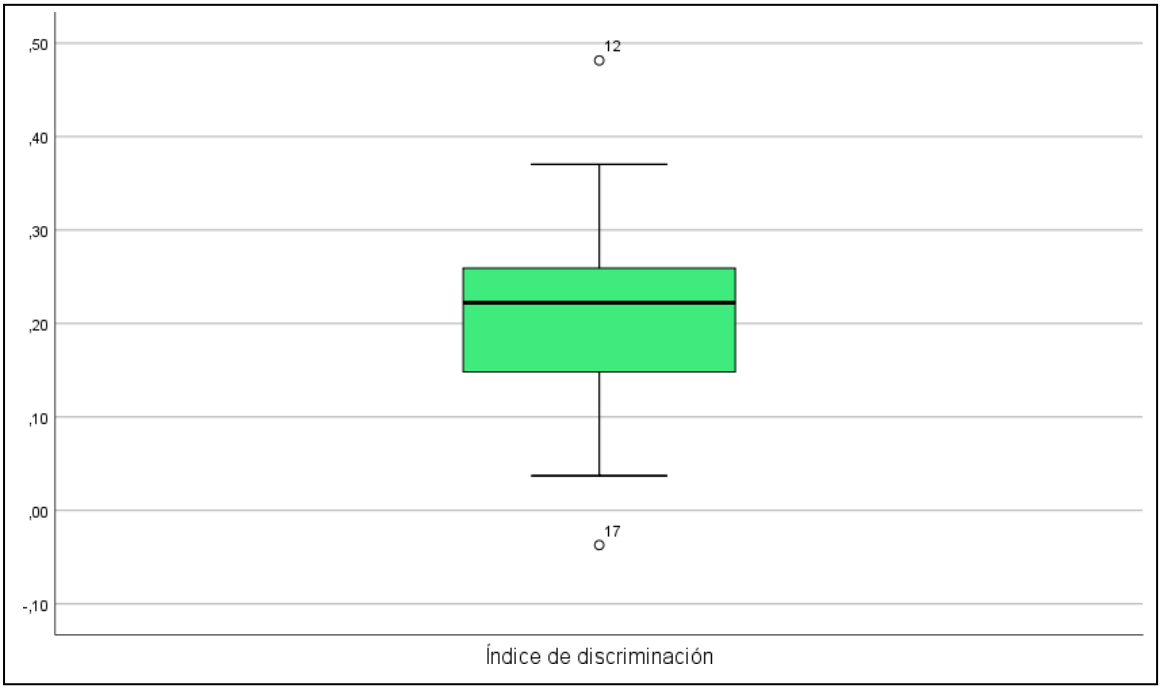


El 40% de las preguntas (8) puede ser catalogado con una calidad Regular y el 35% con una calidad Pobre (7). Solo el 5% fue catalogado como Pésima (1).

Tabla 34. Global Índice de discriminación de Microbiología

	N	Media	Desv. Error	Mínimo - Máximo	Mediana	Calidad
Índice de discriminación	20	.214	.026	-.04 - .48	.222	Regular

Figura 34. Global Índice de Discriminación de Microbiología

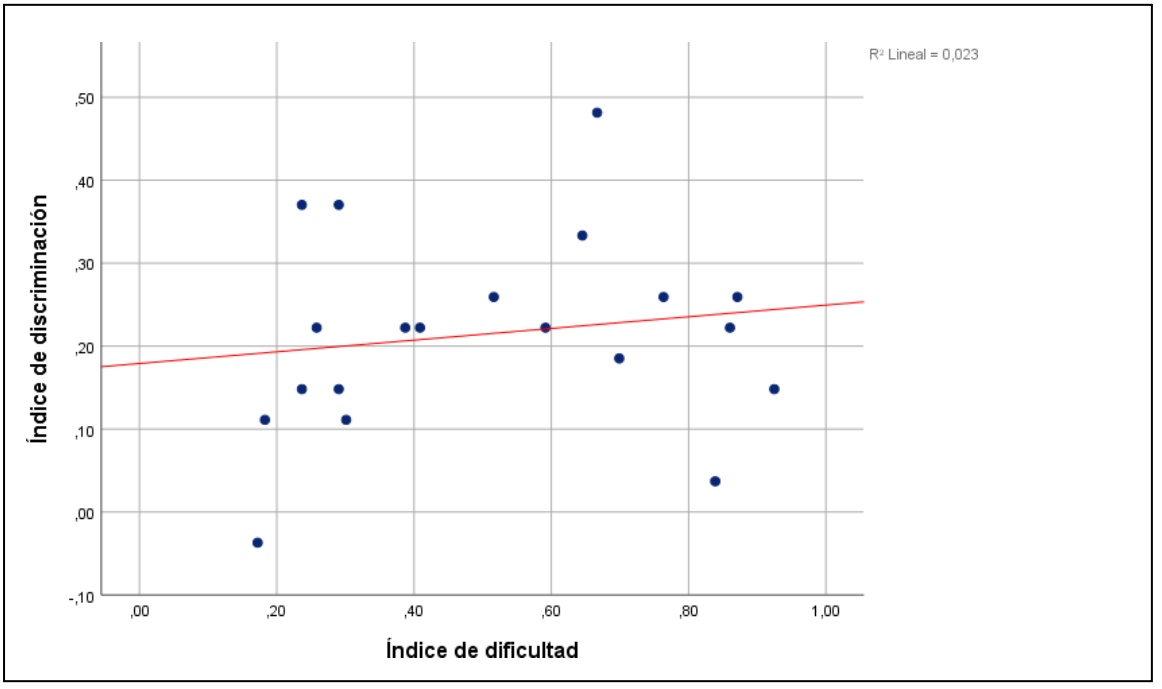


El índice global de discriminación tuvo una media de 0,21±02 un valor que puede ser catalogado con una calidad de Regular. La mediana global fue 0,22.

Tabla 35. Correlación índice dificultad x índice discriminación de Microbiología

			Índice de dificultad	Índice de discriminación
Rho de Spearman	Índice de dificultad	Coeficiente de correlación	1.000	.205
		Sig. (bilateral)	.	.385
		N	20	20
	Índice de discriminación	Coeficiente de correlación	.205	1.000
		Sig. (bilateral)	.385	.
		N	20	20

Figura 35. Correlación índice dificultad x índice discriminación de Microbiología



Existe una correlación no significativa entre el índice de dificultad y el índice de discriminación (r=0,205; p-value=0,385. La mayoría de las preguntas son preguntas con dificultad fácil o alta, pero con discriminación media.

Confiabilidad de Microbiología

Estadísticas de fiabilidad	
Alfa de Cronbach	N de elementos
0,506	20

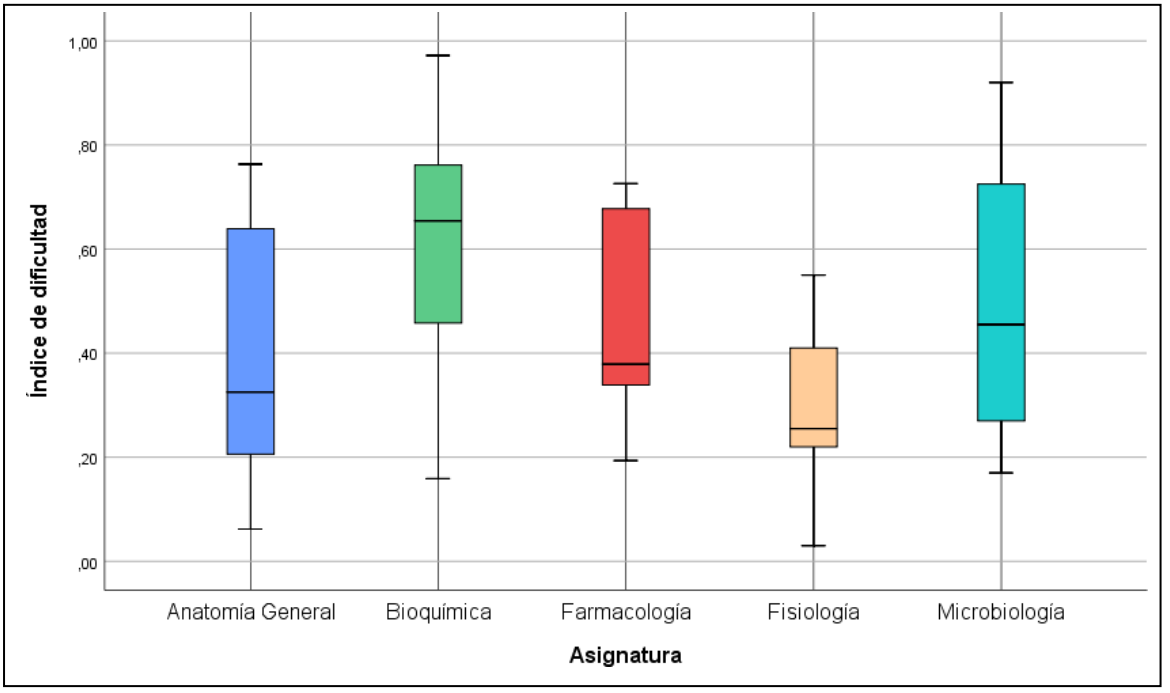
5.1.6 Resultados globales

Índice de dificultad

Tabla 36. Índice de dificultad por asignatura

		N	Media	Desv. Error	Mínimo - Máximo	Mediana
Índice de dificultad	Anatomía	30	.38	.03	.06 -.76	.32
	Bioquímica	20	.62	.04	.16 - .97	.65
	Farmacología	10	.43	.06	.19 - .73	.37
	Fisiología	10	.29	.04	.03 - .55	.25
	Microbiología	20	.50	.058	.17 - .92	.45

Figura 36. Índice de dificultad por asignatura



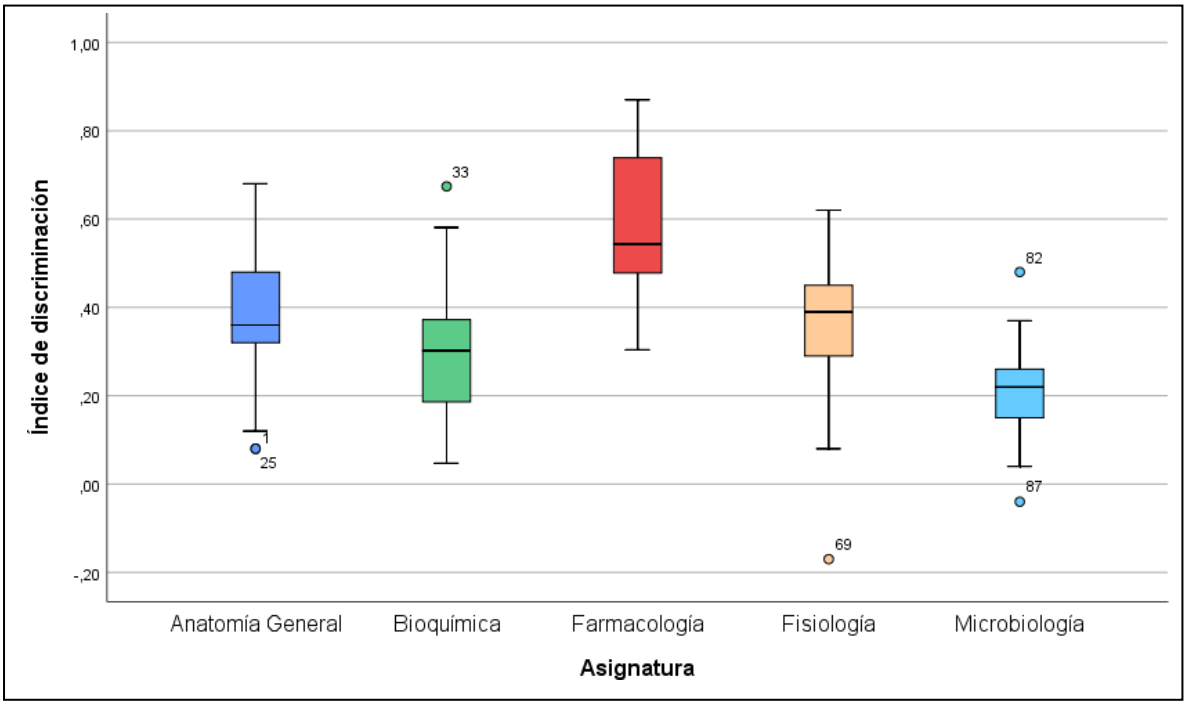
La asignatura con mayor media del índice de dificultad fue Bioquímica (0,62±04), seguido de Microbiología (0,50±05) y Farmacología (0,43±06). La asignatura con menor media del índice de dificultad fue Fisiología (0,29±04) y, luego Anatomía (0,38±03).

Índice de discriminación

Tabla 37. Global índice de discriminación por asignatura

		N	Media	Desv. Error	Mínimo - Máximo	Mediana
Índice de discriminación	Anatomía	30	.37	.02	.08 -.68	.36
	Bioquímica	20	.30	.03	.04 - .67	.30
	Farmacología	10	.58	.05	.30 - .87	.54
	Fisiología	10	.34	.07	-.17 – .62	.39
	Microbiología	20	.21	.026	-.04 - .48	.22

Figura 37. Global índice de discriminación por asignatura



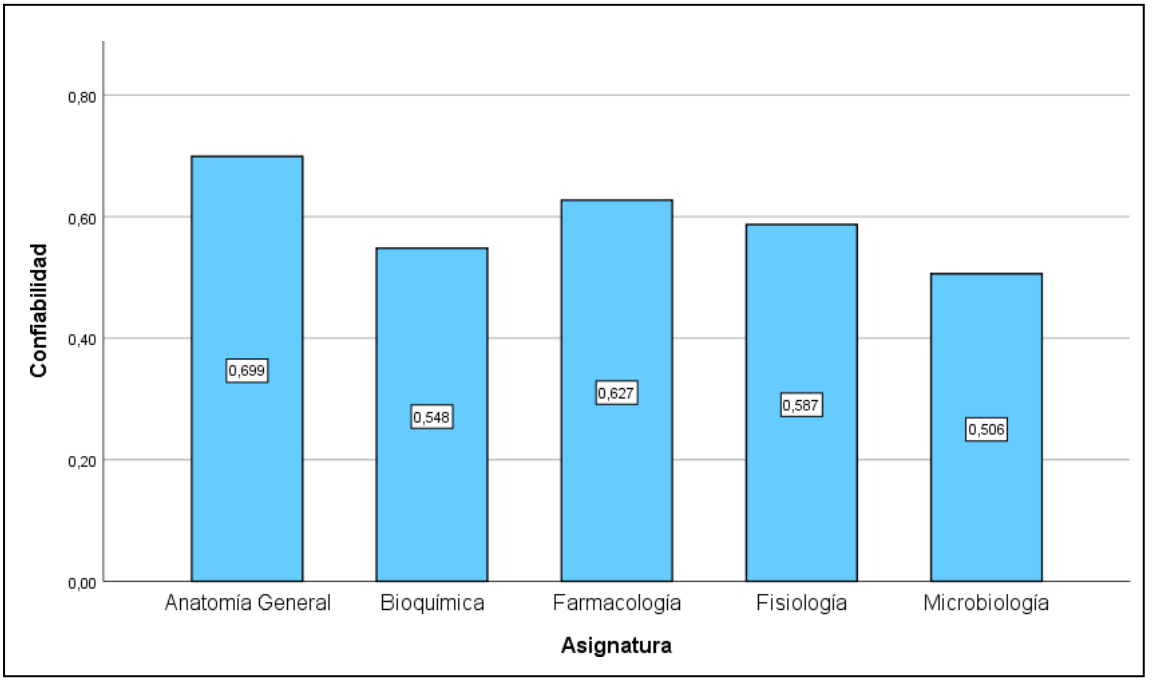
La asignatura con mayor media del índice de discriminación fue Farmacología (0,58±05), seguido de Anatomía (0,37±02) y Fisiología (0,34±07). Los menores índices se presentaron en las asignaturas de Microbiología (0,21±02) y Bioquímica (0,30±03).

Confiabilidad

Tabla 38. Global confiabilidad por asignatura

Confiabilidad	
Anatomía	.699
Bioquímica	.548
Farmacología	.627
Fisiología	.587
Microbiología	.506

Figura 38. Global confiabilidad por asignatura



La asignatura con mayor índice de confiabilidad fue Anatomía (0,699), seguido de Farmacología (0,627) y Fisiología (.587). Los menores índices de confiabilidad se presentaron en las asignaturas de Microbiología (0,506) y Bioquímica (0,548).

5.2 Análisis y resumen de los resultados

5.2.1 Anatomía General

El examen de la asignatura de Anatomía General constó de 30 preguntas. Fue catalogado como un examen Medianamente Fácil (media de 0,38) con 15 preguntas (50%) Altamente Fáciles; 5 preguntas (16.7%) Medianamente Fáciles; 9 preguntas (30%) Dificultad Media; y 1 pregunta (3.3%) Medianamente Difícil. El índice de discriminación fue de 0,37 lo que lo califica como un examen de Buena Calidad con 13 preguntas (43.3%) de Excelente Calidad; 10 preguntas (33.3%) de Buena Calidad; 3 preguntas (10%) de Regular Calidad; y 4 preguntas (13.3%) de Pobre Calidad. Al relacionar los índices de dificultad y discriminación del examen se observa que existe una correlación baja no significativa (r= 0,132 y p- value= 0,486)). El índice de fiabilidad (alfa de Cronbach) fue de 0,699 lo cual lo califica como un examen de Alta Confiabilidad.

5.2.2 Bioquímica General

El examen de Bioquímica General constó de 20 preguntas. Fue catalogado como un examen de Dificultad Media (media de 0,62) con 3 preguntas (15%) Altamente Difíciles; 3 preguntas (15%) Medianamente Difíciles; 8 preguntas (40%) de Dificultad Media; 4 preguntas (20%) Medianamente Fáciles; y 2 preguntas (10%) Altamente Fáciles. El índice de discriminación fue de 0,30 lo que lo califica como un examen de Buena calidad con 4 preguntas (20%) de Excelente Calidad; 7 preguntas (35%) de Buena Calidad; 3 preguntas (15%) de Regular Calidad; y 6 preguntas (30%) de Pobre Calidad. Al correlacionar los índices de dificultad y de discriminación del examen se observa una correlación negativa, baja y no significativa (r = - 0,125 y p-value = 0,599). El índice de fiabilidad (alfa de Cronbach) del examen fue de 0,548, lo que lo cataloga como un examen de Moderada Confiabilidad.

5.2.3 Farmacología II

El examen de Farmacología II estuvo compuesto de 10 preguntas. Fue catalogado como un examen Medianamente Fácil (media de 0,43) con 3 preguntas (30%) de Dificultad Media; 5 preguntas (50%) Medianamente Fáciles; y 2 preguntas (20%) Altamente Fáciles. El índice de discriminación fue de 0,58 lo que lo cataloga comoun Excelente Examen con 9 preguntas (90%) de Excelente Calidad; y una pregunta (10%) de Buena Calidad. Al correlacionar los índices de dificultad y discriminación del examen se observó una correlación negativa no significativa (r = - 0,540 y p- value = 0,107). El índice de fiabilidad (alfa de Cronbach) fue de 0,627 lo que lo cataloga como un examen de Alta Confiabilidad.

5.2.4 Fisiología General

El examen de Fisiología General constó de 10 preguntas. Fue catalogado como un examen Altamente Fácil (media de 0,296) con 1 pregunta (10%) de Dificultad Media; 3 preguntas (30%) Medianamente Fáciles; y 6 preguntas (60%) Altamente Fáciles. El índice de discriminación fue de 0,345 lo que lo califica como un examen de Buena Calidad con 5 preguntas (50%) de Excelente Calidad; 2 preguntas (20%) de Buena Calidad; 1 pregunta (10%) de Regular Calidad; 1 pregunta (10%) de Pobre Calidad; y 1 pregunta (10%) de Pésima Calidad. Al relacionar los índices de dificultad y discriminación del examen se observó una correlación alta y significativa(r = 0,856 y p-value = 0,002). El índice de fiabilidad fue de 0,587 lo que lo cataloga como un examen de Moderada Confiabilidad.

5.2.5 Microbiología General

El examen de Microbiología General constó de 20 preguntas. Fue catalogado como un examen Medianamente Fácil (media de 0,507) con 3 preguntas (15%) Altamente Difíciles; 2 preguntas (10%) Medianamente Difíciles; 4 preguntas (20%) de Dificultad Media; 3 preguntas (15%) Medianamente Fáciles; y 8 preguntas (40%) Altamente

Fáciles. El índice de discriminación fue de 0,214 lo que lo califica como un examen de Regular Calidad con 1 pregunta (5%) de Excelente Calidad; 3 preguntas (15%) de Buena Calidad; 8 preguntas (40%) de Regular Calidad; 7 preguntas (35%) de Pobre Calidad; y 1 pregunta (5%) de Pésima Calidad. Al relacionar los índices de dificultad y discriminación del examen se observó una correlación baja no significativa (r= 0,205 y p-value = 0,385). El índice de fiabilidad del examen fue de 0,506 lo que lo cataloga como un examen de Moderada Confiabilidad.

Al comparar los índices de dificultad, de discriminación y de fiabilidad de los exámenes, se observa lo siguiente:

- En relación al índice de dificultad, el examen de Bioquímica General fue el más difícil con una media de 0,62 (Dificultad Media); el segundo lugar fue para el examen de Microbiología General con una media de 0,50 (Medianamente Fácil); el tercer lugar lo ocupó el examen de FarmacologíaII con una media de 0,43 (medianamente Fácil); el cuarto lugar fue para el examen de Anatomía General con una media de 0,38 (Medianamente Fácil); mientras que el examen de Fisiología General fue el más fácil con una media de 0,29 (Altamente Fácil).

- En relación al índice de discriminación, se observó que el examen que mejor discriminó fue el Farmacología II con una media de 0,58 (Excelente Calidad); el segundo lugar fue para el examen de Anatomía General con una media de 0,37 (Buena Calidad); el tercer lugar lo ocupó el examen de Fisiología General con una media de 0,34 (Buena Calidad); el cuarto lugar fue para el examen de Bioquímica General con una media de 0,30 (Buena Calidad); mientras que el examen que menos discriminó fue el de Microbiología General con una media de 0,21 (Regular Calidad).

- En relación al índice de fiabilidad, se observó que el examen de Anatomía General obtuvo la mayor confiabilidad con una media de 0,69 (Alta Confiabilidad); el segundo lugar fue para el examen de Farmacología II con una media de 0,62 (Alta Confiabilidad); el tercer lugar lo ocupó el examen

de Fisiología General con una media de 0,58 (Moderada Confiabilidad); el cuarto lugar fue para el examen de Bioquímica General con una media de 0,54 (Moderada Confiabilidad); y el examen con menos confiabilidad fue el de Microbiología General con una media de 0,50 (Moderada Confiabilidad).

Los resultados comparativos observados muestran diferencias entre los exámenes de las asignaturas del área de Ciencias Básicas que formaron parte de este estudio. En lo referente al promedio de los índices de dificultad de los exámenes se obtiene un índice de 0,44 lo que calificaría a los exámenes tomados como Medianamente Fáciles. En relación a los índices de discriminación se obtuvo un promedio de 0,36 lo que califica a los exámenes tomados como de Buena Calidad, pero con posibilidades de mejorar. Al promediar los índices de fiabilidad se obtuvo un valor de 0,593 que cataloga a los exámenes tomados como de moderada Confiabilidad.

Por otro lado, al observar la extensión, el número de preguntas de los exámenes tomados se nota que estos oscilaron entre 30 preguntas del examen de Anatomía General y 10 preguntas de los exámenes de Farmacología II y Fisiología General. Aunque no existe una norma en relación a la extensión de un examen, se sugiere que éste debe de tener un número de preguntas que permita evaluar todos los contenidos desarrollados en la asignatura, que el número de preguntas permitan cumplir con los objetivos o competencias planteados y, finalmente que el número de preguntas permita evaluar los distintos tipos de habilidades cognitivas que se buscandesarrollar en los estudiantes.

5.3 Resultados de la Prueba de Hipótesis

Contraste de Hipótesis

- **Hipótesis 1: según Índice de Dificultad**

  a) **Hipótesis estadísticas**

  Ho: Las muestras provienen de poblaciones con medianas iguales

  H1: Al menos una mediana es diferente.

  b) **Nivel de significación:** α= 0.05 Nivel de confianza: 95%.

  c) **Estadístico de prueba**

  Prueba H de Kruskal-Wallis.

  d) **Regla de decisión**

  Si p< 0.05 se rechaza la Ho

  Cálculos:

```
. kwallis IDif, by(Curso)

Kruskal-Wallis equality-of-populations rank test

    ----------------------
    |          Curso |  Obs |  Rank Sum |
    ----------------------
    |       Anatomia |   30 |   1108.00 |
    |     Bioquimica |   20 |   1260.00 |
    |   Farmacología |   10 |    450.00 |
    |     Fisiologia |   10 |    278.50 |
    |  Microbiologia |   20 |    998.50 |
    ----------------------

chi-squared =    17.342 with 4 d.f.
probability =     0.0017

chi-squared with ties =    17.346 with 4 d.f.
probability =     0.0017
```

e) Decisión

Como p=0.0017, rechazamos Ho.

f) Conclusión

Se concluye que existen diferencias estadísticamente significativas respecto a la mediana entre los índices de dificultad.

g) Prueba Post-hoc

Utilizando la prueba de Dunn.

Cálculos:

Dunn's Pairwise Comparison of IDif by Curso (Bonferroni)

Col Mean-Row Mean	Anatomia	Bioquimi	Farmacol	Fisiolog
Bioquimi	-3.456748			
	0.0027			
Farmacol	-0.845700	1.779174		
	1.0000	0.3761		
Fisiolog	0.952286	3.474332	1.468050	
	1.0000	**0.0026**	0.7105	
Microbio	-1.722848	1.582826	-0.486802	-2.181959
	0.4246	0.5673	1.0000	0.1456

Family-wise Error Rate = 0.05
Reject Ho if p = P(Z <= |z|) <= FWER/2

Resultado de la prueba post-hoc de Dunn, se encuentran diferencias significativas respecto al índice de dificultad entre las pruebas de los cursos de: Bioquímica y Anatomía (p=0.0027) y las pruebas de Bioquímica y Fisiología (p=0.0026).

- **Hipótesis 2: según Índice de Discriminación**

a) Hipótesis estadísticas

Ho: Las muestras provienen de poblaciones con medianas iguales

H1: Al menos una mediana es diferente.

b) Nivel de significación: α= 0.05 Nivel de confianza: 95%.

c) Estadístico de prueba:

Prueba H de Kruskal-Wallis.

d) Regla de decisión

Si p< 0.05 se rechaza la Ho.

Cálculos:

```
Kruskal-Wallis equality-of-populations rank test
```

Curso	Obs	Rank Sum
Anatomia	30	1527.50
Bioquimica	20	781.00
Farmacología	10	752.00
Fisiologia	10	508.00
Microbiologia	20	526.50

```
chi-squared =    26.619 with 4 d.f.
probability =     0.0001

chi-squared with ties =    26.645 with 4 d.f.
probability =     0.0001
```

e) Decisión

Como p=0.0001, rechazamos Ho.

f) Conclusión

Se concluye que existen diferencias estadísticamente significativas respecto a la mediana entre los índices de discriminación.

g) Prueba Post-hoc

Utilizando la prueba de Dunn.

Cálculos:

Dunn's Pairwise Comparison of Idisc by Curso
(Bonferroni)

Col Mean- Row Mean	Anatomia	Bioquimi	Farmacol	Fisiolog
Bioquimi	1.574276 0.5771			
Farmacol	-2.546833 0.0544	-3.574573 **0.0018**		
Fisiolog	0.012236 1.0000	-1.161860 1.0000	2.089471 0.1833	
Microbio	3.262421 **0.0055**	1.541059 0.6165	4.832842 **0.0000**	2.420129 0.0776

Family-wise Error Rate = 0.05
Reject Ho if p = P(Z <= |z|) <= FWER/2

Resultado de la prueba post-hoc de Dunn, se encuentran diferencias significativas respecto al índice de discriminación entre las pruebas de los cursos de: Bioquímica y Farmacología (p=0.0018); las pruebas de Anatomía y Microbiología (p=0.0055) y las pruebas de Farmacología y Microbiología (0.0001).

CAPÍTULO 6: DISCUSIÓN

La finalidad del presente estudio fue evaluar y comparar la calidad de los instrumentos de evaluación de aprendizaje de conocimientos tipo cuestionarios de algunas asignaturas que se cursan en el área de Ciencias Básicas de la Facultad de Odontología de la Universidad Nacional Mayor de San Marcos (UNMSM). Para cumplir este objetivo primero se determinaron los índices de dificultad, de discriminación y la fiabilidad de los exámenes a partir del análisis de las respuestas a las preguntas (ítems). Estos análisis a las preguntas que conformaron un determinado instrumento de evaluación (examen) permitieron conocer la idoneidad, pertinencia y calidad de los mismos, una vez que éstos fueron aplicados. (Quiñones, 2016; Sánchez y cols. 2014; Rivera y Díaz, 2014; Rodríguez y cols. 2013; Morales, 2009).

En el presente estudio los resultados señalan que los exámenes aplicados en los cursos de Anatomía General, Bioquímica General, Farmacología II, Fisiología General y Microbiología General de Ciencias Básicas, tuvieron en promedio un índice de dificultad de 0,44 que los califica como exámenes Medianamente Fáciles; un índice de discriminación promedio de 0,36, calificando a los exámenes de Buena Calidad; y, un índice de fiabilidad promedio de 0,59 que los cataloga como exámenes de Moderada Confiabilidad.

Resultados similares a Quiñones (2016) en donde el examen Estatal Ordinario de Licenciatura en Enfermería fue catalogado como Medianamente Fácil; pero distinto en cuanto a la discriminación y confiabilidad que fueron catalogados como pobre y aceptable, respectivamente.

En los resultados de Ortiz y Cols.(2015), al analizar el examen final de la asignatura de Metodología señalaron que la mayoría de ítems eran fáciles, en el presente estudio el grado de dificultad fue Medianamente Fácil; en relación a la discriminación del examen este fue excelente; distinto al presente estudio que presentó una buena calidad.

En la investigación de Rodríguez y Cols. (2013), en Cuba, se concluyó que los exámenes finales de las asignaturas de Metodología de la Investigación y de Estadística clasificaron como fáciles, en el presente estudio es medianamente fácil; igualmente, respecto a la discriminación fue buena, similar a nuestro estudio que también fue de buena calidad; respecto a la confiabilidad presentan una marcada diferencia en la investigación de Quiñonez (2016), similar al presente estudio que hubo diferencia en la confiabilidad en las asignaturas estudiadas.

En el estudio realizado por Carrazana y Cols. (2011) en Cuba, éstos tuvieron como objetivo identificar el nivel de dificultad y poder de discriminación de la asignatura Morfofisiología Humana I, en los resultados se apreció que la dificultad osciló entre media y medianamente fácil, en el presente estudio también fue medianamente fácil; en la investigación de Carrazana obtuvieron un poder de discriminación adecuado y en el presente estudio fueron de buena calidad.

En la investigación realizada por Pérez y Cols.(2008) para determinar el nivel de dificultad y el índice de discriminación del tercer y quinto examen parcial de Cito-Histología, los resultados mostraron que el tercer parcial tuvo una dificultad medianamente difícil, mientras que el quinto parcial tuvo una dificultad media, en ambos parciales respecto al índice de discriminación tuvieron una calidad de excelente; en nuestro estudio, en la asignatura de Anatomía general, fue distinto los resultados, porque en el índice de dificultad fue medianamente fácil y el índice de discriminación fue de calidad buena. Además, en la investigación que realizó Backhoff y Cols. (2000), en su estudio donde expresan que el examen de

habilidades y conocimientos básicos, obtuvieron como resultados que la mayoría de los reactivos tienen una dificultad media y un buen poder discriminativo; diferente al presente estudio que los exámenes presentaron como dificultades medianamente fáciles y el índice de discriminación calificó a los exámenes de buenacalidad.

CONCLUSIONES

Conclusión General

Al evaluar la calidad de los instrumentos de evaluación de las asignaturas de Anatomía General, Bioquímica General, Microbiología General, Fisiología General y Farmacología II del área de Ciencias Básicas de la Facultad de Odontología de la Universidad Nacional Mayor de San Marcos durante el año lectivo 2016, se concluye que los exámenes de Anatomía General tienen un grado de dificultad medianamente fácil (0,38); un grado de discriminación bueno (0,37) y un grado de confiabilidad alto (0,699) ; los exámenes de Bioquímica General presentan un grado de dificultad media (0,62), un grado de discriminación bueno (0,30) y un grado de confiabilidad moderado (0,548); los exámenes de Farmacología II ostentan un grado de dificultad medianamente fácil (0,43), un grado de discriminación excelente (0,58) y un grado de confiabilidad alto (0,627); los exámenes de Fisiología General manifiestan un grado de dificultad altamente fácil (0,29), un grado de discriminación bueno (0,34) y un grado de confiabilidad moderado (0,587) y finalmente, los exámenes de Microbiología poseen un grado de dificultad medianamente fácil (0,50), un grado de discriminación regular (0,21) y un grado de confiabilidad moderado (0,506).

Al comparar los resultados se concluye que existen diferencias entre los índices de dificultad de los instrumentos de evaluación de las asignaturas del área de ciencias básicas; la prueba de Kruskal-Wallis arrojó un nivel de significancia de 0.0017.

Las diferencias significativas en el índice de dificultad se dan entre los exámenes de Anatomía General y Bioquímica General (p=0.0027); y, entre los exámenes Fisiología General y Bioquímica General (p=0.006) (prueba Post- hoc de Dunn). Del mismo modo, al comparar los resultados del índice de discriminación se concluye que existen diferencias entre los instrumentos de evaluación de las asignaturas del área de ciencias básicas; la prueba de Kruskal-Wallis arrojó un nivel de significancia de 0.001.

Las diferencias significativas en el índice de discriminación ocurren entre los exámenes de Farmacología II y Bioquímica General (p=0.0018); entre los exámenes de Microbiología General y Anatomía General (p=0.0055) y, entre los exámenes de Microbiología General y Farmacología II (p=0.0000), (prueba Post-hoc de Dunn).

Conclusiones Específicas

1. Los índices de dificultad de los instrumentos de evaluación de las asignaturas de Anatomía General, Farmacología II y Microbiología General, fueron medianamente fácil; el de Bioquímica General fue de Dificultad media y el de Fisiología General fue altamente fácil.

2. Los índices de discriminación de los instrumentos de evaluación de las asignaturas de Anatomía General, Bioquímica General y Fisiología General fueron buenos; mientras que el de Farmacología II fue excelente y el de Microbiología General fue regular.

3. Los índices de confiabilidad de los instrumentos de evaluación de las asignaturas de Anatomía General y Farmacología II fueron altos; mientras que el de las asignaturas de Bioquímica General, Fisiología General y Microbiología General fueron moderados.

RECOMENDACIONES

Recomendación General

Con respecto a las diferencias de la calidad de los instrumentos de evaluación de las asignaturas de Anatomía General, Bioquímica General, Microbiología General, Fisiología General y Farmacología II del área de Ciencias Básicas de la Facultad de Odontología de la Universidad Nacional Mayor de San Marcos durante el año lectivo 2016, se recomienda tomar en consideración los resultados para que en el Área de Ciencias Básicas se uniformice la dificultad de los exámenes, la discriminación sea mejorada y se eleve la confiabilidad en todas las asignaturas y que la extensión de los exámenes sea similar; con ello se lograría que los alumnos sean mejor evaluados y que los exámenes cumplan con los objetivos o competencias planteadas en los sílabos.

Una manera de ir mejorando la calidad de las preguntas de los exámenes sería la posibilidad de formar en el Departamento Académico de Ciencias Básicas un Comité de Evaluación que se encargaría de orientar acerca de las características de las evaluaciones, como por ejemplo el número de preguntas del examen, el tipo de reactivos, la puntuación, la duración de la evaluación, lasmodalidades de evaluación, etc. Por otro lado, podría encargarse también del análisis técnico de los exámenes, determinando sus indicadores de dificultad, de discriminación y de confiabilidad. En relación con las asignaturas, sería recomendable que el conjunto de la plana docente participe en la construcción del instrumento de evaluación, promoviendo el intercambio de ideas; igualmente la plana docente debería realizar el análisis de los exámenes

aplicados, lo cual permitiría uniformizar criterios, detectar preguntas muy fáciles, muy difíciles, mal redactadas, con claves erróneas, etc.

Se recomienda que el análisis de las evaluaciones forme parte inherente de la labor docente para aprovechar el potencial que tiene la evaluación en el proceso de enseñanza aprendizaje.

Recomendaciones Específicas

1. Referente al índice de dificultad de los instrumentos de evaluación de las asignaturas de Anatomía General, Bioquímica General, Microbiología General, Fisiología General y Farmacología II del área de Ciencias Básicas de la Facultad de Odontología de la Universidad Nacional Mayor de San Marcos durante el año lectivo 2016, se recomienda tomar en cuenta los resultados para que en todas las asignaturas se tenga el grado de dificultad óptimo que vendría a ser la Dificultad Media, con ello se lograría que la evaluación al estudiante sea un reflejo del nivel de conocimientos que éste posee.

2. En cuanto al índice de discriminación de los instrumentos de evaluación de las asignaturas de Anatomía General, Bioquímica General, Microbiología General, Fisiología General y Farmacología II del área de Ciencias Básicas de la Facultad de Odontología de la Universidad Nacional Mayor de San Marcos durante el año lectivo 2016, se recomienda tomar en cuenta los resultados para que en la medida de lo posible al preparar los exámenes todos los ítems planteados tengan la calidad de excelente; se lograría que un ítem bien planteado pueda tener un alto índice de discriminación en beneficio de los estudiantes.

3. Con respecto al índice de confiabilidad de los instrumentos de evaluación de las asignaturas de Anatomía General, Bioquímica General, Microbiología General, Fisiología General y Farmacología II del área de Ciencias Básicas de la Facultad de Odontología de la Universidad

Nacional Mayor de San Marcos durante el año lectivo 2016, se recomienda tomar en cuenta los resultados para lograr que todos los exámenes aplicados se acerquen más a la unidad haciendo del examen más confiable; se lograría que las preguntas de las pruebas se encuentren bien elaboradas y una óptima evaluación de los alumnos.

REFERENCIAS BIBLIOGRÁFICAS

Alonso, A. Ramos, M. & Quiñones, M. (2014). Influencia de la confección de los instrumentos teóricos de evaluación en la calidad de los resultados de los exámenes estatales en la Carrera de Tecnología de la Salud en los cursos 2009-2010 y 2010-2011. Revista Cubana de Tecnología de la Salud, 24(3), 32-36.

Argudín, E., Díaz, P. & Leyva, E. (2011). Índice de dificultad del examen de Morfofisiología Humana I. Educación Médica Superior, 25(2), 97-106.

Backhoff, E. Larrazolo, N. & Rosas, M. (2000). Nivel de dificultad y poder de discriminación del Examen de Habilidades y Conocimientos Básicos (EXCOHBA). Revista de Investigación Educativa, 2(1).

Baladrón, J., Sánchez-Lasheras, F., Villacampa, T., Romeo- Ladrero, J.M., Jiménez-Fonseca, P., Curbelo, J & Fernández- Somoano, A. (2017). El examen MIR 2015 desde el punto de vista de la Teoría de la Respuesta al Item. FEM, 20 (1), 29-38.

Baladrón, J., Curbelo, J., Sánchez-Lasheras, F., Romeo-Ladrero, J.M., Villacampa, T. & Fernández-Somoano, A. (2016). El examen al MIR 2015. Aproximación a la validez estructural a través de la Teoría Clásica de los Test.FEM, 19 (4), 217-226.

Carrazana, A., Salas, R. & Ruíz, A. (2011). Nivel de dificultad y poder de discriminación del examen diagnóstico de la asignatura de Morfofisiología Humana I. Educación Médica Superior, 25(1), 103-114.

Delgado, K. (2004) Evaluación y calidad de la educación. Nuevos Aportes. (3°ed.) Lima: Derrama Magisterial

Delgado-Maldonado, L. & Sánchez, M. (2012). Análisis del examen profesional de la Facultad de Medicina de la UNAM: una experiencia de evaluación objetiva del aprendizaje con la Teoría de la Respuesta al Item. Investigación en Educación Médica, 1 (3), 130-139.

Fermoso, P. (2013). Teoría de la educación. (2°ed.) México: Editorial trillas S.A de C.V.

González, M. (2002). La evaluación del aprendizaje: CEPES Universidad de la Habana, febrero.

González, M. (2001). La evaluación del aprendizaje: tendencias y evaluación crítica. Revista Cubana Superior, 15(1), 85-96.

González, M. (2000). Evaluación del aprendizaje en la enseñanza universitaria. Revista de Pedagogía Universitaria, 5(2).

Guerrero, J. & Faro, M.T. (2012). Breve análisis del concepto de educación superior. Papeles de Psicología, 15(27).

Hurtado, L.L. (2018). Relación entre los índices de dificultad y discriminación. Revista Digital de Investigación en Docencia Universitaria, 1(12), 273-300.

Mateo, J. & Vlachopoulos, D. (2013). Reflexiones en torno al aprendizaje y a la evaluación en la universidad. Educación XXI. 16 (2), 183-208.

Morales, P. (2009). Análisis de ítems en las pruebas objetivas. Universidad Pontificia Comillas, Madrid, Facultad de Ciencias Humanas y Sociales.

Moreno, R., Martínez, J. & Muñiz, J. (2011). Calidad de los ítems de los exámenes PIR. Papeles de Psicología, 32 (3), 254-264.

Moreno, T. (2009). La evaluación del aprendizaje en la universidad. Tensiones, contradicciones y desafíos. Revista Mexicana de Investigación Educativa, 14 (41), 563-591.

Muñiz, J. (2010). Las teorías de los test: Teoría clásica y Teoría de Respuesta a los Ítems. Papeles de Psicología, 32 (11), 57-66.

Muñiz, J. & Fonseca-Pedrero, E. (2008). Construcción de instrumentos de medida para la evaluación universitaria. Revista de Investigación Educativa, 1 (5), 13-25.

Ortiz, G., Díaz, P.A., Llanos, O.R., Pérez, S.M. & González, K. (2015). Dificultad y discriminación de los ítems del examen de Metodología de la Investigación y Estadística. EDUMECENTRO, 7 (2), 19-35.

Ortiz, L. (2006). La nueva concepción de la evaluación de los aprendizajes en la educación de las ciencias de la salud. Revista de Educación en Ciencias de la Salud, 3(1), 6-13.

Páez, R.O & Di Carlo, S. (2012). Aproximación docimológica a la evaluación de competencias digitales y didácticas de profesores universitarios. Revista iberoamericana de Evaluación Educativa, 5(1).

Pérez, J.H., Acuña, N. & Arratia, E.R. (2008). Nivel de dificultad y poder de discriminación del tercer y quinto examen parcial de la cátedra de Citohistología 2007 de la Carrera de Medicina de la UMSA. Revista Cuadernos, 2(1).

Quiñones, D. (2016). Caracterización del examen estatal aplicado en la Carrera de Licenciatura en Enfermería: curso 2012-2013. EDUMECENTRO, 8 (1), 111-124.

Rivera, M. & Díaz, P. (2014). Calidad del instrumento evaluativo en el examen escrito de Medicina Interna 2012-2013. En: III Jornada Científica de la Sociedad Cubana de Educadores en Ciencias de la Salud (SOCECS).

Rodríguez, A., Fernández, M.E. & Vigo, M.R. (2013). Indicadores de calidad en dos exámenes finales de Metodología de la Investigación y Estadística 2010- 2012. En: III Jornada de Educación Médica. Facultad de Odontología, Universidad de Ciencias Médicas de la Habana.

Sánchez, E., Medina, M., Rodríguez, M., Vega, L. & De La Torre, G. (2014). Indicadores de calidad para un examen teórico de la especialidad de Medicina General Integral. MEDISAN, 19 (2), 150-156.

San Martín, S., Jiménez, N. & Jerónimo, E. (2016). La evaluación del alumnado universitario en el Espacio Europeo de Educación Superior. Aula Abierta, 44, 7-14.

Soler, S. & Salas, R. (1997). Calidad de los instrumentos evaluativos. Ministerio de Salud Pública. Centro Nacional de Perfeccionamiento. Vicedirección de Evaluación. Cuba.

Villamarín, R.S. (2017). Indicadores de calidad de los instrumentos de evaluación y el rendimiento académico. Facultad de Educación, UNMSM, Tesis para optar el grado de doctor.

ANEXOS

Anexo N°01

Ficha de Recolección de Datos

Asignatura: .....
Área de Ciencias Básicas
Código de la asignatura .....
Examen de .....
Número de preguntas del examen .....

Estudiante (E)	Respuesta a las Preguntas										Calificación (Nota)
	P1	P2	P3	P4	“	“	“	“	“	Pn	
E1											
E2											
E3											
E4											
E5											
E6											
E7											
“											
“											
“											
“											
“											
“											
“											
“											
“											
“											
“											
“											
En											

Anexo N°02

Ficha resumen del índice de dificultad

Asignatura: .....
Código de la asignatura .....
Nivel académico de la asignatura: .....
Ciclo 2016-1
Examen de .....

Índice de dificultad por pregunta		
Número de preguntas	Índice de dificultad	Valoración
P1		
P2		
P3		
P4		
P5		
P6		
P7		
P8		
P9		
P10		
P11		
P12		
P13		
“		
“		
“		
“		
“		
“		
“		
“		
“		
“		
“		
“		
“		
“		
“		
“		
“		
Pn		

Índice de dificultad del examen: .....

Anexo N°03

Ficha resumen del índice de discriminación

Asignatura: .....
Código de la asignatura .....
Nivel académico de la asignatura .....

Índice de discriminación por pregunta		
Número de pregunta	Índice de discriminación	Valoración
P1		
P2		
P3		
P4		
P5		
P6		
P7		
P8		
P9		
P10		
P11		
P12		
P13		
“		
“		
“		
“		
“		
“		
“		
“		
“		
“		
“		
“		
“		
“		
“		
“		
Pn		

Ciclo 2016-1
Examen de .....

Índice de discriminación del examen .....

Ficha de la estructura de un examen

Asignatura: .....

Código de la asignatura .....

Nivel académico de la asignatura .....

Ciclo 2016-1

Examen de .....

Examen de	
Temario	**Número de preguntas**
1.	
2.	
3.	
4.	
5.	
Total de preguntas	

Confiabilidad de los exámenes	
Asignaturas	**Índice de confiabilidad**
Bioquímica general	
Anatomía general	
Microbiología general	
Fisiología general	
Farmacología II	