

**ANÁLISIS PSICOMÉTRICO DE UNA PRUEBA DE FÍSICA PARA GRADO
DECIMO DEL COLEGIO NICOLÁS GÓMEZ DÁVILA - IED**

**ELIANA ANGÉLICA GALEANO VANEGAS
LILIANA OTERO JIMÉNEZ**

**FUNDACIÓN UNIVERSITARIA LOS LIBERTADORES
DEPARTAMENTO DE CIENCIAS BÁSICAS
ESPECIALIZACIÓN EN ESTADÍSTICA APLICADA
BOGOTÁ D.C.
2016**

**ANÁLISIS PSICOMÉTRICO DE UNA PRUEBA DE FÍSICA PARA GRADO
DECIMO DEL COLEGIO NICOLÁS GÓMEZ DÁVILA - IED**

**ELIANA ANGÉLICA GALEANO VANEGAS
LILIANA OTERO JIMENEZ**

**Trabajo de Tesis, para optar el título de
Especialista en Estadística Aplicada**

**Asesor
WILMER DARIO PINEDA RIOS**

**FUNDACIÓN UNIVERSITARIA LOS LIBERTADORES
DEPARTAMENTO DE CIENCIAS BÁSICAS
ESPECIALIZACIÓN EN ESTADÍSTICA APLICADA
BOGOTÁ D.C.
2016**

Nota de aceptación

Presidente del Jurado

Jurado

Jurado

Bogotá _____ de (mes) _____ de _____

AGRADECIMIENTOS

Manifestamos nuestros más profundos agradecimientos a todas las personas que contribuyeron y colaboraron durante el desarrollo del presente trabajo de investigación; agradecemos su conocimiento, orientación, paciencia, dedicación, tiempo y apoyo moral, ya que sin ellos, los resultados de esta investigación no hubieran sido posibles.

A nuestros padres, por darnos la vida, el amor, la educación, los valores, las buenas costumbres y sobre todo por el apoyo constante que nos han ofrecido en cada etapa de nuestras vidas.

A nuestros hermanos, por ser nuestros confidentes, consejeros y ejemplo a seguir.

A nuestros docentes de la especialización en estadística aplicada, por transmitirnos sus valiosos conocimientos, por la orientación, por la buena disposición, por el tiempo que nos dedicaron al momento de atender todas nuestras inquietudes; como también por las correcciones y sugerencias a nuestra investigación.

CONTENIDO

	pág.
GLOSARIO	9
1. INTRODUCCIÓN	11
1.1 PLANTEAMIENTO DEL PROBLEMA	12
2. MARCO DE REFERENCIA	13
2.1 TEORÍA CLÁSICA DE LOS TEST	14
2.1.1 Alfa de Cronbach	15
2.1.2 Índice de dificultad	16
2.1.3 Índice de discriminación	16
2.1.4 Coeficiente de correlación biserial puntual	18
2.2 TEORIA DE RESPUESTA AL ITEM	18
2.2.1 La Curva Característica del Ítem	20
2.2.2 Dificultad	22
2.2.3 Modelo de Rasch	22
2.2.4 Ajuste cercano y lejano (infit y outfit)	23
3. MARCO METODOLOGICO	24
3.1 TIPO DE ESTUDIO	24
3.2 VARIABLES DE ESTUDIO	24
3.3 PARTICIPANTES	24
3.4 INSTRUMENTO DE RECOLECCIÓN	24
3.5 ESTRUCTURA DEL TEST	25
3.6 DISEÑO ESTADÍSTICO	25
4. RESULTADOS	26
4.1 TEORIA CLÁSICA DE LOS TEST	26
4.1.1 Alfa de Cronbach	26
4.1.2 Índice de dificultad	26
4.1.3 Índice de Discriminación	27
4.1.4 Coeficiente de correlación biserial puntual (ítem – test)	28
4.2 TEORIA DE RESPUESTA AL ITEM	30
4.2.1 Curva característica del Ítem (CCI)	31
CONCLUSIONES	34
REFERENCIAS BIBLIOGRAFICAS	35
ANEXOS	36

LISTA DE TABLAS

	pág.
Tabla 1. Interpretación del Alfa de Cronbach	16
Tabla 2. Interpretación del índice de dificultad	16
Tabla 3. Estructura del test	25
Tabla 4. Resultado Alfa de Cronbach	26
Tabla 5. Estadísticos de los elementos	27
Tabla 6. Resultados e interpretación del índice de dificultad	27
Tabla 7. Índice de discriminación	28
Tabla 8. Resultado ítems con índice de Correlación $\geq 0,25$	28
Tabla 9. Estadísticos total-elemento	29
Tabla 10. Resultados generales de la TRI Sujetos	30
Tabla 11. Resultados generales de la TRI Ítems	30
Tabla 12. Índice de dificultad y discriminación. Infit y Outfit de cada Item	33

LISTA DE GRÁFICAS

	pág.
Gráfica 1. Curva característica del ítem	20
Gráfica 2. Parámetros de la CCI	21
Gráfica 3. CCI de los ítems con mayor habilidad	31
Gráfica 4. CCI de los ítems con menor habilidad	32

LISTA DE ECUACIONES

	pág.
Ecuación 1. Coeficiente Alfa de Cronbach	15
Ecuación 2. Índice de dificultad	16
Ecuación 3. Índice de discriminación	17
Ecuación 4. Coeficiente de correlación biseral puntual	18
Ecuación 5. Función de probabilidad	20

GLOSARIO

CONFIABILIDAD: se refiere a la consistencia de los resultados; en lo referente a un instrumento de medición, hace énfasis al grado de coincidencia que se evidencia en sus resultados, teniendo en cuenta que su aplicación se destina de forma repetitiva a sujetos que se encuentran en condiciones equitativas. Por lo tanto, se podría hablar de resultados consistentes y coherentes, y por ende de la confiabilidad de resultados.

DIFICULTAD: se relaciona con la complejidad que demanda un determinado ítem al sujeto para responder de forma acertada.

DISCRIMINACIÓN: las respuestas obtenidas de un ítem, varían en razón al grado de habilidad. Señala en qué grado se responde el ítem acertadamente por sujetos de superior habilidad y erróneamente por sujetos de baja habilidad.

HABILIDAD: consiste en la destreza, talento, pericia y aptitudes innatas que un determinado sujeto posee, y por lo tanto le permite un mayor éxito en la obtención de óptimos resultados en el desarrollo de alguna actividad o tarea. Encontramos diferentes tipos de habilidades como: las sociales, las cognitivas y las específicas.

ÍTEMS: el término ítem, se usa para aludir a una pregunta o interrogación que hace parte de una prueba subjetiva. Los ítems se clasifican en: ítem de respuesta abierta e ítem de respuesta cerrada, esta clasificación obedece al tipo de respuesta arrojada por un sujeto, concretamente por un estudiante.

PSICOMETRÍA: se ocupa de la medición de los procesos psíquicos, utilizando la estadística como pilar básico para la elaboración y aplicación de teorías y para el desarrollo de métodos y técnicas específicas de medición.

TCT: teoría clásica de los test, es la teoría más utilizada dentro del campo de la evaluación psicológica y educativa; es un modelo que aborda el problema del error de las medidas realizadas mediante la aplicación de un test; determina el grado de dificultad de un ítem y establece si el ítem discrimina de forma correcta. Esta teoría afirma que la precisión de la medida no depende de las destrezas que un individuo demuestre.

TRI: teoría de respuesta al ítem, es un modelo estadístico que relaciona la dependencia de las puntuaciones obtenidas en un ítem con el grado de la habilidad o rasgo medio que posee un sujeto, al igual que la relación existente entre el ítem con la prueba.

VALIDEZ: concepto estadístico que hace referencia a la capacidad de un instrumento de medición para medir y cuantificar la variable que el investigador pretende medir, de tal manera que el instrumento cumpla la función para el cual fue diseñado.

RESUMEN

Se muestran los resultados de la investigación para medir la confiabilidad de una prueba de física previamente diseñada y aplicada a los estudiantes del ciclo V del colegio Nicolás Gómez Dávila IED; el análisis de este trabajo se hizo a través de la perspectiva de la teoría clásica de los test y se complementa con la teoría de respuesta al ítem. Lo anterior, como fundamento práctico que permita a la institución educativa utilizar estos análisis y resultados como herramienta de apoyo en el momento de decidir cómo fortalecer los elementos de evaluación del estudiantado.

La prueba se aplicó a 60 estudiantes del ciclo V, los cuales correspondían al 100% del público objetivo al que se podía aplicar dicho test; los ítems utilizados fueron tomados de un banco de preguntas. Los resultados de las teorías mencionadas, muestran que el test es confiable y que para mejorar el índice de confiabilidad es necesario eliminar dos ítems, por lo tanto el 87% de los ítems evaluados cumplen con un nivel alto de confiabilidad y se recomienda que la prueba pueda seguir siendo utilizada.

Palabras Clave: confiabilidad, TRI, TCT, psicometría.

ABSTRACT

This document shows the results of an investigation to measure the reliability of one physics test, this was designed previously and it was applied to students of cycle V of Nicolas Gómez Dávila IED School. The analysis of this work was done through the perspective of classical test theory and complemented with the theory of item response. Thus, the school can use these analyzes and results as a support tool in deciding how to strengthen the elements of students assessment, based on the previous practical basis.

The test was applied to 60 students of the V cycle, which corresponds to 100 % of the target audience; the items used were taken from a bank of questions. The results of these theories show that the test is reliable. However to improve the reliability index is necessary to remove two items, thus 87% of the evaluated items meet a high level of reliability and it is recommended that the test it may still be used.

Key words: reliability, TRI, TCT, psychometry

1. INTRODUCCIÓN

La preocupación por la calidad de la educación en Bogotá, ha sido uno de los pilares fundamentales en las políticas distritales y nacionales, lo cual se evidencia en el mejoramiento de la planta física y en el fortalecimiento del recurso humano. Dentro del plan de desarrollo Bogotá Humana (2012-2016) del gobierno distrital, en su artículo 9, *la construcción de saberes: educación incluyente, diversa y de calidad para disfrutar y aprender*; se busca reducir las brechas de calidad de la educación entre los niños, niñas y adolescentes del distrito capital; en este plan de desarrollo, en el numeral 4, se señala adicionalmente dentro del mejoramiento de la calidad educativa, el fortalecimiento del sistema de evaluación distrital, teniendo en cuenta, que son estos procesos de evaluación inherentes al proceso educativo en todos sus escenarios y los mismos son importantes para comprender los impactos de la enseñanza en cada actor del escenario educativo. (Secretaría de Educación de Bogotá, 2013)

Dentro de los procesos evaluativos, es importante brindar elementos para el análisis sobre los efectos de los procesos de aprendizaje en los estudiantes, por lo tanto es necesario que las evaluaciones aplicadas a los mismos sean confiables, permitiendo aportes y conclusiones fidedignos, que de igual manera servirán como referentes al momento de tomar decisiones relacionadas con los esquemas de enseñanza del distrito. Existe un amplio consenso en todos los países de la región en que la evaluación sistemática y correctamente orientada constituye una de las estrategias principales para la mejora de la calidad de la enseñanza, y es por eso que no hay que perder el foco en la importancia al momento de abordar una evaluación educativa y la utilización de los resultados que se obtienen al aplicar la misma. (Marchesi)

A veces se piensa que se ha evaluado el aprendizaje, cuando solamente se ha medido; evaluar exige comparar el resultado de la medición con un punto de referencia que establezca el resultado que se debería obtener, para de esta manera, llegar a un juicio sobre lo adecuado o inadecuado de lo medido. Un test es confiable cuando permite hacer mediciones precisas sobre lo que se está evaluando; las mediciones aplicadas a una prueba y específicamente la medición de la confiabilidad permite generar elementos esenciales para producir resultados que describan, con un alto nivel de precisión, las competencias de las personas evaluadas, en este caso, los estudiantes de ciclo V del colegio Nicolás Gómez Dávila IED.

Por lo anterior, se hace necesario determinar la confiabilidad de un test de Física que se aplicó a los estudiantes, del colegio en mención, y para ello, se tendrán en cuenta los fundamentos de la Teoría Clásica (TCT) y la Teoría de Respuesta al Ítem (TRI).

1.1 PLANTEAMIENTO DEL PROBLEMA

La evaluación es una constante en la vida de las personas, en las instituciones educativas, en las trayectorias profesionales y otros ámbitos de la vida. Dentro del ámbito escolar, la evaluación se planea y realiza dentro de las mismas instituciones, en coherencia con lo que se **preveè** alcanzar en cada una de las áreas y ciclos de escolaridad; tradicionalmente, la evaluación se ha centrado en contenidos y en obtener cuanta información se logra almacenar sobre determinada área, sin embargo, hoy además de los contenidos, la evaluación interna debe tener en cuenta el desarrollo de pensamiento, de actitudes y de valores, así como los niveles de aprendizaje del estudiante, donde los contenidos no son el fin sino el medio para lograrlo. (Secretaría de Educación de Bogotá, 2014)

Dentro del esquema de enseñanza en Colombia, las evaluaciones tienen un nivel alto de importancia, toda vez que son elementos decisivos en el momento de permitir la continuidad de un alumno dentro del proceso educativo; adicional a esto, generalmente es el profesor el que realiza y aplica las pruebas inherentes a la materia de la cual son responsables, sin embargo, dichas evaluaciones nunca son sometidas a pruebas de confiabilidad, para garantizar que las mismas aporten mejoras al proceso de enseñanza en la institución y que el foco de estas permita el cumplimiento del objetivo final que es validar las habilidades y competencias adquiridas por los alumnos en el desarrollo de una materia.

El hecho de que haya una evaluación no resuelve el problema del mejoramiento de la calidad de la enseñanza, pero sí permite ayudar a mejorarlo, ya que es una ventana para conocer de manera práctica que aspectos de los esquemas de aprendizaje pueden ser mejorados; la política educativa en Bogotá busca orientar a la comunidad educativa en el análisis, interpretación y comprensión de los resultados de las diferentes evaluaciones y pruebas que se aplican al estudiantado; promoviendo estrategias en pro del fortalecimiento de la cultura educativa y construyendo un esquema educativo de calidad que se enfoque igualmente en el diseño y aplicación de pruebas que permitan adicionalmente medir las transformaciones logradas en la población estudiantil y de esta manera generar mejoras en los procesos educativos a que haya lugar. (Secretaría de Educación Bogotá, 2014)

Es por las razones antes expuestas, que este proyecto realiza un análisis de confiabilidad a una prueba de Física realizada al ciclo V del colegio Nicolás Gómez Dávila IED, aplicando a dicho test el estudio psicométrico de las teorías TCT y TRI, que permitirán concluir si los ítems de la evaluación son coherentes entre sí y guardan relación directa con el objetivo de aplicación de la misma; es de aclarar, que los ítems incluidos en dicha prueba, fueron tomados de un banco de preguntas asociados con el tema en particular y con la materia en mención.

A partir de los resultados de este trabajo, esperamos que este test se pueda seguir aplicando dentro de la institución, y en los grados a los que corresponde la temática de la misma.

Por lo anterior y para efectos del presente estudio, se plantea la siguiente pregunta:

¿Cuál es el grado de confiabilidad desde la TCT y la TRI de una prueba de Física aplicada al ciclo V del colegio Nicolás Gómez Dávila IED de Bogotá?

2. MARCO DE REFERENCIA

La psicometría tiene su origen en la aplicación y medición de test psicológicos y de inteligencia; Sir Francis Galton (1822-1911) fue el pionero de la nueva psicología experimental en la Gran Bretaña del siglo XIX, su carrera intelectual parece haberse visto dominada por la creencia de que podía medirse casi cualquier cosa, y debido a sus esfuerzos en el diseño de medidas como parte de estudio del individuo, es que en general los historiadores de las pruebas psicológicas lo consideran como el fundador de la psicometría. (Barrero, Melissa Judith Ortiz, 2014).

El rápido progreso económico y social en la Europa de finales del siglo XIX planteó la necesidad de evaluar las capacidades y conocimientos de los individuos en contextos educativos, laborales, etc. Las necesidades de la evaluación educativa fueron las primeras demandas sociales con un impacto significativo y duradero en la evolución del estudio de las diferencias individuales y la inteligencia, por ende, de la Psicometría. (Universidad Nacional Abierta y a Distancia, 2010)

Las evaluaciones y/o test se utilizan en casi todos los países con propósitos de orientación psicológica, selección, asignación, comprensión de saberes, definición de objetivos, etc., y los contextos de aplicación son diversos como por ejemplo en colegios, universidades, centros de orientación psicológica, industrias, clínicas y hospitales, entre otros. Prácticamente desde que se ingresa al sistema escolar, se comienzan a aplicar evaluaciones que buscan medir y concluir algo.

La teoría de los test, está interesada en el desarrollo de modelos matemáticos que son útiles para el análisis de los datos proporcionados por las respuestas de los sujetos en determinados test; se busca identificar qué factores influyen en las puntuaciones observadas de un test y ayudan a describir supuestos sobre dichos resultados. Los ítems que se formulan de manera lógica en una evaluación, buscan medir la variable, constructo o rasgo que interesa evaluar dentro del cuestionario, y se busca finalmente medir en los ítems tres indicadores: (Francisco J. Abad, 2006):

- Índice de dificultad
- Índice de homogeneidad o discriminación
- Índice de validez

A la construcción y análisis de los test subyacen teorías que guían su construcción, que condicionan y tiñen los test según los avances teóricos y estadísticos de cada momento. Las teorías estadísticas de los test, permiten la estimación de propiedades psicométricas de los test para garantizar que las decisiones tomadas a partir de ellos sean las adecuadas. Hay dos grandes enfoques o teorías a la hora de construir y analizar los test, son la Teoría Clásica de los Test (TCT) y el enfoque de la Teoría de Respuesta a los Ítems (TRI). (Muñiz, Las teorías de los Test: Teoría Clásica y Teoría de Respuesta a los Ítems, 2010)

2.1 TEORÍA CLÁSICA DE LOS TEST (TCT).

A principios del siglo XX, Spearman propone una formulación matemática para estudiar las propiedades métricas de las puntuaciones que se asignan mediante test elaborados en ese tiempo, para cuantificar el nivel de las personas en funciones psicológicas superiores. La principal preocupación de Spearman fue incorporar en la formulación matemática los errores de medida que se cometen en la aplicación de los test psicológicos; este es el inicio de la Teoría Clásica de los Test (TCT). (Francisco J. Abad, 2006)

La TCT ha sido el modelo dominante en la teoría de test, y tiene aún una vigencia representativa en el campo de la evaluación psicológica y educativa. Esta teoría usa un modelo matemático sustentado en la curva normal que supone que la habilidad de un sujeto es la sumatoria de los puntajes obtenidos al responder una serie de ítems de una prueba. La TCT se centra en la estimación del puntaje de una persona como si esta hubiera respondido al universo total de preguntas posibles, como este universo es infinito, es necesario hacer una estimación de este puntaje, el cual tendrá cierta cantidad de error.

Es importante asegurar que los instrumentos de medida, midan con precisión y con poco error; cuando se aplica un test o evaluación a una o varias personas, lo que se obtiene son puntuaciones empíricas, pero estas puntuaciones no dicen nada sobre el grado de precisión de las mismas, por lo tanto, se hace necesario plantear algunos modelos que subyacen a las puntuaciones para ayudar a entender el grado de precisión de las mismas.

Spearman propone un modelo lineal muy simple, para las puntuaciones de las personas en los test; consiste en asumir que la puntuación que una persona obtiene en un test, denominada 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), que puede ser debido a muchas causas que se escapan a la comprensión inicial y que no se controlan. Lo dicho puede expresarse formalmente así: $X = V + e$. (Muñiz, Las teorías de los Test: Teoría Clásica y Teoría de Respuesta a los Ítems, 2010)

Con lo anterior, se pueden definir los siguientes supuestos en los que esta soportada la TCT:

- $V = E(X)$: el puntaje verdadero es igual a la esperanza matemática o valor esperado de las puntuaciones observadas.
- $r_{V, e} = 0$: la correlación entre el puntaje verdadero en un test y el error en ese test es igual a cero. Es decir, no existe relación entre ambos.
- $r_{e_j, e_k} = 0$: la correlación entre los errores dados en dos tests diferentes es igual a 0; es decir, los errores son independientes.

Los supuestos de la TCT pueden llevar a un gran número de conclusiones. Si los supuestos son razonables, las conclusiones serán ciertas; si no son razonables las conclusiones y sus aplicaciones pueden no reflejar la realidad de forma adecuada.

2.1.1 Alfa de Cronbach (Confiabilidad). Es un coeficiente de correlación entre dos grupos de puntajes e indica el grado en que los individuos mantienen sus posiciones dentro de un grupo. Abarca valores desde 0 hasta 1, cuanto más se acerque el coeficiente hasta 1, más confiable será la prueba. El coeficiente señala la cuantía en que las medidas del test están libres de errores casuales o no sistemáticos. (Aliaga Tovar)

La fórmula para este coeficiente es:

Ecuación 1. Coeficiente Alfa de Cronbach

$$\alpha = \frac{k}{(k - 1)} \left(1 - \frac{\sum_{j=1}^k \sigma_j^2}{\sigma_x^2} \right)$$

Donde

k : Número de ítems

σ_j^2 : Suma de las varianzas de los k ítems

σ_x^2 : Varianza de las puntuaciones en el test

Este índice depende del grado de covariación de los ítems; tendrá un valor alto cuando los ítems covaríen fuertemente entre si y asumirá valores cercanos a cero si los ítems son linealmente independientes.

Por lo tanto, dependiendo del valor de α , se determina la confiabilidad de la prueba:

Tabla 1. Interpretación del Alfa de Cronbach

VALOR	INTERPRETACIÓN
0,00 – 0,29	Baja o Nula (no es confiable)
0,30 – 0,59	Confiabilidad media
0,60 - 1	Confiabilidad alta

2.1.2 Índice de dificultad. El análisis del ítem trata de medir su nivel de facilidad o dificultad, por lo tanto, según la TCT, la dificultad de un ítem se define como la proporción de personas que responden correctamente un ítem de una prueba, entre mayor sea esta proporción menor será su dificultad.

Para calcular la dificultad de un ítem, se divide el número de personas que contestan correctamente el ítem entre el número total de personas que intentan resolver el ítem (correcta o incorrectamente)

Ecuación 2. Índice de dificultad

$$ID = \frac{A_i}{N_i}$$

Dónde:

A_i = Número de personas que aciertan el ítem

N_i = Número total de personas que intentaron resolver el ítem

A medida que ID aumenta, se puede afirmar que el ítem fue respondido por una mayor cantidad de personas; por lo tanto, se puede afirmar que por medio del índice de dificultad se determina y se clasifica un ítem en las categorías de “fácil o difícil”.

Tabla 2. Interpretación del índice de dificultad

GRADO DE DIFICULTAD DEL ÍTEM	INTERPRETACIÓN
0.0 – 0.15	Se considera muy difícil
0.15 – 0.4	Se considera difícil
0.4 – 0.6	Se considera moderado
0.6 – 0.85	Se considera fácil
0.85 – 1.0	Se considera muy fácil

2.1.3 Índice de discriminación. La discriminación es el grado en el cual las respuestas de un ítem varían con respecto al nivel de habilidad. Llamado también índice de homogeneidad, se define como la correlación de Pearson entre las puntuaciones de los N sujetos en el ítem y las puntuaciones en el total del test; informa el grado en que el ítem está midiendo lo mismo que la prueba globalmente, es decir, el grado en que contribuye a la consistencia interna del test.

Este índice P es la diferencia entre la proporción de sujetos competentes que aciertan el ítem (P_c) y la proporción de sujetos incompetentes que también lo aciertan. Entendiendo por competentes aquellos que se puntúan en el test por encima de la mediana, e incompetentes por debajo.

Para hallarlo se hace lo siguiente: se debe tener en cuenta que el grupo de individuos a los cuales se les aplico el test "N", se dividen en dos grupos (con la misma cantidad de individuos " $M=N/2$ "), los del nivel superior (los que obtuvieron mayores puntajes) y los del nivel inferior (los que obtuvieron los menores puntajes).

Morales (López, P. Op. Cit., p. 31) define el índice de discriminación como la diferencia entre dos proporciones; la proporción de aciertos en el grupo superior menos la proporción de aciertos en el grupo inferior. Por consiguiente la expresión matemática de este índice de discriminación es:

Ecuación 3. Índice de discriminación

$$P = \frac{A_c - A_i}{M}$$

Dónde:

A_c : la frecuencia de aciertos en el grupo superior (convenientes).

A_i : la frecuencia de aciertos en el grupo inferior (inconvenientes).

M : el total de individuos en cada grupo.

Ahora bien este índice determina que tanto el ítem contribuye (establece diferencias) a situar un individuo en el nivel superior o inferior.

Conforme con los datos que se emplean en la fórmula, se establece que el índice de discriminación oscila entre 1 y -1. La capacidad discriminativa del ítem aumenta a medida que P se aleja de cero, bien sea hacia 1 o -1. En el caso extremo que fuese 1, indica que todos los competentes aciertan el ítem y todos los incompetentes lo falla, la discriminación sería perfecta. En el caso de -1, $P_c = 0$ y $P_i = 1$, sería el caso paradójico en el que todos los incompetentes lo aciertan y todos los competentes lo falla, indica que el ítem también discrimina perfectamente, pero habría que tener cuidado a la hora de la interpretación.

Sin embargo, el valor mínimo que puede asumir este índice es de 0.30. Cuando el valor es inferior se considera que el ítem no discrimina adecuadamente, es decir que el ítem no diferencia entre los individuos del nivel superior y del nivel inferior. (Cabrera, 2014, pág. 5, 6)

Se concluye que la utilidad del índice de discriminación, radica en determinar la condición o posición de los sujetos que responden el test, de acuerdo a su

ubicación en relación al nivel superior o nivel inferior, niveles que se traducen en puntuaciones altas y puntuaciones bajas. Este índice indica que tan efectivo es el ítem al momento de discriminar.

2.1.4 Coeficiente de correlación biserial puntual (ítem – test). El Coeficiente de Correlación Biserial permite estimar el grado de asociación entre dos variables lineales, entre la respuesta correcta o respuesta incorrecta. Se utiliza cuando se quiere conocer la correlación existente entre dos variables cuantitativas, cuando una de ellas ha sido dicotomizada previamente; se usa sobre todo para hacer inferencias.

El coeficiente de correlación biserial puntual, oscila como la r de Pearson entre ± 1 ; los valores positivos indican que la respuesta correcta al ítem está asociada a altos puntajes en la prueba; valores negativos indican que dicha asociación se da de manera inversa. Se consideran aceptables los ítems con valores superiores a 0.25. (Reyes y Barbosa, 2007, pág. 21)

La fórmula de la correlación biserial puntual está dada por:

Ecuación 4. Coeficiente de correlación biserial puntual

$$r_{bp} = \frac{x_p - \bar{x}}{s_x} \frac{p}{q}$$

Dónde:

p : La proporción de individuos que acertaron.

q : La proporción de individuos que fallaron.

x_p : La media en X de los sujetos cuya proporción es p .

\bar{x} : La media del test.

s_x : La desviación típica del test.

2.2 TEORÍA DE RESPUESTA AL ÍTEM (TRI)

El origen de esta teoría, se puede rastrear en los trabajos pioneros de Thurstone alrededor de los años veinte (Thurstone, 1925) y continua en los cuarenta con aportes de Lawley (1943, 1944) y Tucker (1946); sin embargo no es sino hasta la década de los sesenta que se da un gran impulso y desarrollo a esta teoría a través de Lord y Novick (1968) en el libro *Statistical Theories of Mental Test Scores* donde dedican cinco capítulos al tema. A partir de su libro, las investigaciones sobre los modelos TRI dominarán la psicometría hasta nuestros días. (Muñiz, *Las teorías de los Test: Teoría Clásica y Teoría de Respuesta a los Ítems*, 2010)

Tanto la TCT como la TRI estiman características latentes de individuos y es importante tener en cuenta ambas teorías en el momento de analizar ítems y estimar los niveles de aptitud de diferentes sujetos.

La TRI considera al ítem como unidad de medición, y no contradice ni los supuestos ni las conclusiones fundamentales de la TCT, sino que hace asunciones adicionales que permiten superar las debilidades de esta última, por lo tanto la TRI se enfoca más en el análisis de los ítems en general que en el análisis del test global; la TRI usa funciones matemáticas las cuales relacionan el nivel de probabilidad de contestar a un ítem, y el nivel de habilidad del individuo, esto se conoce como funciones de respuesta al ítem; en resumen la TRI propone:

- El rendimiento de un sujeto puede predecirse o explicarse por un conjunto de habilidades o rasgos.
- La relación entre el rendimiento de un examinado en un ítem y el conjunto de rasgos responsables de dicho rendimiento puede describirse mediante una función monótona creciente denominada función característica del ítem, en esta función, la CCI establece una relación funcional entre la proporción de respuestas correctas a un ítem y el nivel de aptitud en el rasgo.

Las principales ventajas del TRI sobre la TCT son (Terron, 2010):

- Existencia de invarianza de los parámetros de los ítems respecto a la muestra que se calcula, es decir, que los parámetros del ítem no cambian aunque las personas que contesten sean distintas.
- Existencia de invarianza del parámetro del rasgo del sujeto (aptitud) respecto al instrumento utilizado para estimarlo. Es decir, que el nivel de habilidad de la persona no depende del test.
- Aportación de medidas locales de precisión para diferentes niveles de aptitud por medio de funciones de información, y no exclusivamente con un índice global como el empleado por la TCT. Es decir, que ofrece un indicador de la precisión de las estimaciones para niveles concretos.

Las limitaciones de la TCT han llevado a la propuesta de modelos alternativos, de los cuales el más difundido y aplicado es el modelo de Rash (modelo logístico de un parámetro) que permite la medición conjunta de personas e ítems en una misma dimensión, aplicando los supuestos de la TRI. Este es el modelo más simple de todos, y soporta que la probabilidad de acertar un ítem dependa solamente del nivel de dificultad de dicho ítem y del nivel del sujeto en la variable medida (nivel de habilidad).

2.2.1 La curva característica del ítem. Es una función matemática que relaciona la probabilidad de éxito en una pregunta con la habilidad, medida por el conjunto de ítems que la contienen; la probabilidad de éxito se determina como la razón entre el número de respuestas correctas sobre el número total de preguntas:

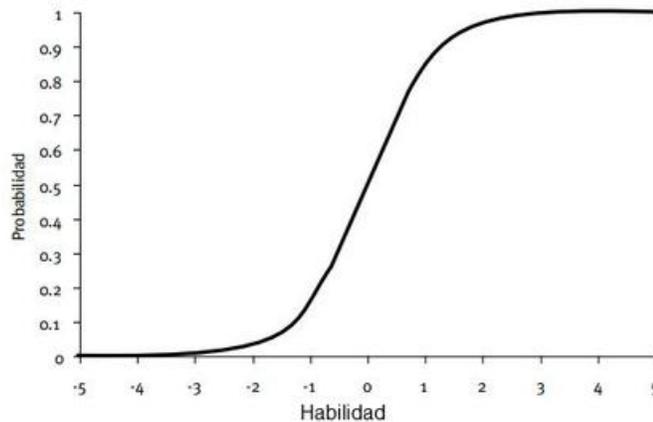
Ecuación 5. Función de probabilidad

$$p = \frac{\text{respuestas correctas}}{\text{total de preguntas}}$$

Teniendo en cuenta lo anterior, la probabilidad de fracaso se denotaría como:
 $q = 1 - p$

En la mayor parte de los test, la variable respuesta es una variable dicotómica, con dos posibles valores: 1 para la respuesta correcta y 0 para la incorrecta.

Gráfica 1. Curva característica del ítem



Fuente. Disponible en <https://wikimetria.wikispaces.com/Teor%C3%ADa+de+la+respuesta+al+item>

La gráfica muestra la relación entre la habilidad y la probabilidad de respuesta correcta a un ítem. Puede observarse que la probabilidad está próxima a cero en los niveles más bajos de la aptitud y se aproxima a 1 en los más altos.

Describe donde está situado el ítem en la escala de aptitud, es decir, que nivel de aptitud se requiere para responder el ítem con éxito; es por lo tanto un índice de posición del ítem en la escala de medida del rango de aptitud (θ).

Cada ítem está caracterizado por una CCI particular y propia, es decir, las CCI de los ítems que miden una determinada variable (θ) no son iguales. La forma particular de la CCI depende de los parámetros o características de cada ítem; bajo esta teoría las características de los ítems son independientes de la distribución de la aptitud en la población de los sujetos.

En la gráfica 1, se tiene una CCI empírica, pero la TRI necesita resumir la información que tiene cada CCI en un modelo que permita dar una descripción del rendimiento de los ítems; los modelos más utilizados, son los logísticos de uno, dos y tres parámetros (Gil & Sevilla):

La función del modelo está dada de la siguiente forma la cual lo valida.

Ecuación 6. Modelo de un parámetro (modelo de Rasch)

$$P_i \theta = \frac{e^{D(\theta-b_i)}}{1 + e^{D(\theta-b_i)}}$$

Donde,

$P_i \theta$: es la probabilidad de responder correctamente el ítem i en determinado nivel θ

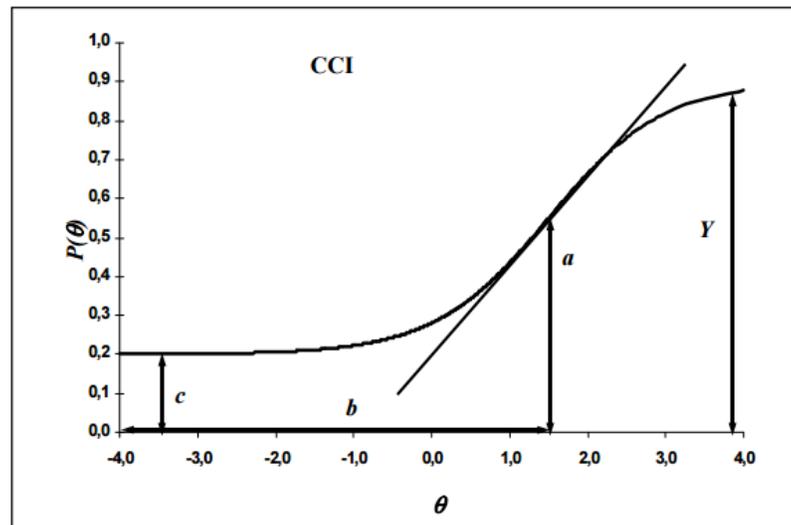
θ : son los valores de la variable medida. Nivel de habilidad del sujeto

b_i : es el índice de dificultad del ítem i

e : es la base de los logaritmos neperianos (2.72)” (Reyes y Barbosa, 2007, pág. 17 y 18)

D : Constante (cuando $D = 1.7$ los valores de la función logística apenas difieren de los de la curva normal acumulada) (Muñiz, 1997, pág. 36)

Gráfica 2. Parámetros de la CCI



Fuente. Disponible en <http://innoevalua.us.es/files/irt.pdf> pag. 9

- El parámetro a se le denomina índice de discriminación del ítem y representa la magnitud del cambio en la probabilidad de acertar el ítem conforme varía el nivel de habilidad. Su valor es proporcional a la pendiente de la recta tangente a la CCI en el punto de inflexión de esta.
- Parámetro b se corresponde con el valor de la abscisa (escala de habilidad (θ)) del punto de máxima pendiente de la CCI. Se le denomina índice de dificultad del ítem, y es un parámetro de localización del ítem que

representa la posición de la CCI en relación al nivel de habilidad (θ) necesario para obtener una probabilidad de acierto. $P(\theta) = \frac{e^{c\theta}}{1 + e^{c\theta}}$ Por lo tanto, el índice de dificultad es aquel valor de θ para el cual $P(\theta) = 0.5$.

- Parámetro c representa la probabilidad de acertar de los sujetos que desconocen la respuesta correcta (la probabilidad de acertar el ítem al azar), es decir, es el valor de $P(\theta)$ cuando θ tiende a su valor más bajo.

2.2.2 Dificultad. La dificultad indica el grado de habilidad de los sujetos para contestar un ítem, este valor es diferente al obtenido en la TCT. La dificultad indica la posición de la CCI a lo largo de la escala de habilidad; entre más difícil es un ítem, su curva estará localizada más a la derecha en la escala de habilidad.

Es uno de los dos parámetros fundamentales en el modelo de dos parámetros de la TRI; es indicador base para la conformación de pruebas y de conjuntos de ítems, así como para establecer comparabilidad de escalas. Se requiere para obtener otros indicadores de ítems (curvas características, función de información).

Los valores de dificultad oscilan entre menos infinito y más infinito en la escala logit, aunque en términos prácticos los ítems asumen valores entre -3.0 y $+3.0$, cuando el promedio de dificultades del grupo de ítems se centra en cero; valores positivos y altos indican alta dificultad y los valores negativos indican baja dificultad. (Reyes y Barbosa, 2007. Pág. 20)

2.2.3. El modelo de Rasch. El análisis Rasch es un método para obtener mediciones lineales objetivas y fundamentales (calificadas por errores estándar y estadísticas de ajuste de control de calidad) a partir de observaciones estocásticas de respuestas categóricas organizadas. Georg Rasch, un matemático Danés, formuló esta aproximación en 1953 con el fin de analizar las respuestas a unas series de pruebas de lectura (Rasch G, Probabilistic Models for Some Intelligence and Attainment Tests, Chicago: MESA Press, 1992).

El modelo de Rasch se fundamenta en los siguientes supuestos (Gerardo Prieto, 2003):

- El atributo que se desea medir puede representarse en una única dimensión en la que se situarían conjuntamente las personas y los ítems.
- El nivel de la persona en el atributo y la dificultad del ítem determinan la probabilidad de que la respuesta sea correcta.

A partir de este modelo se establece la probabilidad de respuesta de una persona ante un estímulo dado, en términos de la diferencia entre la medida del rasgo de la persona y la medida del estímulo aplicado. Se tiene como supuesto que la

respuesta a un ítem solo depende de la interacción entre la habilidad del sujeto y la dificultad del ítem.

La CCI, dada por la función logística, tiene en cuenta como único parámetro de los ítems el índice de dificultad. No se hacen hipótesis relacionadas con el patrón de respuestas de un grupo y si acerca de la adivinación sistemática. Plantea que la adivinación es un patrón personal y se puede detectar a partir del ajuste o desajuste de las respuestas de una persona al modelo. La medida de una persona es independiente de la prueba aplicada y la calibración del ítem es independiente de la población que lo aborda.

De esta manera se observa el modelo de Rasch como un método logístico, mediante el cual, se tiene la posibilidad de predecir la probabilidad de acertar en la respuesta a un ítem, conociendo de antemano la competencia del estudiante y la dificultad de dicho ítem. En el caso de este trabajo, se utiliza el modelo de un parámetro, donde la respuesta a un ítem sólo depende de la competencia del sujeto (θ) y de la dificultad del ítem (b), dado que contamos con el software ministep, una versión libre del software winstep, el cual fue adecuado para los datos obtenidos en la investigación.

2.2.4. Ajuste cercano y lejano (infit y outfit). El ajuste cercano y lejano indica la correspondencia entre un grupo de datos y el modelo estadístico utilizado para representarlos; el ajuste cercano (infit) se refiere a la relación entre los datos que se encuentran cerca del valor de dificultad del ítem y el valor de dificultad; el ajuste lejano se refiere a la relación de los datos que se encuentran lejos de dicho valor de dificultad y esa dificultad. Para las demás opciones de respuesta debe calcularse el ajuste lejano. (Barbosa & Reyes, 2007)

La utilización de un modelo para representar datos debe fundamentarse en la verificación de que dicho modelo en verdad representa el comportamiento de los datos y, por ende, puede inferirse el cumplimiento de los supuestos de dicho modelo para los datos analizados. Los valores posibles se encuentran entre 1 e infinito (positivo); el valor que determina el ajuste perfecto entre los datos y el modelo es 1. Los valores muy inferiores a 1 indican dependencia de los datos; valores superiores a 1 indican un “ruido” en la información; valores superiores a 2 indican que el “ruido” es mayor que la información útil; se aceptan valores de ajuste entre 0.8 y 1.2 (medición productiva). (Reyes y Barbosa, 2007)

Es decir, que los resultados obtenidos en las medidas de ajuste, indican que los ítems con mayor dificultad son aprobados por los estudiantes que se presentan más hábiles y reprobados por aquellos con bajas habilidades.

Según lo dicho anteriormente, en lo que respecta a los señalamientos de la introducción, del marco de referencia y del planteamiento del problema, fue

necesario la formulación de objetivos, para guiar el adecuado desarrollo del proyecto.

Objetivo General

Analizar mediante la TCT y la TRI la confiabilidad y algunos elementos psicométricos de una prueba de Física aplicada a grado décimo del colegio Nicolás Gómez Dávila IED de la localidad de Ciudad Bolívar.

Objetivos específicos

- Determinar los índices de dificultad y discriminación de los ítems que componen una prueba de conocimientos de física, a partir del análisis de las teorías TCT y TRI.
- Determinar el grado de consistencia interna de la prueba para observar el nivel de correlación de los ítems entre sí.
- Analizar el grado de confiabilidad que presenta la estructura de la prueba.

3. MARCO METODOLÓGICO

3.1 TIPO DE ESTUDIO

El tipo de estudio que resultó ser el más pertinente para el desarrollo del presente proyecto, es el estudio de tipo descriptivo, de igual manera se acudió a los fundamentos de la TCT y de la TRI.

3.2 VARIABLES DE ESTUDIO

Los 16 ítems de una prueba conceptual y matemática de la asignatura de Física.

3.3 PARTICIPANTES

Se contó con la participación de setenta estudiantes que conforman el grado décimo, de la jornada tarde, del Colegio Nicolás Gómez Dávila IED; que se encuentra ubicado en el barrio San Francisco de la localidad de Ciudad Bolívar.

La prueba se aplicó al 100% de la población, en el tercer periodo académico del año 2014. De los setenta estudiantes, 35 corresponden al género femenino y 35 al género masculino, con una edad aproximada de 16 años, pertenecientes a estrato dos.

3.4 INSTRUMENTOS DE RECOLECCIÓN

La recolección de la información para determinar la confiabilidad de un test aplicado a grado décimo, del Colegio Nicolás Gómez Dávila IED, se hizo a través de un test de física que incluye ítems tomados de un banco de preguntas sobre el

tema a evaluar: caída libre y lanzamiento vertical; la prueba se aplicó durante el tercer periodo académico del año 2014. Estos temas hacen parte de la malla curricular diseñada para este grado. El análisis se realizó utilizando la teoría clásica de los test y la teoría de respuesta al ítem.

Para el test se tomaron dieciséis ítems de selección múltiple con única respuesta, la cual fue resuelta en un tiempo de dos horas.

3.5 ESTRUCTURA DEL TEST

El test consta de 16 ítems distribuidos de la siguiente forma:

Tabla 3. Estructura del test

TEMA	CANTIDAD DE ÍTEMS	CONTENIDO
CAÍDA LIBRE	Ocho	Cinco conceptuales Tres matemáticos
LANZAMIENTO VERTICAL	Ocho	Cinco conceptuales Tres matemáticos

3.6 DISEÑO ESTADÍSTICO

Los procedimientos estadísticos que se utilizaron, para determinar la confiabilidad y otros elementos psicométricos de los ítems de una prueba de Física, corresponden a la teoría clásica de los ítems y teoría de respuesta al ítem; a continuación se hace mención de los procedimientos, que hacen parte de cada una de las teorías, y que para efectos del presente trabajo fueron aplicados.

TCT

- Alfa de Cronbach
- Índice de dificultad
- Índice de discriminación
- Coeficiente de correlación biseral puntual

Los procedimientos de la TCT, se realizaron por medio del software EXCEL y SPSS.

TRI

- Curva característica del Ítem (CCI)
- Índice de dificultad
- Correlación del punto biseral
- Medidas de ajuste: INFIT y OUTFIT

Los procedimientos de la TRI, se realizaron por medio del software MINISTEP.

4 RESULTADOS

4.1 TEORIA CLÁSICA DE LOS TEST

4.1.1 Alfa de Cronbach. Este índice, es un modelo de consistencia interna, basado en el promedio de las correlaciones entre los ítems. Entre las ventajas de esta medida se encuentra la posibilidad de evaluar cuanto mejoraría (o empeoraría) la fiabilidad del test si se excluyera determinado ítem. Hay que recordar que a mayor valor del indicador, mayor es la fiabilidad, el mayor valor teórico de la medida es 1.

Para calcular este indicador, se utilizó el programa SPSS, generando los siguientes resultados:

Tabla 4. Resultado Alfa de Cronbach
Estadísticos de fiabilidad

Alfa de Cronbach	Alfa de Cronbach basada en los elementos tipificados	N de elementos
,622	,611	16

Teniendo en cuenta la tabla 1 (pag. 16) sobre la interpretación de este indicador, el resultado de 0,622 permite concluir de manera preliminar que el nivel de confiabilidad es alto, es decir los ítems covarían fuertemente entre sí; sin embargo, es importante analizar este resultado generando una validación del mismo ítem a ítem, y teniendo en cuenta los resultados de los demás indicadores de la TCT y la TRI antes de generar una conclusión definitiva sobre la fiabilidad de la prueba.

4.1.2 Índice de dificultad. La tabla 5, muestra los estadísticos de los elementos calculados en SPSS; para el caso del índice de dificultad corresponde a la media.

Tabla 5. Estadísticos de los elementos
Estadísticos de los elementos

	Media	Desviación típica	N
ITEM1	,42	,497	60
ITEM2	,18	,390	60
ITEM3	,80	,403	60
ITEM4	,60	,494	60
ITEM5	,57	,500	60
ITEM6	,62	,490	60
ITEM7	,33	,475	60
ITEM8	,63	,486	60
ITEM9	,50	,504	60
ITEM10	,50	,504	60
ITEM11	,62	,490	60
ITEM12	,35	,481	60
ITEM13	,55	,502	60
ITEM14	,70	,462	60
ITEM15	,53	,503	60
ITEM16	,55	,502	60

Teniendo en cuenta, que el índice de dificultad indica la proporción de personas que no responden correctamente a un ítem, de los resultados de la tabla 5 se puede decir: el 63% de los ítems, tiene un índice de dificultad superior al 0,4, es decir presentan un nivel de dificultad entre moderado y fácil (tabla 2. Interpretación índice de dificultad, pag. 16); mientras que el 25% de los ítems, presenta un resultado por debajo de 0,4, es decir, son ítems que se consideran difíciles.

Con los resultados obtenidos y la valoración dada en la Tabla 2 sobre la interpretación de los resultados, se resume el nivel de dificultad de los ítems en la tabla 6.

Tabla 6. Resultados e Interpretación del índice de dificultad de los ítems

ITEM	INTERPRETACIÓN	ITEM	INTERPRETACIÓN
1	Moderado	9	Moderado
2	Difícil	10	Moderado
3	Fácil	11	Fácil
4	Moderado	12	Difícil
5	Moderado	13	Moderado
6	Fácil	14	Fácil
7	Difícil	15	Moderado
8	Fácil	16	Moderado

Por lo tanto, el 50% de los ítems presentan una dificultad moderada, el 31% fácil y el 19% restante difícil. Todos los ítems varían entre un nivel fácil y difícil, y ninguno se clasifica como demasiado fácil o demasiado difícil.

4.1.3 Índice de discriminación. Un buen ítem debe discriminar entre aquellos que obtuvieron buenas calificaciones en la prueba y aquellos que obtuvieron bajas calificaciones. Cuanto más alto es el índice de discriminación, el ítem diferenciará mejor a las personas con altas y bajas calificaciones; lo deseable, es que este valor sea mayor a 0.30; cuando el valor es inferior se considera que el ítem no

discrimina adecuadamente, es decir que el ítem no diferencia entre los individuos del nivel superior y del nivel inferior.

Los resultados de este índice, son los detallados en la tabla 7, los cuales fueron estimados en Excel mediante la ecuación 3 (pag. 17).

Tabla 7. Índice de discriminación

Ítem	Discriminación
1	0,21
2	0,01
3	0,03
4	0,16
5	0,1
6	0,15
7	0,2
8	0,06
9	0,2
10	0,16
11	0,25
12	0,08
13	0,28
14	0,13
15	0,2
16	0,11

El valor esperado para definir si un ítem discrimina adecuadamente a las personas con altas y bajas calificaciones es 0,30; de acuerdo a los resultados, ninguno de los ítems permite discriminar adecuadamente, lo que da lugar a concluir que el grado de consistencia interna del test es muy bajo, por lo tanto no se puede asegurar que los ítems diferencien a individuos con baja o alta habilidad.

4.1.4 Coeficiente de correlación biserial puntual (ítem – test).

Tabla 8. Resultado índice de Correlación

Ítem	Correlación $\geq 0,25$
1	7, 9, 11, 13, 14, 16
2	-
3	-
4	5, 13
5	4, 8,
6	13
7	1, 10, 15
8	5
9	1, 10, 12, 13
10	7, 9, 12, 13, 16
11	-

12	9, 10,
13	1, 4, 6, 9, 10, 14
14	1, 13
15	7
16	1, 10

De la tabla anterior se puede concluir, que los ítems mejor relacionados son el 1, 9, 10 y 13 porque se relacionan con más ítems que el resto y porque la relación es superior a 0,25; es decir muestran una relación homogénea; por el contrario y soportados en los resultados, los ítems 2, 3 y 11 presentan correlación inversa con 5 o más ítems del test.

Como complemento al análisis de la TCT, se calcula en SPSS los estadísticos totales de los elementos de la prueba, donde se tienen los cálculos del alfa de Cronbach eliminando cada ítem; estos resultados se muestran en la tabla 9. Al analizar los mismos se obtiene que la consistencia interna mejora si se eliminan los ítems 3 y 8, siendo adicionalmente estos ítems los que presentan una relación negativa con el 50% del resto de ítems y adicionalmente, no muestran relaciones superiores a 0,25 con el restante 50% de ítems. Con lo anterior, podemos concluir que la prueba mejora su nivel de fiabilidad, eliminando los ítems 3 y 8 del test.

Tabla 9. Estadísticos total-elemento

	Media de la escala si se elimina el elemento	Varianza de la escala si se elimina el elemento	Correlación elemento-total corregida	Correlación múltiple al cuadrado	Alfa de Cronbach si se elimina el elemento
ITEM1	8,03	7,389	,466	.	,569
ITEM2	8,27	8,606	,060	.	,629
ITEM3	7,65	8,774	-,017	.	,638
ITEM4	7,85	8,231	,148	.	,620
ITEM5	7,88	8,206	,154	.	,620
ITEM6	7,83	7,904	,273	.	,601
ITEM7	8,12	7,732	,355	.	,589
ITEM8	7,82	8,525	,047	.	,635
ITEM9	7,95	7,947	,244	.	,606
ITEM10	7,95	7,845	,282	.	,600
ITEM11	7,83	7,633	,377	.	,585
ITEM12	8,10	8,024	,235	.	,607
ITEM13	7,90	7,447	,437	.	,574
ITEM14	7,75	8,055	,239	.	,607
ITEM15	7,92	7,908	,260	.	,603
ITEM16	7,90	8,024	,218	.	,610

4.2 TEORIA DE RESPUESTA AL ITEM

En la TRI, se utilizan generalmente dos tipos de familias de funciones de distribución para la CCI: la función logística y la curva normal acumulada, que dan lugar a seis modelos generales según se contemple uno, dos o tres parámetros de los ítems para cada una de estas dos funciones. En todos los casos son modelos unidimensionales y se asume que la respuesta a los ítems es dicotómica. En el presente trabajo, el modelo que se utiliza es el de un parámetro y para analizarlo con mayor detalle, se utilizó el modelo de Rasch desarrollado a través del software MiniStep obteniendo los siguientes resultados estadísticos:

Tabla 10. Resultados generales de la TRI para los sujetos

PERSON	60 INPUT		60 MEASURED		INFIT		OUTFIT	
	TOTAL	COUNT	MEASURE	REALSE	IMNSQ	ZSTD	OMNSQ	ZSTD
MEAN	8.4	16.0	.17	.61	.99	.0	1.01	.1
S.D.	3.0	.0	1.01	.12	.18	.9	.30	.9
REAL RMSE	.62	TRUE SD	.80	SEPARATION	1.30	PERSON	RELIABILITY	.63

Los resultados generales nos muestran que el coeficiente de confiabilidad es de 0,63 (RELIABILITY); el nivel de habilidad de respuesta de los estudiantes evaluados es del 0,17 con lo cual se puede concluir que la habilidad de los sujetos tiene un nivel de dificultad medio ya que este indicador se acerca a cero; el número de respuestas correctas observadas tiene una media de 8.4, la desviación estándar es de 3.0; se respondieron 16 ítems de los 16 que tenía el test, es decir, no quedó ningún ítem sin respuesta.

Tabla 11. Resultados generales de la TRI para los ITEM's

ITEM	16 INPUT		16 MEASURED		INFIT		OUTFIT	
	TOTAL	COUNT	MEASURE	REALSE	IMNSQ	ZSTD	OMNSQ	ZSTD
MEAN	31.7	60.0	.00	.30	1.00	-.1	1.01	.0
S.D.	8.8	.0	.77	.03	.11	.9	.15	.7
REAL RMSE	.31	TRUE SD	.71	SEPARATION	2.32	ITEM	RELIABILITY	.84

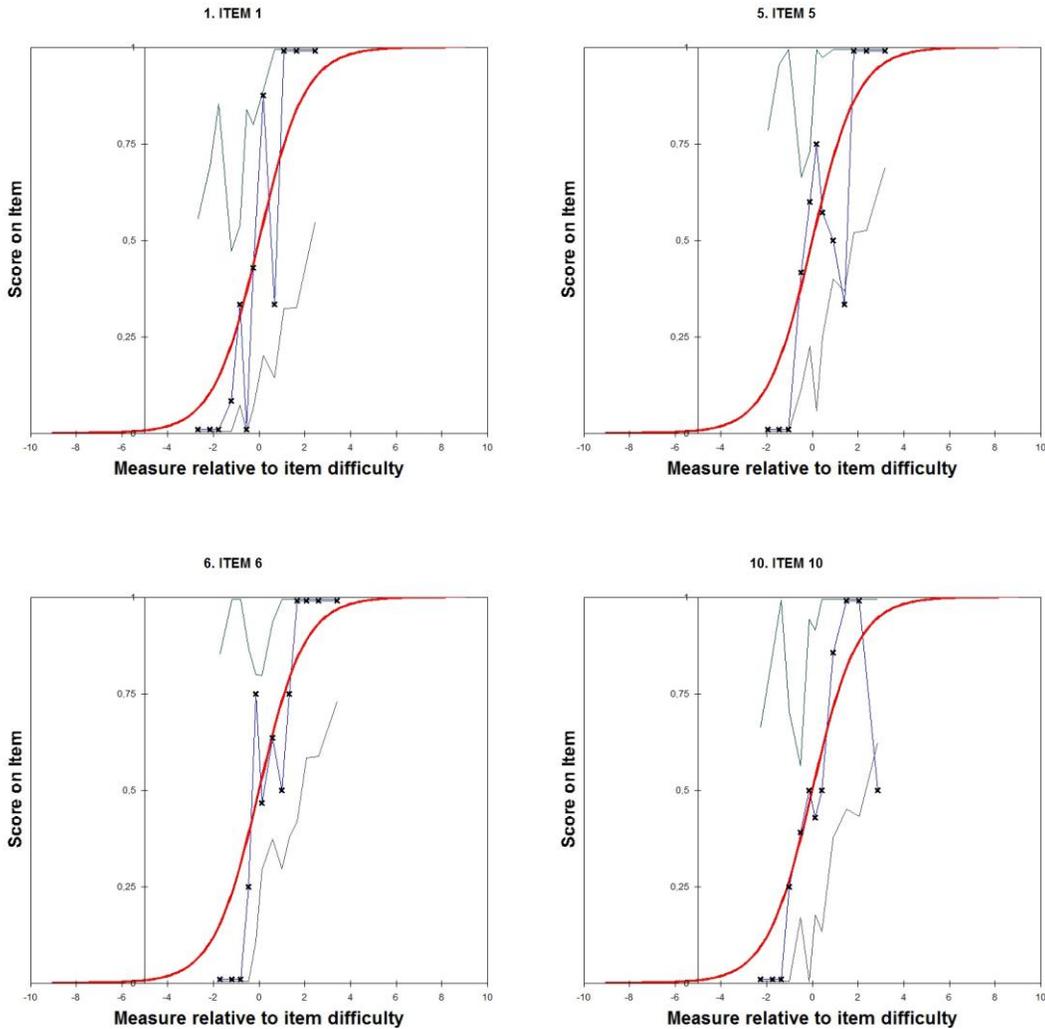
Con respecto a la dificultad del ítem (MEASURE), este es de 0,00; es decir, presentan un nivel de dificultad medio. Los valores de infit y outfit de la media de los ítems se encuentran entre 1.00 y 1.01 respectivamente, mostrándose que se encuentran dentro del rango establecido como deseable de acuerdo a la definición dada en el numeral 2.2.5; por lo tanto, se concluye que el modelo de Rasch es adecuado aplicarlo en la prueba puesto que los sujetos que poseen el conocimiento o la habilidad están aprobando el test.

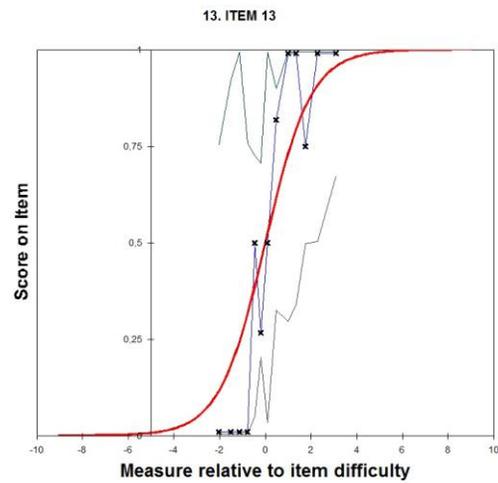
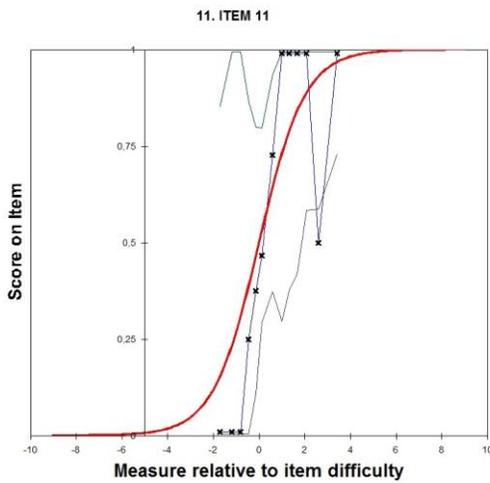
En relación a los valores de la desviación estándar que se muestran en la tabla son 0.11 y 0.15, respectivamente para las medidas de ajuste Infit y Outfit de los Items, encontrándose dentro del rango establecido como aceptable, y se concluye que la aplicación de la prueba, los sujetos que no poseen el conocimiento o la habilidad están en general reprobando los ítems.

De acuerdo a las medidas de ajuste y el conocimiento del grupo al cual se le aplico la prueba, estos resultados indican que en general los ítems difíciles son aprobados por los estudiantes de alta habilidad y reprobados por los estudiantes de baja habilidad o con poco dominio del tema de caída libre y lanzamiento vertical.

4.2.1 Curva característica del Ítem (CCI). Existe una relación funcional entre la variable que miden los ítems y la probabilidad de acertar a estos, la función asociada a esta relación es lo que se muestra en la CCI. Se realizó el análisis de la CCI para cada uno y de acuerdo con los resultados obtenidos, los ítems 1, 5, 6, 10, 11 y 13 son los más representativos

Grafica 3. CCI de los ítems en sujetos con mayor habilidad





Con lo anterior, se puede decir que la dificultad de los ítems es mínima puesto que los resultados de la variable b fueron muy cercanos a cero y negativos. Los ítems que no fueron correctamente contestados por los estudiantes, es decir, lo contestaron erróneamente y por lo tanto, son los que cuentan con menor habilidad, son: 2, 3, 4, 7, 8, 9, 12, 14, 15, 16; estos ítems presentan menos puntos de corte con la CCI esperada (Gráfica 4) se muestran algunos de los ítems que tienen menos aproximación con la CCI.

Gráfica 4. CCI de los ítems con menor habilidad

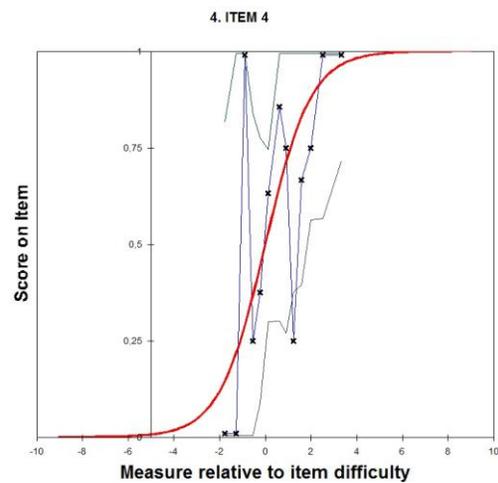
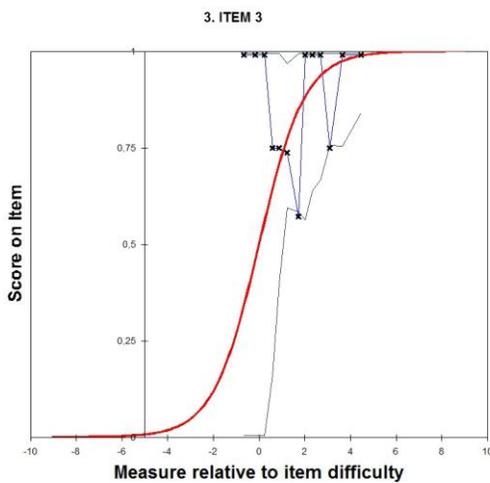


Tabla 12. Índice de dificultad e Índice de discriminación. Infit y Outfit de cada Ítem

ITEM STATISTICS: MEASURE ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT		PTMEASURE-A		EXACT OBS%	MATCH EXP%	ITEM
					MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.			
2	11	60	1.94	.37	1.22	1.0	1.30	.9	.21	.40	81.7	83.6	ITEM 2
7	20	60	.96	.30	.92	-.5	.91	-.4	.49	.42	78.3	73.8	ITEM 7
12	21	60	.87	.30	1.03	.3	1.00	.1	.40	.42	73.3	72.8	ITEM 12
1	25	60	.53	.29	.82	-1.6	.77	-1.5	.58	.42	80.0	69.6	ITEM 1
9	30	60	.12	.28	1.00	.0	1.07	.5	.38	.40	73.3	66.9	ITEM 9
10	30	60	.12	.28	.97	-.3	1.11	.8	.41	.40	66.7	66.9	ITEM 10
15	32	60	-.04	.28	.98	-.1	.99	.0	.41	.39	70.0	66.2	ITEM 15
13	33	60	-.11	.28	.83	-1.9	.79	-1.3	.54	.39	78.3	66.1	ITEM 13
16	33	60	-.11	.28	1.02	.3	.98	-.1	.38	.39	61.7	66.1	ITEM 16
5	34	60	-.19	.28	1.07	.8	1.02	.2	.33	.38	60.0	65.9	ITEM 5
4	36	60	-.36	.29	1.06	.7	1.07	.4	.31	.37	68.3	66.1	ITEM 4
6	37	60	-.44	.29	.96	-.4	.86	-.7	.42	.37	63.3	66.8	ITEM 6
11	37	60	-.44	.29	.86	-1.4	.91	-.4	.48	.37	73.3	66.8	ITEM 11
8	38	60	-.52	.29	1.15	1.4	1.13	.7	.23	.36	58.3	67.5	ITEM 8
14	42	60	-.87	.30	.97	-.2	.92	-.2	.37	.34	75.0	71.9	ITEM 14
3	48	60	-1.47	.34	1.18	1.0	1.29	.9	.12	.29	76.7	80.7	ITEM 3
MEAN	31.7	60.0	.00	.30	1.00	-.1	1.01	.0			71.1	69.9	
S.D.	8.8	.0	.77	.02	.11	.9	.15	.7			7.2	5.3	

La tabla anterior, nos muestra el grado de dificultad de cada ítem (MEASURE) que corresponde al parámetro b . Los ítems 2, 7, 12 y 1 tienen alta dificultad y el resto de ítems presentan baja dificultad, siendo el ítem 3 el que presenta más baja dificultad de todos; hay que recordar que eliminando este ítem el alfa de Cronbach del test mejora.

Los resultados en la columna PTMEASURE-A CORR, se refiere a la correlación del punto biserial que mide la relación entre las puntuaciones de los ítems; el valor esperado debe ser positivo, en caso contrario la prueba presenta un alto desajuste. En los resultados de la prueba, el 88% de los ítems, presenta un índice de discriminación superior a 0.3, y como todos son positivos, no hay por lo tanto ningún ítem que esté generando algún desajuste en el test.

Para INFIT/OUTFIT MNSQ la expectativa del modelo es de valores de 1.00; en el caso de los resultados de este trabajo, se observa en la tabla 12 que no hay ningún ítem con valores mayores a 1.3 siendo este valor un umbral indicativo de falta de ajuste puesto que indican ruido o alta variabilidad aleatoria en los datos (Montesinos), valores mayores a 2 indican que el ruido es mayor que la información útil; los ítems 7, 1, 13 y 11 son significativamente menores a 1 lo cual indica que aparentemente se ajustan demasiado bien al modelo; cuando este indicador toma rangos entre -2 y 2 se dice que están en el intervalo de lógitos aceptable para determinar un ajuste razonable en ítems.

CONCLUSIONES

Los resultados del análisis de la TCT, muestran que la prueba tiene un nivel de confiabilidad del 0,622 que de acuerdo a la tabla 1 (Pag. 16) de valorización de este indicador, estaría en el intervalo de definición de alta confiabilidad y por lo tanto los ítems no son independientes entre sí; la dificultad muestra que todos los ítems varían entre fácil y difíciles (Tabla 6. Pag. 16) y ninguno se encuentra en los límites de demasiado fácil o demasiado difícil; sin embargo y de acuerdo a los resultados del índice de discriminación (Tabla 7, Pag. 17), la prueba no permite diferenciar entre los individuos con alta y baja habilidad.

En los resultados de la TRI, la confiabilidad es del 0.63 siendo muy parecido al indicado en la TCT, al analizar los resultados de este indicador por ítem, se observa que este resultado mejora al pasar a 0,635 y 0,638 eliminando los ítems 8 y 3 respectivamente; con estos resultados, se puede concluir que la prueba si es confiable.

De acuerdo con la TCT, los ítems no discriminan adecuadamente a las personas con alta y baja calificación, puesto que el resultado de este indicador fue inferior a 0.30 en todos los ítems.

Con lo anterior, se recomienda descartar de la prueba los ítems 3 y 8, que son los que más afectan el nivel de confiabilidad de la misma; por lo tanto, la confiabilidad estaría por encima del 0,63 que corresponde a un índice alto de confiabilidad; adicionalmente la prueba está en un rango aceptable de dificultad puesto que todos los ítems varían entre difíciles y fáciles, e inclusive el 50% de ellos presenta una dificultad moderada y teniendo en cuenta las medidas de ajuste, ningún ítem genera desajuste en el test; sin embargo y con el objetivo de mejorar el índice de confiabilidad, se recomienda reemplazar los dos ítems mencionados, de tal manera que el test pueda ser utilizado con mayor confianza en la evaluación de otros estudiantes.

REFERENCIAS BIBLIOGRAFICAS

- Aliaga Tovar, J. (s.f.). *Psicometria: Tests Psicométricos, Confiabilidad y Validez*. Recuperado el 12 de abril de 2015, de aprendeenlinea.udea.edu.co:
http://aprendeenlinea.udea.edu.co/lms/investigacion/pluginfile.php/1909/mod_resource/content/0/ARCHIVOS_2010/PDF/IntPsicometria_aristidesvara_1_.pdf
- Barbosa, Y., & Reyes, J. F. (2007). *Universidad Nacional de Colombia*. Recuperado el 25 de 04 de 2015, de <http://fce.unal.edu.co/uifce/pdf/MiniSteps.pdf>
- Barrero, Melissa Judith Ortiz. (2014). *Introducción a la historia y conceptos básicos de la psicometría*. Recuperado el 17 de 08 de 2014, de http://datateca.unad.edu.co/contenidos/401517/TEXTO_UNIDAD_1.pdf
- Blandón, A. C. (2014).
- Francisco J. Abad, J. G. (2006). *Universidad Autonoma de Madrid*. Recuperado el 17 de 08 de 2014, de http://aprendeenlinea.udea.edu.co/lms/investigacion/file.php/39/ARCHIVOS_2010/PDF/IntPsicometria_aristidesvara_1_.pdf
- Gerardo Prieto, A. R. (2003). *Análisis de un test mediante el modelo de Rasch*. Recuperado el 25 de 04 de 2015, de www.psicothema.com: <http://www.psicothema.com/pdf/1029.pdf>
- Gil, J. A., & Sevilla, U. d. (s.f.). <http://innoevalua.us.es/>. Recuperado el 23 de 05 de 2015, de GrupoInnoevalua - Universidad de Sevilla: <http://innoevalua.us.es/files/irt.pdf>
- Gregorio Suarez Arrieta, S. M. (2010). *Plan de mejoramiento, definición del horizonte institucional para el colegio Bolivariano*. Bogota: Trabajo de grado.
- Marchesi, Á. (s.f.). Avances y desafíos en la evaluación educativa. En E. Martin, & F. Martínez Rizo, *Avances y desafíos en la evaluación educativa* (págs. 6-7). Madrid, España: Fundación Santillama.
- Muñiz, J. (1997). *Introducción a la teoría de respuesta a los ITEMS*. España: Ediciones Pirámide.
- Muñiz, J. (2010). Las teorías de los Test: Teoría Clasica y Teoría de Respuesta a los Items. *Redalyc*, 57-66.
- Secretaria de Educación Bogota. (2014). *Educación Bogotá*. Recuperado el 07 de 08 de 2014, de http://www.educacionbogota.edu.co/index.php?option=com_content&view=article&id=2808
- Secretaria de Educación de Bogotá. (2013). *Evaluación para la calidad educativa*. Bogotá: Secretaria de Educación de Bogotá.
- Secretaria de Educación de Bogota. (16 de 08 de 2014). *Educación Bogotá*. Recuperado el 16 de 08 de 2014, de <http://evaluacion.educacionbogota.edu.co/index.php/evaluacion-de-los-aprendizajes/pruebas>
- Terron, A. M. (2010). [http://riuma.uma.es/Introducción al análisis de la Teoría de Respuesta al Ítem](http://riuma.uma.es/Introducción%20al%20análisis%20de%20la%20Teoría%20de%20Respuesta%20al%20Ítem) . Recuperado el 25 de 04 de 2015, de Aidesoc.net:
http://riuma.uma.es/xmlui/bitstream/handle/10630/4711/TRI_aidesoc_2011.pdf?sequence=1
- Universidad Nacional Abierta y a Distancia. (2010). *Psicodiagnóstico de las Funciones Cognoscitivas*. Recuperado el 16 de 08 de 2014, de http://datateca.unad.edu.co/contenidos/401519/Material_Didactico_para_descargar/Material_e_n_Exe-learning_del_Modulo/leccin_2_surgimiento_y_evolucin_de_la_evaluacin_psicologica.html
- Universidad Nacional Abierta y a Distancia. (s.f.). *UNAD Normas de Investigación*. Recuperado el 12 de 08 de 2014, de http://datateca.unad.edu.co/contenidos/100104/100104_EXE/leccin_6_investigacin_exploratoria_descriptiva_correlacional_y_explicativa.html

ANEXO

Test de física aplicado a los estudiantes y que sirvió de instrumento de recolección de información para este trabajo

NICOLAS GOMEZ DAVILA IED
GRADO DÉCIMO
EVALUACION DE FISICA
CAIDA LIBRE Y LANZAMIENTO VERTICAL

1. Para que una moneda y una pluma caigan exactamente al mismo tiempo se debe considerar un espacio:
- Con luz
 - Sin luz (vacío)
 - Con rozamiento
 - Sin rozamiento (vacío)

Resolver las preguntas de la 2 a la 7 con la siguiente situación.

- Desde el cuarto piso del colegio que se encuentra a una altura de 723 cm. Carlos deja caer un balón de baloncesto.

2. A medida que el balón desciende que pasa con el valor de la aceleración debida a la gravedad.

- Aumenta
- Disminuye
- Permanece constante
- Es cero

3. Cuando el balón esta cayendo el valor de la velocidad:

- Aumenta
- Disminuye
- Permanece constante
- Es cero

4. El tiempo que demora descendiendo es:

- 1.21 h
- 1.21 s
- 12.1 h
- 12.1 s

5. La velocidad con que llega al suelo es de:

- 1186 m/s
- 1186 km/h
- 11.86 m/s
- 11.86 km/h

6. Si en el ejercicio anterior se lanza el balón con una velocidad inicial de 10 m/s, la velocidad final es:

- Mayor que la de lanzamiento
- Menor que la de lanzamiento
- Igual a la de lanzamiento
- Cero

7. El tiempo de caída con respecto al ejercicio anterior donde la velocidad inicial es igual a cero es:

- Igual
- Mayor
- Menor
- No se puede determinar

8. La velocidad final de esta situación es:

- 21.86 m/s
- 2186 m/s
- 21.86 km/h
- 2186 km/h

Resolver las preguntas de la 9 a la 16 con la siguiente situación.

- Para un experimento de Física Juan diseña un cohete el cual alcanza una altura de 6223 m hasta que agota su combustible, la velocidad de lanzamiento fue de 115 m/s.

9. Desde el momento en que el cohete inicia su viaje hasta que lo finaliza, regresando al sitio de lanzamiento su velocidad cambia de la siguiente forma:

- a. Cuando asciende la velocidad va aumentando y cuando desciende la velocidad va disminuyendo.
- b. Cuando asciende la velocidad va disminuyendo y cuando desciende la velocidad va aumentando.
- c. Cuando asciende la velocidad va aumentando y cuando desciende la velocidad va aumentando.
- d. Cuando asciende la velocidad va disminuyendo y cuando desciende la velocidad va disminuyendo.

10. Con respecto al tiempo de viaje del cohete podemos concluir que:

- a. El tiempo de subida es mayor que el de bajada.
- b. El tiempo de subida es menor que el de bajada.
- c. El tiempo de subida es igual que el de bajada.
- d. El tiempo de bajada es mayor que el de subida.

11. La velocidad vertical en el punto donde comienza a descender es:

- a. Igual a la de lanzamiento.
- b. Cero.
- c. Igual a la velocidad final.
- d. No se puede determinar.

12. La aceleración debida a la gravedad cuando el cohete asciende es _____ que cuando el cohete desciende:

- a. Mayor
- b. Menor
- c. Igual
- d. Variables

13. El tiempo total de viaje (ascendiendo y descendiendo) es de:

- a. 23.46 s
- b. 23.46 m
- c. 234.6 s
- d. 234.6 m

14. La máxima altura alcanzada es:

- a. 62.23 m
- b. 62.23 km
- c. 6223 m
- d. 6223 km

15. Cuando han transcurrido 8 segundos de viaje, la altura alcanzada es de:

- a. 920 m
- b. 6223 cm
- c. 35.63 s
- d. 606.4 m

16. La velocidad de regreso al sitio de lanzamiento es:

- a. El doble de la velocidad inicial.
- b. La mitad de la velocidad inicial.
- c. Cero.
- d. La misma velocidad inicial.

ECUACIONES CAIDA LIBRE:

(Recuerde que la velocidad inicial es igual a cero)

$$y = v_0 t + \frac{1}{2} g t^2 ; v_f = v_0 + g t ; t = \frac{v_f - v_0}{g} ; t = \frac{\sqrt{2y}}{g}$$

ECUACIONES LANZAMIENTO VERTICAL

$$y = v_0 t - \frac{1}{2} g t^2 ; v_f = v_0 - g t$$

