

**UNIVERSIDAD PEDAGÓGICA NACIONAL
FACULTAD DE CIENCIA Y TECNOLOGIA
DEPARTAMENTO DE FISICA
LÍNEA ENSEÑANZA DE LAS CIENCIAS DESDE UNA PERSPECTIVA
CULTURAL**



**ANÁLISIS PSICOMÉTRICO DE LA PRUEBA ESPECÍFICA DE ADMISIÓN A LA
LICENCIATURA EN FÍSICA DE LA UNIVERSIDAD PEDAGÓGICA NACIONAL**

PRESENTADO POR:

**WILSON ENRIQUE PATIÑO SAINEA
2010246044**

TUTOR:

NESTOR MENDEZ

**BOGOTÁ D.C. COLOMBIA
2018**

**UNIVERSIDAD PEDAGÓGICA NACIONAL
FACULTAD DE CIENCIA Y TECNOLOGIA
DEPARTAMENTO DE FISICA
LÍNEA ENSEÑANZA DE LAS CIENCIAS DESDE UNA PERSPECTIVA
CULTURAL**



**ANÁLISIS PSICOMÉTRICO DE LA PRUEBA ESPECÍFICA DE ADMISIÓN A LA
LICENCIATURA EN FÍSICA DE LA UNIVERSIDAD PEDAGÓGICA NACIONAL**

PRESENTADO POR:

**WILSON ENRIQUE PATIÑO SAINEA
2010246044**

ASESOR:

NESTOR MENDEZ

**BOGOTÁ D.C. COLOMBIA
2018**


AGRADECIMIENTOS

Agradezco a mi familia por estar conmigo durante los 7 años de esfuerzos, sacrificios, donde se presentaron diferentes dificultades durante el camino y fueron ellos quienes con su apoyo y fortaleza me dieron las fuerzas suficientes para poder superarlas y nunca dejarme desfallecer, fueron esas cuatro personas quienes me dieron la posibilidad de cumplir este sueño, también este sueño va por mi hijo David Patiño quien desde su llegada a este mundo fue fuente de mi fuerza para salir adelante y querer culminar con éxito este sueño que nos dará la posibilidad de estar más tiempo juntos y aprovechar mi rol como padre.

Gracias a la persona que me creyó en la idea que hoy se plasma como trabajo de grado mi profesor y asesor de trabajo de grado Néstor Méndez Hincapié quien durante muchos semestres estuvo ahí apoyándome, colaborándome con el trabajo de grado siendo esa persona que desde el ante proyecto estuvo ahí dándome un respaldo con las diferentes adversidades que se nos presentaban en el trabajo por ayudarme a salir adelante y poder cumplir una meta mas en mi vida como persona y como profesional

A los profesores que tuve durante los diferentes semestres que me ayudaron a formarme como persona y profesional y fueron pieza importante para darle forma y cuerpo a lo que es mi trabajo de grado y a la persona que soy hoy como profesional

GRACIAS...

	FORMATO	
	RESUMEN ANALÍTICO EN EDUCACIÓN - RAE	
Código: FOR020GIB	Versión: 01	
Fecha de Aprobación: 10-10-2012	Página 4 de 45	

1. Información General	
Tipo de documento	Trabajo de Grado
Acceso al documento	Universidad Pedagógica Nacional. Biblioteca Central
Título del documento	Análisis Psicométrico de la Prueba Específica de Admisión a la Licenciatura en Física de la Universidad Pedagógica Nacional
Autor(es)	Patiño Sainea, Wilson Enrique
Director	Méndez Hincapié, Néstor
Publicación	Bogotá, Universidad Pedagógica Nacional, 2017,46 p.
Unidad Patrocinante	Universidad Pedagógica Nacional
Palabras Claves	PSICOMETRÍA, ANÁLISIS, PRUEBA, TEORIA CLÁSICA DE LOS TEST.

2. Descripción
<p>A partir de la Teoría Clásica de los Test (TCT), se evaluó la Prueba Específica aplicada en el Departamento de Física de la Universidad Pedagógica Nacional a los aspirantes a entrar al programa de Licenciatura en Física. Este Trabajo tuvo en cuenta un rango de 30 ítems de la prueba que se emplearon entre los años 2011 y 2012 y que se aplicaron a 197 sujetos, desde las cuales se desarrolló el análisis.</p> <p>Se calcularon los índices de dificultad, homogeneidad y discriminación de la Prueba Específica de la Licenciatura en Física realizada en los años 2011 y 2012 verificando si se cumple con el criterio de fiabilidad bajo la Teoría Clásica de los Test</p>

3. Fuentes

- Abad, F. et.al. (2006). *Introducción a la Psicometría Teoría Clásica de los Tests y Teoría de la Respuesta al Ítem*. Facultad de Psicología Universidad Autónoma De Colombia. Recuperado de http://aprendeenlinea.udea.edu.co/lms/investigacion/file.php/39/ARCHIVOS_2010/PDF/Int_Psicometria_aristidesvara_1_.pdf
- Barbero García, I. Prieto Marañón, P. Suárez Falcón, J y Concepción San Luis Costas. (2001). Relaciones empíricas entre los estadísticos de la teoría clásica de los tests y los de la teoría de respuesta a los ítems. Vol. 13, nº 2, pp. 324-329. Recuperado de <http://www.psicothema.com/psicothema.asp?id=455>
- Ding, L; Chabay, R; Sherwood, B &, Beichner, R. (2006). Evaluating an electricity and magnetism assessment tool: Brief electricity and magnetism assessment. Physical review special topics - physics education research **2**, 010105. Recuperado de <https://pdfs.semanticscholar.org/6be5/349b6b209bd53945a7cb87cb374fba768578.pdf>
- García Cueto, E. (1996). Aplicación de modelos psicométricos para test y medidas alternativas en psicometría. Revista psicothema. Vol. 8, nº 2. Recuperado de: <http://www.redalyc.org/pdf/727/72780205.pdf>
- García, L. (2012). Construcción de un test para medir los conocimientos y aptitudes en matemáticas de estudiantes de primer año de universidad. En *Cuadernos de la Maestría en Docencia e Investigación Universitaria. Tomo II. Artículos de grado 2012*. Carmen Cecilia Suárez (Compiladora). Universidad Sergio Arboleda
- Martínez, R. (1995). *Psicometría: Teoría de los Tests Psicológicos y Educativos*. Universidad Complutense de Madrid. Editorial Síntesis S.A. España.
- Matas, A. (2011). Introducción al análisis de la teoría de respuesta al ítem. Aidesoc. Recuperado de: https://riuma.uma.es/xmlui/bitstream/handle/10630/4711/TRI_aidesoc_2011.pdf
- Montero, E. (2000). La teoría de respuesta a los ítems: una moderna alternativa para el análisis psicométrico de instrumentos de medición. Revista de Matemática: 217–228. Recuperado de: <https://revistas.ucr.ac.cr/index.php/matematica/article/viewFile/191/171>
- Muñiz. J (2010). Las teorías de los test: teoría clásica y teoría de respuesta a los ítems. Papeles del psicólogo. Vol. 31(1), pp. 57-66. Recuperado de <http://www.papelesdelpsicologo.es/pdf/1796.pdf>
- Navas, M. (1994). Teoría clásica de los test versus teoría de respuesta al ítem. U.N.E.D. Madrid. Pag: 175-180. Recuperado de: <http://www2.uned.es/490015/CV/TCTTRI94.pdf>

Reyes, F. Importancia de la Psicometría en la Psicología. Universidad Santo Tomás Psicología. recuperado de: <https://es.scribd.com/doc/58386407/Importancia-de-la-psicometria-en-psicologia>

4. Contenidos

Teniendo en cuenta que este trabajo se dividió en tres capítulos los cuales están estructurados de tal forma en la que se puede observar los pasos que se emplearon para poder desarrollar este trabajo, a continuación se hará una breve reseña de lo que comprende cada capítulo.

A partir del capítulo uno se da cuenta de la base fundamental del trabajo realizado puesto que trata de un Trabajo de Grado donde se realiza un análisis psicométrico de la Prueba Específica realizada a los aspirantes previos a su ingreso al programa Licenciatura en Física ofertado por la Universidad Pedagógica Nacional. Esto con el fin de resolver la incógnita generada a través del desarrollo de este, la cual es: ¿Cuáles fueron los índices de dificultad, homogeneidad y discriminación de la Prueba Específica de la Licenciatura en Física realizada en los años 2011 y 2012? ¿Cumple esta Prueba con el criterio de fiabilidad bajo la Teoría Clásica de los Test

Una vez conocidas las bases de datos de las preguntas empleadas en cada semestre desde el 2006 al 2014 se realiza un análisis de las mismas mirando cómo era aplicada esta prueba encontrando que se efectuaba de manera aleatoria y que solo consta de 157 preguntas. A partir de esto se halló que algunas pruebas se reiteraron entre los semestres 2011 y 2012. Con esto se da inicio al desarrollo del trabajo de grado teniendo claro primero la teoría la cual se sustenta el mismo.

Finalmente, es necesario recalcar que este trabajo de investigación tiene como objetivo, retroalimentar y potenciar de manera significativa la base de preguntas que es empleada para realizar un diagnóstico acertado en cuanto al ingreso de la población, a partir Teoría Clásica de los Test, es necesario abordar de manera previa el concepto de Psicometría la cual tuvo como finalidad medir con exactitud o diferenciar de una manera precisa el error en alguna prueba en específico, generando conceptos claros como facilidad o discriminación, los cuales determinan qué tan precisa es o no la prueba que se está aplicando, evaluando pregunta por pregunta, obteniendo así respuestas aplicadas a unos índices que se desarrollarán posteriormente y dan cuenta de la efectividad del test.

Ahora bien, durante el capítulo dos este cuenta de los referentes que se emplearon para poder desarrollar el trabajo de grado, puesto que se tomó los principales exponentes que hicieron parte de de la Teoría Clásica de los Tests como lo son Spearman , Francis Galton y Catell entre otros, de igual forma se hace referencia a los índices que se emplearon para poder dar una explicación al (TCT) como los son el índice de facilidad, homogeneidad y discriminación su aplicación en la prueba específica aplicada en la Licenciatura de Física de la Universidad Pedagógica Nacional y de la metodología que se

emplea durante el desarrollo del trabajo la cual fue una metodología cuantitativa ya que desde la Teoría Clásica de los Test se puede hacer un análisis estadístico de un instrumento de medición como lo es un test.

Finalmente, en el capítulo tres explica como el análisis psicométrico realizado a la prueba específica empleada en la Licenciatura en Física de la Universidad Pedagógica Nacional, los pasos necesarios para poder llegar a interpretar los resultados obtenidos en las pruebas aplicadas entre los años 2011 y 2012. A partir de la Teoría Clásica de los Tests, aplicando una metodología cuantitativa haciendo un análisis estadístico de la prueba

5. Metodología

Este ejercicio investigativo se encuentra enmarcado una metodología cuantitativa ya que desde la Teoría Clásica de los Test se puede hacer un análisis estadístico de un instrumento de medición como lo es un test. Esto se evidenciará a partir de gráficas donde se mostrará el banco de preguntas empleados en las pruebas específicas del semestre comprendidos entre los años 2011 y 2012.

6. Conclusiones

A partir del trabajo realizado con la prueba específica del Departamento de Física de la Universidad Pedagógica Nacional que se aplicó durante los años 2006-2014, se observó que la prueba consta un banco de preguntas de las cuales se aplicaban entre 20 y 30 escogidas de manera aleatoria durante los semestres comprendidos entre los años 2006-2 al 2014-1. Se encontró que entre los semestres del 2011-1 al 2012-2 se reiteraron las mismas 30 preguntas y que fueron aplicadas a un total de 197 sujetos y son estas preguntas las que se analizaron.

Bajo la Teoría Clásica de los Test y según los criterios que presenta Lin Ding, R. et.al. (2006), la prueba que se realizó entre el 2011-1 y el 2012-2 cumple con los índices de facilidad, homogeneidad, discriminación y discriminación total (Delta de Ferguson) como se aprecia en la tabla 1. Sin embargo, no se cumple con el índice de fiabilidad, aunque estuvo su valor de 0.645, cerca del valor deseado de ≥ 0.7 . Esto puede deberse a que la Prueba Específica de la Licenciatura en Física de la Universidad Pedagógica Nacional no mide necesariamente un solo rasgo psicológico sino varios: razonamiento matemático, comprensión lectora y conocimientos básicos de física. En ese sentido sería necesario realizar el análisis dividiendo el test en estas tres partes.

La primera pregunta (P3 del banco original de preguntas) arrojó valores negativos en los índices de homogeneidad y discriminación lo cual no es un resultado deseado porque quiere decir que está midiendo algo muy diferente a lo deseado. Esto enfatiza lo

mencionado en el párrafo anterior frente a diferentes rasgos psicológicos que se están midiendo. Además, se encontró que otras cinco (5) preguntas presentan simultáneamente índices de facilidad, homogeneidad y discriminación, fuera del rango deseado. Son las preguntas 4, 11, 13, 17 y 24, que corresponden al banco original de preguntas P19, P53, P63, P83 y P118 respectivamente, que hicieron que, aunque se lograron valores aceptados para los promedios de los índices de homogeneidad y discriminación, estos estuvieron muy cerca del límite permitido. Principalmente se debe revisar la primera pregunta, P3 del banco de preguntas original, ya que sus índices de homogeneidad y discriminación arrojaron valores negativos.

Hacer un estudio a la Prueba Específica del Departamento de Física de la Universidad Pedagógica Nacional, con el fin de saber cómo se encontraba constituida de qué forma estaba compuesta y cómo era su margen de calificación, conociendo esto se empezó hacer el estudio a la prueba basado en una teoría específicamente en la Teoría Clásica de los Test, haciendo un estudio a la misma interpretando y analizando cada dato que arrojaba la prueba y evaluándola a partir de la Teoría Clásica de los Test, con el fin de hacer un aporte a la prueba con el ánimo de contribuir para tener una prueba que tenga mejores estándares de “calidad”.

Se quiere dar un aporte a la Prueba Específica del Departamento de Física de la Universidad Pedagógica Nacional a partir del trabajo de grado realizado. Se deja preparada una Hoja de Cálculo con la cual es posible continuar este estudio que aporte elementos para hacer los ajustes necesarios a este análisis y a la prueba en general, y que además permita realizar una caracterización de los estudiantes que están ingresando a la Licenciatura en Física.

A futuro también sería deseable realizar un análisis bajo la Teoría de Respuesta al Ítem TRI teniendo en cuenta que su fundamentación matemática es mucho más compleja y requiere software especializado para realizar este tipo de trabajo, pero que puede aportar nuevos elementos de análisis ya que es factible realizar comparaciones aun cuando diferentes grupos respondan diferentes preguntas.

Elaborado por:	Wilson Enrique Patiño Sainea
Revisado por:	Néstor Méndez Hincapié

Fecha de elaboración del Resumen:	29	01	2018
--	----	----	------

TABLA DE CONTENIDO TABLA DE CONTENIDO

Introducción.....	10
1. CAPITULO I.....	12
1.1 Planteamiento del problema.....	12
1.2 Pregunta problema.....	13
1.3 Justificación.....	14
1.4 Objetivos.....	16
1.4.1 Objetivo General.....	16
1.4.2 Objetivos Específicos.....	16
2. CAPITULO II.....	17
2.1 Antecedentes.....	17
2.2 Marco Teórico.....	18
2.3 Metodología.....	26
3. CAPITULO III.....	28
3.1 Análisis de la Prueba.....	28
3.2 Base de datos: Bancos de preguntas y Respuestas periodos 2011-1, 2011-2 y 2012-1.....	29
3.2.1 Tabla de Respuestas de Selección Múltiple.....	30
3.2.2 Tabla de Consolidación de Respuestas.....	31
3.2.3 Tabla de Consolidación de Índices 1.....	32
3.3 Análisis de Ítems o Reactivos.....	33
3.3.1 Gráfica 1: Índice de Facilidad.....	33
3.3.2 Gráfica 2: Índice de Homogeneidad.....	34
3.3.3 Gráfica 3: Índice de Discriminación.....	35
3.3.4 Tabla de Consolidación de Índices 2.....	37
3.3.5 Gráfica 4: Índice de Facilidad.....	38
3.3.6 Gráfica 5: Índice de Homogeneidad.....	38
3.4 Tabla 1 Valores Deseados a partir de los Resultados.....	40
4. Conclusiones.....	42
5. Referentes Bibliográficos.....	44

INTRODUCCIÓN

Para ingresar al programa de Licenciatura en Física de la Universidad Pedagógica Nacional, los aspirantes deben realizar tres (3) pruebas. Una *Prueba de Potencialidad Pedagógica* PPP (que realizan todos los aspirantes a cualquier programa de la Universidad); las otras dos (2) pruebas las realiza Departamento de Física que son una *Entrevista* y una *Prueba Específica*. Esta última consiste en responder un test o cuestionario que oscila entre 20 y 30 preguntas aproximadamente de elección múltiple con única respuesta. Las preguntas son seleccionadas al azar de un conjunto de 157 preguntas que elaboraron varios de los docentes de la Licenciatura en Física.

En general los test son empleados por los psicólogos para analizar el comportamiento de las personas en diferentes ámbitos dando una observación desde su comportamiento a nivel personal basado en la aptitud. En particular, el test o cuestionario que realiza el Departamento de Física es denominado *test de rendimiento óptimo*. Consta de preguntas de razonamiento matemático, conocimientos básicos de física y comprensión lectora, y busca seleccionar los mejores estudiantes para que ingresen a la Licenciatura.

Para que esto sea factible, la prueba debe seguir un análisis estadístico que se ha desarrollado desde finales del siglo XIX, llamado *psicometría*, rama de la psicología que utiliza los métodos estadísticos para analizar algún rasgo psicológico.

La psicometría nos brinda un cuerpo de teoría y métodos para la medición de constructos en ciencias sociales. Uno de sus propósitos principales es el desarrollo de técnicas de aplicación empírica que permitan construir instrumentos de medición, indicadores, de alta confiabilidad y validez. Estas técnicas y métodos se basan en enfoques cuantitativos y utilizan conceptos, procedimientos y medidas derivadas de la estadística y la matemática. (Montero, 2000)

Existen dos enfoques teóricos a partir de los cuales se analizan los test: la Teoría Clásica de los Test (TCT) y la Teoría de Respuesta al Ítem (TRI). Si suponemos que frente a un constructo o rasgo psicológico (por ejemplo, conocimientos en física y matemáticas al nivel

de la Educación Media) existe un valor numérico verdadero que tiene un sujeto, entonces la TCT se basa en el supuesto de que la puntuación obtenida en un test es la misma puntuación verdadera más un pequeño error que no se correlaciona con la puntuación. La TRI se basa en la suposición de que a una pregunta (o ítem) del test, el sujeto la responde con una probabilidad previamente definida. Este segundo enfoque es mucho más complejo que el primero, pues requiere del uso de un software especializado, por lo que aquí se ha escogido para el análisis el primer enfoque (TCT).

En el estudio de la TEORIA CLASICA DE LOS TEST, surgió la psicometría, que tuvo como finalidad medir con exactitud o diferenciar de una manera precisa el error en alguna prueba en específico, generando conceptos claros como índices de facilidad, homogeneidad y discriminación, los cuales determinan qué tan precisa es o no una prueba que se está aplicando, evaluando pregunta por pregunta o en conjunto, obteniendo así respuestas aplicadas a los índices mencionados que se desarrollarán posteriormente y dan cuenta de la efectividad del test.

1. CAPITLO I

1.1 PLANTEAMIENTO DEL PROBLEMA

Para garantizar la selección adecuada de aspirantes a ingresar al programa de Licenciatura en Física, a través de la Prueba Específica, se debe garantizar que esta conste de preguntas adecuadas de razonamiento matemático, de conocimientos de física y de comprensión de lectura, que garanticen la correcta discriminación entre los sujetos que realizan el test.

El presente Trabajo de Grado busca hacer un análisis psicométrico de la Prueba Específica que se les realiza a los aspirantes a ingresar a la Licenciatura en Física de la Universidad Pedagógica Nacional bajo la TCT, donde la prueba se debe ajustar a unos criterios de calidad dado que es una prueba de elección múltiple con única respuesta en el cual este test pretende hacer la selección a los aspirantes que ingresarían a la Universidad. A través de la Teoría Clásica de los Tests se quiere examinar la misma a partir de los indicadores que se nombran a continuación para poder identificar si es una prueba discriminatoria y si los criterios de fiabilidad y validez se cumplen.

Para lograr identificar los indicadores nombrados en el párrafo anterior, es necesario realizar una exploración de las diferentes pruebas que se han venido aplicando desde el segundo semestre del año 2006 al primer semestre del año 2014, para poder reducir el número de preguntas que fueron tenidas en cuenta en cada una de las pruebas aplicadas.

Una vez identificadas las preguntas se dio inicio al proceso de análisis de resultados referentes a los siguientes cuestionamientos: ¿cuántos estudiantes lograron responder de manera acertada cada una de las preguntas?, ¿cuántos lo hicieron de manera incorrecta?, con base a las treinta preguntas, se realizó un promedio general, concibiendo índices bajos en cuanto al conglomerado total de respuestas, permitiendo así, aplicar los índices (homogeneidad, facilidad y confiabilidad) según la Teoría Clásica de los Test (TCT), que arrojaron resultados relevantes, evidenciando que la prueba cuenta con algunas preguntas que presentan algún tipo de dificultad, debido a su bajo índice de solución, las cuales serán

evaluadas para poder identificar el motivo por el cual se presenta un índice poco favorable, cotejando el motivo por el cual se presenta un margen negativo y el proceso a seguir para formular una posible reestructuración de las mismas, permitiendo ser evaluadas por los docentes encargados en las pruebas y posterior a ello, ser tomadas en cuenta en un próximo estudio de la prueba, donde posiblemente se evidencie una mejoría en la aplicación de la misma.

1.2 Pregunta problema

¿Cuáles fueron los índices de facilidad, homogeneidad y discriminación de la Prueba Específica de la Licenciatura en Física realizada en los años 2011 y 2012? ¿Cumple esta Prueba con el criterio de fiabilidad bajo la Teoría Clásica de los Test?

1.3 JUSTIFICACIÓN

El presente Trabajo de Grado hace un análisis psicométrico de la Prueba Específica realizada a los aspirantes al programa Licenciatura en Física ofertado por la Universidad Pedagógica Nacional, con el fin de saber con certeza que era lo que se estaba evaluando en la prueba específica del departamento y como era evaluada. A partir de la Teoría Clásica de los Tests, se desarrolló el interrogante de cómo se llevan a cabo los índices de evaluación de la prueba que era dada por el Departamento del mismo programa. Esto con el fin de resolver la incógnita generada para el desarrollo de este trabajo, la cual es: ¿Cuáles son los índices de dificultad, homogeneidad y discriminación de la prueba específica de la Licenciatura en Física? ¿Cumple con el criterio de fiabilidad bajo la Teoría Clásica de los Test?

Para dar solución al cuestionamiento planteado anteriormente, se abordó a partir de la Teoría Clásica de los Test (TCT), pues es importante identificar a través de este si la prueba actualmente realizada cuenta con adecuados índices de facilidad y si los criterios de fiabilidad y discriminación se cumplen.

Para llevar a cabo lo anterior, se pone en marcha un estudio que permite evaluar cada una de las pruebas ejecutadas durante los periodos comprendidos entre el segundo semestre del año 2006 y el primer semestre del año 2014, encontrando una serie de resultados aleatorios obtenidos a partir de la poca, escasa o nula reiteración en la empleabilidad del banco de preguntas durante la aplicación de la prueba creando aún más incógnitas y generando una serie de preguntas como: ¿cuántos estudiantes lograron responder de manera acertada cada una de las preguntas?, ¿cuántos lo hicieron de manera incorrecta?, ¿Por qué no eran reiteradas las preguntas durante los primeros años que se empezó aplicar en el departamento? Preguntas como estas incentivan el desarrollo de la investigación.

Finalmente, es necesario recalcar que este trabajo de investigación tiene como objetivo, retroalimentar y potenciar de manera significativa la base de preguntas que es empleada para

realizar un diagnóstico acertado en cuanto al ingreso de la población. En el cual se puede destacar que todas las pruebas que se emplean se hace a través de pruebas psicométricas, pues estas pruebas tiene como objetivo contextualizar la población a la cual es aplicada, para esto es necesario poder conocer el banco de preguntas la cuales fueron empleadas y poder dar un criterio preciso de cuál fue la problemática.

Debido a cuestiones de seguridad no fue posible obtener el banco de preguntas lo cual generaría la posibilidad de ser un poco más exacto en la problemática que presente en la pregunta en específico, puesto que puede obedecer a diferentes posibilidades que dichas preguntas no sean contestadas de manera correcta bien sea por su formulación o el contexto de la misma.

El objetivo de este trabajo durante el proceso de formación como docente contribuyo en la forma de como se es posible implementar una prueba “test” teniendo claro unos indicadores que ratifiquen la viabilidad de la prueba teniendo trazado un objetivo claro a la hora de implementarla, ya que muchas veces el educador realiza este tipo de pruebas no con el fin de evaluar los conceptos de un estudiante “ un resultado veraz” si no el fin que la prueba arroje unos resultados que demuestre la eficacia del mismo en un plantel educativo haciéndolo de manera inconsciente, debido a que no se tiene un conocimiento que ayude a atender la estructura que esta tiene para poder estructurar una prueba. Teniendo un ejemplo como este se puede ir llevando al estudiante que hacia un futuro se desempeñará como docentes de diversas instituciones del país llevando el nombre de la Universidad Pedagógica Nacional en alto por su compromiso en la formación de excelentes y calificados profesionales.

1.4 OBJETIVOS

1.4.1 OBJETIVO GENERAL

- Hacer un análisis psicométrico de la Prueba Específica de admisión a la Licenciatura en Física de la UPN por medio de la Teoría Clásica de los Test, TCT, que se realizó en los años 2011 y 2012, para determinar los índices de facilidad, homogeneidad, discriminación y fiabilidad, con el fin de comprobar que la prueba se ajusta a los criterios de calidad dada que es una prueba de elección múltiple de única respuesta para elegir de cierta manera los aspirantes más preparados para la Licenciatura en Física.

1.4.2 OBJETIVOS ESPECIFICOS

- Realizar un seguimiento conciso de la aplicación de la prueba específica de la Licenciatura en Física para construir la matriz de trabajo.
- Calcular los índices de dificultad, homogeneidad y discriminación e interpretar los resultados.
- Calcular el índice de fiabilidad y discriminación de la prueba total para conocer el funcionamiento de la prueba y su intención al momento de utilizarla.

2. CAPITULO II

2.1 ANTECEDENTES

Al indagar acerca de la Teoría Clásica de los Test (TCT), se encuentra que esta no ha sido objeto de estudio que aporte al proceso de mejora y reestructuración de metodologías o filtros que faciliten el proceso de evaluación de una prueba en específico generando conceptos claros como facilidad o discriminación, los cuales determinarían la precisión de la prueba. Por tal razón los aspirantes que pretendan ingresar estarán sometidos a un proceso de evaluación que determine los conceptos nombrados anteriormente y de acuerdo a los resultados obtenidos posterior aceptación de aspirantes a la Universidad Pedagógica Nacional, por tal motivo los recursos e información con la cual se sustenta toda esta investigación, son sustraídos de autores referentes y calificados tomados de fuentes ajenas a la institución, los cuales son presentados a continuación:

Según Muñiz. J (2010), para una interpretación y utilización adecuada de las propiedades psicométricas de los test es necesario ir más allá del mero cálculo empírico, y conocer los fundamentos en los que se basan esos cálculos. Con el fin de contribuir a esta comprensión más allá del mero manejo superficial de las fórmulas psicométricas, el objetivo fundamental de este trabajo es presentar de una manera no excesivamente técnica y especializada las dos grandes teorías que guían la construcción y análisis de la mayoría de los test: la Teoría Clásica de los Test y la Teoría de Respuesta a los Ítems. En primer lugar, se hace un apunte histórico sobre los test, indicando cómo surgen y evolucionan al hilo de los avances técnicos y estadísticos. Tras razonar acerca de la necesidad de utilizar teorías psicométricas para el análisis y construcción de los test, se expone la lógica que subyace a la Teoría Clásica de los Test, así como sus dos variantes más granadas, la Teoría de la Generalizabilidad y los Test Referidos al Criterio. Luego se subrayan las limitaciones más importantes del enfoque clásico y se exponen los fundamentos de la Teoría de Respuesta a los Ítems, en cuyo marco encuentran una solución satisfactoria algunos de los problemas que el enfoque clásico no

había sido capaz de resolver de forma satisfactoria. Finalmente se comparan ambos enfoques, y se concluye indicando la necesidad de conocer las teorías de los test para una mejor comprensión y utilización de los instrumentos de medida.

2.2 MARCO TEÓRICO

A continuación se encontrará la teoría general sobre la cual soporta la Teoría Clásica de los test, basándose inicialmente en los autores que dieron inicio la psicometría y en que ámbitos era utilizada. Posteriormente se definirán los índices de facilidad, discriminación y homogeneidad, los cuales permite entender cómo se está evaluando la prueba específica de la Licenciatura en Física de la Universidad Pedagógica Nacional. Durante el proceso de investigación realizado en la Universidad Pedagógica Nacional en la Licenciatura en Física, se llevó a cabo un análisis a la prueba específica de admisión implementada en la licenciatura en los períodos comprendidos entre los años 2011-1 al 2012-2 por medio de un análisis psicométrico. Esta prueba consta de un banco de preguntas de 157 las cuales fueron distribuidas de manera aleatoria durante los semestres, que fueron iguales durante los periodos mencionados anteriormente. Con base en la prueba y la TCT se realizó un análisis psicométrico desarrollando un estudio de la prueba que dictamina la eficacia que esta tiene y su desempeño al momento de ser aplicada.

Para lograr esto, es necesario empezar a hablar de la psicología, la psicometría y de los diferentes exponentes que contribuyeron al desarrollo de la psicometría. Teniendo como base fundamental un artículo o guía para el desarrollo del este trabajo de grado como: *Evaluating an electricity and magnetism assessment tool: Brief electricity and magnetism assessment.* (Ding, Chabay, Sherwood & Beichner, 2006)

Al momento de llevar a cabo un análisis psicométrico es obligatorio hablar de dónde nace y para ello es necesario hablar de la psicología como ciencia que estudia la conducta y los procesos mentales, que describe algunos aspectos del pensamiento, específicamente en tres

dimensiones: cognitiva, afectiva y comportamental. A partir de esta ciencia surge la psicometría que se encarga de medir por medio de una escala los atributos psicológicos del individuo.

Por otro lado, la psicometría es el estudio de las pruebas realizadas a partir de test de rendimiento óptimo que ayuda a medir de manera objetiva y estandarizada, un rasgo psicológico, por medio de instrumentos arrojando un diagnóstico de la prueba en específico.

La psicometría, engloba la teoría y la construcción de pruebas, test y otros procedimientos de medición válidos y confiables. Incluye, por tanto, la elaboración y aplicación de procedimientos estadísticos que permitan determinar si una prueba o (test) es válido o no para la medición de una variable o conducta psicológica previamente definida. Sin embargo, la medida de estos aspectos no directamente observables es difícil, y gran parte de la investigación y técnicas acumuladas en esta disciplina están diseñadas para definirlos de manera fiable antes de cuantificarlos. (Reyes, s.f)

En esta investigación es necesario tener en cuenta los principales exponentes de la psicometría para poder contextualizar la época en la cual se empezaron a desarrollar las pruebas psicométricas y a partir de estos los diferentes términos que se emplean para el desarrollo de las pruebas.

La psicología aparece como una disciplina independiente en 1860 con ayuda de Fechner, quien se preocupó por medir la intensidad de las sensaciones desarrollando la primera metodología cuantitativa proponiendo procedimientos de medida, siendo este personaje uno de los exponentes más importante para la psicología moderna.

Sir Francis Galton (1822-1911) fue quien fundó el laboratorio antropométrico de Kensington, además de ser el primero en aplicar la tecnología estadística que utilizó para aplicar en sus datos realizados de sus test.

Observó que los sujetos retrasados mentales mostraban deficiencias en la capacidad para discriminar la temperatura y en su reacción al dolor y fue esto, entre otras cosas, lo que le llevo a pensar que la capacidad de discriminación sensorial sería más alta entre los más capaces intelectualmente y a intentar medir las capacidades intelectuales por medio de dichos procesos básicos. (Martínez, 1995)

Martínez (1995), cita a J. Mckeen Cattell (1860 -1944) como unos de los primeros psicólogos del siglo XIX y principios del siglo XX *“escribió en 1890 un artículo en el que apareció por primera vez de la denominación de test mental en el que describió una serie de test que se aplicaban a estudiantes de “college” para determinar su nivel intelectual”*. (Martínez, 1995)

Con esto comenzó a estudiar de manera práctica y experimental el desarrollo de diferentes pruebas de memoria lo cual ayudó a evaluar los tiempos de desarrollo de una actividad, diferencia sensorial, con el fin de poder evaluar el desempeño que lograrse tener una estudiante en sus diferentes actividades.

Ademas de esto, el mismo Martínez (1995) menciona a Karl Pearson (1857 -1936) (fundador de la estadística moderna) como un personaje que contribuyó en el desarrollo de metodos y teorías como usos de test y de cuestionarios. Ya que fue una de las personas que desarrolló la estadística de manera integra, uno de sus mas grandes aportes fue la correlación lineal.

Charles Spearman (1863 – 1945) tambien es mencionado por Martínez, quien indica lo siguiente:

Desarrolló la teoría de la fiabilidad de los test y el modelo estadístico de las puntuaciones conocido como teoría clásica de los tests, así como la técnica de los análisis de factorial que sigue siendo una de las técnicas más utilizadas en la validación de los test (Spearman,1904a,1904b, 1907 1910, 1913)

Cada una de las técnicas que creó las llevó a las aulas escolares; un ejemplo fue el diseño una técnica que valorara cómo un estudiante obtenía buenos puntajes en una asignatura en particular, este mismo individuo debería de obtener los mismos resultados en las diferentes materias, denominado teoría bifactorial de la inteligencia. Así como esta técnica, empleó muchas otras a favor no solo de la psicología si no su implementación en la educación basandose en la estadística.

El principal exponente de la Teoría Clásica de los tests fue Spearman, quien introdujo la psicología como un factor importante para el desarrollo de la inteligencia humana, comenzando con la conducta y el desarrollo específico de cada ser, de acuerdo con lo mencionado, a partir del texto realizado por Muñiz, este último comenta el modelo que propuso Spearman, y lo que pretendía evaluar al ser aplicada.

Propone un modelo muy simple, de sentido común, para las puntuaciones de las personas en los test, y que ha dado en llamarse modelo lineal clásico. Consiste en asumir que la puntuación que una persona obtiene en un test, que denominamos su puntuación empírica, y que suele designarse con la letra X , está formada por dos componentes, por un lado, la puntuación verdadera de esa persona en ese test (V), sea la que sea, y por otro un error (e). (Muñiz, 2010).

Spearman plantea tres supuestos que se incorporaran en la Teoría Clásica de los Test, el primero de ellos es el definir la puntuación verdadera de un test implementada en la siguiente ecuación.

Supuestos de la TCT

$$V = X + E$$

V = Es la puntuación verdadera, desconocida, del individuo

X = Es la puntuación obtenida en el test

E = Es el error de la medición.

El segundo supuesto

El valor de la puntuación verdadera de una persona no tiene nada que ver con el error que afecta esa puntuación, es decir, puede haber puntuaciones verdaderas altas con errores bajos, o altos, no hay conexión entre el tamaño de la puntuación verdadera y el tamaño de los errores. De nuevo se trata de un supuesto en principio razonable, que formalmente puede expresarse así: $r(v,e) = 0$. (Muñiz, 2000)

Tercer supuesto

Establece que los errores de medida de las personas en un test no están relacionados con los errores de medida en otro test distinto. Es decir, no hay ninguna razón para pensar que los errores cometidos en una ocasión vayan a covariar sistemáticamente con los cometidos en otra ocasión. Formalmente este supuesto puede expresarse así: $r(e_j, e_k) = 0$. (Muñiz, 2000)

Evaluación estadística de una prueba

Para determinar si los ítems miden lo que se pretende en una prueba, se definen los índices de facilidad, homogeneidad y discriminación.

- *Índice de facilidad* del ítem j , (P_j) se define como la razón entre el número de sujetos que han acertado en este ítem, (A_j) y el número total de sujetos que han intentado resolverlo N_j (Abad, F. et. al. 2006).

$$P_j = \frac{A_j}{N_j}$$

Los N_j que han intentado resolverlo son los que seleccionaron una de las opciones; quienes no seleccionaron una opción y dejaron en blanco la respuesta no son considerados dentro de los N_j . Los valores posibles de este índice varían de 0 a 1, siendo 0 un ítem muy difícil (nadie respondió acertadamente) y, 1 muy fácil (todos los que lo intentaron lo respondieron acertadamente).

- *Índice de homogeneidad H_j* hace referencia a una prueba que es aplicada a cierto tipo de población, la cual tiene conocimiento y manejo de esta. “*Índice de homogeneidad de un ítem H_j se define como la correlación de Pearson r entre las puntuaciones en el ítem j de los N sujetos y las puntuaciones X en el total del test*”. (Abad, F. et. al. 2006)

$$H_j = r_{jX}$$

Este índice toma valores entre -1 y 1. Valores negativos indican que el ítem mide algo diferente a lo que pretende medir la prueba total y por tanto este ítem o reactivo debería descartarse.

- *Índice de discriminación D_j* también mide la consistencia del test y se define como la diferencia entre el número de aciertos del ítem j del 25% de los sujetos que obtuvieron el máximo puntaje total $NH(25\%)$, y el número de aciertos del mismo ítem del 25% de sujetos que obtuvieron los puntajes más bajos $NL(25\%)$:

$$D_j = \frac{NH(25\%) - NL(25\%)}{N/4}$$

Este índice también toma valores entre -1 y +1. Los valores negativos indican que hay más sujetos en el 25% inferior que respondieron correctamente este ítem que sujetos del 25% superior. El caso ideal extremo +1 se obtiene si el 25% superior contestó correctamente este ítem y ninguno del 25% inferior. En el peor de los casos sucede lo contrario.

- Índice de fiabilidad de Kuder-Richardson mide la auto-consistencia del test completo. Si por ejemplo el test se aplicara en dos momentos sucesivos al mismo grupo de sujetos, se esperaría que los resultados fueran altamente correlacionados. Hay que suponer, por supuesto, que los dos test se realizaron en las mismas condiciones ambientales. La correlación dará el índice de fiabilidad. Pero este método no resulta apropiado ya que los estudiantes pueden recordar y estudiar las preguntas lo que hará que las condiciones para el segundo momento no serán idénticas a las del primer momento. Podrían diseñarse dos test que midan el mismo rasgo, con los mismos o muy similares índices de facilidad, homogeneidad y discriminación, pero esto también resulta muy difícil. Kuder - Richardson se basaron en la idea de suponer que el test completo se puede dividir en partes iguales y calcular la correlación entre las partes. Luego llevaron la idea al extremo, suponiendo que cada ítem es un test en sí y para una prueba de opción múltiple donde cada elemento solo se califica como "correcto" o "incorrecto", la fórmula se puede escribir como

$$r_{test} = \frac{K}{K-1} \left(1 - \frac{\sum P(1-P)}{\sigma_X^2} \right)$$

Donde P es el índice de facilidad, σ_X es la desviación estándar de los puntajes de la prueba total y K es el número total de ítems que tiene la prueba. Expresa una relación entre la varianza de error y la varianza verdadera, de acuerdo a un porcentaje pequeño de error en una prueba en el especial, se entiende por este concepto el grado de estabilidad, precisión o consistencia que manifiesta el test como instrumento de medición de un rasgo determinado. Este índice toma valores entre 0 y 1. “Un criterio ampliamente aceptado es que las pruebas con índices de confiabilidad superiores a 0,7 son confiables para la medición grupal y las pruebas con índices de confiabilidad superiores a 0,8 son confiables para la medición individual.” (Lin Ding, R. et.al. 2006).

- El delta de Ferguson δ es otra estadística de la prueba completa y mide el poder discriminatorio del test completo porque muestra qué tanto se distribuyen las puntuaciones totales entre los sujetos dentro de un rango posible. Si se diseña y aplica una prueba para discriminar entre unos estudiantes, una distribución amplia del puntaje total sería lo ideal (Lin Ding, R. et.al. 2006). El cálculo del delta de Ferguson se basa en la relación entre los puntajes totales de dos sujetos. Estos puntajes pueden ser diferentes o iguales. Si una muestra tiene N sujetos, entonces el número total de parejas que se pueden formar es $N(N - 1)/2$, y el número total de parejas de puntajes iguales es $\sum f_i(f_i - 1)/2$. Aquí f_i representa el número de sujetos que obtuvieron de puntaje total i donde $i = 0, 1, \dots, K$. La resta entre el número total de parejas y el número total de parejas con puntajes iguales da como resultado el número total de parejas de puntajes diferentes:

$$\frac{N(N - 1)}{2} - \sum \frac{f_i(f_i - 1)}{2} = \frac{N^2 - \sum f_i^2}{2}$$

La máxima distribución se logra si $f_i = N/(K + 1)$ que al reemplazar en la fórmula anterior dará $[N^2 - N^2/(K + 1)]/2$. Entonces el delta de Ferguson se define como la razón entre el número de parejas de puntajes totales diferentes y el máximo número posible de parejas de puntajes totales diferentes:

$$\delta = \frac{N^2 - \sum f_i^2}{N^2 - N^2/(K + 1)}$$

Toma valores entre 0 (cuando todos los sujetos obtienen el mismo puntaje) y 1 cuando $f_i = N/(K + 1)$ para todo i . Se considera (Lin Ding, R. et.al. 2006) que test con valores δ superiores a 0.9 ofrecen una buena discriminación.

Debido que la prueba específica de la Licenciatura en Física de la Universidad Pedagógica Nacional está diseñada básicamente para medir una puntuación desde cero a una puntuación

máxima (igual al número total de preguntas o ítems = 30), esta se considera buena para realizar un análisis psicométrico visto desde la Teoría Clásica de los Test (TCT), la cual requiere calcular para cada ítem o pregunta los índices de facilidad, homogeneidad, confiabilidad. Y los índices de toda la prueba: K 21, Ferguson.

2.3 METODOLOGIA

Este ejercicio investigativo se encuentra enmarcado en una metodología cuantitativa ya que desde la Teoría Clásica de los Test se puede hacer un análisis estadístico de un instrumento de medición como lo es un test (García, 2012).

Este instrumento de medición “psicometría” será empleada en la prueba específica de la Licenciatura en Física realizada en los años 2011 y 2012, la cual está compuesta por 30 ítems o preguntas de elección múltiple con única respuesta, resueltas por 197 aspirantes a dicho programa. Los ítems o preguntas arrojan información desde la parte cognitiva valorado en una escala de 0 a 1 siendo cero una respuesta errada y uno una respuesta correcta susceptible de ser medido utilizando la estadística como herramienta de análisis.

Para esto se realizó una revisión a cada una de las pruebas que se realizaron en el departamento de Física de la Universidad Pedagógica Nacional comprendido entre los años 2006-2 al 2012-2, encontrando que esta presenta varios vacíos en su formato puesto que no en todos los semestres se evidenciaba una consolidación de preguntas de respuestas por parte de los aspirantes que realizaron las pruebas en dichos semestres, hallando que solo se reiteraban las preguntas en los últimos dos semestres y en los cuales se basó la investigación realizada para el desarrollo del trabajo de grado.

Haciendo uso de la Teoría Clásica de los Test se hará un análisis estadístico logrando responder al grado de dificultad y discriminación de la prueba; para esto se tendrá en cuenta los parámetros desarrollados anteriormente en el marco teórico como son: fiabilidad,

homogeneidad y discriminación para todos y cada uno de los ítems, y fiabilidad y discriminación de la prueba total, lo que dará un diagnóstico concreto del por qué se hace esta prueba y si en realidad se cumple con los parámetros establecidos al momento de realizarse la misma. Esto se evidenciará a partir de gráficas donde se mostrará el banco de preguntas empleadas durante las pruebas realizadas las cuales se analizaron, como también se omitieron algunos bancos con el fin de observar el comportamiento que esta tenía al momento de excluir esas preguntas que presentaban bajos índices demostrando que la prueba presentaba una leve mejoría con respecto a los índices, dando aun mayor importancia a la investigación puesto que le da aun mas validez poder evidenciar que al lograr hacer una reestructuración con respecto a esas preguntas la prueba específica tendrá aun mejores resultados.

La investigación se desarrolló en las siguientes etapas:

1. Revisión bibliográfica
2. Selección de ítems del Banco de preguntas y respuestas 2006 – 2014
3. Codificación de los ítems seleccionados
4. Análisis de ítems y de la prueba total (cálculo de los índices)

3. CAPITULO III

3.1 ANALISIS DE LA PRUEBA

Se evaluó la Prueba Específica aplicada en el Departamento de Física de la Universidad Pedagógica Nacional a los aspirantes a ingresar al programa de Licenciatura en Física. Se hizo una revisión de las diferentes pruebas que se han venido aplicando desde el segundo semestre del año 2006 al primer semestre del año 2014, donde se logró evidenciar diferentes factores encontrando una base de preguntas denominado banco de preguntas con un total aproximado de 157 ítems, aplicándose de manera aleatoria en cada uno de los semestres nombrados anteriormente, lo cual trajo consigo el primer inconveniente al momento de ser evaluada desde (TCT), pues no eran reiteradas las preguntas que se empleaban durante los mismos periodos,

Lo anterior dio paso a realizar una revisión de las pruebas aplicadas, encontrando un rango de 30 ítems de la prueba que se emplearon entre los años 2011 y 2012 y que se aplicaron a 197 sujetos, desde las cuales se desarrolló el análisis de este trabajo, bajo la Teoría Clásica de los Test (TCT), los cuales se tomaron para realizar el análisis de la prueba partiendo primordialmente de una base como la que se mostrara a continuación:

3.2 Base de datos: Bancos de preguntas y Respuestas periodos 2011-1, 2011-2, 2012-1 y 2012-2

	2011-1			2011-2			2012-1			2012-2		
PREGUNTA	BANCO	RESPUESTA	PREGUNTA	BANCO	RESPUESTA	PREGUNTA	BANCO	RESPUESTA	PREGUNTA	BANCO	RESPUESTA	
1	P3	e	1	P3	e	1	P3	e	1	P3	e	
2	P8	a	2	P8	a	2	P8	a	2	P8	a	
3	P13	c	3	P13	c	3	P13	c	3	P13	c	
4	P19	a	4	P19	a	4	P19	a	4	P19	a	
5	P23	b	5	P23	b	5	P23	b	5	P23	b	
6	P28	c	6	P28	c	6	P28	c	6	P28	c	
7	P33	a	7	P33	a	7	P33	a	7	P33	a	
8	P38	d	8	P38	d	8	P38	d	8	P38	d	
9	P43	c	9	P43	c	9	P43	c	9	P43	c	
10	P48	d	10	P48	d	10	P48	d	10	P48	d	
11	P53	b	11	P53	b	11	P53	b	11	P53	b	
12	P58	c	12	P58	c	12	P58	c	12	P58	c	
13	P63	d	13	P63	d	13	P63	d	13	P63	d	
14	P68	a	14	P68	a	14	P68	a	14	P68	a	
15	P73	a	15	P73	a	15	P73	a	15	P73	a	
16	P78	a	16	P78	a	16	P78	a	16	P78	a	
17	P83	d	17	P83	d	17	P83	d	17	P83	d	
18	P88	d	18	P88	d	18	P88	d	18	P88	d	
19	P93	d	19	P93	d	19	P93	d	19	P93	d	
20	P98	c	20	P98	c	20	P98	c	20	P98	c	
21	P103	b	21	P103	b	21	P103	b	21	P103	b	
22	P108	d	22	P108	d	22	P108	d	22	P108	d	
23	P113	a	23	P113	a	23	P113	a	23	P113	a	
24	P118	b	24	P118	b	24	P118	b	24	P118	b	
25	P123	c	25	P123	c	25	P123	c	25	P123	c	
26	P128	d	26	P128	d	26	P128	d	26	P128	d	
27	P133	a	27	P133	a	27	P133	a	27	P133	a	
28	P138	d	28	P138	d	28	P138	d	28	P138	d	
29	P143	a	29	P143	a	29	P143	a	29	P143	a	
30	P148	c	30	P148	c	30	P148	c	30	P148	c	

Teniendo en cuenta lo anterior, se toma la base de datos de cada prueba realizada en dichos semestres donde se pudo obtener un compendio total de treinta preguntas que se realizaron y las respuestas que se obtuvieron por parte de los aspirantes, las cuales arrojaron dos bases, en una mostraba las respuestas de selección múltiple comprendidas entre (a-e) y la otra la consolidación de las respuestas de la prueba dadas de la siguiente manera: cero (0) tomado como respuesta errada y uno (1) como respuesta acertada.

A continuación, se mostrarán dos ejemplos que hacen relación a lo mencionado anteriormente.

3.2.1 Tabla de Respuestas Elección Múltiple

P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20	P21	P22	P23	P24	P25	P26	P27	P28	P29	P30	
d	a	e	b	d	a	a	b	c	d	b	c	d	c	a	a	a	d	d	c	d	a	a	b	c	d	a	d	a	c	
d	a	e	a	d	n	a	d	c	d	b	a	b	a	a	a	b	d	d	d	b	a	d	b	c	d	a	c	n	c	
c	a	e	c	b	c	a	d	c	d	c	a	d	a	b	a	c	d	d	n	b	d	a	a	c	d	a	d	b	c	
d	a	c	d	b	c	a	d	c	c	c	b	d	a	a	b	b	d	d	c	b	d	b	c	a	c	d	d	d	c	
c	n	c	c	n	c	c	d	c	d	c	b	d	n	a	a	b	d	b	c	b	d	a	b	c	d	a	n	a	c	
b	b	d	b	a	c	c	c	a	d	c	b	b	b	a	a	d	c	c	c	b	d	c	c	n	a	b	b	b	c	
d	a	c	a	b	c	a	d	c	a	c	c	b	a	b	c	a	d	c	c	b	d	b	c	d	c	c	d	b	c	
c	a	d	b	d	c	a	d	c	a	d	b	e	b	a	a	b	d	d	c	b	a	a	d	c	d	a	d	c	c	
d	d	c	c	c	d	c	d	c	c	c	c	c	a	b	c	d	d	b	c	b	b	n	b	c	d	a	c	c	c	
n	n	n	n	n	n	n	n	n	n	n	a	n	n	n	n	n	n	n	n	n	n	n	n	n	d	n	n	n	a	n
d	a	c	n	n	d	b	n	n	d	c	a	n	a	d	n	b	d	b	n	b	a	n	n	n	n	a	d	d	c	
b	a	c	a	b	c	d	d	c	a	c	b	b	a	d	a	a	d	d	c	b	d	a	c	c	d	a	b	a	c	
b	a	c	d	c	c	b	c	d	a	c	b	d	b	d	a	a	b	d	c	b	d	c	b	d	a	a	d	c	c	
d	d	c	b	b	c	c	d	c	d	d	b	d	a	a	a	b	d	a	c	b	d	c	a	c	d	a	a	a	c	
b	a	a	b	d	d	a	d	c	b	a	b	b	a	c	b	b	a	b	d	d	a	b	d	a	d	b	c	a	c	
b	a	c	c	b	c	e	d	c	a	b	b	d	a	a	a	d	b	d	c	b	a	b	a	c	a	n	c	b	c	
c	a	c	a	b	d	a	d	c	d	c	b	d	a	b	d	b	a	d	c	d	d	c	b	b	a	a	d	d	c	
d	a	c	c	b	a	c	d	c	d	c	c	b	b	a	a	c	d	d	d	b	a	c	b	c	c	a	d	d	c	

3.2.2 Tabla Consolidación de Respuestas

P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20	P21	P22	P23	P24	P25	P26	P27	P28	P29	P30
0	1	0	0	0	0	1	0	1	1	1	1	1	0	1	1	0	1	1	1	0	0	1	1	1	1	1	1	1	1
0	1	0	1	0	0	1	1	1	1	1	0	0	1	1	1	0	1	1	0	1	0	0	1	1	1	1	0	0	1
0	1	0	0	1	1	1	1	1	1	0	0	1	1	0	1	0	1	1	0	1	1	1	0	1	1	1	1	0	1
0	1	1	0	1	1	1	1	1	0	0	0	1	1	1	0	0	1	1	1	1	1	0	0	0	0	0	1	0	1
0	0	1	0	0	1	0	1	1	1	0	0	1	0	1	1	0	1	0	1	1	1	1	1	1	1	1	0	1	1
1	0	0	0	0	1	0	0	0	1	0	0	0	0	1	1	1	0	0	1	1	1	0	0	0	0	0	0	0	1
0	1	1	1	1	1	1	1	1	0	0	1	0	1	0	0	0	1	0	1	1	1	0	0	0	0	0	1	0	1
0	1	0	0	0	1	1	1	1	0	0	0	0	0	1	1	0	1	1	1	1	0	1	0	1	1	1	1	0	1
0	0	1	0	0	0	0	1	1	0	0	1	0	1	0	0	1	1	0	1	1	0	0	1	1	1	1	0	0	1
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
0	1	1	0	0	0	0	0	0	1	0	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0	1	1	0	1
1	1	1	1	1	1	0	1	1	0	0	0	0	1	0	1	0	1	1	1	1	1	1	0	1	1	1	0	1	1
1	1	1	0	0	1	0	0	0	0	0	0	1	0	0	1	0	0	1	1	1	1	0	1	0	0	1	1	0	1
0	0	1	0	1	1	0	1	1	1	0	0	1	1	1	0	1	0	1	1	1	1	0	0	1	1	1	0	1	1

Una vez obtenidas todas las bases de datos (ubicadas en formato Excel) se desarrolló la Teoría Clásica de los Test implementando los índices que en ella se encuentra tales como: índice de facilidad, índice de homogeneidad e índice de confiabilidad nombrados a lo largo del trabajo, con estos aplicados a la prueba, se obtuvo una serie de datos que permitieron observar que la prueba tiene algunas deficiencias en ciertas preguntas aunque al sacar el promedio de cada una de los índices trabajados se observa que cumplía con valores deseados por la (TCT), visualizados a continuación con la tabla de datos que se obtuvo a partir de aplicar cada una de las fórmulas empleada en cada índice.

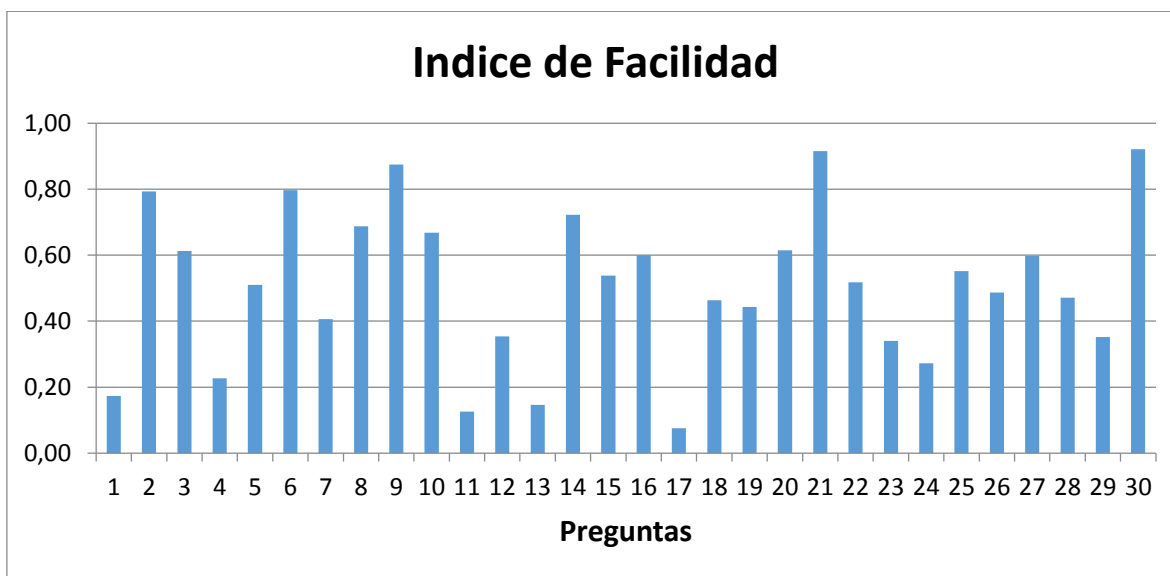
3.2.3 Consolidación de los índices 1

	P3	P8	P13	P19	P23	P28	P33	P38	P43	P48	P53	P58	P63	P68	P73	P78	P83	P88	P93	P98	P103	P108	P113	P118	P123	P128	P133	P138	P143	P148		
	b	a	c	a	b	c	a	d	c	d	b	c	d	a	a	a	d	d	d	c	b	d	a	b	c	d	a	d	a	c		
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20	P21	P22	P23	P24	P25	P26	P27	P28	P29	P30		
Facilidad	0,17	0,79	0,61	0,23	0,51	0,80	0,41	0,69	0,88	0,67	0,13	0,35	0,15	0,72	0,54	0,60	0,08	0,46	0,44	0,61	0,92	0,52	0,34	0,27	0,55	0,49	0,60	0,47	0,35	0,92	0,50887	
Homogeneidad	-0,05	0,31	0,33	0,16	0,35	0,30	0,22	0,43	0,37	0,20	0,15	0,09	0,18	0,33	0,30	0,41	0,14	0,41	0,42	0,37	0,33	0,30	0,24	0,15	0,53	0,37	0,46	0,19	0,35	0,36	0,30204	
Discriminación	-0,10	0,35	0,37	0,10	0,41	0,29	0,24	0,53	0,31	0,20	0,10	0,06	0,18	0,37	0,39	0,51	0,04	0,59	0,53	0,53	0,24	0,35	0,27	0,20	0,76	0,55	0,61	0,27	0,47	0,24	0,33197279	
Desv. Estandar	0,37	0,43	0,49	0,41	0,5	0,43	0,49	0,47	0,36	0,48	0,33	0,48	0,35	0,46	0,5	0,5	0,26	0,5	0,5	0,49	0,32	0,5	0,47	0,44	0,5	0,5	0,49	0,5	0,48	0,3	3,96185061	
Desv. Esta. Cuad	0,14	0,18	0,24	0,17	0,25	0,18	0,24	0,23	0,13	0,23	0,11	0,23	0,12	0,21	0,25	0,25	0,07	0,25	0,25	0,24	0,1	0,25	0,22	0,2	0,25	0,25	0,24	0,25	0,23	0,09	15,6962602	
Indice de fiabilidad																															0,63666885	
																																0,64667546
	0,14	0,16	0,24	0,18	0,25	0,16	0,24	0,21	0,11	0,22	0,11	0,23	0,13	0,2	0,25	0,24	0,07	0,25	0,25	0,24	0,08	0,25	0,22	0,2	0,25	0,25	0,24	0,25	0,23	0,07	0,64504777	
																																0,65528569
Deltade Ferg	0	1	1	0	0	2	2	2	6	10	4	10	13	24	19	18	16	18	15	16	14	2	1	2	1	0	0	0	0	0	0,95444012	
	0	1	1	0	0	4	4	4	36	100	16	100	169	576	361	324	256	324	225	256	196	4	1	4	1	0	0	0	0	0	0	2963

Como se puede observar en la tabla de datos anterior se encuentran todos los índices que se utilizaron para desarrollar la (TCT) y que explicaran de una manera más clara y detallada los valores obtenidos en ella. A partir de un gráfico de barras que ayudará a los datos observados en la trabajo anterior.

3.3 ANÁLISIS DE ÍTEMS O REACTIVOS

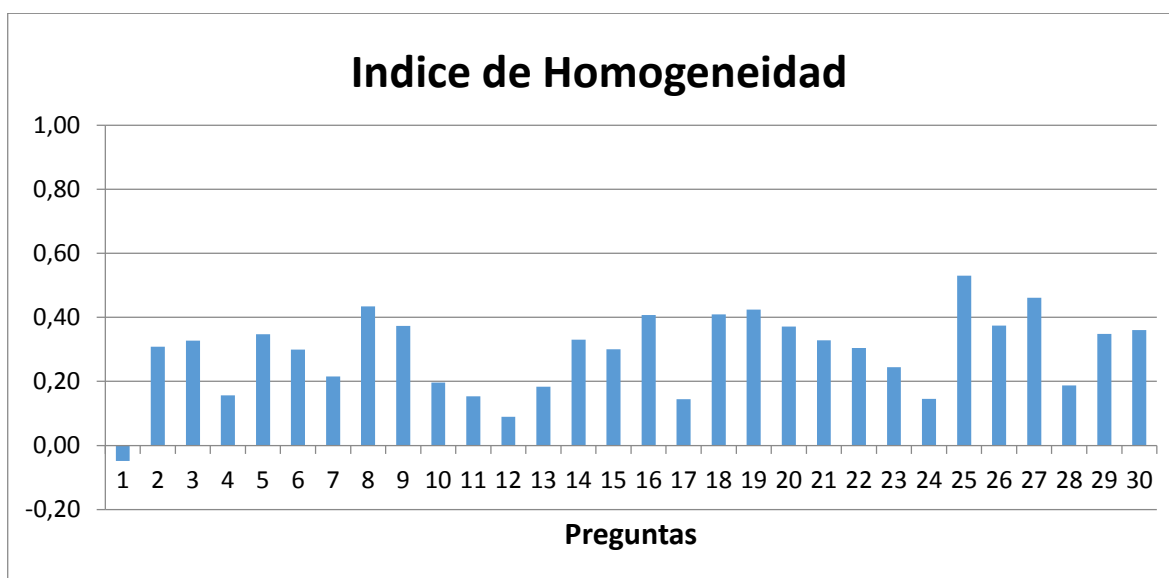
3.3.1 Grafica 1



Teniendo en cuenta estos valores y obteniendo un promedio presentado en la prueba que sería alrededor de 0.5 no sería un resultado negativo, pero sí obliga a analizar de manera detallada aquellas preguntas que presentan bajos índices de respuesta, claros ejemplos de estos son las preguntas P1, P4, P11, P13, P17, P24., como las de mayor impacto negativo obteniendo así promedio considerablemente bajo.

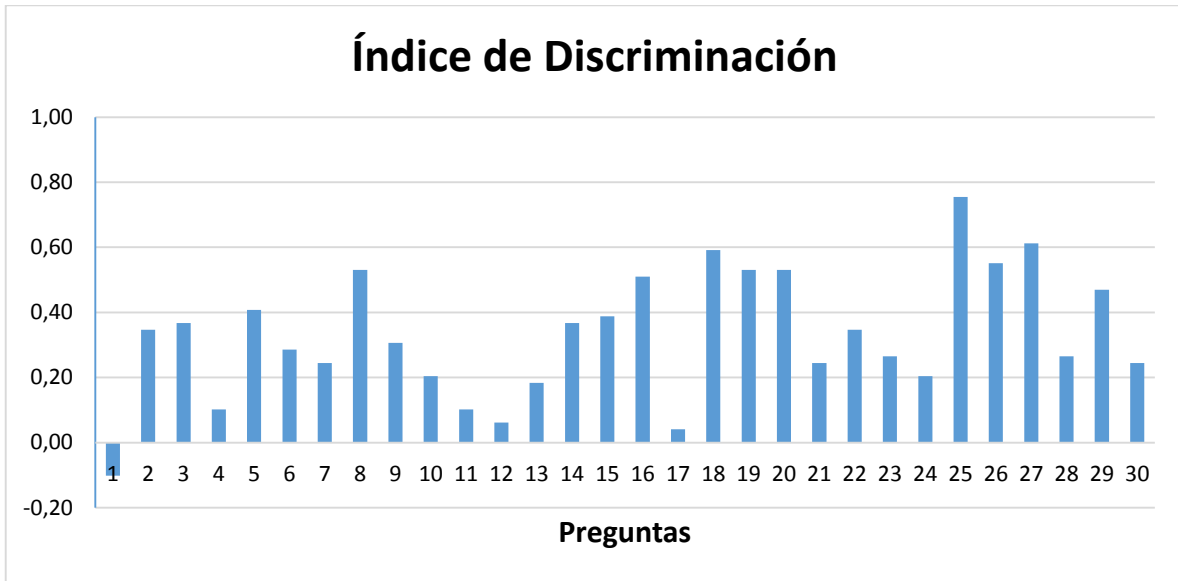
El índice de facilidad arroja unos promedios por pregunta que oscilan entre 0.13 a un máximo de 0.92 pero visto desde la Teoría Clásica de los Test, se cuenta con promedios considerablemente bajos, si se analiza de manera individual por preguntas, hay seis (6) que se encuentran por debajo del promedio establecido, el cual sería un valor deseado de 0.3 pero si se realiza un ponderado general del índice de facilidad se puede observar que se encuentra por encima del valor deseado que es de un 0.5.

3.3.2 Grafica 2



Por otro lado, el índice de homogeneidad mostró algunas deficiencias en ocho (8) preguntas en una cantidad mayor de las que se presentaba en el índice de facilidad, puesto que se encontraron rangos que oscilan entre -0,05 y un máximo de 0,53 lo cual indicó que la prueba no discriminaba bien, pues arroja una cantidad muy alta de preguntas con pocos estudiantes que lograron resolverla de manera correcta. Un ejemplo de esto es la pregunta P1 en donde se puede observar como el valor obtenido por el índice de homogeneidad es negativo lo cual nos estaría diciendo que el ítem estaría midiendo algo diferente a lo que está midiendo la prueba total, por lo tanto debería ser tomada en cuenta para poder verificar el motivo por el cual está arrojando dicho valor. Pero, tal como se presentó con el índice de facilidad al realizar un ponderado de la prueba muestra que si cumpliría con cada uno de los valores deseados que sería de 0,3, aunque no se discrimina de la manera esperada que este tipo de pruebas lo hiciera.

3.3.3 Grafica 3



Entrando a evaluar la gráfica de índice de discriminación donde se expresa índice vs preguntas, este índice en particular preocupa bastante, ya que es un ítem que se aplica para personas que tienen cierto dominio del tema, aunque se obtiene un promedio relativamente favorable en comparación al artículo sobre el cual se apoya esta investigación, pero se logra observar detalladamente, a partir de la gráfica de barras, este índice presenta demasiadas preguntas que demuestran un grado de complejidad debido al bajo índice de personas que los respondieron acertadamente lo que conlleva a pensar lo siguiente: 1. la pregunta presenta un grado de dificultad muy alto generando que el índice no logre discriminar de manera adecuada algunos resultados con márgenes bajos, ó, 2. la pregunta puede presentar otro tipo de dificultad como su estructura ocasionando índices poco favorables evidenciado particularmente en las preguntas P1, P4, P11, P13, P17, P24. Retomando particularmente lo sucedido con la pregunta P1, al ser evaluado particularmente por este índice nos dice que la totalidad de preguntas se van a segmentar en cuatro partes iguales 25% de los sujetos que obtuvieron el máximo puntaje total y el 25% de sujetos que obtuvieron los puntajes más bajos, teniendo en cuenta que las personas que respondieron de manera acertada, en dicho ítem el 25% de los sujetos que obtuvieron los índices más bajos, lo cual no corroboraría con

los puntajes obtenidos, ya que lo más normal es que este tipo de preguntas sean resueltas por los sujetos que obtuvieron mayor puntaje.

Visto desde el índice de confiabilidad se logró calcular a partir del KR-21 como lo muestra la imagen anterior de consolidación de los índices, donde nos mostró dos valores se optó por tomar el que discriminaba de mejor manera el KR-21 el cual nos dice que aunque no está arrojando los valores deseados, puesto que debería arrojar un valor aproximado o entre los parámetros establecidos entre 0.7 y 0.8 entendido como el grado de estabilidad y precisión de un instrumento de medición, aclarando que cada porcentaje mide algo en particular cada uno tiene una medición en específico, debido a que el valor de 0.7 es usado para medir a un grupo de personas y en comparación al valor obtenido por la prueba que era aplicada a un número de sujetos que presentaron la prueba con un valor obtenido de 0.645. El cuál aunque no es el valor deseado tampoco se estaría por debajo en cantidades grandes pero si llega a ser una voz de alerta para poder entrar a evaluar con precisión. Por qué se está presentando este índice tan bajo teniendo en cuenta las preguntas nombradas anteriormente.

Teniendo claro cada uno de los índices explicados anteriormente se pudo observar con el Delta de Ferguson que la prueba discrimina en su totalidad según el valor obtenido en la prueba de 0.954. Puesto que se encuentra por encima del valor esperado que es de 0.9, con lo cual muestra que aunque la prueba discrimina en su totalidad es aconsejable entrar a evaluar esas preguntas que presentan un grado de dificultad y con esto poder contar con una prueba aún con mayor discriminación.

A partir del análisis y del desarrollo de cada uno de los índices representados en las gráficas de barras surge un interrogante que tanto afectan las preguntas que presentan índices bajos y además que son concurrentes en los mismos (facilidad, homogeneidad y discriminación). Para esto se optó hacer una prueba omitiendo algunas de las preguntas nombradas anteriormente con el fin de poder observar el comportamiento que iba a tener la

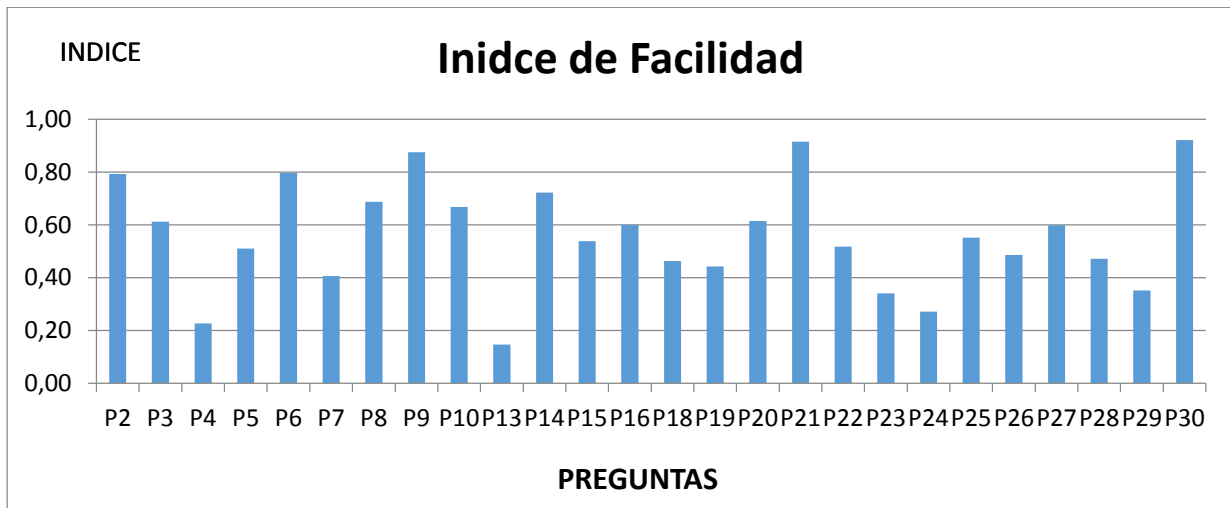
prueba y como estas preguntas afectan el instrumento de medición de los índices. Estos se desarrollan con el fin de observar como estas preguntas de una u otra manera afecta de forma directa el índice de discriminación total de la prueba vista a partir de la siguiente tabla:

3.3.4 Consolidación de Índices 2

	P8	P13	P19	P23	P28	P33	P38	P43	P48	P63	P68	P73	P78	P88	P93	P98	P103	P108	P113	P118	P123	P128	P133	P138	P143	P148	
	a	c	a	b	c	a	d	c	d	d	a	a	a	d	d	c	b	d	a	b	c	d	a	d	a	c	
	P2	P3	P4	P5	P6	P7	P8	P9	P10	P13	P14	P15	P16	P18	P19	P20	P21	P22	P23	P24	P25	P26	P27	P28	P29	P30	
facilidad	0,79	0,61	0,23	0,51	0,80	0,41	0,69	0,88	0,67	0,15	0,72	0,54	0,60	0,46	0,44	0,61	0,92	0,52	0,34	0,27	0,55	0,49	0,60	0,47	0,35	0,92	0,55912
Homogeneidad	0,30	0,35	0,16	0,38	0,31	0,21	0,43	0,38	0,20	0,20	0,34	0,30	0,41	0,42	0,41	0,37	0,35	0,31	0,24	0,14	0,53	0,38	0,47	0,21	0,35	0,38	0,32754
desv. Estand	0,43	0,49	0,41	0,5	0,43	0,49	0,47	0,36	0,48	0,35	0,46	0,5	0,5	0,5	0,5	0,49	0,32	0,5	0,47	0,44	0,5	0,5	0,49	0,5	0,48	0,3	3,910695
desv. Est. C	0,18	0,24	0,17	0,25	0,18	0,24	0,23	0,13	0,23	0,12	0,21	0,25	0,25	0,25	0,25	0,24	0,1	0,25	0,22	0,2	0,25	0,25	0,24	0,25	0,23	0,09	15,293536
Indice de fiabilidad																											0,6628031
																											0,6635921
	0,16	0,24	0,18	0,25	0,16	0,24	0,21	0,11	0,22	0,13	0,2	0,25	0,24	0,25	0,25	0,24	0,08	0,25	0,22	0,2	0,25	0,25	0,24	0,25	0,23	0,07	0,6721121
																											0,6729122

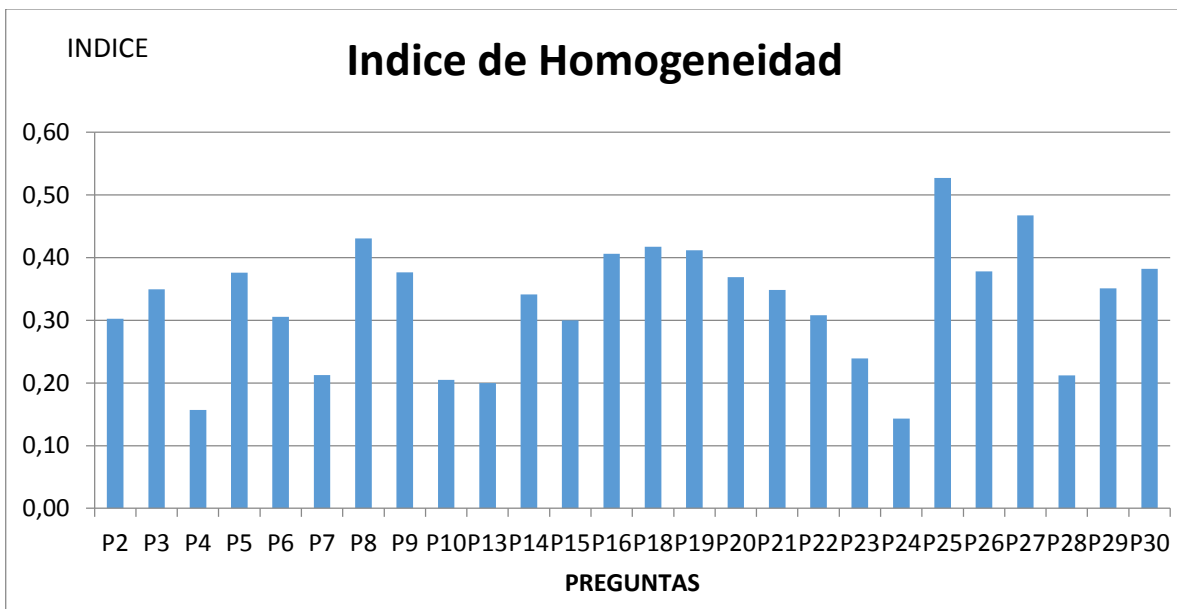
Lo que se logra observar es como con solo omitir tres preguntas el ponderado de los índices aumenta, no de una manera considerable, pero si se vislumbra una mejoría y el punto al cual se busca llegar, si al omitir genera esto con la base de datos y empieza a discriminar mejor que pasaría si las preguntas que se omitieron fueran estructuradas y aplicadas y/o tuviesen mejor discriminación en la prueba. Esto generaría que la prueba tuviese mayor fiabilidad y validez al momento de ser aplicada generando unos rangos de favorabilidad a la prueba. A partir de ello, se plantea de manera gráfica la mejora ya mencionada para los índices de la TCT

3.3.5 Grafica 4



Lo que se puede observar con anterior gráfico de barras, como omitiendo tres preguntas esta presenta una leve mejoría mostrando una menor cantidad de preguntas por debajo del valor deseado a comparación de la figura 1 mencionada anteriormente.

3.3.6 Grafica 5



Al igual que en la gráfica anterior se notó una leve mejoría en su índice, es evidente como con solo tres preguntas se puede hacer una diferencia, se evidencia una mejoría, ayudando a discriminar acertadamente aún más la prueba, a comparación con la gráfica 2

Es aún más favorable su cambio al omitir esas tres preguntas debido a que su ponderado cambia satisfactoriamente generando un aumento en el contenido total de la prueba concibiendo una validez y fiabilidad a la prueba aplicada en la Universidad Pedagógica Nacional.

Con base a los conceptos nombrados anteriormente y basados en el artículo BEMA escrito por varios autores, entre los cuales se encuentra Lin Ding, R. (2006), arroja una serie de parámetros los para cuantificar cada uno de los índices nombrados anteriormente, en relación con la prueba específica aplicada en la Universidad Pedagógica Nacional en la Licenciatura en Física.

3.4 Tabla 1. VALORES DESEADOS A PARTIR DE LOS RESULTADOS

Estadística de la prueba	Posibles valores	Valores deseados	Valores promedio obtenidos N =197 estudiantes
Índice de facilidad	[0, 1]	0.3 – 0.9	0.503
índice de homogeneidad	[-1, 1]	≥ 0.2	0.303
Índice de discriminación	[-1, 1]	≥ 0.3	0.335
Índice de confiabilidad KR 21	[0, 1]	≥ 0.7 o ≥ 0.8	0.645
Delta de Ferguson	[0, 1]	≥ 0.9	0.954

La tabla presentada anteriormente, muestra los valores obtenidos del análisis de la prueba aplicada en el Departamento de Física, que se a compararon con el artículo del BEMA, tomado como base para poner en marcha y dar una estructura al análisis de la misma, partiendo de la base o ejemplo que allí nos brinda y teniendo claro los resultados a obtener a través de la Teoría Clásica de los Test.

Tal como se mencionó con anterioridad, esta es una prueba que tiene dos parámetros claros y previamente establecidos con dos valores numéricos (0, 1), donde cero demuestra la representación de los estudiantes que no lograron responder la pregunta acertadamente y, el valor numérico 1 como la representación de los estudiantes que lograron responder la pregunta acertadamente.

Teniendo esto claro se estiman valores óptimos o deseados como es nombrado en el recuadro anterior, los cuales oscilan según el índice que se esté estudiando entre 0.2 y un máximo de 0.9, visto desde los valores obtenidos en la prueba específica realizada en el

Departamento de Física de la Universidad Pedagógica Nacional, cuyos resultados son favorables puesto que se encontraría entre los rangos o valores deseados según la (TCT), pero aun así, invita a analizar de manera detallada índice por índice para lograr un nivel de favorabilidad aún más alto, al tomar en cuenta algunas preguntas que no ayudan a discriminar del todo la prueba

A partir de la Teoría Clásica de los Test y de ser evaluados en la prueba específica de la Licenciatura en Física de la Universidad Pedagógica Nacional, se pudo observar que la prueba presenta un grado de desconcierto gracias a esas preguntas que no cumplen con los puntajes óptimos para cada uno de los índices, puesto que se encuentran por debajo de las condiciones que arroja la (TCT), alguna de las preguntas serán tomadas como base para realizar el estudio del trabajo logrando hacer una reestructuración de las mismas y poder así mirar su desempeño al momento de ser aplicada nuevamente, encontrando una respuesta satisfactoria mejorando los índices de facilidad, homogeneidad y confiabilidad.

A partir del análisis total y detallado a la prueba específica realizada en el Departamento de Física en la Universidad Pedagógica Nacional, y basado en la Teoría Clásica de los Test (TCT), se pudo evidenciar que: la prueba tiene una serie de falencias presentes en algunas preguntas usadas en la misma. Las cuales fueron empleadas en las pruebas realizadas durante los periodos de (2011-1, 2012-1 y 2012-2). Haciéndose evidentes en las preguntas P3, P53, y P83, las cuales se estudiaron para encontrar el por qué se presenta un índice tan bajo en las respuestas presentadas en la prueba del Departamento, para ello se estudió pregunta por pregunta para poder concebir la respectiva estructuración y generar que estas incógnitas tengan respuesta favorable con los aspirantes de nueva admisión.

De no ser posible reestructurar las preguntas sería necesario diseñar nuevas preguntas para poder ser tenidas en cuenta para una posible aplicación.

4. CONCLUSIONES

A partir del trabajo realizado con la prueba específica del Departamento de Física de la Universidad Pedagógica Nacional que se aplicó durante los años 2006-2014, se observó que la prueba consta un banco de preguntas de las cuales se aplicaban entre 20 y 30 escogidas de manera aleatoria durante los semestres comprendidos entre los años 2006-2 al 2014-1. Se encontró que entre los semestres del 2011-1 al 2012-2 se reiteraron las mismas 30 preguntas y que fueron aplicadas a un total de 197 sujetos y son estas preguntas las que se analizaron.

Bajo la Teoría Clásica de los Test y según los criterios que presenta Lin Ding, R. et.al. (2006), la prueba que se realizó entre el 2011-1 y el 2012-2 cumple con los índices de facilidad, homogeneidad, discriminación y discriminación total (Delta de Ferguson) como se aprecia en la tabla 1. Sin embargo, no se cumple con el índice de fiabilidad, aunque estuvo su valor de 0.645, cerca del valor deseado de ≥ 0.7 . Esto puede deberse a que la Prueba Específica de la Licenciatura en Física de la Universidad Pedagógica Nacional no mide necesariamente un solo rasgo psicológico sino varios: razonamiento matemático, comprensión lectora y conocimientos básicos de física. En ese sentido sería necesario realizar el análisis dividiendo el test en estas tres partes.

La primera pregunta (P3 del banco original de preguntas) arrojó valores negativos en los índices de homogeneidad y discriminación lo cual no es un resultado deseado porque quiere decir que está midiendo algo muy diferente a lo deseado. Esto enfatiza lo mencionado en el párrafo anterior frente a diferentes rasgos psicológicos que se están midiendo. Además, se encontró que otras cinco (5) preguntas presentan simultáneamente índices de facilidad, homogeneidad y discriminación, fuera del rango deseado. Son las preguntas 4, 11, 13, 17 y 24, que corresponden al banco original de preguntas P19, P53, P63, P83 y P118 respectivamente, que hicieron que, aunque se lograron valores aceptados para los promedios de los índices de homogeneidad y discriminación, estos estuvieron muy cerca del límite

permitido. Principalmente se debe revisar la primera pregunta, P3 del banco de preguntas original, ya que sus índices de homogeneidad y discriminación arrojaron valores negativos.

Hacer un estudio a la Prueba Específica del Departamento de Física de la Universidad Pedagógica Nacional, con el fin de saber cómo se encontraba constituida de qué forma estaba compuesta y cómo era su margen de calificación, conociendo esto se empezó hacer el estudio a la prueba basado en una teoría específicamente en la Teoría Clásica de los Test, haciendo un estudio a la misma interpretando y analizando cada dato que arrojaba la prueba y evaluándola a partir de la Teoría Clásica de los Test, con el fin de hacer un aporte a la prueba con el ánimo de contribuir para tener una prueba que tenga mejores estándares de “calidad”.

Se quiere dar un aporte a la Prueba Específica del Departamento de Física de la Universidad Pedagógica Nacional a partir del trabajo de grado realizado. Se deja preparada una Hoja de Cálculo con la cual es posible continuar este estudio que aporte elementos para hacer los ajustes necesarios a este análisis y a la prueba en general, y que además permita realizar una caracterización de los estudiantes que están ingresando a la Licenciatura en Física.

A futuro también sería deseable realizar un análisis bajo la Teoría de Respuesta al Ítem TRI teniendo en cuenta que su fundamentación matemática es mucho más compleja y requiere software especializado para realizar este tipo de trabajo, pero que puede aportar nuevos elementos de análisis ya que es factible realizar comparaciones aun cuando diferentes grupos respondan diferentes preguntas.

5. REFERENTES BIBLIOGRÁFICOS

- Abad, F. et.al. (2006). *Introducción a la Psicometría Teoría Clásica de los Tests y Teoría de la Respuesta al Ítem. Facultad de Psicología Universidad Autónoma De Colombia.* Recuperado de http://aprendeenlinea.udea.edu.co/lms/investigacion/file.php/39/ARCHIVOS_2010/PDF/IntPsicometria_aristidesvara_1_.pdf
- Barbero García, I. Prieto Marañón, P. Suárez Falcón, J y Concepción San Luis Costas. (2001). Relaciones empíricas entre los estadísticos de la teoría clásica de los tests y los de la teoría de respuesta a los ítems. Vol. 13, nº 2, pp. 324-329. Recuperado de. <http://www.psicothema.com/psicothema.asp?id=455>
- Ding, L; Chabay, R; Sherwood, B &, Beichner, R. (2006). Evaluating an electricity and magnetism assessment tool: Brief electricity and magnetism assessment. Physical review special topics - physics education research **2**, 010105. Recuperado de <https://pdfs.semanticscholar.org/6be5/349b6b209bd53945a7cb87cb374fba768578.pdf>
- García Cueto, E. (1996). Aplicación de modelos psicométricos para test y medidas alternativas en psicometría. Revista psicothema. Vol. 8, nº 2. Recuperado de: <http://www.redalyc.org/pdf/727/72780205.pdf>
- García, L. (2012). Construcción de un test para medir los conocimientos y aptitudes en matemáticas de estudiantes de primer año de universidad. En *Cuadernos de la*

Maestría en Docencia e Investigación Universitaria. Tomo II. Artículos de grado 2012. Carmen Cecilia Suárez (Compiladora). Universidad Sergio Arboleda

- Martínez, R. (1995). *Psicometría: Teoría de los Tests Psicológicos y Educativos*. Universidad Complutense de Madrid. Editorial Síntesis S.A. España.
- Matas, A. (2011). Introducción al análisis de la teoría de respuesta al ítem. *Aidesoc*. Recuperado de: https://riuma.uma.es/xmlui/bitstream/handle/10630/4711/TRI_aidesoc_2011.pdf
- Montero, E. (2000). La teoría de respuesta a los ítems: una moderna alternativa para el análisis psicométrico de instrumentos de medición. *Revista de Matemática*: 217–228. Recuperado de: <https://revistas.ucr.ac.cr/index.php/matematica/article/viewFile/191/171>
- Muñiz, J (2010). Las teorías de los test: teoría clásica y teoría de respuesta a los ítems. *Papeles del psicólogo*. Vol. 31(1), pp. 57-66. Recuperado de <http://www.papelesdelpsicologo.es/pdf/1796.pdf>
- Navas, M. (1994). *Teoría clásica de los test versus teoría de respuesta al ítem*. U.N.E.D. Madrid. Pag: 175-180. Recuperado de: <http://www2.uned.es/490015/CV/TCTTRI94.pdf>
- Reyes, F. *Importancia de la Psicometría en la Psicología*. Universidad Santo Tomás Psicología. recuperado de: <https://es.scribd.com/doc/58386407/Importancia-de-la-psicometria-en-psicologia>