

UNIVERSIDAD NACIONAL AUTÓNOMA DE HONDURAS

FACULTAD DE CIENCIAS SOCIALES

ESCUELA DE CIENCIAS PSICOLÓGICAS

MAESTRÍA EN PSICOMETRÍA Y EVALUACIÓN EDUCATIVA



ANÁLISIS PSICOMÉTRICO DE LAS ÁREAS BÁSICAS DE LA PRUEBA DE
ADMISIÓN A ESCUELAS NORMALES EN LA MUESTRA APLICADA A
ESTUDIANTES HONDUREÑOS ASPIRANTES EN EL AÑO 2011.

PRESENTADA POR: ANAYANCY DELCID RAMÍREZ

PARA OPTAR AL GRADO DE MÁSTER EN PSICOMETRÍA Y EVALUACIÓN
EDUCATIVA

ASESOR TEMÁTICO: DENNIS FERNANDO CÁCERES VALLADARES

ASESOR METODOLÓGICO: GERMAN EDGARDO MONCADA GODOY

TEGUCIGALPA, M.D.C., HONDURAS

ENERO, 2023

DEGT-UNAH

Universidad Nacional Autónoma de Honduras

Facultad de Ciencias Sociales

Escuela de Ciencias Psicológicas

Maestría en Psicometría y Evaluación Educativa



Análisis psicométrico de las áreas básicas de la Prueba de Admisión a Escuelas Normales en la muestra aplicada a estudiantes hondureños aspirantes en el año 2011.

Presentada por: Anayancy Delcid Ramírez

Para optar al grado de Máster en Psicometría y Evaluación Educativa

Asesor Temático: Dennis Fernando Cáceres Valladares

Asesor Metodológico: German Edgardo Moncada Godoy

Tegucigalpa, M.D.C., Honduras

Enero, 2023

Tegucigalpa, M.D.C., 17 de enero de 2023

Señores

Universidad Nacional Autónoma de Honduras

Coordinadores Maestría en Psicometría y Evaluación Educativa

Estimados señores:

Por este medio autorizo a la UNAH para que publique en su REVISTA DIGITAL, WEB, CD Y CUALQUIER OTRO MEDIO, y uso de beneficio sociales y académicos de la sociedad hondureña mi trabajo de tesis titulado: ***“Análisis Psicométrico de las áreas básicas de la Prueba de Admisión a Escuelas Normales en la muestra aplicada a estudiantes hondureños aspirantes en el año 2011”***, el cual he presentado como parte de los requisitos para optar por la Maestría en Psicometría y Evaluación Educativa.

Atentamente,

Anayancy Delcid Ramírez

Cuenta No. PSE100210

AUTORIDADES UNIVERSITARIAS

Doctor Francisco José Herrera Alvarado

Rector Interino

Máster Belinda Flores de Mendoza

Vice-Rectora Académica

Doctora Jessica Patricia Sánchez Medina

Secretaria General

Doctora Leonarda del Carmen Andino

Directora de Docencia

Doctor Armando Euceda

Director Sistema de Estudios de Postgrado

Máster Carmen Julia Fajardo

Decana Facultad de Ciencias Sociales

Máster María José Irías Escher

Coordinadora General Postgrados Facultad de Ciencias Sociales

Máster Lina María Mendoza Recarte

Coordinadora Académica Maestría en Psicometría y Evaluación Educativa

Resumen

La Prueba de Admisión a Escuelas Normales fue elaborada con el propósito de evaluar conocimientos básicos y precisar el proceso de ingreso de estudiantes hondureños aspirantes a las Escuelas Normales en el año 2011. La estructura de la prueba comprende 80 ítems de selección única de las áreas de español, matemáticas, ciencias sociales y ciencias naturales organizadas en dos formas: forma A y forma B. Este estudio de investigación se orientó con un enfoque cuantitativo de Tipo Instrumental, utilizando una muestra conformada por 5,523 aspirantes en su mayoría estudiantes de noveno grado, de los cuales 2,776 fueron evaluados con la forma A y 2,747 fueron evaluados con la forma B.

El objetivo de este estudio fue analizar los parámetros psicométricos de las áreas básicas de la prueba de Admisión utilizando los datos de los aspirantes del año 2011. Este trabajo de investigación consistió primeramente en evaluar la consistencia interna de la prueba de admisión donde se lograron obtener índices aceptables de confiabilidad para ambas formas de la prueba. Seguidamente se calcularon los parámetros psicométricos de los ítems desde la Teoría Clásica de los Test (TCT), para efectuar los análisis se utilizaron programas estadísticos como Jamovi, RStudio y SPSS; se obtuvieron en general para los ítems de la prueba índices bajos de dificultad y discriminación.

Continuando con el estudio de investigación se evaluó la dimensionalidad de la prueba mediante un Análisis Factorial Exploratorio (AFE) haciendo uso de una matriz de correlaciones tetracóricas, inicialmente se aplicó el estadístico de Henze-Zirkler para evaluar la normalidad multivariante de los datos, corroborando que no existe normalidad multivariante en ninguna de las dos formas de la prueba ($p = < 0.05$), se calcularon los índices de la prueba de Esfericidad de Barlett y el Índice de Medida de la adecuación muestral Kaiser-Meyer-Oikin (KMO).

El Análisis Paralelo de Horn, sugiere la presencia de un factor en la prueba, se hizo uso de los estimadores mínimos cuadrados ordinarios (MCO) como método de extracción y del estadístico MINRES basados en el criterio de los MCO; como método de rotación se utilizó un método de rotación ortogonal: Varimax. En el desarrollo del análisis se descartaron ítems debido a que su peso factorial era negativo, quedando en la solución final 51 de 80 ítems en

la forma A y 49 de 80 ítems en la forma B, la Varianza común explicada del factor general obtenida fue de 0.34.

Finalmente se realizó la estimación de parámetros desde la Teoría de Respuesta al Ítem (TRI) aplicando el modelo logístico de un parámetro, mediante el programa RStudio. Tomando en cuenta los resultados obtenidos se concluye que la prueba de admisión a las escuelas normales cuenta con pocos ítems que cumplen con los criterios métricos de calidad por lo cual se sugiere la revisión de la misma para mejorar los índices y de esta forma no comprometer el constructo que se está midiendo.

Palabras clave: Psicometría, Evaluación, Prueba de Admisión, Dificultad, Confiabilidad, Discriminación, Análisis Factorial Exploratorio.

Dedicatoria

Dedico este trabajo de investigación a mi querida madre y hermano que siempre me han apoyado y animado a seguir adelante en este proceso de estudio que ha sido muy difícil y que en momentos pensé no podría lograr.

DEGT-UNAH

Agradecimiento

Primeramente, agradezco a Dios quien ha sido mi fortaleza y guía en todo este camino recorrido, me ha brindado la sabiduría e inteligencia que han regido cada una de las fases que he tenido que sobrellevar. En momentos que pensé que sería imposible alcanzar esta meta siempre logre la fortaleza de mi querido padre y agradezco por cada una de las bendiciones que ha reflejado en mi vida.

Agradezco al Dr. Dennis Cáceres por ser un excelente asesor y brindarme las herramientas necesarias para poder presentar este trabajo académico.

Agradezco al Dr. German Moncada por compartir sus conocimientos y orientarme en el desarrollo de este estudio siendo sus orientaciones un pilar fundamental en este proceso.

Agradezco a la Msc. Lina Mendoza quien ha desempeñado una labor invaluable en el direccionamiento académico de la maestría y ha contribuido a que pueda alcanzar esta meta profesional.

Finalmente, agradezco a mis compañeros de maestría por su apoyo, ánimo y por su amistad que han sido una guía que me ha ayudado a superar obstáculos que se me presentaron.

Contenido

Dedicatoria.....	i
Agradecimiento.....	ii
Contenido.....	iii
Índice de Tablas.....	vi
Índice de Figuras.....	vii
Acrónimos.....	ix
Introducción.....	1
Capítulo 1. Planteamiento del problema.....	2
1.1 Situación Problema.....	2
1.2. Objetivos.....	5
1.2.1 General.....	5
1.2.2. Específicos.....	5
1.3. Justificación.....	6
Capítulo 2. Marco Teórico.....	8
2.1. Bases Conceptuales y definiciones Básicas.....	8
2.1.1. Psicometría.....	8
2.1.2. Evaluación Estandarizada de los aprendizajes.....	9
2.1.3. Prueba de Admisión o conocimientos.....	10
2.2. Áreas Básicas del Currículo.....	11
2.2.1. Área de las Matemáticas.....	11
2.2.2. Área de Español.....	14
2.2.3. Área de las Ciencias Sociales.....	16
2.2.4. Área de las Ciencias Naturales.....	18
2.3. Teorías de construcción y análisis de los Test.....	20

2.4. Teoría Clásica de los Test (TCT).....	21
2.4.1. Características Generales.....	22
2.4.2. Supuestos	23
2.4.3. Dificultad	25
2.4.4. Discriminación.....	27
2.4.5. Confiabilidad	30
2.5. Teoría de Respuesta al Ítem (TRI).....	33
2.5.1. Características Generales.....	34
2.5.2. Supuestos	36
2.5.3. Modelos	37
2.6. Análisis de Dimensionalidad	38
2.6.1. Características Generales.....	38
2.6.2. Supuestos	39
2.6.3. Etapas del Análisis Factorial Exploratorio	39
Capítulo 3. Marco Contextual.....	43
3.1. Evaluación Estandarizada de los Aprendizajes	43
3.1.1. Procesos en América Latina	43
3.1.2. Procesos en Honduras.....	45
3.2. Procesos de Admisión en los sistemas educativos en la región.....	49
3.3. Oferta formativa de Educación Media.....	50
Capítulo 4. Metodología	53
4.1. Tipo de investigación.....	53
4.2. Población y muestra.....	54
4.2.1. Participantes.....	54
4.3. Variables	55

4.4. Instrumento de Medición.....	56
4.5. Plan de Análisis	57
4.6. Procedimientos	58
4.6.1. Logísticos.....	58
4.6.2. Metodológicos.	58
4.6.3. Procedimientos éticos	59
Capítulo 5. Resultados de la Investigación.....	60
5.1. Análisis de parámetros psicométricos desde la TCT.....	60
5.2. Análisis de Dimensionalidad (Análisis Factorial Exploratorio).....	68
5.3. Cálculo de parámetros psicométricos con la TRI.	71
Capítulo 6. Discusión de los Resultados.....	77
6.1. Limitaciones	81
6.2. Conclusiones.....	82
6.3. Recomendaciones generales para estudios futuros.....	83
Referencias.....	84
Declaración ética de la Investigación	90
Apéndice	91

Índice de Tablas

Tabla 1.	Formulación del modelo de la Teoría Clásica de los Test (TCT)	23
Tabla 2.	Nivel de dificultad y evaluación de los reactivos	27
Tabla 3.	Valores e interpretaciones del Índice de Discriminación (ID)	28
Tabla 4.	Parámetros del ítem en los modelos de TRI	37
Tabla 5.	Programas y valuaciones Internacionales	44
Tabla 6.	Principales áreas de trabajo del Proyecto MIDEH	46
Tabla 7.	Estructura Curricular de la Educación Media en Honduras	51
Tabla 8.	Número de aspirantes que contestaron la prueba de admisión	54
Tabla 9.	Matriz de Operacionalización de las Variables	55
Tabla 10.	Número de preguntas por área temática	56
Tabla 11.	Coeficientes de Confiabilidad (Forma A y B) por subgrupo temático.....	61
Tabla 12.	Resúmenes estadísticos KMO y Prueba de Esfericidad de Barlett	69
Tabla 13.	Ítems mejor evaluados de la prueba de admisión	75
Tabla 14.	Confiabilidad de los ítems utilizando Alpha de Cronbach y Omega de McDonald	93
Tabla 15.	Dificultad y Discriminación de los ítems de Ciencias Sociales con TCT.....	95
Tabla 16.	Dificultad y Discriminación de los ítems de Matemáticas con TCT.....	95
Tabla 17.	Dificultad y Discriminación de los ítems de español con TCT.....	96
Tabla 18.	Dificultad y Discriminación de los ítems de Ciencias Naturales con TCT....	96
Tabla 19.	Carga Factorial rotada de los ítems de la Forma A	98

Tabla 20.	Carga Factorial rotada de los ítems de la Forma A eliminando ítems	100
Tabla 21.	Carga Factorial rotada de los ítems de la Forma B	101
Tabla 22.	Carga Factorial rotada de los ítems de la Forma B eliminando ítems	103
Tabla 23.	Índice de Dificultad basado en la TRI	104

Índice de Figuras

Figura 1.	Tipos de momentos didácticos esenciales de la organización matemática	14
Figura 2.	Curva Característica de un ítem	35
Figura 3.	Etapas del Plan de Análisis de datos	58
Figura 4.	Índice de Dificultad de ítems de Matemáticas (Forma A y B) con TCT.....	62
Figura 5.	Índice de Dificultad de los ítems de Español (Forma A y B) con TCT.....	63
Figura 6.	Índice de Dificultad Ciencias Sociales (Forma A y B) con TCT.....	64
Figura 7.	Índice de Dificultad Ciencias Naturales (Forma A y B) con TCT.....	64
Figura 8.	Índice de Discriminación de ítems de Matemáticas (Forma A y B) con TCT...	66
Figura 9.	Índice de Discriminación de los ítems español (Forma A y B) con TCT.....	66
Figura 10.	Índice de Discriminación de los ítems de SS.CC. (Forma A y B) con TCT.....	67
Figura 11.	Índice de Discriminación de los ítems de CC.NN. (Forma A y B) con TCT.....	68
Figura 12.	Gráficos de Análisis Paralelo de la Forma A y Forma B	70
Figura 13.	Gráfico de Dificultad aplicando TRI a ítems de Matemáticas y Español.....	72
Figura 14.	Gráfico de Dificultad aplicando TRI a ítems de CC. SS y CC. NN	73
Figura 15.	Curva Característica del Item19	74
Figura 16.	Curva Característica del Item20	74

Figura 17. Curva Característica del Item9	74
Figura 18. Curva Característica del Item 15	74
Figura 19. Porcentaje de resultados obtenidos por estudiantes en prueba de admisión	91
Figura 20. Porcentaje de resultados obtenidos en prueba de admisión Forma A	91
Figura 21. Porcentaje de resultados obtenidos en prueba de admisión Forma B	92
Figura 22. Matriz de Correlación inicial y final de la Forma A	97
Figura 23. Matriz de Correlación inicial y final de la Forma B	97

Acrónimos

AFE	Análisis Factorial Exploratorio
AP	Análisis Paralelo
CENEVAL	Centro Nacional de Evaluación para la Educación Superior
CNB	Currículo Nacional Básico
CONAEVA	Comisión Nacional para la Evaluación de la Educación Superior
IEA	International Association for the Evaluation of Educational Achievement
KMO	Índice de Medida de la adecuación muestral Kaiser-Meyer-Oikin.
MCO	Mínimos Cuadrados Ordinarios
MIDEH	Proyecto Mejorando el Impacto al Desempeño Estudiantil de Honduras
OCDE	Organización para la Cooperación y Desarrollo Económico
PIRLS	Estudio Internacional de Progreso en Comprensión Lectora.
PISA	Programa para la Evaluación Internacional de Alumnos (siglas en ingles)
SEDUC	Secretaría de educación
TCT	Teoría Clásica de los Test
TIMSS	Estudio Internacional de Tendencias en Matemáticas y Ciencias.
TRI	Teoría de Respuesta al Ítem
UMCE	Unidad de Medición de la Calidad de la Educación
UNAH	Universidad Nacional Autónoma de Honduras
UNESCO	Organización de Naciones Unidas para la Cultura, las Ciencias y la Educación
UPNFM	Universidad Pedagógica Nacional Francisco Morazán
USAID	Agencia de Estados Unidos para el Desarrollo Internacional (siglas en ingles)

Introducción

Es muy importante considerar que más allá de la evaluación de alumnado para procesos de selección o certificación, la estandarización ha llegado a instaurarse en diversos procesos de evaluación como la evaluación de docentes, instituciones y sistemas educativos y es frecuente observar que planes de evaluación toman como eje del análisis evaluativo el uso de pruebas estandarizadas de desempeño o logros educativos (Jornet Meliá, 2017).

Este estudio de investigación enmarca su proceder en el análisis de una de las pruebas de conocimientos para la selección de estudiantes que se aplicó en el marco de uno de los procesos más importantes del sistema educativo como ser el acceso a la carrera de magisterio en las escuelas normales de Honduras.

El capítulo 1 de este documento da a conocer los elementos en los que se fundamenta el problema que da origen al estudio y señala los objetivos que se desean cumplir. El capítulo 2 por su parte desarrolla toda una unidad teórica que sustenta el estudio, las cuales han sido consultadas a través de fuentes académicas y científicas que abordan desde las Teorías de construcción y análisis de los Test hasta elementos como la evaluación de aprendizajes desde las áreas de matemáticas, español, ciencias sociales y ciencias naturales. El capítulo 3 brinda un panorama contextual acerca de la Evaluación Estandarizada en la región, el Sistema Educativo en Honduras, proceso de transformación de las Escuelas Normales y los efectos de la pandemia del COVID-19 en educación.

En el capítulo 4 se describen los procedimientos metodológicos empleados para el análisis, así mismo se mencionan los elementos más importantes relacionados con la muestra y el instrumento utilizado en el estudio. El capítulo 5 detalla de manera puntual los resultados obtenidos a través de los análisis psicométricos en tanto que en el capítulo 6 da a conocer los principales planteamientos en cuando a la discusión de los resultados y principales hallazgos determinados en el estudio.

Capítulo 1. Planteamiento del problema

En el desarrollo del presente capítulo se dan a conocer los argumentos que sustentan la necesidad, conveniencia e importancia del Análisis Psicométrico de las áreas básicas de la prueba de Admisión a Escuelas Normales aplicado en el año 2011. Este capítulo menciona la debilidad del sistema educativo en el impulso de la Psicometría como área del conocimiento que puede contribuir a mejorar la calidad educativa en sus diferentes niveles. Este espacio también resalta la importancia de evidenciar que a pesar de que las Escuelas Normales dejaron de funcionar en el país, el examen aplicado a los aspirantes a la Carrera de Magisterio es un instrumento que fue elaborado y aplicado pero que requiere de un estudio que evidencie y respalde las cualidades psicométricas del mismo y pueda ser utilizado para evaluar las áreas básicas del conocimiento con estudiantes que aspiran a otras carreras de educación media en el país.

1.1 Situación Problema

Históricamente desde el año 1875 fueron las Escuelas Normales los centros que se ocuparon de la formación de docentes en Honduras hasta el año 2015 cuando entra en vigencia la nueva Ley Fundamental de Educación, que es cuando se comienza el proceso de Transformación de las Escuelas Normales y estas dejan de funcionar como tal. Durante los últimos años de su existencia, “las escuelas normales como centros de formación de maestros, para su ingreso aplicaron una prueba de admisión como principal requisito de ingreso, el cual evaluaba conocimientos básicos generales, vocación profesional y habilidad cognitiva” (Viquez, 2016, p. 214).

En la misma línea de aplicación de exámenes o pruebas de aptitud para ingresar a estudiar, en el nivel de educación superior la Universidad Nacional Autónoma de Honduras (UNAH) y Universidad Pedagógica Nacional Francisco Morazán (UPNFM) en los últimos años se han caracterizado por ser instituciones que aplican un proceso de admisión para que sus postulantes puedan ingresar a las diversas carreras que ofertan, ambas instituciones “en 1989 con la reforma de educación Superior se les designa la responsabilidad exclusiva de formación de docentes a nivel universitario” (Viquez, 2016, p. 220).

Bon y Pini (2019) mencionan que si bien las escuelas normales ya no existen como instituciones formadoras de docentes, los procesos de admisión si siguen marcando una brecha importante en la identificación de resultados alcanzados por los estudiantes en áreas como matemáticas y español; prueba de ello es que en el estudio realizado por la UPNFM denominado “Las Claves de la desigualdad educativa en Honduras”, se indica que parte de las desigualdades en el logro educativo asociadas a las desigualdades sociales se reflejan en la Prueba de Aptitud Académica (PAA) que aplica la UNAH y que mide aptitudes relacionadas con el idioma español y las Matemáticas, uno de los datos relevantes es que en la última década se han evaluado casi 300,000 estudiantes de los cuales aproximadamente 3 de cada 4 son admitidos.

Al momento de ingresar al nivel de educación superior se reflejan datos que indican como se encuentran los estudiantes en áreas básicas. El rendimiento académico de los estudiantes en el nivel de educación básica indica que las evaluaciones realizadas entre los años 2007-2016 de primero a noveno grado en español y Matemáticas dejan claras dos tendencias: la primera es que los resultados en cuanto al acceso a la educación secundaria son bastante bajos, alrededor del 60% del total de estudiantes evaluados se encuentran clasificados en los niveles de desempeño* “Debe Mejorar” e “Insatisfactorio”; y en segunda instancia es que los resultados muestran una tendencia ligera pero de continua mejora hasta el año 2014, a partir de ese año los datos se mantienen estancados (Alas Solís, 2017).

La evaluación de los aprendizajes se traduce en el “proceso que posibilita determinar los modos y actuación en que los estudiantes se incorporan en sus procesos formativos”; es por ello que tanto los exámenes de admisión aplicados para el ingreso a las Escuelas Normales como la Prueba de Aptitud o de admisión que se aplica en la UNAH, UPNFM y otros instrumentos de evaluación facilitan el reconocimiento de resultados académicos que en Honduras no son muy alentadores pero que posibilitan argumentos en la toma de decisiones para mejorar las falencias del sistema educativo (Díaz y Leyva, 2013, p. 270).

La evaluación es un proceso que permite obtener datos para fortalecer los aspectos que sean necesarios (Flores, 2017), y la psicometría se consideran como un espacio de trabajo metodológico en torno al desarrollo y administración de test (Meneses et al., 2013, p. 32). Ambas son complementarias y de vital importancia para la mejora continua de la calidad

educativa. Sin embargo, en Honduras no se cuenta con una amplia trayectoria que permita evidenciar de manera científica los aportes en esta área, han sido pocas las instituciones que se han preocupado por desarrollar instrumentos que cumplan con criterios psicométricos válidos, entre ellas se encuentra el Proyecto Mejorando el Impacto al Desempeño Estudiantil de Honduras (MIDEH), UPNFM, UNAH y algunas unidades de la secretaria de educación.

En la actualidad las Escuelas Normales ya no funcionan, y los instrumentos que fueron utilizados para direccionar el proceso de admisión y evaluación de aprendizajes en las mismas ha quedado en el olvido, sin embargo, al revisar su estructura y contenido siguen teniendo vigencia, ya que los logros y cambios en educación no han sido tan significativos. Es más (Hervis, 2018, p. 2) menciona que, a causa de la pandemia del Coronavirus SARS-CoV-2, se ha retrocedido en varios aspectos, el año 2020 marca un antes y un después en la dinámica mundial, regional y nacional, y esto asociado al ámbito educativo contribuye al recrudecimiento de problemas ya existentes como la insatisfacción y la crítica a los sistemas educativos por la débil pertinencia expresada en la inoperancia de la formación que se brinda en las instituciones educativas, las necesidades que señala la vida social, el desarrollo de la ciencia, la tecnología y el propio ser humano en sus actividades vitales.

Con este estudio de investigación se pretende realizar un análisis de las propiedades psicométricas de la Prueba de Admisión aplicada en las Escuelas Normales en el año 2011, para utilizar los análisis como respaldo y evidencia de que se cuenta con un instrumento que cumple o no con las propiedades psicométricas básicas, y verificar si este puede ser utilizado para valorar los conocimientos, capacidades y competencias de estudiantes que deseen ingresar a los bachilleratos que se ofertan en el nivel de educación media. Siendo este un instrumento que ya no es de admisión, si no de diagnóstico en las áreas básicas que evalúa tomando en consideración las peculiaridades, características y dinámica que identifican al perfil académico en este nivel de enseñanza donde se comparten conocimientos similares en áreas básicas como matemáticas, español, ciencias naturales y ciencias sociales que ya evalúa esta prueba de admisión.

1.2. Objetivos

1.2.1 General

Analizar los parámetros psicométricos de las áreas básicas de la Prueba de Admisión a Escuelas Normales aplicada a estudiantes hondureños aspirantes en el año 2011.

1.2.2. Específicos

Determinar el Nivel de Confiabilidad de los ítems que forman parte de la prueba por medio de la Teoría Clásica de los Test (TCT).

Calcular los Índices de Dificultad de los ítems que forman parte de la Prueba mediante la TCT.

Calcular los Índices de Discriminación de los ítems que forman parte de la Prueba mediante la TCT.

Evaluar la Dimensionalidad de la Prueba por medio de Análisis Factorial Exploratorio (AFE).

Calcular los Índices de Dificultad de los ítems que forman parte de la Prueba mediante la TRI.

1.3. Justificación

En este apartado se dan a conocer los principales elementos que se valoran como aporte de este estudio de Análisis Psicométrico de las áreas básicas de la Prueba de Admisión a Escuelas Normales aplicado en el año 2011, considerando su utilidad dentro del área para el fortalecimiento del sistema de evaluación, contribuyendo de esta manera en la mejora continua de la calidad educativa, y reforzando de manera científica el uso de instrumentos de evaluación que cumplan con los parámetros psicométricos básicos.

La utilidad de este estudio radica en que se conocerán las estimaciones de las propiedades psicométricas de un instrumento que fue elaborado con fines de valoración e identificación de competencias básicas de estudiantes de magisterio, pero que al evidenciar sus resultados puede ser utilizado para conocer el nivel académico de estudiantes que aspiran a ingresar a otras carreras en el nivel de educación media, ya que las áreas que evalúa corresponden a conocimientos generales de educación básica en matemáticas, español, ciencias naturales y ciencias sociales.

Es destacable mencionar que la Fundación para la Educación Ricardo Ernesto Maduro Andreu (FEREMA, 2017) menciona que aun cuando las pruebas o exámenes de evaluación pueden ser mejorados tanto en su construcción como en los procesos de aplicación, este estudio pretende, a nivel social, contribuir a resaltar que es un logro de gran importancia que alumnos, docentes, padres de familia y comunidad en general, perciban la evaluación como un proceso “normal” y conveniente para saber cómo está funcionando el sistema; ya que la elaboración y aplicación de instrumentos prácticos y con criterios validos contribuirá a manejar un concepto más claro, no coercitivo ni negativo, de lo que es la evaluación.

Este estudio es relevante, ya que comenzará a formar parte de una serie de estudios que puedan darse a conocer en función de mostrar las características básicas con las que debe contar un instrumento de evaluación, ya que “un instrumento o prueba de admisión juega un rol importante en una institución, son instrumentos útiles para realizar procesos confiables, eficientes, transparentes” (Backhoff, 2018), siempre y cuando estén bien elaborados y cumplan patrones psicométricos básicos.

En el aspecto teórico y metodológico, este estudio contribuirá con el desarrollo de procedimientos claves para conocer las propiedades psicométricas de un instrumento de evaluación de conocimientos, y con ello fortalecer el campo del conocimiento en cuanto a la psicometría, definida como el área que aglutina las bases y los modelos para medir fenómenos psicológicos y educativos de forma adecuada (Martínez Arias et al., 2006).

El sistema educativo hondureño se encuentra en una etapa inicial en cuanto a la creación y ejecución de procesos de evaluación basados en la psicometría, por ello es de gran importancia el aspecto metodológico, ya que aporta al conocimiento existente sobre el uso de nuevos instrumentos de evaluación con criterios psicométricos válidos y aceptables, cuyos resultados son proporcionados a partir de instrumentos creados con una fundamentación conceptual, actitudinal y procedimental de carácter científico.

Capítulo 2. Marco Teórico

En este capítulo se describe las fundamentaciones teóricas, que son el soporte del estudio de Análisis Psicométrico de las áreas básicas de la Prueba de Admisión a Escuelas Normales aplicado en el año 2011. Cada enunciado que se desarrolla en este capítulo resalta la información más importante que ha sido consultada en la literatura científico-académica. Se describen cada una de las bases conceptuales y definiciones básicas de la psicometría y la evaluación, seguidamente se abordan los elementos básicos de las teorías de construcción y análisis de Test como lo son la Teoría Clásica de los Test y Teoría de Respuesta al Ítem. Finalmente se menciona y describe la teoría que sustenta los procedimientos teóricos implementados en el estudio: confiabilidad, dificultad, discriminación y dimensionalidad.

2.1. Bases Conceptuales y definiciones Básicas

La bases conceptuales y definiciones básicas que se desarrollan a continuación tienen como finalidad conocer la concepción académica de los principales términos que acompañan el desarrollo de procedimientos, para calcular los parámetros psicométricos de instrumentos de evaluación.

2.1.1. Psicometría

La psicometría es la definición que recoge un conjunto de métodos, técnicas y teorías que implica la medición de variables psicológicas; así mismo hay autores que limitan la temática psicométrica en cinco grandes bloques: 1) Teoría de la medición, incluye lo relativo a la fundamentación teórica de la medida; 2) Teoría de los test, explicita la lógica y modelos matemáticos subyacentes a la construcción y uso de test; 3) Escalamiento psicológico, aborda la problemática inherente al escalamiento de estímulos psicológicos; 4) Escalamiento psicofísico, hace lo propio con estímulos físico y 5) Técnicas multivariadas, junto con el resto de tecnología estadística resultan imprescindibles para la construcción y análisis de instrumentos de medida (Muñiz Fernández, 2018).

La psicometría también se define como: “El área de la psicología que se ocupa de los procedimientos de medición del comportamiento humano, incluyendo los test psicométricos” (Ramos Vega, 2018). Es así como la psicometría dentro del campo científico guía su

funcionamiento para dar respuesta a tres preguntas: 1) ¿Qué es lo que se debe de medir?, 2) ¿Cómo medirlo? y finalmente 3) ¿Con que medirlo? Esto permite que el desarrollo de esta rama de la psicología tenga un auge y progreso muy importante.

“El objeto de medición de la psicometría es el atributo psicológico y no las personas” (Muñiz Fernández, 2018). Teniendo claro esto, es importante mencionar que nos referimos a un constructo teórico, al sistema de proposiciones que se relacionan entre si formalmente, y se relacionan con los fenómenos observables por medio de reglas de correspondencia entre la teoría y la conducta (Ramos Vega, 2018).

Finalmente, González Llanea, (2007) menciona que la Psicometría también garantiza que los instrumentos de medida sean debidamente estandarizados, y posean la validez requerida que haga posible dichas medidas y posibiliten la diferenciación de unas personas en relación a otras en determinada población

2.1.2. Evaluación Estandarizada de los aprendizajes

Para comenzar a indagar sobre evaluación estandarizada de los aprendizajes es importante definir qué es la evaluación en su conjunto, y de esta manera poder digerir mucho más amigablemente el mundo de la psicometría, y es que la “evaluación es un proceso en el cual se determina hasta qué punto los objetivos educativos han sido actualmente alcanzados, mediante los programas y currículos de enseñanza” (Castillo y Cabrerizo, 2010).

Desde el punto de vista normativo, en Honduras el Reglamento de Evaluación Docente se deriva de la Ley Fundamental de Educación, que en su Artículo 9, hace referencia a los términos a emplear para efectos de comprensión y aplicación de este definiendo la evaluación como:

“Una práctica fundamental que se desarrolla en el Sistema Nacional de Educación como un medio que garantiza la calidad de los servicios que ofrece. Es un proceso dinámico, permanente, sistemático, continuo, flexible, científico, participativo, integral e inherente al quehacer educativo. Permite formular juicios de valor para la toma de decisiones con el propósito de mantener o cambiar el curso de la acción” (Reglamentos Ley Fundamental de Educación, 2014).

Adentrándonos un poco a ilustrar la evaluación estandarizada de los aprendizajes, este es un enfoque de la evaluación que concibe el proceso que puede y debe potenciar el aprendizaje del educando, pero que para hacerlo requiere que se cumplan condiciones que la evaluación sea una experiencia de aprendizaje (Moreno Olivos, 2016). Conociendo el objetivo que persigue este enfoque es innegable que toma en consideración la rendición de cuentas del aprendizaje del alumno y las pruebas de alto impacto para comprobar la calidad de la escuela.

De manera general, Pimienta Prieto, (2008) menciona que se puede precisar que el aprendizaje es como “el conjunto de productos obtenidos por los estudiantes como resultado de la incidencia de la educación, siendo estos productos mediatos o inmediatos”; e incluso algunos de ellos no se manifiestan como conductas observables. Para ser más específicos los productos inmediatos de la educación, son aquellos resultados de la intervención educativa de las escuelas, como los conocimientos declarativos, procedimentales y actitudinales.

La dificultad de llevar el concepto de estandarización a programas o planes de evaluación radica en las características del objeto de evaluación y en el concepto de calidad que quiera comprobarse; con lo expresado anteriormente podemos afirmar que diseñar instrumentos de evaluación estandarizada es un poco complejo. Sin embargo, la estandarización ha llegado a instaurarse en diversos procesos de evaluación con docentes, instituciones y sistemas educativos (Jornet Meliá, 2017).

2.1.3. Prueba de Admisión o conocimientos

La clasificación de pruebas se da de acuerdo con su contenido o proceso, por lo tanto, hay diversos tipos de pruebas o exámenes entre ellos los cognoscitivos. Este tipo de prueba intenta cuantificar procesos y productos de la actividad mental y pueden clasificarse como mediciones de rendimiento y aprovechamiento. “Una prueba de rendimiento evalúa el conocimiento de algún tema u ocupación académica y se centra en el comportamiento pasado del examinando en lo que ya ha aprendido o logrado” (Aiken, 2003).

Un test, prueba o examen de conocimientos es un instrumento compuesto por una serie de ítems (preguntas, tareas, problemas entre otros), diseñados para evaluar conocimientos o aptitudes en el ámbito psicológico y educativo. La prueba es entonces un instrumento cuya función radica en proporcionar información sobre características de un candidato, como la

amplitud de conocimientos, de forma tal que las características puedan medirse (Doval & Pinsach, 2019).

2.2. Áreas Básicas del Currículo.

“La evaluación a lo largo de la historia, ha tenido diversas corrientes, teorías o posturas epistemológicas siendo algunos de los autores relevantes en este hilo histórico: Comenio, Comte, Durkheim, Wundt, Foucault, House, Gimeno Sacristán, Santos Guerra, J. I. Pozo, Jesús Palacios, Casanova, Hernández Rojas entre otros, esta riqueza histórica identifica dos períodos; 1) las posturas epistemológicas definidas y que impactan en cada época y 2) teorías de diferentes campos del conocimiento como el psicológico y pedagógico, tomando en cuenta también a la sociología y la filosofía” (Monzón Troncoso, 2015).

2.2.1. Área de las Matemáticas.

Currículo Nacional Básico (CNB). Según el Currículo Nacional del Tercer Ciclo de Educación Básica la matemática es una disciplina que sistematiza la capacidad intuitiva del ser humano de poder encontrar ideas necesarias para resolver problemas; el conocimiento matemático es un conocimiento intuitivo que precisa de la demostración para poder ser explicado y explicitado. La matemática es una disciplina vinculada al desarrollo de estructuras del pensamiento lógico, capacidad de abstracción a los procesos deductivos e inductivos y a la capacidad de síntesis y análisis (Secretaría de Educación, 2006).

La Secretaría de Educación (2006), identifica en el Currículo Nacional Básico (CNB) como Bloques del Campo del Conocimiento de Matemáticas los siguientes:

- Bloque 1: Geometría.
- Bloque 2: Números y Operaciones.
- Bloque 3: Medidas.
- Bloque 4: Estadística Descriptiva y Probabilidad Discreta.
- Bloque 5: Álgebra.

El CNB indica que para el área de matemáticas se desarrollan los siguientes procesos: a) proceso de resolución de problemas, b) proceso de comunicación, c) proceso de

razonamiento y d) proceso de conexión. Esto basándose en que la utilidad de la matemática se debe a que, a pesar de su abstracción, sus conceptos y resultados tienen su origen en el mundo real y encuentra muchas y diversas aplicaciones en otras ciencias.

Se deben de realizar aplicaciones permanentes en esta área, desde el desarrollo de saberes significativos como: números reales, estudio de figuras y formas, números fraccionarios y decimales, cálculo de longitudes, áreas y volúmenes, organización, extracción e interpretación de información estadística para promover en los alumnos una base de conocimientos que sirva de soporte en la promoción de un nivel a otro (Secretaría de Educación, 2006).

Marco de Matemáticas de PISA para el Desarrollo. Según este marco PISA evalúa la competencia matemática de los jóvenes de 15 años, valorando hasta qué punto los alumnos de dicha edad son capaces de manejar con destreza las matemáticas cuando se enfrentan a situaciones y problemas, la mayoría de los cuales están presentes en contextos del mundo real (OCDE, 2017).

En cuanto a competencia matemática PISA la define de la siguiente manera:

“Capacidad del individuo para formular, emplear e interpretar las matemáticas, incluye razonar y utilizar conceptos, procedimientos, herramientas y hechos para describir, explicar y predecir fenómenos, esto ayuda a emitir juicios y decisiones fundamentadas que necesitan los ciudadanos constructivos y reflexivos” (OCDE, 2017).

Estudio Internacional de Tendencias en Matemáticas y Ciencias (TIMSS) de la IEA. Sobre el rendimiento en Matemáticas y Ciencias TIMSS utiliza el currículo como principal concepto organizador, ya que tiene en cuenta oportunidades educativas ofrecidas a los alumnos, así como factores que influyen en cómo se emplean. El modelo curricular de TIMSS tiene tres aspectos: el currículo pretendido, el aplicado y el obtenido; estos aspectos representan, las Matemáticas y las Ciencias que la sociedad pretende que aprendan los alumnos y cómo debería organizarse el sistema educativo para facilitar este aprendizaje (Mullis, Martín, et al., 2012).

En el marco teórico para las especificaciones de evaluación de Matemáticas de TIMSS elaborado por Mullis, Martín, et al., (2012) se menciona que en educación primaria y secundaria este marco teórico se organiza en torno a dos dimensiones, una de contenido especificando dominios o asignaturas que han de evaluarse dentro de Matemáticas (números, álgebra, geometría, y datos y probabilidades) y otra dimensión cognitiva, especificando dominios de procesos de pensamiento a evaluar (conocer, aplicar y razonar).

TIMSS hace énfasis en que el dominio de los números incluye la comprensión del valor posicional de las cifras, de las maneras de representar los números y de las relaciones entre los números; así mismo los estudiantes deben haber desarrollado sentido numérico, fluidez de cálculo, ser capaces de comprender los significados de las operaciones y cómo se relacionan entre sí, además de, ser capaces de usar números y operaciones (sumar, restar, multiplicar y dividir). Deben estar familiarizados con una serie de modelos de números, explorando las relaciones entre los números que están en el modelo o que se utilizan para deducirlo (Mullis, Martín, et al., 2012).

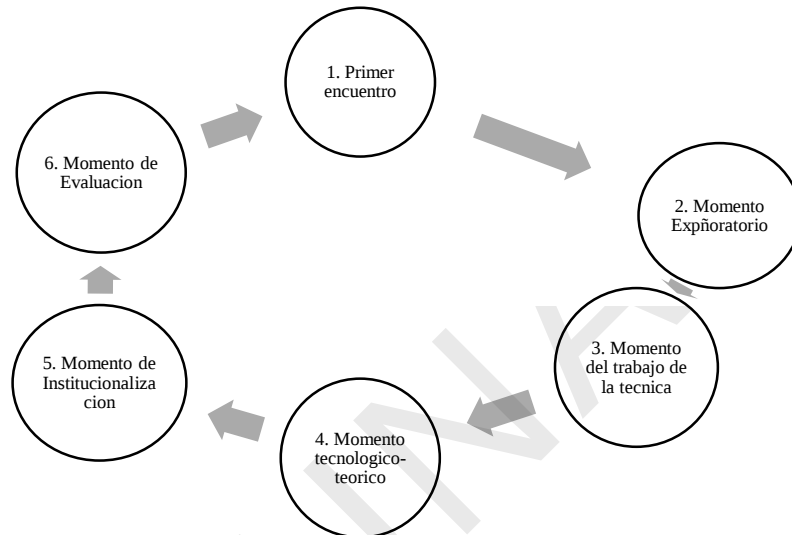
Teoría de los momentos didácticos. Propone un modelo del proceso de estudio de las matemáticas desde el punto de vista de la corriente antropológica. Dentro de este modelo, hacer matemáticas consiste en activar una organización matemática, resolver determinados tipos de problemas con determinados tipos de técnicas, de manera inteligible, justificada y razonada (Godino, 2010).

Finalmente enseñar y aprender matemáticas corresponde a la actividad de reconstruir organizaciones matemáticas para poderlas utilizar en nuevas situaciones y bajo distintas condiciones. La enseñanza o tarea docente consiste básicamente en dirigir dicha reconstrucción, mientras que el aprendizaje puede considerarse como el fruto de la reconstrucción, ya sea individual o en grupo.

El objetivo de un proceso de enseñanza - aprendizaje puede formularse en términos de los componentes de organizaciones matemáticas que se quieren reconstruir: qué tipos de problemas hay que ser capaz de resolver, con qué tipos de técnicas, sobre la base de qué elementos descriptivos y justificativos, en qué marco teórico, entre otros (Godino, 2010). Así mismo hay autores como (Barajas, 2010), que señalan que las Ciencias Matemáticas en el

marco del proceso enseñanza aprendizaje se enmarcan en el desarrollo de 6 momentos didácticos organizados y planificados tal como se observa en la figura 1 que se presenta a continuación.

Figura 1. Tipos de momentos didácticos esenciales de la organización matemática



Nota. Fuente: Elaboración propia, basada en información de (Barajas, 2010).

2.2.2. Área de Español

Currículo Nacional Básico (CNB). El CNB del Tercer Ciclo de la Educación Básica conceptualiza el lenguaje, en el Área de Español, como: “la capacidad del ser humano que le ofrece los recursos para abstraer, conceptualizar, representar e interpretar la realidad”, y es gracias a esta capacidad que el individuo construye, por una parte, un universo conceptual, por otra, distintos sistemas sígnicos (la lengua, la música, la pintura, la escritura, los gestos, entre otros) y la principal finalidad es posibilitar formas diversas de comunicación entre los seres humanos (Secretaría de Educación, 2006).

El CNB identifica como Bloques de Conocimiento en el área de español los siguientes:

- Bloque 1: Lengua oral
- Bloque 2: Lengua escrita (escritura y lectura)
- Bloque 3: Reflexión sobre la lengua
- Bloque 4: Creación y expresión literaria (Secretaría de Educación, 2006)

Marco de Español de PISA para el Desarrollo. PISA para el Desarrollo define la lectura como el reconocimiento de palabras y la comprensión lingüística y que ambas acciones son necesarias, pero no suficientes para leer; así mismo establece que la Competencia lectora es el proceso de “comprender, utilizar, reflexionar e interactuar” con los textos escritos para alcanzar los propios objetivos, desarrollar el conocimiento, potencial personal y participar en la sociedad (OCDE, 2017).

La OCDE hace énfasis en que existen otros componentes de lectura esenciales, incluyendo el reconocimiento de los elementos impresos del alfabeto, la descodificación de palabras a sonidos y la comprensión oral básica, estos no se incluyen en el marco de PISA-D, ya que se da por hecho que los jóvenes de 15 años han adquirido estas competencias si asisten al curso que les corresponde en la escuela, así como los jóvenes de 15 años no escolarizados que han adquirido unos niveles básicos de alfabetización (OCDE, 2017).

Estudio internacional de progreso en comprensión lectora (PIRLS). Este estudio evalúa la comprensión lectora de los alumnos, este momento es clave en el desarrollo de los niños como lectores, puesto que ya han aprendido a leer y están ya leyendo para aprender. PIRLS se basa en un concepto de la lectura que “consiste en la capacidad de reflexionar sobre textos escritos y utilizarlos como herramientas para lograr metas individuales y sociales, también conocido como «leer para actuar” (Montoya Gómez, 2016).

Según Gómez Viera, (2008) PIRLS desarrolla definiciones precisas: en el estudio hay presentes dos tipos de textos: informativos y literarios (narrativos); este estudio indica que cuando se leen textos informativos, el lector entra en contacto con el mundo real, lo que le permite conocerlo, tal cual es en el presente, cómo era en el pasado, y le permite también comprender el funcionamiento de la sociedad y las instituciones. Por otra parte, en el caso de los textos literarios, el lector entra en un mundo de personajes, ambientes y eventos ficticios, que le permiten desarrollar su imaginación y mejorar su dominio de la lengua.

En el Informe en español de PIRLS 2016: Estudio Internacional de progreso en comprensión lectora de la IEA se define la comprensión lectora como esa habilidad para comprender y utilizar las formas lingüísticas requeridas y/o valoradas por el individuo. Los lectores son capaces de construir significado a partir de una variedad de textos. leen para

aprender, para participar en las comunidades de lectores del ámbito escolar y de la vida cotidiana, y para su disfrute personal (Montoya Gómez, 2016).

El Estudio internacional de progreso en comprensión lectora (PIRLS) trabaja con cuatro competencias: 1) Comprender: supone procesar la información en forma automática e inmediata; 2) Inferir: el lector debe construir sentido estableciendo relaciones entre la información explícita y su sentido no dicho, este es el primer paso para acceder a la estructura profunda del texto; 3) Interpretar: la comprensión necesaria va más allá del nivel superficial, el lector debe relacionar las diferentes informaciones para construir un sentido global que se refiera a un párrafo, varios, o a todo el texto; 4) Evaluar: un lector capaz de desarrollar todas las tareas antes descritas, estará también en condiciones de juzgar el texto, tanto la información que contiene como su estructura, así como establecer comparaciones con otros textos y con su propia experiencia (Gómez Viera, 2008).

2.2.3. Área de las Ciencias Sociales

Currículo Nacional Básico. El CNB del Tercer Ciclo de la Educación Básica indica que las Ciencias Sociales tienen como sujeto de estudio al hombre viviendo en sociedad, actor de procesos que se desarrollan en un contexto Tiempo y Espacio determinado, dicho esto se entiende que el sujeto de estudio es complejo en sí mismo, pues las múltiples interrelaciones que se establecen entre él y su entorno natural y social obligan a pensar en formas de abordaje que articulen los conocimientos que conforman el área, estableciendo entre ellos relaciones significativas (Secretaría de Educación, 2006).

En las Ciencias Sociales se integran disciplinas como la Geografía y la Historia que constituyen su columna vertebral, en tanto espacio y tiempo componen dos categorías básicas para pensar la vida del hombre en sociedad, ambas junto al aporte de la Sociología, Antropología, Economía, Filosofía y Ciencias Políticas que permiten conocer, comprender y organizar las complejas relaciones que se dan entre el espacio natural y la cada vez más dinámica actividad antrópica denominada “cultura” (Secretaría de Educación, 2006).

En esta disciplina la Formación Ética y Ciudadana son fundamentales para preparar al hondureño de tal modo que se pueda desenvolver con éxito en un mundo globalizado que exige de cada persona conocimientos, habilidades, actitudes y valores para ser ciudadanos

productivos, creativos y críticos, capaces de transformar la realidad (Secretaría de Educación, 2006). El CNB identifica como Bloques de Conocimiento en el área de Ciencias Sociales los siguientes:

- Bloque 1: la persona y su ser social
- Bloque 2: las sociedades y los espacios geográficos
- Bloque 3: las sociedades y el tiempo social
- Bloque 4: las sociedades organizadas y las actividades humanas.

Según Tapia, (1997) la evaluación se realiza en el contexto de un planteamiento curricular que define los objetivos que los alumnos han de conseguir trabajando los contenidos y por otra parte ayudar al alumno, calificar su progreso o ambas cosas. En el caso de las Ciencias Sociales, Geografía e Historia, se persigue la construcción de un conjunto de representaciones sobre las sociedades humanas y el medio físico en que viven, sobre las relaciones que se establecen entre ellos, el modo en que cambian a lo largo del tiempo y las causas y consecuencias de tales cambios.

La evaluación en ciencias sociales se efectúa en forma fraccionada y se basa principalmente en procesos memorísticos; estas circunstancias han limitado el impacto que el área debe jugar en la formación del ser humano; otra característica es la retención de hechos datos y fechas o la elaboración de resúmenes y mapas; lo cual no permite correlacionar hechos históricos ni situaciones geofísicas con realidades contemporáneas negando la posibilidad de que los niños y jóvenes, realicen análisis que viabilicen la comprensión de las circunstancias locales, distritales, nacionales y mundiales (Tarazona y Zapata, 2017).

El propósito primario de la evaluación debe ser facilitar a los alumnos las ayudas que pueden permitirles progresar, los profesores deben evaluar, por un lado, el grado en que los alumnos van construyendo, a partir de los contenidos trabajados en clase, representaciones de la realidad que constituye el mundo social, sus relaciones con el medio físico, la forma en que interactúan y el modo en que cambian en el tiempo (Tapia, 1997).

De acuerdo con Munar y Riveros, (2016), la mayor dificultad que se encuentra al evaluar los aprendizajes en ciencias sociales es que se evalúa para evaluar: para calificar y nada más; no hace parte de un proceso que construya oportunidades de promoción de las y los

estudiantes como personas independientes y autónomas con posibilidad de generar cambios personales y sociales. La evaluación no es entendida como un proceso que se utiliza como punto de partida para avanzar, detenerse o retomar sino para sancionar o promover; no se utiliza para que las y los evaluados cambien su actitud o posición frente al conocimiento y al saber, sino que es el punto de llegada, el fin último de los aprendizajes

Tarazona y Zapata, (2017) hace énfasis en que las Ciencias Sociales se visualizan cómo la “comprensión del conocimiento y del mundo social que permite dar sentido y significado tanto a eventos y fenómenos sociales como al aprendizaje escolar”, lo cual representa, en los estudiantes, cambios cognitivos, procedimentales y afectivos en los aprendizajes.

El sentido del conocimiento en las ciencias sociales y su evaluación en la escuela entonces debe apuntar a esta aproximación conceptual: la idea del conocimiento social como herramienta para transformar el entorno y al sujeto implica entender la complejidad del conocimiento social atravesado por las diferentes disciplinas y su enseñanza por tópicos generadores o problemas como lo plantean los Lineamientos curriculares (Munar y Riveros, 2016)

2.2.4. Área de las Ciencias Naturales

Currículo Nacional Básico. Según el CNB de la Educación Básica, Ciencias Naturales es la disciplina que aborda el estudio de distintos elementos naturales del ambiente, sus relaciones y los fenómenos que resultan de dichas relaciones, en esta área del conocimiento convergen otras disciplinas; el estudio de los seres vivos se denomina biología, el estudio de los seres no vivos con la geología y la química, el estudio de las relaciones entre los diferentes elementos naturales del ambiente, vivos y no vivos abordado por la ecología y el estudio de los fenómenos físicos, derivados de las propiedades de la materia, es objeto de la física (Secretaría de Educación, 2006).

La diferenciación entre las ramas de las Ciencias Naturales se inicia a partir del tercer ciclo de educación básica, en el que se estudia una parte de conceptos y procesos desde el punto de vista de la física y de la química, sin que esto produzca una división del área en asignaturas. El CNB identifica como Bloques de Conocimiento en el área de Ciencias Naturales los siguientes:

- Bloque 1: Los seres vivos en su ambiente
- Bloque 2: El ser humano y la salud
- Bloque 3: La tierra y el universo
- Bloque 4: Materia, energía y tecnología (Secretaría de Educación, 2006).

Marco de Ciencias de PISA para el Desarrollo. De acuerdo a PISA para el Desarrollo OCDE, (2017), Competencia Científica es “la habilidad para interactuar con cuestiones relacionadas con la ciencia y con las ideas de la ciencia. Como un ciudadano reflexivo, una persona con conocimientos científicos está dispuesta a participar en un discurso razonado sobre ciencia y tecnología”, lo cual requiere competencias como las siguientes:

- Explicar fenómenos científicamente: reconocer, ofrecer y evaluar explicaciones para una serie de fenómenos naturales y tecnológicos.
- Interpretar datos y pruebas científicamente: analizar y evaluar datos, alegaciones y argumentos en una variedad de representaciones y sacar conclusiones científicas adecuadas.
- Evaluar y diseñar la investigación científica: describir y evaluar las investigaciones científicas y proponer formas de abordar científicamente las cuestiones (OCDE, 2017).

En evaluaciones de competencia científica de PISA 2015 y PISA-D solo se puede evaluar una muestra de contenidos del campo de la ciencia, se utilizan criterios para guiar la selección de conocimientos que se evalúan, los criterios se aplican a los conocimientos de los principales campos de física, química, biología, ciencias de la tierra y ciencias del espacio; estos requieren que el conocimiento tenga relevancia en situaciones de la vida real, representa un importante concepto científico o teoría explicativa que tenga utilidad duradera y sea apropiado al nivel de desarrollo del alumnado de 15 años (OCDE, 2017).

Estudio Internacional de Tendencias en Matemáticas y Ciencias (TIMSS) de la IEA. Según Mullis, Martín, et al., (2012) el marco teórico de evaluación de Ciencias para TIMSS consta de una dimensión de contenido, que especifica dominios de asignaturas que han de ser evaluadas dentro de la ciencia (Biología, Química, Física y Ciencias de la Tierra) y una

dimensión cognitiva que especifica dominios o capacidades y comportamientos (conocer, aplicar y razonar); la Física y Química se evalúan como dominios separados y reciben más importancia donde se evalúan como un dominio de contenido.

El marco teórico cognitivo de TIMSS incluye una gama de procesos que están implicados en el aprendizaje de conceptos de ciencias y de investigación científica en primaria y secundaria; en ciencias se tienen una curiosidad natural por el mundo, así que es apropiado comenzar a aprender elementos básicos de esta disciplina. Este conocimiento y comprensión deben cimentarse a lo largo de toda la escolarización, de manera que, se vea la necesidad de tomar decisiones, como pueden ser el tratamiento de enfermedades, el calentamiento global y las aplicaciones de la tecnología desde una sólida base científica (Mullis, Martín, et al., 2012).

2.3. Teorías de construcción y análisis de los Test

Las técnicas psicométricas quedan englobadas dentro del proceso de evaluación psicológica. El proceso evaluativo utiliza las pruebas psicométricas como instrumentos de medida con el objetivo de alcanzar una información más amplia del sujeto, que le permita al investigador, crear hipótesis de trabajo que orienten su proceder evaluativo y diagnóstico (González Llana, 2007). La tabla de especificación del test suele ser la estrategia más utilizada en el ámbito cognitivo para representar empíricamente el constructo o característica que se desea medir. (Navas et al., 2001).

Las teorías de la medición son los modelos o paradigmas de medida que la psicometría ha desarrollado de manera formal a partir del debate iniciado en los años cuarenta sobre la viabilidad de la medida de los fenómenos psicológicos (Meneses et al., 2013).

Para construir un test hay que cubrir cuatro fases: definición del objetivo del test, especificación del test, confección de los ítems y confección de la prueba. (Navas et al., 2001). Además de ello, la construcción de buenos instrumentos de medición requiere de dos etapas: la elaboración propiamente dicha y el análisis posterior de la misma a partir de los resultados observados una vez que fue administrada a un conjunto de sujetos (Ramos Vega, 2018).

Para poder explicar el objeto de estudio de la psicometría, determinar o establecer condiciones y propiedades numéricas de los atributos, se cuenta con Teorías como la Teoría Clásica de los Test (TCT) y la Teoría de Respuesta al Ítem (TRI). El problema central que aborda cualquier teoría de los test es precisamente la relación existente entre el nivel del sujeto en la variable inobservable que se desea estudiar y su puntuación observada en el test (Navas et al., 2001).

Las teorías de la medición son el fundamento conceptual que hace posible asociar números a los objetos que nos interesa medir (Manzi y San Martín, 2003).

La regla de asignación de números o puntuaciones a los ítems varía según sea el tipo de ítem considerado. La mayoría de las pruebas cognitivas y buena parte de los test no cognitivos utilizan ítems de elección o de respuesta cerrada (escalas de valoración), entre otras razones, por la simplicidad y economía de su procedimiento de puntuación. (Navas et al., 2001). Dentro del ámbito cognitivo hay dos tipos de ítems: de elección (emparejamiento, elección múltiple y binaria) y de construcción (de respuesta corta y extensa).

Finalmente, en los apartados subsiguientes se definen características básicas de cada teoría y se indica cómo procede cada una de ellas para dar respuesta a los dos problemas fundamentales a los que se enfrenta toda teoría de los test: 1) la cuantificación o estimación del error de medida y 2) la determinación o estimación de la característica de interés que supuestamente está midiendo el test (Navas et al., 2001).

2.4. Teoría Clásica de los Test (TCT)

Lo que proporciona una teoría de los test es precisamente un modelo, una representación de las puntuaciones de los test. Este modelo da cuenta de la información contenida en esas matrices de datos $N \times n$, a partir de: 1) la característica de interés que supuestamente está midiendo el test y 2) el error que se introduce en el proceso de medición (Navas et al., 2001).

La TCT fue expuesta por vez primera de forma completa en la obra de Guilford (1936) *Psychometric methods* (Navas et al., 2001). La síntesis por antonomasia de la teoría clásica de los test fue realizada por Gulliksen (1950) en su más que clásico libro *Theory of Mental*

Test, que constituye, sin duda, la exposición mejor estructurada del corpus clásico (Muñiz Fernández, 2018).

La TCT se basa en el modelo lineal propuesto por Spearman en el que las variables se relacionan de forma directa, la puntuación observable de una persona en un test es una función lineal de dos componentes, su puntaje verdadero (inobservable, el atributo en tanto tal) y el error de medición implícito en la prueba (Ramos Vega, 2018).

Desde esta teoría se estiman los posibles errores que se pueden inducir al usar un test para medir variables psicológicas con el objetivo de lograr la exactitud en la medida, es decir, que el resultado de la prueba de cuenta de manera muy cercana a lo que verdaderamente es el atributo (Ramos Vega, 2018).

2.4.1. Características Generales

Esta teoría de los tests no está interesada en el sistema de relaciones empíricas, sino que centra su atención en el análisis de las puntuaciones obtenidas para valorar los errores cometidos en el proceso de medida indirecta de los fenómenos psicológicos (Meneses et al., 2013). De acuerdo con esta teoría, responde a una relación lineal de dos componentes fundamentales: $X = V + e$.

En la TCT la función que relaciona la actuación observable en el test — X — y el nivel del sujeto en la variable inobservable — V — es una función lineal, ya que la relación establecida en la ecuación entre X y V es una relación lineal que puede ser descrita gráficamente por medio de una recta (Navas et al., 2001).

Muñiz Fernández, (2018) indica que, se asume que la puntuación empírica de una persona en un test (X) consta de dos componentes aditivos: uno, la «verdadera» puntuación de la persona en el test (V), y otro, el error (e) que inevitablemente va asociado a todo proceso de medición; es decir, según el modelo.

La teoría clásica desarrollará todo un conjunto de deducciones encaminadas a estimar la cuantía del error que afecta a las puntuaciones de los test (Muñiz Fernández, 2018). En la tabla 1, se muestra la formula que se designa para el Modelo de la Teoria Clasica y los cuatro supuestos que se derivan de esta importante teoria.

Tabla 1. Formulación del modelo de la Teoría Clásica de los Test (TCT)

Modelo	$X = V + e$
Supuestos	1. $E(E_i)=0$
	2. $\sigma_{(V_i E_i)}=0$
	3. $\sigma_{(E_i E_j)}=0$
	4. $\sigma_{(V_i E_j)}=0$

Nota. Fuente: Elaboración propia

Donde:

X = Puntuación empírica

V = Puntuación verdadera

e = Error de medida

E = Esperanza matemática de los errores de medida

2.4.2. Supuestos

Según Martínez Arias et al., (2006), los supuestos del modelo de la TCT se resumen en los siguientes:

1- El valor esperado de la variable aleatoria *error de medida* es igual a cero, para una población de personas medidas con el mismo test o para una repetición infinita de medidas realizadas sobre la misma persona.

$$E(E_i)=0$$

2- La falta de relación entre la puntuación verdadera y el error, ello implica que los sujetos con diferentes valores en el atributo se ven afectados por cantidades de error que no tienen que ver con su estatus en el atributo.

$$\sigma_{(V_i E_i)}=0$$

3- Si se aplican dos test distintos, a una población de sujetos, los errores que se cometan al aplicar el primer test no estarán correlacionados con los errores que se cometan al aplicar el segundo test.

$$\sigma_{(E_i E_j)}=0$$

4- Los errores de medida de un test tampoco están correlacionados con las puntuaciones verdaderas de otro test, ya que en otro caso no se trataría de un error aleatorio, sino de un error relacionado con alguna dimensión medida por segundo test.

$$\sigma_{(V_i E_j)} = 0$$

Estos supuestos son *a priori* perfectamente razonables y sensatos, pero *a posteriori* resulta que no se pueden comprobar empíricamente (recuerden que la puntuación verdadera y el error nunca se conocen). Lo único que se puede hacer es derivar de ellos deducciones y contrastar o someter a prueba éstas para ver si el modelo lineal se confirma o se ve falsado (Navas et al., 2001).

De acuerdo a Meneses et al., (2013), de este modelo con sus supuestos también se deriva un conjunto de deducciones inmediatas importantes que son las siguientes :

- $\mathbf{e} = \mathbf{X} - \mathbf{V}$. El error de medida sería la diferencia entre la puntuación empírica y la puntuación verdadera.

- $\mathbf{E}(\mathbf{e}) = \mathbf{0}$. La esperanza matemática de los errores de medida es cero. por lo que, si fuera posible administrar un test un número infinito de veces, estos errores aleatorios o no sesgados se compensarían o anularían entre ellos.

- $\mu_x = \mu_v$. La media de las puntuaciones empíricas (μ_x) es igual a la media de las puntuaciones verdaderas (μ_v).

- $\text{cov}(\mathbf{V}, \mathbf{e}) = \mathbf{0}$. Las puntuaciones verdaderas no covarían con los errores de medida.

- $\text{cov}(\mathbf{X}, \mathbf{V}) = \text{var}(\mathbf{V})$. La covarianza entre las puntuaciones empíricas y verdaderas es igual a la varianza de las puntuaciones verdaderas.

- $\text{cov}(\mathbf{X}_j, \mathbf{X}_k) = \text{cov}(\mathbf{V}_j, \mathbf{V}_k)$. La covarianza entre las puntuaciones empíricas de dos tests ($\mathbf{X}_j$ y $\mathbf{X}_k$) es igual a la covarianza de sus puntuaciones verdaderas ($\mathbf{V}_j$, $\mathbf{V}_k$).

- $\text{var}(\mathbf{X}) = \text{var}(\mathbf{V}) + \text{var}(\mathbf{e})$. La varianza de las puntuaciones empíricas es el producto de la suma de la varianza de las puntuaciones verdaderas y la de los errores de medida.

- $\rho(\mathbf{X}, \mathbf{e}) = \sigma_e / \sigma_x$. La correlación entre las puntuaciones empíricas y los errores de medida es el resultado de la división de la desviación típica de los errores entre la de las puntuaciones empíricas.

- $\mu_1 = \mu_2 = \dots = \mu_k$. Para k tests paralelos, las medias son idénticas.

- $\sigma^2(X_1) = \sigma^2(X_2) = \dots = \sigma^2(X_k)$. Las varianzas de k tests paralelos son idénticas.
- $\rho(X_1, X_2) = \rho(X_1, X_3) = \dots = \rho(X_i, X_k)$. Las correlaciones entre k tests paralelos son también idénticas.

Ni las puntuaciones verdaderas ni los errores de medida son observables directamente a partir de las puntuaciones empíricas, por lo tanto, ninguno de los supuestos ni las ocho primeras deducciones inmediatas son demostrables (Meneses et al., 2013). Las ocho primeras deducciones son meras tautologías, no se pueden contrastar empíricamente, mientras que sí se pueden someter a prueba empírica las tres últimas (Muñiz Fernández, 2018).

A partir de lo indicado, Ramos Vega, (2018) menciona que las dificultades de la TCT, radican esencialmente en tres puntos importantes:

1- Se queda corta en dar respuesta, porque un test no puede medir con la misma precisión a todos los sujetos, puesto que las características del test no se pueden separar de las del sujeto y la precisión de la medida dependerá del nivel del sujeto en el atributo medido.

2- La invariabilidad respecto del instrumento utilizado, porque teniendo dos test diferentes que miden un mismo atributo, la medición será diferente si se usa uno u otro test, aunque esta medición sea realizada en el mismo sujeto.

3- La propiedad del test podrá variar según el sujeto evaluado, porque un conjunto de ítems del mismo test puede ser altamente difíciles para un sujeto, pero para otro pueden ser sencillos y esto dependerá más de las características de los sujetos que de las propiedades de los test.

De acuerdo con los cálculos de los parámetros desde la TCT los principales índices utilizados son la dificultad y la discriminación que se describen a continuación.

2.4.3. Dificultad

Se entiende por índice de dificultad (ID) de un ítem la proporción de personas que lo aciertan de aquellas que han intentado resolverlo (Muñiz Fernández, 2018). Se han propuesto diferentes índices para estimar la dificultad del ítem, el más utilizado y sencillo de calcular es el índice (p). Según este parámetro, la dificultad de un ítem es la proporción de sujetos que

responden correctamente a ese ítem (Navas et al., 2001), el cual se expresa matemáticamente de la siguiente forma:

$$ID = \frac{A}{N}$$

Donde:

A = Número de personas que aciertan el ítem.

N = Número de personas que han intentado resolver el ítem.

Al índice de dificultad sería semánticamente más apropiado denominarlo «índice de facilidad» (Muñiz Fernández, 2018). Valores altos implican ítems fáciles mientras que valores bajos ítem difíciles, así mismo el índice de dificultad (p) depende tanto de las características del ítem como de la muestra de sujetos utilizada (Navas et al., 2001).

Este parámetro oscila entre un valor mínimo de $p = 0$, lo que implica un ítem muy difícil debido a que ningún sujeto lo ha respondido correctamente, y un valor máximo de $p = 1$, tratándose, en este caso, de un ítem muy fácil que ha sido respondido correctamente por todos los sujetos. Los ítems con estos valores extremos ($p = 0$ y $p = 1$) son descartados debido a que no contribuyen a medir las diferencias existentes entre los sujetos, todos los sujetos responden de la misma manera: incorrecta o correctamente (Navas et al., 2001).

La dificultad de los ítems de la prueba debe ser apropiada para la población de sujetos a la que ésta va destinada, ya que solo así permitirá al test diferenciar entre sujetos que poseen en distinto grado la característica que se está midiendo (Navas et al., 2001).

Para evitar que sujetos con poca o nula aptitud respondan al azar, se puede corregir el efecto azar en el cálculo de la dificultad de ítems de elección múltiple. Esta corrección se utiliza cuando se percibe que un considerable número de individuos responden por pura adivinación (Navas et al., 2001). De acuerdo con la valoración del nivel de dificultad que tengan los ítems se interpretan sus índices, tal como se muestra en la tabla 2, siendo los índices más aceptables los que presentan valores intermedios.

Tabla 2. Nivel de dificultad y evaluación de los reactivos

Dificultad (dj)	Evaluación del reactivo
$dj \geq 0.8$	Altamente fácil
$0.6 \leq dj < 0.8$	Medianamente fácil
$0.4 \leq dj < 0.6$	Dificultad media
$0.2 \leq dj < 0.3$	Medianamente difícil
$dj < 0.2$	Altamente difícil

Nota. Fuente: Creación propia basado en (López Meza, 2019)

Una característica interesante del parámetro (p) es que proporciona una medida de dificultad que permite la comparación de ítems que miden diferentes dominios o constructos, aplicados a los mismos sujetos; así mismo informa de los ítems que presentan una dificultad o facilidad extrema y que deben ser revisados o eliminados debido a que no contribuyen a la medida del constructo (Navas et al., 2001).

Un ítem que todos aciertan o que todos fallan no sirve para nada más que para contestarlo. Si todo aquel que responde acierta, es como si regaláramos a todos los alumnos una parte de la puntuación y si todos lo fallan, es como si los penalizáramos. En una prueba que mide contenidos diversos debería tener ítems de todas las dificultades y éstas deberían corresponderse a la dificultad de los conceptos medidos (Meneses et al., 2013).

Una limitación del índice de dificultad de la TCT es su dependencia directa de la muestra de personas en la que se calcula, es decir, el índice de dificultad no constituye una propiedad intrínseca del ítem, su valor depende del tipo de personas a las que se aplique (Muñiz Fernández, 2018).

2.4.4. Discriminación

El índice de discriminación se define como la correlación entre las puntuaciones de las personas en el ítem y sus puntuaciones en el test. Un ítem tiene poder discriminativo si distingue, entre aquellas personas que puntúan alto en el test y las que puntúan bajo, es decir, “si discrimina entre los eficaces en el test y los ineficaces” (Muñiz Fernández, 2018). A diferencia del índice de dificultad, el índice de discriminación se calcula para ítems de pruebas cognitivas y no cognitivas (Navas et al., 2001).

Navas Ara, (2001) indica que el Índice de discriminación (D), se basa en la comparación de las respuestas al ítem de dos grupos extremos de sujetos, conforme a su puntuación alta o baja en el test. Un ítem tendrá un buen índice de discriminación si las personas que aciertan el ítem son en su mayoría los mismos individuos que tienen una alta puntuación total. A continuación, en la tabla 3 se detallan las interpretaciones que corresponden a los valores del índice de Discriminación (ID).

Tabla 3. Valores e interpretaciones del Índice de Discriminación (ID)

Índice de Discriminación (ID)	Interpretación
$ID \geq 0.40$	El ítem presenta gran poder discriminativo
$0.30 \leq ID \leq 0.39$	La discriminación del ítem es aceptable
$0.20 \leq ID \leq 0.29$	El ítem discrimina poco y necesita una revisión
$0.00 < ID \leq 0.19$	El ítem no es adecuado y debe ser modificado sustancialmente o eliminado del test
$ID < 0.00$	El ítem debe eliminarse directamente

Nota. Fuente: elaboración propia basada en (Navas et al., 2001).

Llamado también índice de homogeneidad J. Pérez et al., (2008) indica que la discriminación es una característica en términos de la eficiencia del reactivo y permite diferenciar entre quienes saben y no. “Un reactivo discriminará de manera eficaz si lo responden correctamente más sujetos con puntuación alta que sujetos con puntuaciones bajas; por el contrario, un reactivo será ineficaz si quienes más aciertan son los que obtienen bajas puntuaciones” (López Meza, 2019).

Aparte del Índice de Discriminación, hay varios posibles coeficientes de correlación para la estimación de la discriminación: 1) correlación biserial-puntual, 2) correlación biserial, 3) coeficiente ϕ , 4) correlación tetracórica y 5) Índice basado en las proporciones de aciertos (Muñiz Fernández, 2018).

Correlación biserial-puntual (ρ_{bp}). Es una mera aplicación de la correlación de Pearson cuando una de las variables es dicotómica y la otra cuantitativa continua, o eventualmente discreta. Suele usarse con frecuencia, dado que es habitual que los ítems sean dicotómicos (o se aciertan o se fallan) y el test constituya una medida cuantitativa discreta (Muñiz Fernández, 2018).

Ccorrelación biserial (ρ_p). Se utiliza cuando una de las variables a correlacionar, no es dicotómica por naturaleza, pero por alguna razón se dicotomiza y se asume que bajo esa dicotomización subyace una variable continua distribuida según la curva normal (Muñiz Fernández, 2018).

La correlación del punto biserial se emplea para saber si los estudiantes aplicados son los que obtienen las respuestas correctas, qué tanto poder predictivo tiene el reactivo, y cómo puede contribuir a las predicciones (Díaz y Leyva, 2013). El punto de correlación biserial tiene la ventaja que toma en cuenta todas y cada una de las personas evaluadas (J. Pérez et al., 2008).

Coeficiente ϕ (ϕ). Se utiliza cuando las variables a correlacionar, en el ítem y test, son ambas dicotómicas, que es una mera aplicación del coeficiente de correlación de Pearson (Muñiz Fernández, 2018).

Correlación tetracórica. Se utiliza cuando ambas variables (ítem y test) están dicotomizadas y se asumen distribuidas normalmente (Muñiz Fernández, 2018).

Índice basado en las proporciones de aciertos. Es un método elemental y directo no basado en la correlación ítem-test. Este índice (d) es la diferencia entre la proporción de personas competentes que aciertan el ítem (P_c) y la proporción de incompetentes que también lo aciertan, entendiendo por competentes aquellos que puntúan en el test por encima de la mediana, e incompetentes por debajo (Muñiz Fernández, 2018).

Muñiz Fernández, (2018) indica que la interpretación de d se refiere a que la capacidad discriminativa del ítem aumenta a medida que d se aleja de cero, bien sea hacia 1 o hacia -1 y se pueden dar los siguientes casos:

- Si es 1, significa que todos los competentes aciertan el ítem ($P_c = 1$) y todos los incompetentes lo fallan ($P_i = 0$); la discriminación es perfecta.
- Si es -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 fallan; luego el ítem también discrimina perfectamente, pero habría que tener cuidado a la hora de la interpretación.

Debe tomarse en consideración que el máximo valor del índice de discriminación de un ítem tiene estrecha relación con su índice de dificultad, ya que un reactivo con valor muy alto o bajo de índice de dificultad en realidad no discrimina, pues es solucionado por la mayoría de los estudiantes o por la minoría de ellos respectivamente (Díaz y Leyva, 2013).

2.4.5. Confiabilidad

Podríamos definir también a la confiabilidad como la proporción de la variabilidad verdadera respecto de la variabilidad obtenida (Argibay, 2006). La confiabilidad es una propiedad de los puntajes y no del correspondiente del test (Soler Cárdenas, 2008).

La confiabilidad es la que garantiza estadísticamente que los resultados obtenidos tras la aplicación de un test / prueba son consistentes y estables en el tiempo y en diferentes sujetos (Ramos Vega, 2018). La confiabilidad se refiere a los **errores aleatorios**, específicamente de dos tipos: errores aleatorios que, no pueden ser controlados y no se pueden predecir y errores sistemáticos que son controlables y pueden ser explicados por alguna fuente de variación sistemática. De ambos errores los que interesan a la teoría de la fiabilidad son los errores aleatorios. (Argibay, 2006).

Mientras la validez se refiere a que se mide lo que se desea medir, la confiabilidad se refiere a la exactitud con que un instrumento de medida mide lo que mide (Virla Quero, 2010). Dentro de los métodos basados en la covarianza de los ítems el más utilizado es el coeficiente Alpha de Cronbach (Argibay, 2006). La confiabilidad es la consistencia en un conjunto de medidas de un atributo sujetos (Ramos Vega, 2018).

Desde la TCT la fiabilidad es definida como la correlación al cuadrado entre las puntuaciones verdaderas y las puntuaciones empíricas, en otras palabras, es el cociente entre la varianza de las puntuaciones verdaderas y empíricas (Navas et al., 2001).

Existen varias alternativas para evaluar la confiabilidad de los test tomando en cuenta la estabilidad temporal entre ellos los diseños clásicos: Test repetido y Formas alternas. Mientras que para evaluar la fiabilidad considerando la consistencia interna hay métodos como: división del test en dos mitades y análisis de covarianza (Soler Cárdenas, 2008).

El análisis de covarianza evalúa la proporción de varianza que los ítems tienen en común, en otras palabras, la proporción de la variabilidad de las respuestas que se atribuye al rasgo que está siendo medido, respecto a la variabilidad total de los puntajes (Frías Navarro, 2021). Las características de un instrumento que se pueden explorar con una única aplicación de una escala son la consistencia interna (homogeneidad), la dimensionalidad (estructura factorial o interna) y la confiabilidad del constructo (Arias y Oviedo, 2008).

La psicometría clásica dispone de fórmulas para estimar la fiabilidad en términos de la consistencia interna del test, entre ellas: Kuder y Richardson (KR20, KR21), Rulon, Guttman, Flanagan, Alfa de Cronbach, Omega de McDonald's, Alfa Ordinal o Nominal entre otros (Muñiz Fernández, 2018). El coeficiente alfa de Cronbach es indudablemente el recurso numérico más utilizado para evaluar la consistencia interna, y es de las fórmulas propuestas cuando el test se aplica solamente una vez (Soler Cárdenas, 2008).

Alfa de Cronbach. Es el coeficiente de confiabilidad vinculado a la homogeneidad o consistencia interna, con este coeficiente los investigadores fueron capaces de evaluar la confiabilidad o consistencia interna de un instrumento constituido por una escala Likert o cualquier escala de opciones múltiples (Virla Quero, 2010).

El Alfa de Cronbach se utiliza en situaciones en que la escala está conformada por diversos ítems que pretenden evaluar una misma habilidad y es un coeficiente que sirve para estimar la confiabilidad de una escala utilizando TCT, refleja el grado en el que covarían los ítems que constituyen el test; por tanto, es un indicador de consistencia interna (Muñiz Fernández, 2018).

Virla Quero, (2010) indica que, de manera “natural” este coeficiente de consistencia interna expresa el grado en que los ítems miden la misma variable: homogeneidad. Si el instrumento consta o mide varias variables, tiene implícitamente subinstrumentos o subpruebas individuales, la recomendación es disponer de medidas de confiabilidad para cada subprueba.

Existen algunos factores que influyen en la estimación del α de Cronbach como la longitud del test, y entre más ítems tenga el test el valor del coeficiente sería más alto. La fiabilidad depende de las características de la muestra a la que se aplicó el test, esta puede aumentar

cuando la muestra es heterogénea, ya que, a mayor variabilidad, mayor correlación. La muestra debería ser numerosa, por lo que desde este enfoque la fiabilidad es una propiedad de las puntuaciones del test en una muestra determinada (Navas et al., 2001).

Dentro de las limitaciones del alfa de Cronbach, primeramente, asume que los ítems miden un mismo constructo (tau-equivalencia) y que están correlacionados entre sí. También supone que las respuestas son politómicas en escala ordinal o continuas. Además, la independencia o no correlación de los errores (Frías Navarro, 2021).

En el caso del supuesto de la tau-equivalencia si no se cumple, el valor del alfa es infraestimado, por el contrario, si el supuesto de no correlación entre los errores se viola, el valor del alfa se sobreestima. Aunado a esto, el alfa de Cronbach también es dependiente del número de alternativas de respuesta (Ventura-León y Caycho-Rodríguez, 2014).

Omega de McDonald. Este coeficiente es calculado a través de los resultados obtenidos en análisis factoriales. El coeficiente omega tiene limitaciones, sus supuestos dependen del tamaño de la muestra y sobre todo es para “iniciados” ya que no existe en paquetes estadísticos amistosos como SPSS (Oyanedel, 2017).

Alfa Ordinal. Sperman desarrolló la técnica estadística conocida como análisis factorial, como complemento indispensable de su teoría. También aportó el coeficiente de correlación ordinal que lleva su nombre, que permite correlacionar dos variables por rangos en lugar de medir el rendimiento separado en cada una de ellas (Martínez et al., 2009).

Otra propiedad importante del alfa ordinal es que es un estimador insesgado de la confiabilidad teórica para datos ordinales. Al trabajar con observaciones cualitativas con pocas alternativas de respuesta, se debe usar el alfa ordinal y no el alfa de Cronbach (Espinoza y Novoa, 2018).

Omega Ordinal. El coeficiente omega, a diferencia del coeficiente de alfa trabaja con cargas factoriales, que son la suma ponderada de las variables estandarizadas; transformación que hace más estable los cálculos y refleja el verdadero nivel de fiabilidad, además de ello no depende del número de ítems (Ventura-León y Caycho-Rodríguez, 2014).

Rho de Sperman. Este nombre es utilizado para designar la correlación de Spearman, mide el grado de asociación entre dos cantidades, pero no mira el nivel de acuerdo o concordancia. Es recomendable utilizarlo cuando los datos presentan valores extremos, ya que dichos valores afectan mucho el coeficiente de correlación de Pearson, o ante distribuciones no normales (Martínez et al., 2009).

La escala de valoración de la fiabilidad puede tomar valores entre 0 y 1, considerándose convencionalmente que confiabilidades superiores a 0,8 son aceptables. En pruebas de alto impacto en donde los resultados implican decisiones directas sobre los examinados se sugiere que dichas pruebas deben exhibir una confiabilidad de al menos 0.9, en instrumentos que van a ser utilizados solamente para procesos de investigación, se puede ser más flexible en el criterio, en ese caso se consideran aceptables valores de Alfa iguales o mayores a 0,7 (Barbero García, 2015).

Valores de consistencia interna menores a 0.70 muestran una baja correlación entre los ítems y valores por encima de 0.95 son considerados indicadores de redundancia o duplicación de ítems, pues por lo menos un par de ítems miden exactamente el mismo aspecto de constructo y uno de ellos debería eliminarse (Frías Navarro, 2021). Un coeficiente de fiabilidad bajo indica que hay errores de medición, puede ser producto de una inadecuada formulación de los ítems, causando confusión en los individuos, también puede deberse a que la muestra es muy homogénea (Mora Vargas, 2013).

“Un coeficiente de fiabilidad muy alto es contraproducente desde el punto de vista de la investigación, indica redundancia y baja validez de constructo estos son en su mayoría multidimensionales” (Oyanedel, 2017). La fiabilidad juega un rol trascendental en el uso de test con fines de investigación, una baja fiabilidad limitaría el valor de los análisis estadísticos, asimismo, en la utilización de test para la toma de decisiones, una baja fiabilidad afectaría una toma de decisión acertada eliminarse (Frías Navarro, 2021).

2.5. Teoría de Respuesta al Ítem (TRI)

Los orígenes de la TRI se remontan a los trabajos de Thurstone, pero el auténtico padre de la TRI es Frederick M. Lord (Navas et al., 2001). La denominación TRI agrupa líneas de

investigación psicométricas independientes iniciadas en (1960) por Rasch y en (1968) por Birnbaum (Attorresi et al., 2009).

En 1968 aparece el libro “*Statistical Theories of Mental Test Scores*” de Lord y Novick, que presenta la síntesis y reanálisis crítico de todo lo hecho en la TCT; este libro incluye el trabajo de Birnbaum sobre los modelos de rasgo latente, que abrieron una línea nueva en la teoría de los test, conocida como Teoría de Respuesta al ítem (Muñiz Fernández, 2018).

El enfoque de la TRI surge desde la TCT y representa una extensión de ella, son complementarias y en la práctica pueden ser usadas conjuntamente; el enfoque de la TRI busca resolver algunas limitaciones importantes de la TCT (Manzi y San Martín, 2003).

La TRI se desarrolla en un contexto diferente al de la TCT, su desarrollo no se vincula a teorías de la inteligencia, sino a problemas técnicos en la construcción de test y en la estadística matemática (Navas et al., 2001). “El factor común de este modelo es que establece una relación entre el comportamiento de un sujeto frente a un ítem y el rasgo responsable de esta conducta llamado rasgo latente” (Attorresi et al., 2009).

La hegemonía de la TRI supone la muerte del enfoque clásico, pero no es así, pues la parsimonia y sencillez del modelo clásico lo hacen apropiado en situaciones en las que la maquinaria pesada de la TRI no puede maniobrar con eficacia (Muñiz Fernández, 2018).

2.5.1. Características Generales

“La TRI aporta una fundamentación probabilística para medir constructos latentes, y considera al ítem como la unidad básica de medición y se plantea como respuesta al problema de la TCT para separar propiedades de los test de propiedades de sujetos evaluados” (Ramos Vega, 2018). Esta teoría se basa en el cálculo para cada ítem de una serie de parámetros, que asumen un modelo matemático concreto.

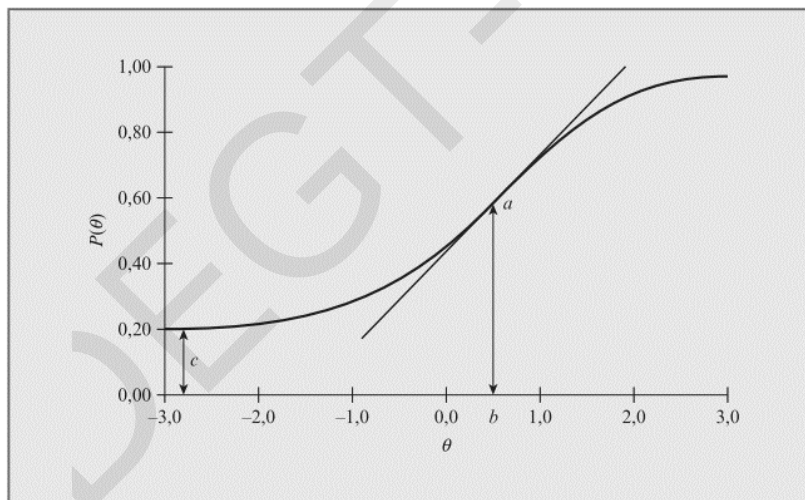
El objetivo fundamental es la medición del rasgo latente a partir de tres parámetros: discriminación, dificultad y acierto al azar (Meneses et al., 2013). Esta teoría toma al ítem como unidad de análisis y se modeliza directamente sobre el rasgo latente. (Attorresi et al., 2009).

La TCT como la TRI persiguen el mismo objetivo: estimar el error que se comete al intentar medir un fenómeno psicológico específico. La TRI enfatiza en la construcción de instrumentos de medición con propiedades invariantes entre poblaciones. “Si dos individuos presentan idéntico nivel de rasgo medido ambos tendrán igual probabilidad de dar la misma respuesta, independientemente de la población de pertenencia” (Attorresi et al., 2009).

Curva Característica del Ítem. Los modelos de TRI asumen que existe una relación funcional entre los valores de la variable que miden los ítems y la probabilidad de acertar estos y denominan a dicha función «curva característica del ítem» (CCI) (Muñiz Fernández, 2018).

Al dibujar en una gráfica la $P_i(\theta)$ como función de la aptitud, el resultado sería una curva en forma de S, tal como la que se aprecia en la figura 2, a la que se le ha denominado Curva Característica del Ítem (CCI), todo ítem puede describirse por su CCI y es esta curva la unidad conceptual básica de la TRI, ya que los restantes conceptos de la teoría se encuentran en ella (Martínez Arias et al., 2006).

Figura 2. Curva Característica de un ítem.



Nota. Fuente: (Muñiz Fernández, 2018, p. 190).

En la Figura anterior se observa la relación entre el rasgo latente que está siendo evaluado y la actuación observada en un ítem, indicando la probabilidad que tiene el individuo de responder correctamente al ítem según el nivel del rasgo que posee (Hidalgo Montesinos y French, 2016). Para su representación se utiliza la ojiva logística creciente, esta curva tiene

dos asíntotas una superior y otra inferior, con una inflexión en el centro de la curva, que es el punto de máxima pendiente (Navas et al., 2001). La CCI se suele representar en una escala situada entre -3 y $+3$ puntos de habilidad (Matas Terrón, 2010).

2.5.2. Supuestos

La fuerza de esta teoría se sostiene en un conjunto de supuestos exigentes a los que la mayoría de los datos empíricos difícilmente se acomodan y que, por ende, condicionan su aplicabilidad (Attorresi et al., 2009).

Los supuestos básicos que exige el modelo de rasgo latente son básicamente tres: 1) Asume la existencia de una variable no observada, latente, que explicaría respuestas de las personas a una prueba o ítem, 2) Unidimensionalidad y 3) Independencia local (Matas Terrón, 2010).

Unidimensionalidad. Se supone la unidimensionalidad como que una única aptitud o rasgo es suficiente para explicar los resultados de los sujetos y las relaciones entre los ítems (Martínez Arias et al., 2006).

Independencia local. Significa que no existe relación entre las respuestas de los examinados a diferentes ítems, dentro del mismo nivel de aptitud (Martínez Arias et al., 2006).

De los supuestos como la unidimensionalidad e independencia local de los ítems, el primero supone que los ítems miden esencialmente uno y sólo un rasgo latente y el segundo asume que las respuestas de un individuo para cualquier pareja de ítems en el test no están relacionadas (Hidalgo Montesinos y French, 2016).

En ciertas ocasiones es difícil comprobar la independencia local de los ítems, por tal motivo, es habitual comprobar solamente la unidimensionalidad a través de un análisis factorial o alguna prueba similar (Matas Terrón, 2010).

En conclusión, existe una complementariedad conceptual y práctica entre la TCT y la TRI, ambas teorías se construyen a partir de dos hipótesis fundamentales: 1) que los instrumentos miden un rasgo unidimensional no observable y 2) el axioma de independencia local (Manzi y San Martín, 2003).

2.5.3. Modelos

El análisis usando TRI a través de modelos matemáticos, nos proporciona una visión de la relación entre el nivel en el rasgo de un individuo (por ejemplo, nivel de depresión) y las características de los ítems (Hidalgo Montesinos y French, 2016).

La TRI proporciona una amplia gama de modelos que permite trabajar con test tanto unidimensionales como multidimensionales y con distintos formatos de respuesta como dicotómico, politómico, continuo entre otros (Hidalgo Montesinos y French, 2016). El Modelo de la TRI define tres los Parámetros básicos en los ítems los cuales se describen en la tabla 4 que se presenta a continuación.

Tabla 4. Parámetros del ítem en los modelos de TRI

Parámetro	Significado	Descripción	Rango	Valores Típicos
a	Discriminación	Valor proporcional a la pendiente de la curva en el punto $\theta = b$ (Pendiente $\theta = b = 0,425$ a $(1-c)$)	$(-00, 00)$	$(0.5, 2)$ se descartan los valores negativos
b	Dificultad	$b = \theta$ donde $P(u = 1/\theta) = (1 + c) / 2$	$(-00, 00)$	$(-2, 2)$
c	Pseudo / adivinación	Valor de la asíntota inferior	$(0, 05)$	Deseable $c \leq 0.2$

Nota. Fuente: (Navas et al., 2001).

Los modelos de TRI aportarán todo un conjunto de avances tecnológicos complementarios para la construcción y análisis de los test. (Muñiz Fernández, 2018). En la TRI se suelen distinguir tres modelos distintos en función del número de parámetros que se estiman:

– **De un parámetro (Modelo de Rasch):** Se estima la dificultad del ítem (b), es un modelo elegante y simple, es posible ajustarlo cuando se asume que no hay azar en las respuestas y los ítems presentan la misma capacidad discriminativa, es decir, igual discriminación (Hidalgo Montesinos y French, 2016).

– **De dos parámetros:** se estima la dificultad (b) y discriminación del ítem (a), este modelo se puede utilizar, dependiendo de la naturaleza de los datos. Si la adivinación no es posible o no se puede asumir, el modelo de dos parámetros (Hidalgo Montesinos y French, 2016).

– **De tres parámetros (Modelo de Birnbaum):** puede considerarse como un enfoque alternativo a la TRI, estima la dificultad (b), discriminación (a) y pseudo-azar o adivinación (c) (Matas Terrón, 2010). El modelo logístico de 3 parámetros (3PL) es un modelo muy popular (común) que se utiliza con ítems de respuesta dicotómica (Hidalgo Montesinos y French, 2016).

Al igual que con ítems puntuados de forma dicotómica, existen modelos de TRI que pueden ser seleccionados en función del tipo de datos, los elementos claves para identificar el modelo más apropiado son: (a) atractivo teórico, (b) tamaño apropiado de la muestra y (c) ajuste del modelo (Hidalgo Montesinos y French, 2016).

2.6. Análisis de Dimensionalidad

Un supuesto importante desde la TRI es que los test son unidimensionales (existen modelos multidimensionales), es decir, que un único rasgo latente es responsable explicar las respuestas de un sujeto ante un conjunto de ítems. Este supuesto es muy importante para obtener buenas mediciones y obtener evidencias de validez de estructura interna del test (Martínez Arias et al., 2006).

El estudio de dimensionalidad consiste en analizar la configuración interna del test, evaluando el nivel en que los ítems encajan con el constructo medido en el test (Navas et al., 2001).

2.6.1. Características Generales

Los modelos logísticos más utilizados actualmente asumen que los datos son unidimensionales, aunque es bien sabido que en la práctica una unidimensionalidad absoluta es difícil de conseguir, dado que hay muchos factores no controlados que afectan las respuestas de los sujetos (Cardeñoso Martínez et al., 1996).

Con el siglo XX, el análisis factorial (AF) se ha desarrollado considerablemente a lo largo de más de 100 años. El sencillo modelo inicial propuesto por Spearman (1904) para validar su teoría de inteligencia ha dado lugar a una amplia familia de modelos (Pere y Anguiano, 2010).

El AF es un modelo estadístico que representa las relaciones entre un conjunto de variables, plantea que estas relaciones pueden explicarse a partir de una serie de variables no observables (latentes) denominadas factores, siendo el número de factores substancialmente menor que el de variables (Pere y Anguiano, 2010).

El análisis factorial exploratorio es una técnica que busca determinar el número de dimensiones que subyacen a un conjunto de ítems, a partir del análisis de la matriz de Inter correlaciones entre ellos (Manzi y San Martín, 2003).

2.6.2. Supuestos

El análisis factorial trabaja en base a ciertos supuestos que al seguirlos se podrá obtener mediciones significativas:

- Normalidad: los datos siguen una distribución normal.
- Linealidad: la relación entre las variables debe ser lineal.
- Multicolinealidad: Los ítems no deben tener correlaciones elevadas o redundantes (E. Pérez y Medrano, 2010).

Es útil considerar el AF como un análisis a dos niveles, en el primer nivel se calculan las correlaciones entre una serie de medidas y en el segundo se analiza la estructura de dichas correlaciones (Pere y Anguiano, 2010).

2.6.3. Etapas del Análisis Factorial Exploratorio

En 1947 Thurstone publica su clásico texto *Análisis factorial múltiple*, técnica estadística con orígenes en el campo psicométrico (Muñiz Fernández, 2018). El modelo general más básico de AF es el modelo lineal, basado en correlaciones, que analiza medidas obtenidas en un solo grupo de sujetos y en una sola ocasión (Pere y Anguiano, 2010).

Aparte de reducir variables o determinar factores, el AFE puede ser de utilidad para identificar variables que son más relevantes de un conjunto de variables, es utilizado también en la construcción de escalas y cuestionarios para detectar variables o ítems redundantes o que aporten poca información, con el fin de ser eliminados (Martínez y Sepúlveda, 2012).

Se denomina exploratorio porque sólo podemos determinar el número de factores que esperamos, pero no su composición ni las relaciones que cada uno de los factores mantiene con el resto (Lloret-Segura et al., 2014). El análisis factorial exploratorio consiste en desarrollar las siguientes fases principales: Preparación inicial de datos, Determinación del número de factores, Extracción de factores, Rotación de factores y finalmente la interpretación de factores (Martínez Arias et al., 2006).

Se recomiendan como pasos básicos para una correcta aplicación del AFE: objetivos, diseño, supuestos, derivación de los factores y evaluación del ajuste global, interpretación de los factores y juzgamiento de la significancia de los factores (Martínez y Sepúlveda, 2012).

Preparación inicial de datos. Es un análisis previo para depurar errores, verificar la existencia de puntuaciones marginales, comprobación de supuestos y determinación de matriz de correlaciones a utilizar (Martínez Arias et al., 2006).

La idea original se basa en el supuesto de normalidad de las variables, en caso de no cumplir con el supuesto de normalidad, se espera que al menos las variables originales tengan moderados grados de correlación entre sí. Antes de realizar un AFE, se debe hacer una evaluación del supuesto de correlación entre las variables con el fin de establecer si se justifica o no su aplicación (Martínez y Sepúlveda, 2012).

La varianza total de una variable se debe a factores compartidos con otras variables (comunalidad), factores específicos de la variable (especificidad) y varianza residual. El análisis factorial busca reproducir únicamente la varianza común del ítem, no la varianza residual, para así agrupar los ítems conforme a las relaciones encontradas (Pere, 1996).

Prueba de Esfericidad de Bartlett. En esta prueba evalúa la hipótesis nula de que no existe correlación entre las variables; es decir, que la matriz de correlación es la identidad (Martínez y Sepúlveda, 2012). El test de esfericidad de Bartlett pone a prueba la hipótesis nula de que las variables no están correlacionadas en la población. Si los resultados son significativos ($p < .05$) se rechaza la hipótesis nula y se considera que las variables están suficientemente correlacionadas para realizar el AFE (Pere y Anguiano-Carrasco, 2010).

Índice Kaiser Meyer Olkin (KMO). Toma valores entre 0 y 1. Este índice toma el valor de 1 solo en el caso de que una variable sea perfectamente predicha, una desventaja es el de ser muy sensible, pues tiende a incrementarse cuando el tamaño de la muestra o el número de variables aumenta, y a decrecer cuando las correlaciones o el número de factores disminuye (Martínez y Sepúlveda, 2012).

La medida de adecuación muestral de Kaiser-Mayer-Olkin (KMO), evalúa hasta qué punto las puntuaciones en cada una de las variables son predecibles desde las demás. El rango de valores es de 0 a 1, si $KMO > .90$ el test es muy bueno, es notable si $KMO > 0.80$ y medianamente aceptable si $KMO > 0.70$ valores menores que 0.60 sugerirán interrelaciones bajas entre los ítems (E. Pérez y Medrano, 2010).

Determinar número de factores. Para la determinación del número de factores por retener, se debe revisar que los autovalores aleatorios no sean superiores a los reales.

Análisis Paralelo (Horn, 1965). El análisis paralelo (AP) es un método de simulación que consiste en generar autovalores aleatorios con similar número de variables y casos que la matriz original. El AP es el método más preciso, por mostrar poca variabilidad y sensibilidad entre los factores y poca tendencia a la sobreestimación, además de ser recomendado por algunos editores de revistas científicas. Por ende, “el AP es un método poderoso para determinar la cantidad de factores que subyace a un conjunto de indicadores” (Martínez y Sepúlveda, 2012).

Extracción de factores. Se determina el número de factores y el método de extracción, después de haber verificado los supuestos, se procede a seleccionar un método de extracción de factores para estimar una solución factorial inicial, los más utilizados son el de Componentes Principales, Ejes Principales y Máxima Verosimilitud (E. Pérez y Medrano, 2010).

Rotación de factores. Se elige el procedimiento de rotación de factores. Para test unidimensionales la rotación no es tan imprescindible, ya que se puede interpretar directamente, existen varios métodos y se clasifican en rotaciones ortogonales (varimax, quartimax y equimax) u oblicuas (oblimin direct, promax, orthoblique, procrustes) (Pere y Anguiano, 2010).

Interpretación de factores. Esta etapa consiste en interpretar y denominar los factores, se analiza en la matriz rotada, los patrones de correlaciones altas y bajas en los distintos factores (E. Pérez y Medrano, 2010).

El AFE no permite al investigador definir qué ítems miden qué factores, ni tampoco las relaciones que se suponen entre los propios factores, más allá de si están o no relacionados entre sí (Lloret-Segura et al., 2014).

DEGT-UNAH

Capítulo 3. Marco Contextual

En este capítulo se concentra información que está relacionada con el contexto en que se desarrolla el estudio de investigación sobre el Análisis Psicométrico de las áreas básicas de Prueba de Admisión a Escuelas Normales aplicado en el año 2011. Se comienza con un contexto sobre la evaluación estandarizada de los aprendizajes en América Latina y en Honduras, seguidamente se mencionan algunas experiencias de proceso de admisión en sistemas educativos de la región, el proceso de transformación de las escuelas normales y finalmente se dan a conocer generalidades del sistema educativo y el impacto que la pandemia del COVID-19 ha generado en la educación.

3.1. Evaluación Estandarizada de los Aprendizajes

En el escenario académico internacional y regional el desempeño académico de los estudiantes se evalúa mediante la aplicación de pruebas internacionales que miden la comprensión lectora y las competencias en matemática y ciencias, para realizar estas evaluaciones se conoce de la aplicación de programas como el Programa para la Evaluación Internacional de Alumnos de la OCDE (PISA por sus siglas en inglés) para países de ingresos medios, y bajos, TIMSS, PIRLS entre otras evaluaciones.

3.1.1. Procesos en América Latina

Es a partir de los años noventa, surgen una serie de reformas en los sistemas educativos enmarcadas en la importancia de medir el desempeño académico de los estudiantes con la intención de evidenciar el cumplimiento de los estándares educativos, principalmente en la educación básica y media; estas reformas tienen su génesis en recomendaciones de Organismos Internacionales y en especial por resultados de pruebas internacionales de evaluación (Rojas, 2020).

Los países de América Latina se vieron en la necesidad de establecer reformas vinculadas a la calidad educativa, bajo la premisa de que la escolarización no es garantía de aprendizaje, y es desde hace más de 20 años que la UNESCO trabaja de forma articulada con la mayoría de los países de América Latina en evaluaciones que miden el logro de aprendizaje de

estudiantes en la educación primaria. A continuación, en la Tabla 5 se presentan los programas y evaluaciones internacionales más importantes en los cuales han participado varios países de América Latina y el Caribe.

Tabla 5. Programas y valuaciones Internacionales

No.	Nombre del Programa	Aspecto que evalúa	Institución que lo realiza
1	TIMSS	Valora los conocimientos en matemáticas y ciencias	Asociación Internacional para la Evaluación del Rendimiento Educativo (IEA)
2	PIRLS	Evalúa la comprensión lectora	(OREALC/UNESCO)
3	Estudios regionales comparativos y explicativos: (PERCE, SERCE y TERCE).	Matemática, lenguaje (lectura y escritura) y ciencias naturales en tercer y sexto grado de primaria y otras variables vinculadas al desempeño de los estudiantes	(OREALC/UNESCO)
4	(ERCE-2019): Cuarto Estudio Regional Comparativo y Explicativo		

Nota. Fuente: Elaboración propia

La UNESCO a través de dependencias como la OREALC/UNESCO ha desarrollado en varios países de América Latina incluido Honduras una serie de estudios regionales comparativos y explicativos los cuales se detallan a continuación:

Primer Estudio Internacional Comparativo (PERCE-1997). Como lo indica el sitio web oficial de la UNESCO, este estudio refleja uno de los mayores logros políticos en educación en el continente, ya que surgió del acuerdo de 13 países de la región: Argentina, Bolivia, Brasil, Chile, Colombia, Costa Rica, Cuba, Honduras, México, Paraguay, Perú, República Dominicana y Venezuela. Estos países aplicaron entre junio y noviembre de 1997 pruebas de Lenguaje y Matemática. Junto con las pruebas se administraron cuestionarios a alumnos, tutores, profesores y directores, para reunir antecedentes sobre factores asociados a niveles de logro.

Segundo Estudio Internacional Comparativo (SERCE-2006). En este estudio participaron Argentina, Brasil, Chile, Colombia, Costa Rica, Cuba, Ecuador, El Salvador, Guatemala, México, Nicaragua, Panamá, Paraguay, Perú, República Dominicana y Uruguay, así como el estado mexicano de Nuevo León. En todos los países se evaluaron aprendizajes en lenguaje y matemáticas, mientras que la evaluación de ciencias se llevó a cabo en sexto

grado en Argentina, Colombia, Cuba, El Salvador, Nuevo León, Panamá, Paraguay, Perú, República Dominicana y Uruguay según datos del sitio web oficial de la UNESCO. Sin embargo, es importante señalar que Honduras no acompañó la secuencia de estas evaluaciones, al no ser partícipe de este segundo estudio regional comparativo y explicativo.

Tercer Estudio Internacional Comparativo (TERCE-2013). En este estudio Honduras participó junto a 15 países: Argentina, Brasil, Colombia, Costa Rica, Ecuador, Guatemala, México, Nicaragua, Panamá, Paraguay, Perú, República Dominicana, Uruguay y el Estado de Nuevo León, México. El TERCE utilizó dos tipos de instrumentos de recolección de información: pruebas de evaluación de aprendizaje y cuestionarios de contexto para evaluar logros de aprendizaje de estudiantes de tercer y sexto grado e identificar factores asociados a dichos logros (UNESCO, 2015).

Cuarto Estudio Internacional Comparativo (ERCE-2019). Este estudio es la cuarta versión del estudio regional comparativo y explicativo, mide logros de aprendizaje de estudiantes de sistemas educativos de América Latina y el Caribe. Se aplicó desde mayo de 2019 en 18 países: Argentina, Bolivia, Brasil, Chile, Colombia, Costa Rica, Cuba, Ecuador, El Salvador, Guatemala, Honduras, México, Nicaragua, Panamá, Paraguay, Perú, República Dominicana y Uruguay. El ERCE evaluó por primera vez las habilidades socioemocionales.

3.1.2. Procesos en Honduras

La historia alude que en Honduras de 1990 a 1994 tuvieron lugar los primeros intentos de llevar a cabo evaluaciones en gran escala como parte del Programa de Mejora de la Educación Primaria apoyado por la USAID, y fue en 1995 que se creó la Unidad de Medición de la Calidad de la Educación (UMCE) de la UPNFM. El primer levantamiento de datos tuvo lugar en 1997, cubriendo los grados 3° y 6° de primaria en las áreas de Español y Matemáticas y en 1998 se realizó el segundo levantamiento de datos, extendiéndose a los grados 2°, 3°, 4° y 6° de primaria; en las mismas áreas curriculares ya en 1999 se evaluaron los grados 2°, 3°, 4° y 5°; ya para el año 2000, 2002 y 2004, sólo 3° y 6° fueron evaluados y se incorporó el área de Ciencias Naturales (Martínez Rizo, 2008).

Como ya se mencionó en el inciso anterior dentro de los estudios regionales comparativos y explicativos desarrollados por la OREALC/UNESCO, Honduras participó en cuatro de los

tres estudios realizados: PERCE en el año 1997, TERCE en el año 2013 y ERCE en el año 2019, Honduras no participó en el SERCE en el año 2006. Sin embargo, estos no son los únicos programas o evaluaciones internacionales y nacionales de los cuales el país ha formado parte, ya que desde la década de los noventa hasta la fecha se pueden mencionar los siguientes:

Proyecto Mejorando el Impacto al Desempeño Estudiantil de Honduras (MIDEH-2004). El Proyecto Mejorando el Impacto al Desempeño Estudiantil en Honduras (MIDEH), fue financiado por los Estados Unidos a través de la Agencia de los Estados Unidos para el Desarrollo Internacional (USAID) y ejecutado en Honduras por American Institutes for Research (AIR), comenzó a operar desde el año 2004 en el país apoyando los procesos educativos, con énfasis en el nivel básico (MIDEH, 2013). En la Tabla 6 se identifican las principales líneas de trabajo del proyecto MIDEH entre las cuales se encuentra la elaboración de materiales, evaluación de aprendizajes, capacitación, fortalecimiento comunitario e investigación educativa, entre otros.

Tabla 6. Principales áreas de trabajo del Proyecto MIDEH

Elaboración de Materiales educativos de apoyo	Evaluación de Aprendizajes
Apoyo técnico a la SEDUC con: - Estándares de Contenido (español, matemáticas, CC.SS. y CC.NN. de 1ero-11mo grado) - Estándares de desempeño para estudiantes - Programaciones Educativas Mensuales - Pruebas Formativas Mensuales - Pruebas Diagnósticas e instructivos - Estándares de Contenido para padres de familia	- Desarrollo de Pruebas de Fin de Grado para evaluar de manera externa y estandarizada niveles de aprendizaje alcanzados por estudiantes de 1° a 9° grado en español y Matemáticas. - Estudio de Factores Asociados al Rendimiento Académico
Capacitación a personal docente y directivo	Fortalecimiento de la participación comunitaria en procesos educativos
- Apoyo al trabajo docente y al uso e interpretación de resultados de evaluación de aprendizajes. - Diseño de estrategias de mejora de aprendizajes de estudiantes con Planes de Mejora a nivel Distrital, Municipal, de Centro y de Aula.	- Apoyo a proyectos presentados por instancias nacionales, para favorecer el conocimiento, colaboración activa y valoración de procesos educativos con padres de familia, municipalidades y sociedad civil.
Investigación educativa	
- Estudios sobre la realidad educativa para apoyar a la SEDUC en su labor de dirección del sistema. - Investigaciones acerca del uso del CNB y materiales alineados al mismo en las aulas de clases	

Nota Fuente: Elaboración propia basada en (MIDEH, 2013).

Estudios TIMS y PIRLS (2011). El Estudio Internacional de Tendencias en Matemáticas y Ciencias (TIMSS) y Estudio Internacional de Progreso en Comprensión lectora (PIRLS)

patrocinado por la International Association for the Evaluation of Educational Achievement (IEA) se aplicó en Honduras entre octubre y noviembre del 2010 a estudiantes de sexto y noveno grado. Los estudiantes contestaron pruebas de conocimientos y adicionalmente un cuestionario referido con antecedentes personales y familiares; a los profesores y director del centro educativo se les aplicó un cuestionario de Matemáticas y de Ciencias en TIMSS y para el estudio PIRLS una encuesta a padres de familia. La muestra estuvo compuesta por 152 centros educativos, llegándose a aplicar pruebas a casi 4,000 estudiantes en 6° grado, el 97% de lo previsto y a cerca de 4,500 estudiantes de 9° grado, el 96% del número previsto (Mullis et al., 2012).

Los resultados obtenidos en TIMSS 2011 para Honduras en el área de Matemáticas indican que: en la categoría de 4° grado, participo con 6° grado y obtuvo 396 puntos ubicándose a 104 puntos del promedio de la escala y ocupando el lugar 45 de 52 participantes en este grado y asignatura. Mientras que en la categoría de 8° grado participo con 9° grado y alcanzó un promedio de 338 puntos ubicándolo en el penúltimo lugar, 44 de 45 países participantes. Mientras que, en el área de Ciencias, Honduras con la participación de 6° grado obtuvo 432 puntos ocupando el lugar número 42 entre los 52 países participantes, este es el mejor resultado logrado por el país en todas las materias evaluadas. Mientras que, con 9° grado, obtuvo 369 puntos alcanzando el lugar 43 entre 45 países, este resultado es más bajo que el punto medio de la escala TIMSS y sólo es más alto que el de países como Ghana (Rápalo Castellanos, 2012).

Mientras que en PIRLS 2011 los resultados obtenidos por Honduras indican que en lectura entre los países que participaron en 6° grado, Honduras ocupó el primer lugar entre los cuatro países participantes con un resultado promedio en la escala de 450 puntos. Este resultado lo ubicaría en el lugar 39 de los 49 países en total. En el caso de Honduras, hay un 26% de estudiantes que tienen serios problemas de lectura ya que no alcanzan ni el nivel bajo de la escala PIRLS, del resto el 38% alcanza el nivel intermedio de lectura básica el 10% y el 1% alcanzan los niveles de lectura, alto y avanzado respectivamente (Rápalo Castellanos, 2012).

PISA D: Programa para la Evaluación Internacional de Alumnos de la Organización para la Cooperación y el Desarrollo Económicos (OCDE- 2014). PISA evalúa hasta qué punto los estudiantes de 15 años cerca del final de su educación obligatoria han adquirido

conocimientos y habilidades clave que son esenciales para la participación plena en las sociedades modernas. La evaluación se centra en asignaturas básicas como lectura, matemáticas y ciencias.

Para atender una gama más amplia de países la OCDE lanzó en 2014 la iniciativa PISA para el Desarrollo (PISA-D) con la participación de 7 países: Camboya, Ecuador, Guatemala, Honduras, Paraguay, Senegal y Zambia. Este proyecto piloto tenía como objetivo hacer que la evaluación sea más accesible y relevante para países de ingresos bajos y medios. La encuesta escolar evaluó a estudiantes en lectura, matemáticas y ciencias. Entre algunos resultados se puede mencionar que Camboya, Honduras y Senegal, el número de niñas elegibles para el examen PISA superó el número de niños en un 10% o más y en Camboya, Ecuador, Guatemala, Honduras y Paraguay, los niños tenían más probabilidades que las niñas de repetir un grado (OECD, 2018).

Algo novedoso en el estudio de PISA- D es que incluyó un enfoque de “prosperidad educativa”. Este enfoque considera las condiciones necesarias para que los sistemas educativos ayuden a estudiantes a tener éxito en la escuela y en la vida e identifica un conjunto de 4 resultados clave, llamados resultados de prosperidad: 1) logro educativo, 2) desempeño académico, 3) salud y bienestar y 4) actitudes hacia la escuela y el aprendizaje. El enfoque también identifica un conjunto de factores familiares, institucionales y comunitarios llamados “fundamentos para el éxito”, que influyen en los resultados: recursos, entornos inclusivos, tiempo de aprendizaje, instrucción de calidad, apoyo familiar y comunitario (OECD, 2018).

Finalmente podemos considerar que en evaluaciones internacionales en las cuales ha participado Honduras, los resultados obtenidos son consistentes con el desempeño observado cuyo rendimiento se encuentra distante de lograr que los estudiantes alcancen un nivel de desempeño académico “Avanzado”, en un escenario donde mayoría de los alumnos aparecen en los niveles de desempeño “Debe mejorar” e “Insatisfactorio”; y muy por debajo de la media regional y mundial, en el caso de las pruebas internacionales (Rojas, 2020).

3.2. Procesos de Admisión en los sistemas educativos en la región.

Es en los años 90 en América Latina se introdujo un conjunto de políticas que transformaron principalmente la educación superior, estableciendo mecanismos de evaluación institucional como forma de regulación. En el caso de México, la evaluación constituye el eje de la reforma de los 90; la misma se institucionalizó a través del Programa para la Modernización Educativa (1989-1994), ejecutado por el Gobierno Federal y que organiza la Comisión Nacional para la Evaluación de la Educación Superior (CONAEVA) en el marco de funcionamiento del Sistema Nacional de evaluación y autoevaluación del sistema interinstitucional. Posteriormente en 1994 se crea el Centro Nacional de Evaluación para la Educación Superior (CENEVAL) cuyo objetivo es contribuir a mejorar la calidad de la educación media y superior evaluando aprendizajes logrados en los procesos educativos mediante la aplicación de exámenes estandarizados (Chiroleu, 2011).

En Honduras no hay mucha evidencia de procesos de admisión en los primeros niveles de enseñanza, pero a nivel superior desde el año 2006 hasta la fecha la UNAH aplica la Prueba de Aptitud Académica (PAA) como herramienta para admitir a estudiantes a la universidad. Esta prueba evalúa dos componentes: Razonamiento Verbal que mide el nivel o capacidad de interpretar la lectura, desde la comprensión del texto hasta el análisis de ideas fundamentales y el Razonamiento Matemático que mide la habilidad para procesar, analizar y utilizar información de la aritmética, el álgebra, la geometría y la estadística. Además de utilizarse en Honduras la PAA se aplica en once países de Latinoamérica: Bolivia, Costa Rica, Ecuador, El Salvador, Guatemala, México, Panamá, Puerto Rico, República Dominicana, Uruguay y en varias universidades de los Estados Unidos (Cortés, 2019).

Finalmente se puede hacer mención que las principales barreras institucionales presentes en la mayoría de los sistemas de educación radican en las tasas de matrícula y los exámenes de ingreso o de admisión. Los exámenes de ingreso y los requisitos aparecen como un medio justificable para evaluar si los estudiantes están equipados para participar en un curso en particular, pero en muchos casos actúan para privilegiar a los estudiantes de escuelas de alta calidad y a aquellos que han podido pagar los cursos preparatorios. Al aplicar los principios meritocráticos de las admisiones universitarias es difícil identificar las ventajas y desventajas sociales, por otro lado, la introducción de exámenes de ingreso nacionales estandarizados y

acciones afirmativas en muchos países como, Brasil y Tayikistán, han reducido esta brecha de entrada (IESALC, 2020).

3.3. Oferta formativa de Educación Media

Según el Reglamento de la Ley Fundamental de Educación No. 1358-SE-2014 define en su artículo 63 que la Educación Media es el tercer nivel de educación formal, constituye el nivel siguiente al de educación básica y previo al de educación superior, con los cuales estará articulado curricular y administrativamente. El nivel de educación media es obligatorio y se ofrece en los centros educativos oficiales de forma gratuita, teniendo como único requisito para su ingreso haber aprobado la Educación Básica (Reglamentos Ley Fundamental de Educación, 2014).

La normativa específicamente en el acuerdo No. 1366-SE-2014 que corresponde al Reglamento de Educación Media, establece que el propósito de la educación media es ofrecer a los educandos la experiencia formativa para incorporarse al mundo del trabajo o proseguir estudios en el nivel superior, mediante la adquisición y construcción de conocimientos, habilidades y actitudes relevantes para su vida personal y social; así como para el desarrollo económico, sociocultural, científico y tecnológico del país (Reglamentos Ley Fundamental de Educación, 2014).

En cuanto al tema de Evaluación, la Secretaría de Educación (SEDUC) en normativa indica que realizara evaluaciones nacionales, regionales, departamentales, distritales, municipales y locales en los niveles y modalidades; y que dicha evaluación incluirá al personal de centros educativos, educandos, materiales didácticos, recursos para el aprendizaje, condiciones del ambiente escolar y otros elementos del proceso educativo que permitan mejorar el nivel de rendimiento y la calidad de la educación. En el nivel específico de la Educación Media se indica en la Ley Fundamental de Educación, que los procesos de evaluación en este nivel se realizarán conforme a la ley especial que regulará la evaluación, acreditación y certificación de la calidad y equidad educativa (Reglamentos Ley Fundamental de Educación, 2014).

A pesar de lo que establece la normativa y los alcances de la educación media, hay diversos factores que contribuyen a que gran parte de la población estudiantil no pueda

acceder al mismo, Honduras muestra grandes limitantes en tres elementos fundamentales que inciden en la oferta educativa nacional como ser: cobertura, equidad y calidad. En la Tabla 7 se presenta un resumen de la estructura curricular de la Educación Media, que se centra en dos grandes modalidades.

Tabla 7. Estructura Curricular de la Educación Media en Honduras.

Modalidad	Objetivo	Requisito de Graduación	1 año	2 año
Ciencias y Humanidades en sus diferentes especialidades (2 años)	Preparará al educando para continuar estudios en el nivel superior universitario o no universitario Titulo obtenido “Bachiller en Ciencias y Humanidades” en la respectiva especialidad.	Trabajo Educativo Social (T.E.S.) Edades de referencia Quince (15) a los diecisiete (17) años	Estudio común a ambas modalidades denominado: <i>Formación de Fundamento</i> • Retoma con mayor nivel de complejidad y profundidad los conocimientos alcanzados por los educandos durante la educación básica.	Estudio denominado: Formación orientada para culminar la Modalidad y especialidad
Modalidad	Objetivo	Requisito de Graduación	2 año	3 año
Técnico Profesional en sus diferentes especialidades (3 años)	Brinda a los educandos las competencias necesarias para acceder al mundo laboral o continuar estudios en el nivel superior Titulo obtenido “Bachiller Técnico Profesional” en la respectiva especialidad.	Práctica Profesional Edades de referencia Quince (15) a los dieciocho (18) años	• Consolida una base de competencias comunes y articuladoras de las dos modalidades del nivel medio.	Estudio Denominado <i>Formación orientada dirigida a determinar las áreas del conocimiento y especialidad</i> Estudio denominado: <i>Formación Específica que desarrolla competencias de la especialidad</i>

Nota. Fuente: Elaboración propia basada en (Reglamentos Ley Fundamental de Educación, 2014).

Algunos datos relevantes del Sistema educativo en Honduras principalmente en lo que se refiere a la Educación Básica en el IV ciclo que comprende las edades de 15-17 años, nos indica que la cobertura promedio en Educación Media que se registra a nivel nacional en 18 años del (2001-2018) es de 23.5%, lo que significa que de cada 100 jóvenes entre las edades de 15-17 años, solamente ingresan 23, quedando excluidos 77 jóvenes, esta cobertura osciló

entre el 7.1% a 29.9% entre los años 2001- 2018, con un leve crecimiento anual, de aproximadamente el 1% (Padilla, 2019).

En cuanto a presupuesto, uno de los más afectados es el tercer ciclo de la educación básica y al nivel medio, que es donde menor cobertura existe como ya se mencionaba anteriormente, ya que la cobertura en los dos primeros ciclos de la educación básica ya casi se ha universalizado. La UPNFM menciona que es importante valorar la inversión en modalidades alternativas, para reducir la brecha educativa en la atención a los jóvenes en edad de estudiar, principalmente en la educación media y de esta manera apalea un poco las razones por las cuales no la población estudiantil de este nivel no se incorpora en el sistema educativo (Hernández Rodríguez et al., 2018b).

Finalmente es importante recalcar que Honduras tiene una deuda social muy grande para poder atraer y dar oportunidades educativas al 70% de los jóvenes de entre 15 y 17 años excluidos del sistema, ya sea por rezago educativo o en la mayoría de los casos porque no están matriculados. El desafío importante aquí recae en el riesgo social al que se enfrentan los jóvenes fuera del sistema educativo entre 15 y 17 años al no asistir a los centros de enseñanza nacionales, pero también ante el escenario laboral acompañado de pocas oportunidades de empleo dada su acotada formación, capacitación y experiencia (Rojas, 2020).

Capítulo 4. Metodología

En este capítulo se dan a conocer los procedimientos metodológicos que fueron utilizados en el estudio de Análisis Psicométrico de las áreas básicas de la Prueba de Admisión a Escuelas Normales aplicado en el año 2011. Esta sección contiene el tipo de investigación, características de la muestra, operacionalización de variables, presentación del instrumento aplicado y finalmente los procedimientos ejecutados para realizar el estudio.

4.1. Tipo de investigación.

El estudio de Análisis Psicométrico de la Prueba de Admisión a Escuelas Normales se orienta con un **Enfoque Cuantitativo** ya que se realiza un análisis a la base de datos de los puntajes obtenidos en la aplicación a aspirantes en el año 2011, se asiste en la recolección de datos para probar mediante la medición numérica y el análisis estadístico pautas de comportamiento y comprobación de la teoría (Hernández Sampieri et al., 2010).

El diseño es de **Tipo Instrumental** ya que los objetivos se enmarcan en el análisis psicométrico de un Instrumento. La investigación de este tipo es donde se incluyen trabajos que analizan propiedades tanto psicométricas de instrumentos psicológicos nuevos o ya existentes que son traducidos, adaptados o creados (Ato et al., 2013).

La estrategia utilizada es **Descriptiva** ya que da a conocer lo que existe, determinando la frecuencia con que ocurren los hechos y clasificando la información (Sousa et al., 2007). Así mismo es de corte **Transversal**, al utilizar los datos recabados en la aplicación del examen de admisión a escuelas normales en el año 2011, obteniendo información de la muestra una única vez (Bernal, 2010).

Finalmente, este estudio se enmarca en un **Diseño No Experimental**, sin manipulación de variables, en este caso solo se realiza un proceso de análisis de los datos, observación y recopilación de información de manera natural sin intervenir de manera alguna en los resultados obtenidos en el análisis psicométrico (Sousa et al., 2007).

4.2. Población y muestra

Este estudio hace uso de una muestra no probabilística dirigida ya que el subgrupo de la población utilizado son estudiantes aspirantes para ingresar a la carrera de magisterio, es decir este grupo ya posee características específicas y un objetivo en común que es lograr una puntuación mediante la aplicación de un instrumento de admisión a la educación media en el año 2011. Esto indica que es un subgrupo de la población en la que la elección de los elementos no depende de la probabilidad, sino de las características de la investigación (Hernández Sampieri et al., 2010).

4.2.1. Participantes

La población objeto de estudio está integrada por 5,523 jóvenes estudiantes que se postularon a la obtención de un cupo para cursar la carrera de Magisterio ofertada en el nivel de educación media en el año 2011. Estos aspirantes se encontraban en un rango de edad de entre los 13 y los 45 años; siendo una mayoría los jóvenes de 15 años con 52.9%. De la totalidad de los participantes la mayoría pertenecen al sexo femenino siendo un 72.1 % y un 27.9% pertenecen al sexo masculino, la muestra se obtuvo en la aplicación del examen de admisión en 13 Escuelas Normales ubicadas en los departamentos de Atlántida, Colón, Comayagua, Cortés, Choluteca, El Paraíso. Francisco Morazán, Intibucá, La Paz, Lempira, Ocotepeque, Olancho y Santa Bárbara. En la Tabla 8 se muestra la cantidad de estudiantes participantes de acuerdo a las formas aplicadas, que en este caso fueron dos (A y B).

Tabla 8. Número de aspirantes que contestaron la prueba de admisión.

Forma	Cantidad de aspirantes
Forma A	2,776
Forma B	2,747
Número Total de Aspirantes	5,523

Nota. Fuente: Elaboración propia

4.3. Variables

Tabla 9. Matriz de Operacionalización de las Variables

Variable / Factor	Definición Conceptual	Definición Operacional	Dimensión	Indicador (Ítems)
Conocimientos básicos en Matemáticas.	Disciplina que sistematiza la capacidad intuitiva del ser humano de poder encontrar ideas medias y necesarias para resolver problemas, el conocimiento matemático es intuitivo que precisa de la demostración para poder ser explicado y explicitado. (Secretaría de Educación, 2006).		Conocimiento Comprensión Aplicación Subtotal	7 ítems 7 ítems 6 ítems 20 ítems
Conocimientos básicos en Español.	Capacidad del ser humano que ofrece recursos para abstraer, conceptualizar, representar e interpretar la realidad, gracias a esta capacidad el individuo construye, un universo conceptual y distintos sistemas signícos (la lengua, música, pintura, escritura, gestos, entre otros), la principal finalidad es posibilitar formas diversas de comunicación entre los seres humanos (Secretaría de Educación, 2006).	Proceso por el cual el hombre refleja por medio de conceptos, leyes, categorías, etc. en su cerebro las condiciones características del mundo circundante. El conocimiento es una de las formas que tiene el hombre para otorgarle un significado con sentido a la realidad (Sánchez Sánchez, 2015)	Conocimiento Comprensión lectora Aplicación Subtotal	7 ítems 7 ítems 6 ítems 20 ítems
Conocimientos básicos en Ciencias Naturales.	Disciplina que aborda el estudio de elementos naturales del ambiente, sus relaciones y fenómenos que resultan de dichas relaciones, incluye el estudio de los seres vivos a través de la biología; estudio de los seres no vivos a través de la geología y la química; el estudio de las relaciones (Secretaría de Educación, 2006).		Conocimiento Comprensión Aplicación Subtotal	7 ítems 7 ítems 6 ítems 20 ítems
Conocimientos básicos en Ciencias Sociales.	Dominio sistematizado de conocimientos muy complejo y sumamente diverso que tiene como sujeto de estudio al hombre viviendo en sociedad, actor de procesos que se desarrollan en un contexto de tiempo-espacio determinado (Secretaría de Educación, 2006).		Conocimiento Comprensión Aplicación Subtotal Total	7 ítems 7 ítems 6 ítems 20 ítems 80 ítems

Nota. Fuente: Elaboración propia.

4.4. Instrumento de Medición.

Se utilizó como instrumento la Prueba de admisión a Escuelas Normales aplicado a estudiantes aspirantes a la Carrera de Magisterio en el año 2011. Este instrumento evalúa 4 áreas del conocimiento básicas como ser matemáticas, español, ciencias naturales y ciencias sociales.

La prueba está estructurada en tres formas: Forma “A”, Forma “B” y Forma “C”, cada una de ellas contiene ítems de selección única donde se presenta el enunciado con cuatro posibles opciones de respuesta, de las cuales una es correcta y las restantes son distractores. La diferencia en cada una de las formas radica en el orden en que aparecen las preguntas de cada área del conocimiento, para efectos de los análisis psicométricos que se realizan en este estudio de investigación se utilizaron dos formas: Forma A y Forma B.

Cada uno de los campos del conocimiento evaluados en esta prueba se califican en base a un puntaje de 100 puntos, que luego son sumados en su totalidad para obtener un puntaje final que coloca a los estudiantes en una posición dentro un ranking, la cantidad de preguntas por área temática se mencionan en la Tabla 10. Los puntajes superiores ocupan los primeros cupos dentro del ranking de ingreso.

Tabla 10. Número de preguntas por área temática

Área Temática	Cantidad de preguntas
Matemáticas	20
Español	20
Ciencias Naturales	20
Ciencias Sociales	20
Cantidad total de preguntas	80

Nota. Fuente: Elaboración propia basado en la Prueba de Admisión a Escuelas Normales

De manera general se puede mencionar que el instrumento fue elaborado en aproximadamente tres meses por un equipo de técnicos de la Dirección de Currículo y Evaluación de la Secretaría de Educación, este instrumento cuenta con ítems de opción única y tarda alrededor de dos a tres horas en poderlo responder ya que a la par de la prueba al estudiante se le aplicó un instrumento con preguntas referentes a la habilidad cognitiva y a la

vocación; la aplicación se realizó una única vez de manera simultánea en todas las escuelas normales del país.

4.5. Plan de Análisis

Esencialmente para cumplir con cada una de las etapas de análisis psicométrico de la prueba se realizó lo siguiente:

1. La confiabilidad de los ítems de la prueba sustentados en la TCT se calculó mediante el uso del Coeficiente Alfa de Cronbach y Omega de McDonald; luego se realizaron los cálculos haciendo uso del Alfa Ordinal Tetracórico y el Omega Ordinal Tetracórico.

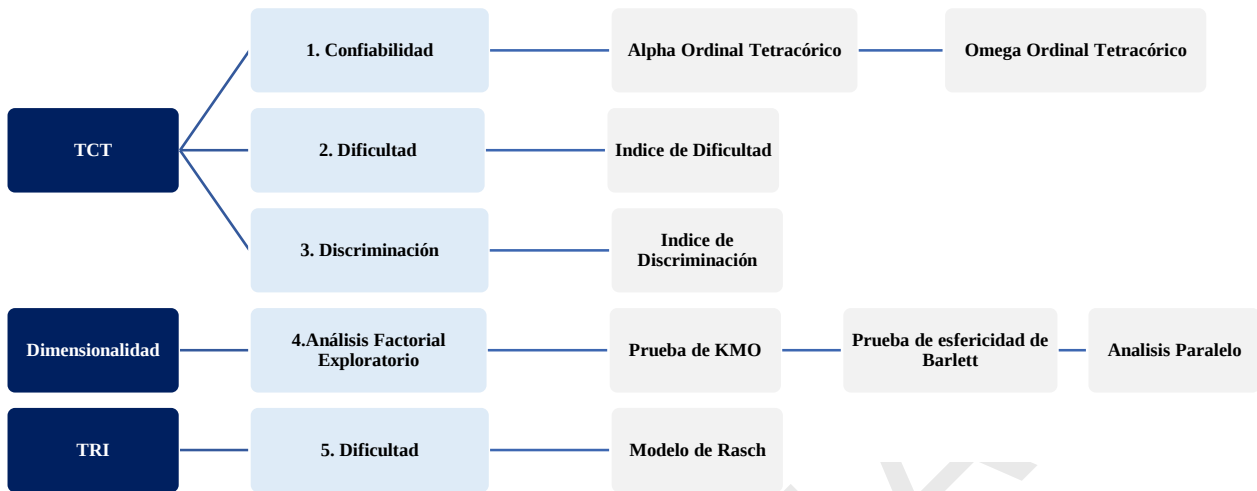
2. El grado de dificultad de los ítems de la prueba sustentados en la TCT se determinó mediante la aplicación del cálculo de Índice de Dificultad de los ítems.

3. El poder de discriminación de los ítems de la prueba sustentados en la TCT se calcularon mediante la estimación del Índice de Discriminación.

4. Para el proceso de evaluación de la Dimensionalidad de la prueba se realizó el Análisis Factorial Exploratorio para determinar si la prueba cumplía con el supuesto de unidimensionalidad y así poder aplicar los modelos de la TRI, se tomó en cuenta la prueba KMO, prueba de esfericidad de Barlett y el porcentaje de varianza.

5. Finalmente, en cuanto a la estimación de parámetros desde la TRI: se aplicó el modelo logístico de un parámetro (Modelo de Rash) donde se hizo un análisis del parámetro b (dificultad).

En resumen, en la Figura 3 se muestran cada uno de los pasos indicados en el Plan de Análisis realizado tomando como base la muestra de estudiantes que contestaron la Prueba de Admisión a Escuelas Normales aplicado en el año 2011.

Figura 3. Etapas del Plan de Análisis de datos

Nota. Fuente: Creación Propia

Para realizar el cálculo de los análisis aplicados se utilizaron los programas de SPSS, R Studio y Jamovi.

4.6. Procedimientos

Los procedimientos logísticos, metodológicos y éticos que guiaron el desarrollo de este estudio de investigación fueron los que se enuncian a continuación:

4.6.1. Logísticos.

Para la etapa de la recolección de los datos, se utilizó una muestra de estudiantes en su mayoría de noveno grado, sin embargo, también participaron una cantidad significativa de sujetos que realizaron noveno grado en años anteriores, cuyo fin común era posicionarse en un lugar dentro del ranking de acceso a las escuelas normales. La aplicación de la prueba se realizó en 13 escuelas normales del país en el año 2011.

4.6.2. Metodológicos.

La prueba de Admisión a Escuelas Normales fue aplicada utilizando dos tipos de formas en los cuadernillos: Forma “A” y Forma “B”. La aplicación se realizó en los centros donde funcionaron las 13 escuelas normales en diferentes zonas de Honduras, logrando una muestra

de 5,523 estudiantes en total de los cuales a 2,776 se les aplico el cuadernillo con la forma “A” y a un total de 2,747 se les aplico el cuadernillo de la Forma “B”.

4.6.3. Procedimientos éticos

Se publicó la fecha de aplicación de la prueba de admisión. Cada una de las escuelas normales organizó el proceso de inscripción de los aplicantes y se contó con todas las medidas de seguridad necesarias para el traslado, aplicación y revisión de resultados de las pruebas. Posteriormente se indicó en cada una de las escuelas normales participantes una fecha para publicación de los resultados y tabla de rankings de cupos obtenidos, donde cada estudiante podría verificar la posición ganada de acuerdo con el número de identidad de cada participante.

Capítulo 5. Resultados de la Investigación

En este apartado se da a conocer la descripción e interpretación de los resultados obtenidos en el desarrollo del estudio de investigación sobre el Análisis Psicométrico de las áreas básicas la Prueba de Admisión a Escuelas Normales aplicado en el año 2011. En función de los objetivos planteados se dan a conocer los resultados alcanzados al analizar los parámetros psicométricos de los ítems basándose en la TCT entre ellos su confiabilidad, dificultad y discriminación, seguidamente se presentan los resultados obtenidos al aplicarse el Análisis de Dimensionalidad mediante el AFE y finalmente se dan a conocer los análisis estadísticos basados en la TRI enmarcados en el Modelo de Rasch (1 Parámetro), tal como se indica en los objetivos de estudio de investigación.

5.1. Análisis de parámetros psicométricos desde la TCT

Confiabilidad. Para trabajar el primer objetivo se realizó el análisis de confiabilidad de las dos formas de la prueba de acuerdo con los componentes temáticos (matemáticas, español, ciencias naturales y ciencias sociales). Se calcularon los índices por medio del Alpha de Cronbach (α) y Omega de McDonald (ω) haciendo uso del programa SPSS y Jamovi respectivamente.

Sin embargo, al realizar la revisión de la literatura y por la naturaleza de los datos se indica que de forma más radical, el alfa de Cronbach tiene una utilidad bastante limitada y se recomienda abandonar su uso (Espinoza y Novoa, 2018). Además se ha demostrado que el alfa de Cronbach subestima la magnitud de la confiabilidad cuando se tienen ítems con menos de cinco alternativas y se propone utilizar el alfa ordinal para escalas de respuesta dicotómica y ordinal (Espinoza y Novoa, 2018).

Es por lo que, además, se calculó el Alpha Ordinal y Omega Ordinal a partir de una matriz de correlación tetracórica al tener solamente dos opciones de respuesta (1=correcta y 0=incorrecta), con el propósito de conocer y corroborar que tan consistente es la prueba y de esta forma poder continuar el proceso de aplicación de procedimientos psicométricos.

La tabla 11 muestra los coeficientes que se obtuvieron al estimar la confiabilidad general de la prueba en sus dos formas y por subgrupo temático; ya que para instrumentos formados

por veinte ítems se recomienda que se calcule la consistencia interna de diferentes subescalas, subgrupo de ítems, que generalmente hacen parte de una escala (Arias y Oviedo, 2008). La escala de valoración de la fiabilidad puede tomar valores entre 0 y 1, considerándose confiabilidades los valores superiores a 0.7 se denominan aceptables (Barbero García, 2015).

Tabla 11. Coeficientes de Confiabilidad (Forma A y B) por subgrupo temático.

Forma de la prueba	Subgrupo	Alpha de Cronbach (α) usando	Omega de McDonald's (ω) usando	Alpha Ordinal Tetracórico usando	Omega Ordinal Tetracórico usando
		SPSS	Jamovi	RStudio	RStudio
Forma A	Matemáticas	0.78	0.79	0.40	0.54
	Español	0.42	0.43	0.54	0.62
	Ciencias Naturales	0.85	0.85	0.36	0.40
	Ciencias Sociales	0.61	0.60	0.33	0.38
	Prueba Total	0.86	0.86	0.74	0.751
Forma B	Matemáticas	0.71	0.71	0.43	0.48
	Español	0.40	0.40	0.53	0.59
	Ciencias Naturales	0.61	0.61	0.35	0.41
	Ciencias Sociales	0.84	0.84	0.38	0.43
	Prueba Total	0.83	0.82	0.71	0.73

Nota. Fuente: Creación Propia haciendo uso de SPSS, Jamovi y RStudio

Mediante el coeficiente omega se puede considerar un valor aceptable de confiabilidad, si el resultado se encuentra entre .70 y .90 aunque en algunas circunstancias pueden aceptarse valores superiores a 0.65 (Ventura-León y Caycho-Rodríguez, 2014).

En este caso la prueba de admisión a escuelas normales refleja una consistencia interna aceptable en sus dos formas (A y B) ya que al utilizar los estadísticos correctos: Alpha Ordinal y Omega Ordinal a partir de una matriz de correlación tetracórica oscilan entre el 0.71 y 0.75 lo que significa que la prueba tiene una aceptable confiabilidad. Existe una variabilidad diferente al aplicar estadísticos psicométricos para estimar la confiabilidad a nivel de subgrupo temático obteniendo índices en la forma A, entre el 0.33 al 0.62 y en la forma B entre el 0.35 al 0.59 considerados valores con una confiabilidad Muy baja o Nula.

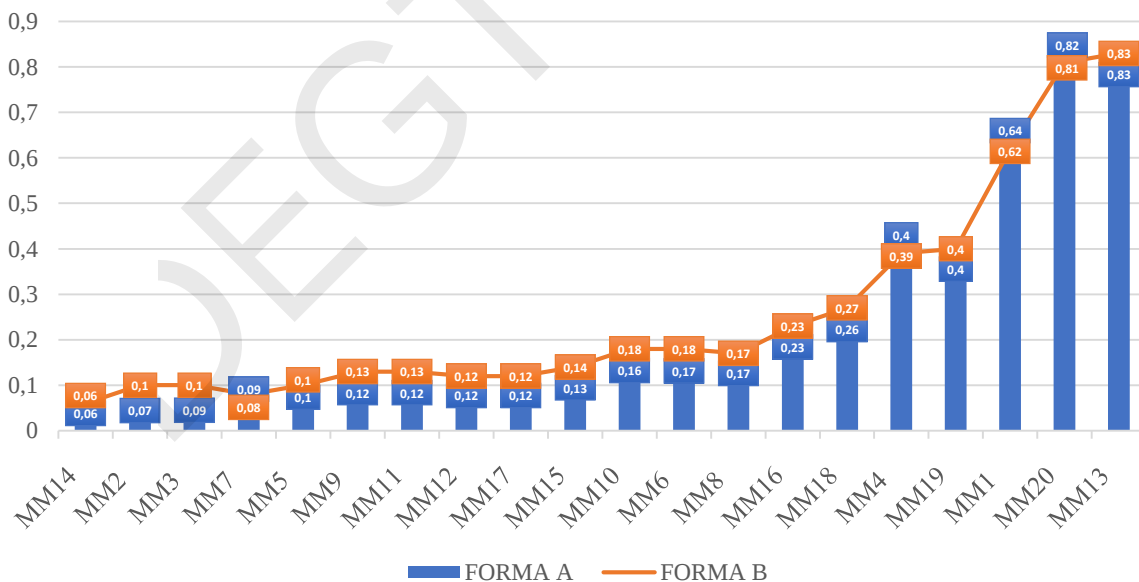
Al utilizar diferentes coeficientes de confiabilidad la prueba presento resultados aceptables en cuanto al índice de confiabilidad, a pesar de que basándose en la literatura una

de las principales recomendaciones es diferenciar muy bien las bondades y características que deben tomarse en cuenta para utilizar un estadístico en específico.

Dificultad. Respecto al cumplimiento del segundo objetivo de este estudio de investigación se estimó el Índice de Dificultad de los ítems de las dos formas de la prueba, de acuerdo con los componentes temáticos (matemáticas, español, ciencias naturales y ciencias sociales); con el propósito de conocer la proporción de personas que acertaron los ítems de aquellas que intentaron resolverlo. Se calculó el Índice de Dificultad (ID) de los 80 ítems considerando como fundamento la TCT, haciendo uso del programa RStudio.

La literatura indica que la dificultad oscila entre un valor mínimo de $p = 0$, lo que implica un ítem muy difícil debido a que ningún sujeto lo ha respondido correctamente, y un valor máximo de $p = 1$, tratándose, en este caso, de un ítem muy fácil que ha sido respondido correctamente por todos los sujetos. Valores altos implican ítems fáciles mientras que valores bajos ítem difíciles (Navas et al., 2001). En la figura 4 se puede observar que de los 20 ítems del área de matemáticas 2 son altamente fáciles, 1 medianamente fácil, 2 son de dificultad media, 2 son medianamente difícil y 13 de ellos son altamente difíciles.

Figura 4. Índice de Dificultad de ítems de Matemáticas (Forma A y B) con TCT.

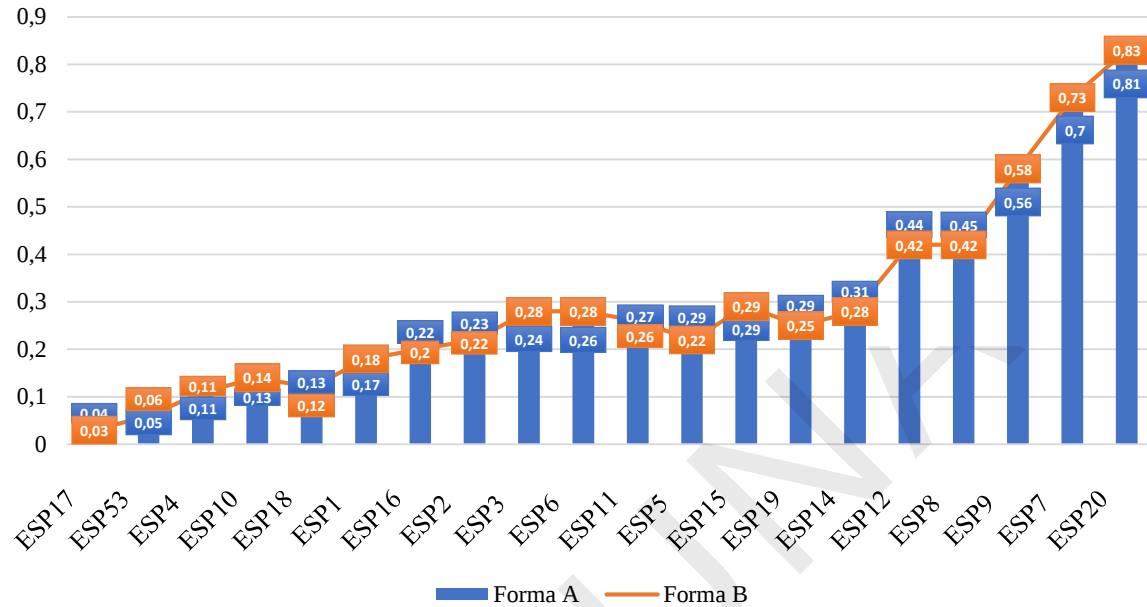


Nota. Fuente: Creación propia utilizando Excel.

En la figura 5 se concibe que en el área de español uno de los ítems es altamente fácil, 1 medianamente fácil, 3 tienen una dificultad media, 9 medianamente difíciles y 6 son

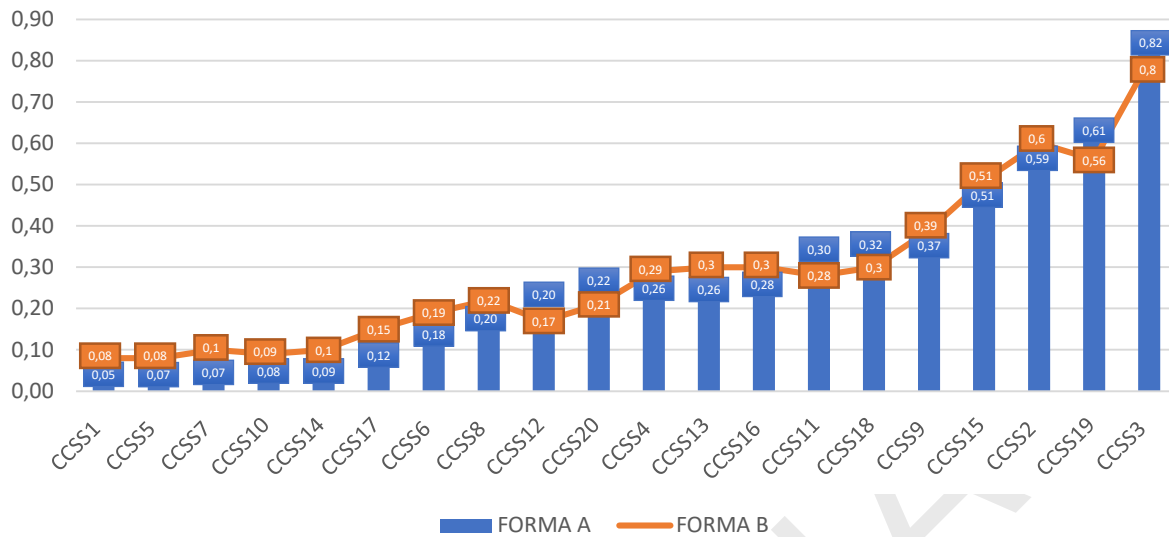
altamente difíciles. Podemos indicar que en esta área hay una distribución más pronunciada entre ítems difíciles y fáciles.

Figura 5. Índice de Dificultad de los ítems de español (Forma A y B) con TCT.



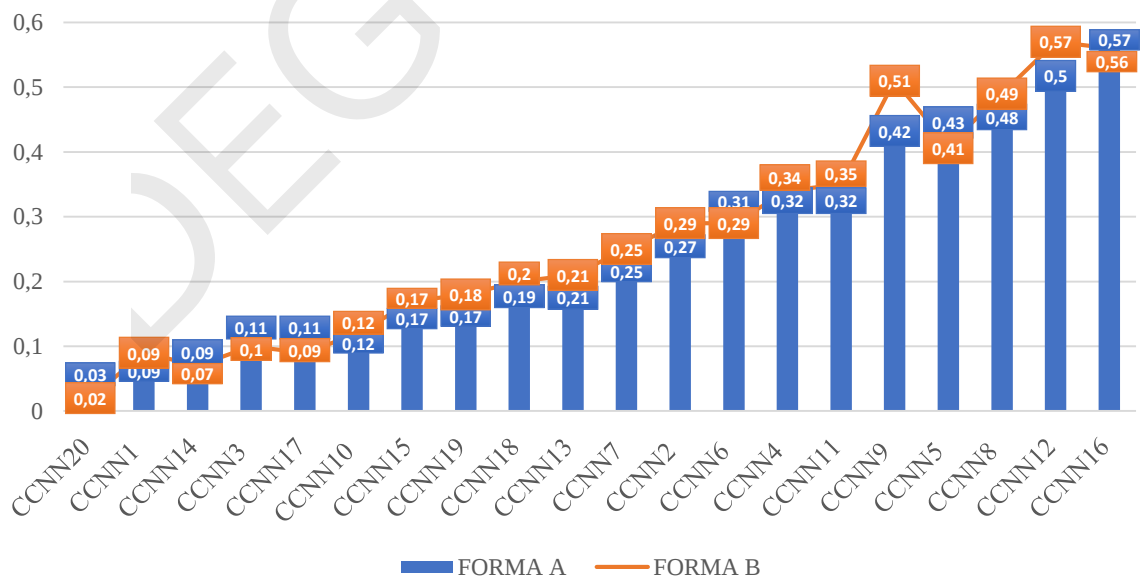
Nota. Fuente: Creación propia utilizando Excel.

En la figura 6 se observa que de los 20 ítems del área de CC.SS.: 1 es altamente fácil, 1 es medianamente fácil, 2 presentan una dificultad media, 9 de ellos son medianamente difíciles y 7 son altamente difíciles.

Figura 6. Índice de Dificultad Ciencias Sociales (Forma A y B) con TCT.

Nota. Fuente: Creación propia utilizando Excel.

En la figura 7 se muestra que de los 20 ítems de Ciencias Naturales 4 son de dificultad media, 6 medianamente difíciles y 10 son altamente difíciles. Es decir, la mayoría de los ítems se consideran con alto grado de complejidad y dificultad para poder ser contestados y en el caso en específico de ciencias naturales no se puntúa ningún ítem que sea fácil o medianamente fácil.

Figura 7. Índice de Dificultad Ciencias Naturales (Forma A y B) con TCT.

Nota. Fuente: Creación propia utilizando Excel.

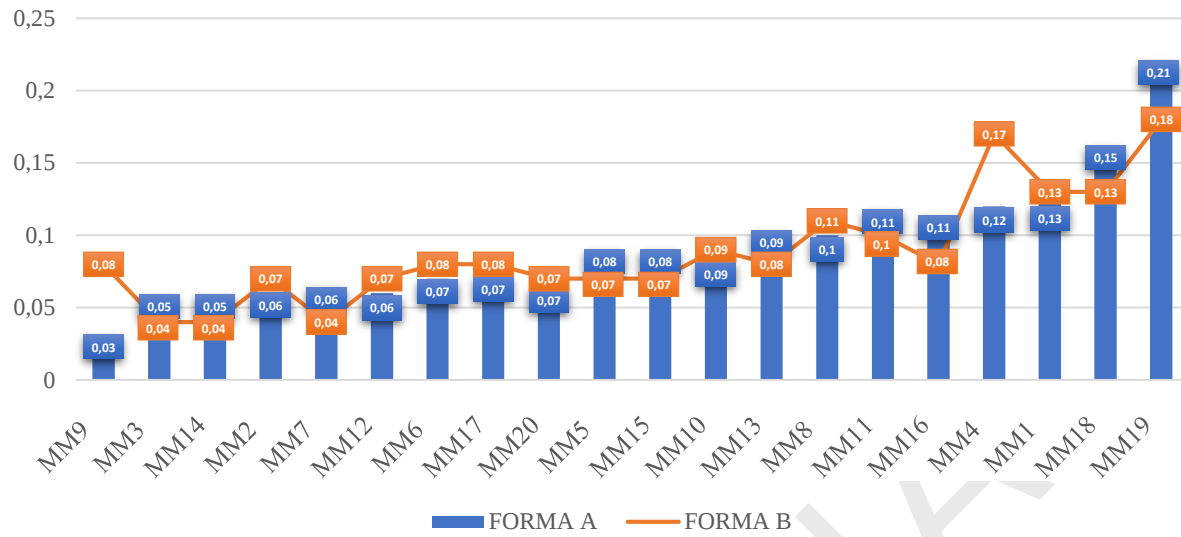
Ahora bien, al hacer un análisis del comportamiento de la dificultad de los 80 ítems en total que forman parte de la prueba de admisión, se puede evidenciar que un gran número de reactivos se agrupan entre el Índice de Dificultad igual o menor a 0.2, lo que indica que hay un número mucho mayor de reactivos que son mediana o altamente difíciles, y son muy pocos los que se encuentran con un Índice de Dificultad entre 0.4 y 0.8 con una dificultad media o alta.

Los ítems con estos valores extremos ($p = 0$ y $p = 1$) deben de ser descartados debido a que no contribuyen a medir las diferencias existentes entre los sujetos, todos los sujetos responden de la misma manera: incorrecta o correctamente (Navas et al., 2001). La prueba presenta alrededor de 36 ítems cercanos al extremo $p = 0$ considerados altamente difíciles y 4 ítems cercanos al extremo $p = 1$ considerados altamente fáciles, estos ítems deben de ser revisados para ser eliminados de la prueba, ya que un ítem que todos aciertan o que todos fallan no sirve para nada más que para invertir tiempo contestándolo (Meneses et al., 2013).

Discriminación. El tercer objetivo de este estudio de investigación se enmarca en la estimación de los índices de Discriminación de los ítems, con el propósito de identificar los ítems que distinguen entre aquellas personas que puntúan alto en el test y las que puntúan bajo, es decir, si discrimina entre los eficaces y los ineficaces (Muñiz Fernández, 2018).

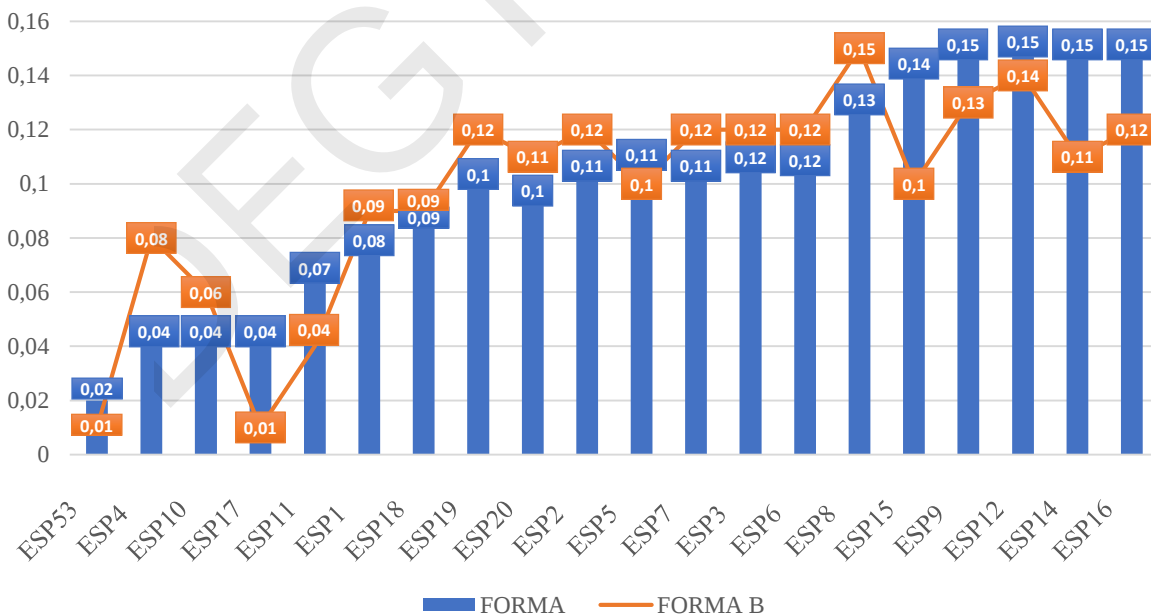
Se calculó el Índice de Discriminación (ID) de cada uno de los 80 ítems que forman parte de las dos formas de la prueba de admisión a las Escuelas Normales: Forma A y Forma B, de acuerdo con los componentes temáticos (matemáticas, español, ciencias naturales y ciencias sociales), considerando como fundamento la TCT, haciendo uso del programa RStudio. La interpretación del ID se refiere a que la capacidad discriminativa del ítem aumenta a medida que se aleja de cero, bien sea hacia 1 o hacia -1 . Si es 1, significa que todos los competentes aciertan el ítem ($P_c = 1$) y todos los incompetentes lo fallan ($P_i = 0$) (Muñiz Fernández, 2018).

En la figura 8 aparecen los 20 ítems del área de matemáticas: 1 discrimina poco y necesita revisión, 6 ítems no son adecuados y deben ser modificados o eliminados y 13 ítems según sus índices deben eliminarse directamente ya que sus valores son menores a 0.00.

Figura 8. Índice de Discriminación de los ítems de Matemáticas (Forma A y B) con TCT.

Nota. Fuente: Creación propia utilizando Excel.

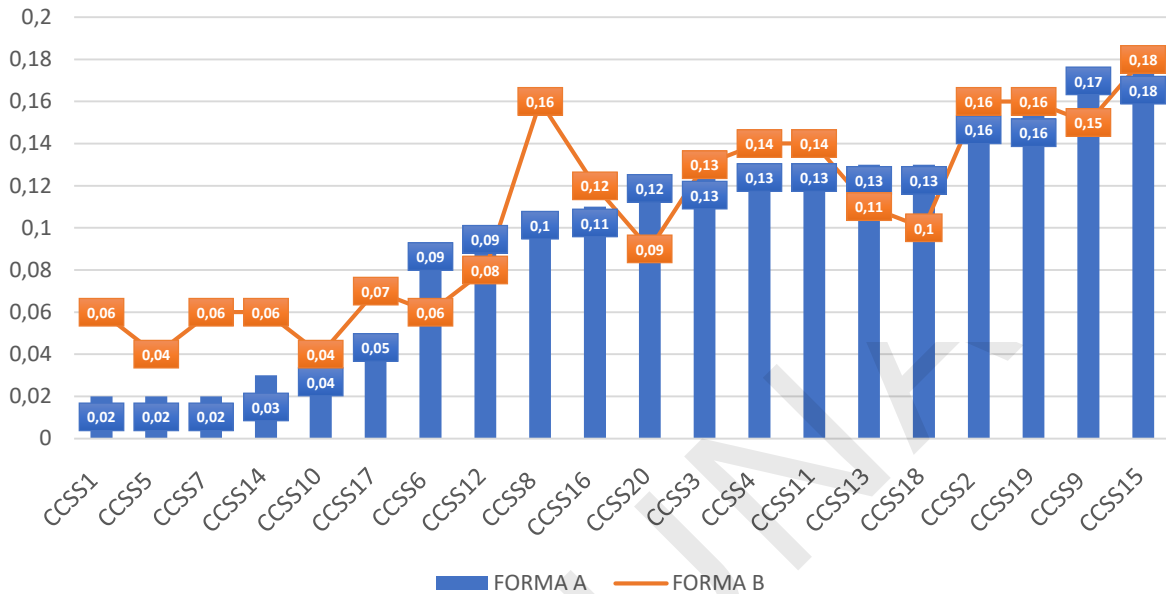
En la figura 9 se muestran los ítems del área de español 13 ítems no son adecuados y deben ser modificados o eliminados y 7 ítems deben eliminarse directamente. Tanto en el área de español y Matemáticas no se encontraron ítems con gran poder discriminativo, ni con una discriminación aceptable.

Figura 9. Índice de Discriminación de los ítems español (Forma A y B) con TCT.

Nota. Fuente: Creación propia utilizando Excel.

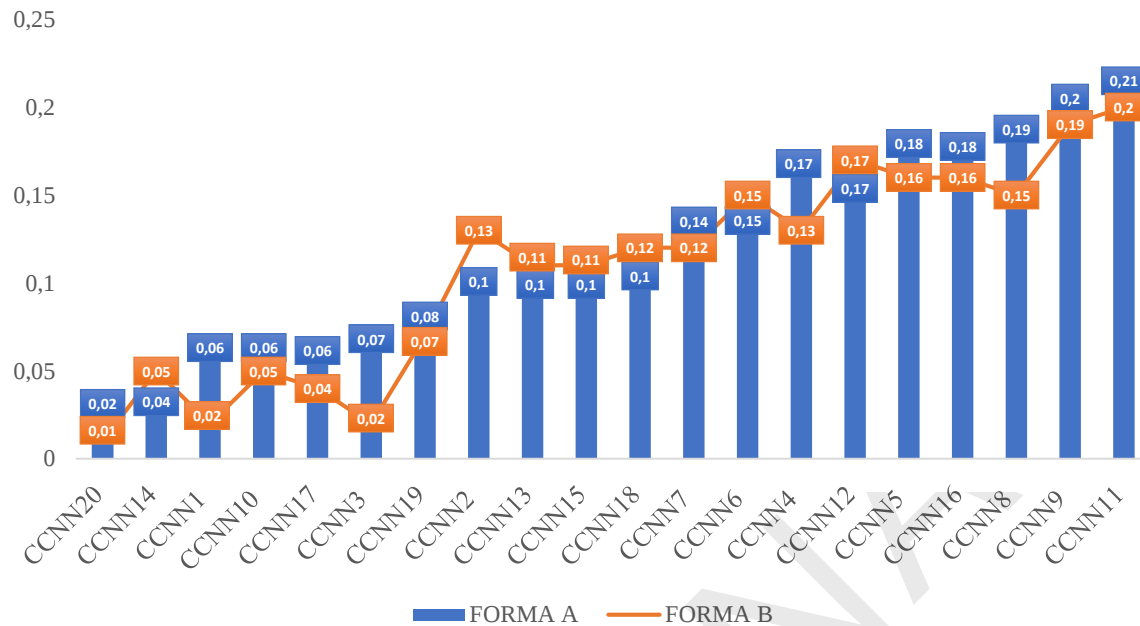
En la figura 10 respecto a los ítems del área de Ciencias Sociales, 12 ítems no son adecuados y deben ser modificados o eliminados y 8 ítems deben eliminarse directamente.

Figura 10. Índice de Discriminación de los ítems de SS.CC. (Forma A y B) con TCT.



Nota. Fuente: Creación propia utilizando Excel.

Finalmente, en la figura 11 se presentan los ítems del área de Ciencias Naturales, de los cuales 2 ítems discriminan un poco y necesitan revisión, 11 ítems no son adecuados y deben ser modificados o eliminados y 7 ítems deben eliminarse directamente. Al igual que en las áreas de español y matemáticas, en ciencias naturales y ciencias sociales no se encontraron ítems con índices aceptables de gran poder discriminativo o con una discriminación aceptable.

Figura 11. Índice de Discriminación de los ítems de CC.NN. (Forma A y B) con TCT.

Nota. Fuente: Creación propia utilizando Excel.

Al analizar el comportamiento de la discriminación de los 80 ítems en total que forman parte de la prueba de admisión, se puede demostrar que solamente 3 ítems discriminan poco y el resto no son adecuados y deben ser revisados o eliminados directamente de la prueba.

Lo anterior no nos abona resultados significativos ya que nos indica que la mayoría de personas que respondieron los ítems obtuvieron muy bajas puntuaciones ya que en términos claros un reactivo discriminará de manera eficaz si lo responden correctamente más sujetos con puntuación alta que sujetos con puntuaciones bajas; por el contrario, un reactivo será ineficaz si quienes más aciertan son los que obtienen bajas puntuaciones (López Meza, 2019); y esta prueba según los resultados obtenidos no cuenta con ítems que tengan un buen poder discriminativo.

5.2. Análisis de Dimensionalidad (Análisis Factorial Exploratorio).

Para cumplir con el objetivo número cuatro de este estudio de investigación, se corroboró el supuesto de la Unidimensionalidad de los ítems mediante un Análisis Factorial Exploratorio con el fin de evaluar la configuración interna de la prueba, el AFE fue aplicado a ambas formas de la prueba de Admisión (Forma A y Forma B).

Mediante el Programa RStudio se realizaron los cálculos de las pruebas estadísticas más adecuadas para el AFE con el propósito de evaluar la varianza explicada de cada una de las formas que se desprenden de la prueba. Primeramente, se evaluaron los supuestos del análisis factorial, para ello se aplicó el estadístico de Henze-Zirkler para evaluar la normalidad multivariante de los datos y se corroboró que no existe normalidad multivariante en ninguna de las dos formas de la prueba ($p = < 0.05$).

En la tabla 12 se muestran los índices obtenidos al corroborar la adecuación muestral de los datos en ambas formas de la prueba (Forma A y B), donde se utilizó la prueba de Esfericidad de Barlett cuyo resultado ($p = < 0.05$) contribuyó a rechazar la hipótesis nula de no correlación entre las variables ya que puede observarse que en ambas formas se obtuvieron valores inferiores al nivel de significancia ($,001$). El Índice de Medida de la adecuación muestral Kaiser-Meyer-Oikin (KMO) obtenido fue en la Forma A de 0.62 y en la Forma B de 0.66 lo cual es medianamente aceptable indicando que es un Modelo Regular; esto nos indica que los datos y la matriz de correlaciones si es adecuada para que pueda ser aplicado el Análisis Factorial Exploratorio.

Tabla 12. Resúmenes estadísticos KMO y Prueba de Esfericidad de Barlett

	Forma de la prueba	Prueba de KMO	Prueba de Esfericidad de Barlett
1	Forma A	0.62	0.001
2	Forma B	0.66	0.001

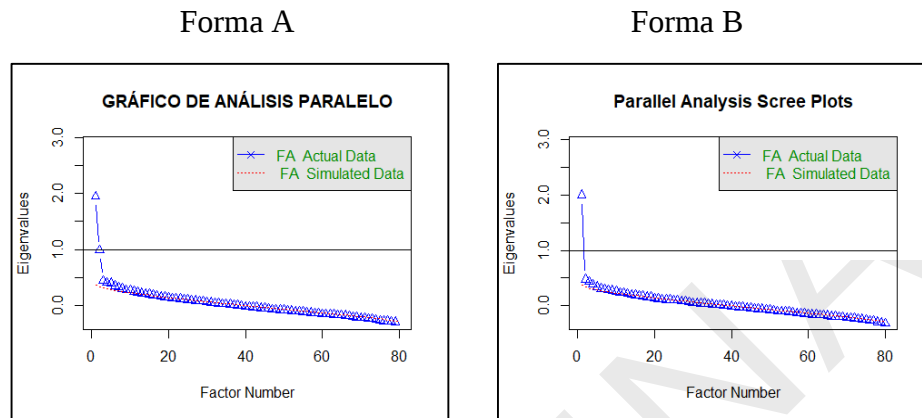
Nota. Fuente: Creación Propia

Horn propuso el análisis paralelo, que parece ser una de las mejores alternativas para decidir el número de factores a extraer. El análisis paralelo genera autovalores de una matriz de datos aleatorios, pero con el mismo número de variables y casos que la matriz original (E. Pérez y Medrano, 2010). Por lo cual se decidió realizar el análisis con la finalidad de conocer el número de factores a extraer de la matriz de correlación, haciendo uso del Análisis Paralelo de Horn, el Análisis Factorial realizado sugiere la presencia de 1 factor en la forma A y en la Forma B.

En la figura 12 de la Forma A y B se observa un autovalor igual a 2 lo que indica la presencia de un solo factor en ambas formas. Cabe señalar que existen varios métodos

recomendables para estimar el modelo AF entre ellos el de Mínimos Cuadrados Ordinarios (MCO) y Máxima Verosimilitud (MV); para desarrollar las etapas del análisis factorial exploratorio se hizo uso de los estimadores mínimos cuadrados ordinarios como método de extracción.

Figura 12. Gráficos de Análisis Paralelo de la Forma A y Forma B



Nota. Fuente: Creación Propia utilizando RStudio

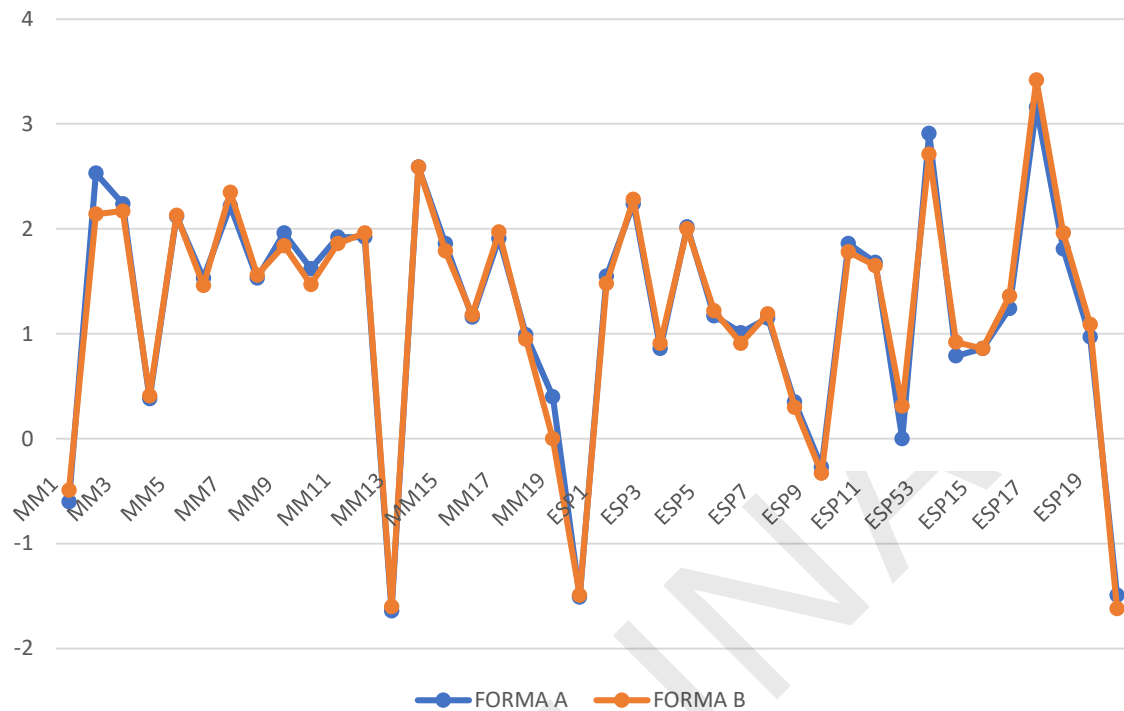
Los principales métodos basados en el criterio MCO son: (a) AF de ejes principales, (b) MINRES de Harman (Harman y Jones, 1966), (c) ULS de Jöreskog (1977), y (d) Residual Mínimo de Comrey (1962). Se hizo uso del estadístico MINRES porque no requiere la estimación inicial de comunilidades y es eficiente en términos de computación (Pere y Anguiano-Carrasco, 2010). El Análisis Factorial se realizó a partir de una matriz de correlación tetracórica debido a la naturaleza dicotómica de los ítems obteniendo un índice RMSR de 0.02 medidas de cuantía de error del modelo cuyos indicadores de un buen ajuste se verifican con valores inferiores a .05.

Existen varios métodos de rotación y se clasifican en rotaciones ortogonales (varimax, quartimax y equimax) u oblicuas (oblimin direct, promax, orthoblique, procrustes) (Pere y Anguiano, 2010), para efectos del análisis factorial de la prueba de admisión se utilizó un método de rotación ortogonal: Varimax. Finalmente, en el desarrollo del análisis se descartaron ítems debido a que su peso factorial era negativo, al descartar estos ítems en la solución final quedaron 51 de los 80 ítems analizados en la forma A y en la forma B quedaron 49 de los 80 ítems analizados, la Varianza común explicada del factor general obtenida fue de 0.34.

5.3. Cálculo de parámetros psicométricos con la TRI.

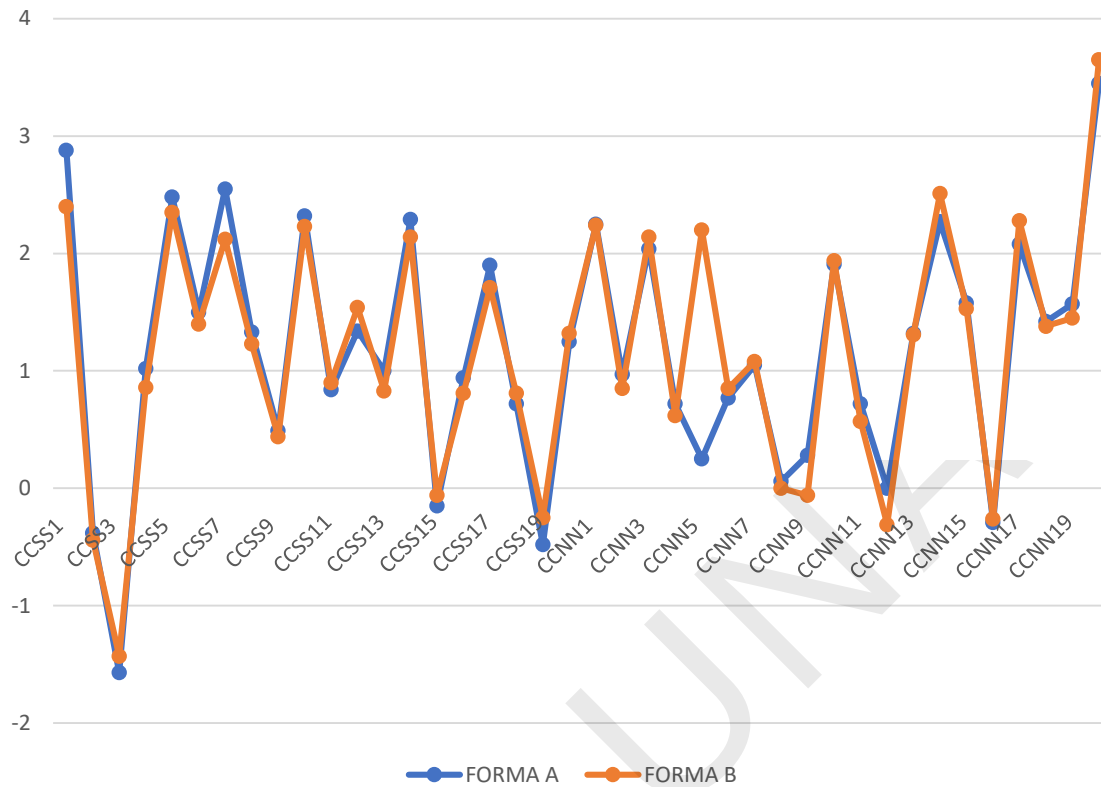
Para cumplir con el quinto objetivo que plantea calcular los índices de dificultad de los ítems de la prueba de admisión mediante la aplicación de la Teoría de Respuesta al Ítem (TRI), buscando de esta forma determinar su calidad métrica, primeramente, se utilizó el Modelo de un parámetro o Modelo de Rasch.

En la figura 13 se observan los valores de dificultad obtenidos a través del Modelo de Rasch aplicando la TRI para los primeros 20 ítems de la prueba correspondientes a las áreas de Matemáticas y español en ambas formas: A y B. El parámetro θ de dificultad, puede oscilar entre $-\infty$ y $+\infty$ (Attorresi et al., 2009). En la práctica la CCI se suele representar en una escala situada entre -3 y $+3$ puntos de habilidad. (Matas Terrón, 2010). Para los ítems de matemáticas y español se aprecian valores favorables considerando que los ítems de buena calidad no deben ser demasiado fáciles o difíciles, puede apreciarse una diferencia notoria en ítems como el 13 y el 20 de matemáticas, y el 19 y 20 de español que se encuentran con índices muy bajos es decir catalogados como muy difíciles (esto puede verse más a detalle en las tablas que se adjuntan en el apéndice).

Figura 13. Gráfico de Dificultad aplicando TRI a ítems de Matemáticas y Español

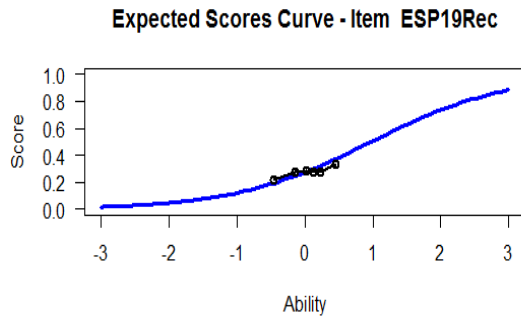
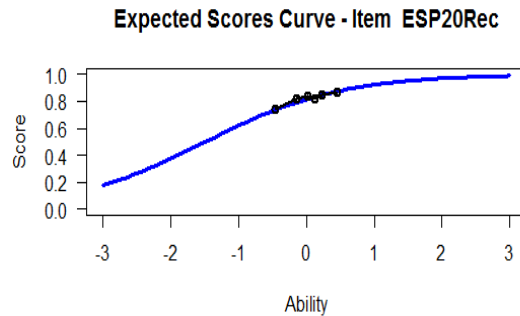
Nota. Fuente: Creación propia utilizando Excel.

En la figura 14 se muestra la dificultad tanto de la forma A como de la forma B para el segundo grupo de 20 ítems correspondiente a las áreas de Ciencias Sociales y Ciencias Naturales, la tendencia refleja valores satisfactorios a pesar de que algunos son bajos. De manera general los ítems 2, 3 y 19 de Ciencias Sociales y el 12 y 16 de Ciencias Naturales presentan una menor dificultad con índices muy bajos. Cabe destacar que en estas áreas se muestra que hay una gran cantidad de ítems que son demasiado difíciles y superan en gran medida a los ítems con una dificultad media ya que son muy pocos.

Figura 14. Gráfico de Dificultad aplicando TRI a ítems de CC. SS y CC. NN

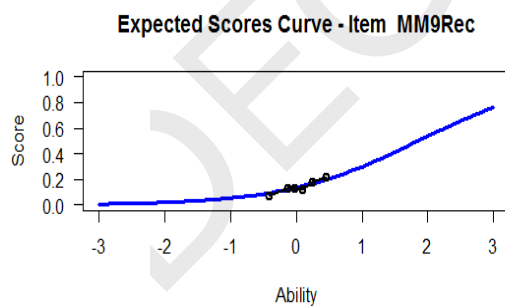
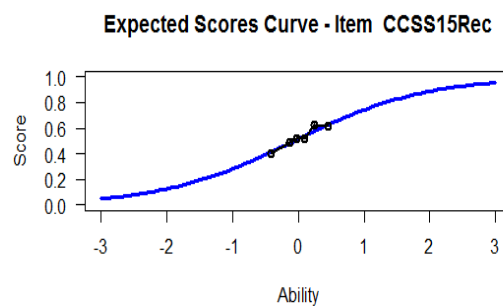
Nota. Fuente: Creación Propia utilizando Excel

En la figura 15 y 16 se puede observar una diferencia entre los ítems 19 y 20 del área de español. En la figura correspondiente al ítem 19 la curva muestra como aun un estudiante con un nivel de habilidad 3 (alta) no alcanzaría la probabilidad de 1 para acertar este ítem ya que el mismo denota una dificultad muy alta (1.09), de esta forma se observa el proceder de varios ítems con un comportamiento similar que puede observarse en el cuadro de resultados del aprendice. Mientras tanto en el ítem 20 de español se percibe como más fácil, reflejando en la curva como un estudiante con un nivel de habilidad 3 tendría la probabilidad de responder correctamente a un ítem al tener una dificultad inferior al (-1.49).

Figura 15. Curva Característica del Item19**Figura 16.** Curva Característica del Item20

Nota: La curva azul representa el valor esperado y la de color negro representa el valor observado.

En la figura 17 y 18 se puede observar que entre los ítems 9 y 15 del área de matemáticas y ciencias sociales respectivamente ocurre lo siguiente: en la figura correspondiente al ítem 9 de matemáticas se refleja en la curva como aun un estudiante con un nivel de habilidad 3 (alta), no tendría la probabilidad de 1 de responder correctamente a un ítem, este ítem muestra tener una dificultad de (-0.33); en cuanto al ítem 15 de Ciencias Sociales se percibe un comportamiento más fácil, reflejando en la curva como un estudiante con un nivel de habilidad 3 tendría la probabilidad más elevada de responder correctamente a un ítem al tener una dificultad inferior al ((-0.15) estos resultados pueden apreciarse mejor en las tablas que se adjuntan en el apéndice.

Figura 17. Curva Característica del Item9**Figura 18.** Curva Característica del Item15

Nota: La curva azul representa el valor esperado y la de color negro representa el valor observado.

En base a los resultados obtenidos al aplicar los estadísticos psicométricos en la tabla 13 se muestra que de los 80 ítems que contiene la prueba solamente 45 ítems alcanzan resultados mínimos que pueden indicar que son los ítems con mejores resultados psicométricos y pueden ser utilizados en una nueva versión de la prueba de admisión que pueda ser utilizada para

evaluar conocimientos en las cuatro áreas básicas en estudiantes que deseen ingresar a carreras del nivel medio en Honduras.

Tabla 13. Ítems mejor evaluados de la prueba de admisión.

Área	Ítem	Confiabilidad	Dificultad	Discriminación
Ciencias Sociales	CCSS2	0.57	0.59	0.16
	CCSS4	0.58	0.26	0.13
	CCSS8	0.60	0.20	0.10
	CCSS9	0.58	0.37	0.17
	CCSS11	0.58	0.30	0.13
	CCSS12	0.57	0.20	0.09
	CCSS13	0.58	0.26	0.13
	CCSS15	0.59	0.51	0.18
	CCSS16	0.60	0.28	0.11
	CCSS18	0.59	0.32	0.13
	CCSS19	0.61	0.61	0.16
	CCSS20	0.59	0.22	0.12
Total de ítems del área		12 ítems		
Matemáticas	MM1	0.78	0.64	0.13
	MM4	0.78	0.40	0.12
	MM8	0.78	0.17	0.10
	MM11	0.78	0.12	0.11
	MM16	0.78	0.23	0.11
	MM18	0.78	0.26	0.15
	MM19	0.77	0.40	0.21
Total de ítems del área		7 ítems		
Español	ESP2	0.42	0.23	0.11
	ESP3	0.41	0.24	0.12
	ESP5	0.39	0.29	0.11
	ESP6	0.43	0.26	0.12
	ESP7	0.42	0.70	0.11
	ESP8	0.41	0.45	0.13
	ESP9	0.42	0.56	0.15
	ESP11	0.43	0.27	0.07
	ESP12	0.42	0.44	0.15
	ESP14	0.43	0.31	0.15
	ESP15	0.41	0.29	0.14
	ESP16	0.40	0.22	0.15
	ESP19	0.42	0.29	0.10
Total de ítems del área		13 ítems		
Ciencias Naturales	CCNN2	0.85	0.27	0.10
	CCNN4	0.85	0.32	0.17
	CCNN5	0.84	0.43	0.18
	CCNN6	0.85	0.31	0.15
	CCNN7	0.85	0.25	0.14
	CCNN8	0.84	0.48	0.19
	CCNN9	0.84	0.42	0.20
	CCNN11	0.84	0.32	0.21
	CCNN12	0.84	0.50	0.17
	CCNN13	0.84	0.21	0.10
	CCNN15	0.84	0.19	0.10
	CCNN16	0.84	0.57	0.18
	CCNN18	0.85	0.20	0.10
Total de ítems del área		13 ítems		
Total de ítems en la prueba		45 ítems		

Nota. Fuente: Creación Propia

La tabla 13 también presenta la cantidad de ítems con mejores resultados psicométricos por área y se concluye que el área con menor cantidad de ítems con buenos resultados es el área de matemáticas donde solo 7 de 20 ítems pueden ser utilizados en una nueva versión de la prueba, seguidamente el área de Ciencias Sociales con 12 de 20 ítems y finalmente español y Ciencias Naturales pueden utilizar 13 de los 20 ítems aplicados en la prueba de admisión inicial.

Capítulo 6. Discusión de los Resultados

Este apartado contiene los enunciados más precisos que pretenden explicar los resultados obtenidos en el estudio de investigación denominado “Análisis Psicométrico de las áreas básicas de la prueba de Admisión a Escuelas Normales aplicado en el año 2011” y comparar estos con datos obtenidos de manera crítica, coherente y objetiva. Las posturas aquí expuestas toman como base la fundamentación teórica referente a la estimación de análisis psicométricos, de la cual se ha respaldado este estudio de investigación. En este espacio también se mencionan las limitantes que formaron parte del desarrollo de este estudio, así también de las recomendaciones que pueden ser tomadas en cuenta para futuros proyectos de investigación.

Este estudio tuvo como objetivo principal analizar los parámetros psicométricos de las áreas básicas de la prueba de Admisión a Escuelas Normales. Esta prueba está diseñada en tres formas: A, B y C, cada una consta de 80 ítems de selección única donde se presenta el enunciado con cuatro posibles opciones de respuesta, de las cuales una es correcta y las restantes son distractores, estas evalúan 4 áreas del conocimiento básicas como ser matemáticas, español, ciencias naturales y ciencias sociales, y fue aplicada a una población integrada por 5,523 jóvenes estudiantes que se postularon a la obtención de un cupo para cursar la carrera de Magisterio ofertada en el nivel de educación media en el año 2011.

Se aplicaron una serie de procedimientos psicométricos basados en la Teoría Clásica de los Test (TCT) y la Teoría de Respuesta al Ítem (TRI) logrando plasmar los siguientes enunciados de gran relevancia:

Confiabilidad. En cuanto a el cumplimiento del primer objetivo enmarcado en la estimación de la confiabilidad de la prueba de admisión, se realizó el cálculo haciendo uso de diferentes programas estadísticos y coeficientes entre ellos: Alpha de Cronbach, Omega de McDonald's, Alpha Ordinal y Omega Ordinal Tetracórico. En todos los coeficientes estimados la prueba muestra valores aceptables de confiabilidad, entre 0.71 y 0.86; siendo los valores más altos los reflejados a través del Alpha de Cronbach, Omega de McDonald's (0.82 a 0.86) y los coeficientes más bajos el Alpha Ordinal y Omega Ordinal Tetracórico (0.71 a 0.75).

En base a lo anterior, hay autores que señalan que el hecho de considerar que un índice no cumpla los supuestos necesarios para ser utilizado con el fin de tomar decisiones respecto a la confiabilidad de una dimensión, puede desembocar en conclusiones erróneas y afectar de manera significativa el desarrollo de una investigación (Espinoza y Novoa, 2018). Es por ello por lo que tomando en cuenta la No Normalidad de los datos y las características dicotómicas de los mismos, se considera que la estimación del Alpha Ordinal y Omega Ordinal Tetracórico son los estadísticos más adecuados a utilizar en el análisis de este estudio de investigación. No se descarta el uso del Alfa de Cronbach ya que es un estadístico de mucha utilidad, pero si se toma en consideración que una de sus limitaciones es que indica que está afectado por el número de ítems, el número de alternativas de respuesta y la proporción de la varianza del test (Ventura-León y Caycho-Rodríguez, 2014).

Finalmente se puede hacer mención que la consistencia interna no es una propiedad inherente de una escala, sino la medida del patrón de respuesta del instrumento en el grupo que responde la escala (Arias y Oviedo, 2008); esto se considera ya que como se puede apreciar en los gráficos que se encuentran en el Apéndice. Los resultados obtenidos en la prueba no fueron muy satisfactorios y esto puede repercutir en los resultados obtenidos al estimar los coeficientes de confiabilidad y otros estadísticos aplicados en el estudio.

Dificultad. En cuanto a el cumplimiento del segundo objetivo enmarcado en la estimación del Índice de Dificultad de la prueba de Admisión se puede mencionar que esta prueba está constituida en su mayoría por ítems altamente difíciles de responder, y esto se traduce en que la prueba es compleja para resolver.

Autores como (Navas et al., 2001) mencionan que la dificultad de los ítems de la prueba debe ser apropiada para la población de sujetos a la que ésta va destinada ya que solo así, permitirá al test diferenciar entre sujetos que poseen en distinto grado la característica que se está midiendo. Para efectos de los análisis realizados en este estudio, los resultados obtenidos por los aspirantes que respondieron el examen de admisión son en su mayoría muy bajos, y esto puede estar asociado al grado que poseen acerca de las características básicas que se están midiendo en la prueba como ser conocimientos básicos en las áreas de matemáticas, español, ciencias naturales y ciencias sociales.

Otra característica interesante de la estimación de la dificultad de los ítems es que proporciona una medida que permite la comparación de ítems que miden diferentes dominios o constructos, aplicados a los mismos sujetos. Así mismo informa de los ítems que presentan una dificultad o facilidad extrema, y que deben ser revisados o eliminados debido a que no contribuyen a la medida del constructo (Navas et al., 2001). En este caso como ya se ha reportado en los resultados obtenidos al estimar el (*ID*), la mayoría de los ítems deben ser revisados o eliminados ya que sus Índices de dificultad son muy altos (altamente fáciles) o muy bajos (altamente difíciles).

Discriminación. En cuanto a el cumplimiento del tercer objetivo enmarcado en la estimación del Índice de Discriminación de los ítems de la prueba de Admisión se puede mencionar que, esta prueba está constituida en su mayoría por ítems que no tienen un poder de discriminación aceptable porque sus índices reportados se encuentran por debajo del 0.3 que es el valor que posiciona a un ítem en un nivel de discrimina poco o aceptable, o que necesita una revisión.

Debe tomarse en consideración que el máximo valor del índice de discriminación de un ítem tiene estrecha relación con su índice de dificultad, ya que un reactivo con valor muy alto o bajo de índice de dificultad en realidad no discrimina, pues es solucionado por la mayoría de los estudiantes o por la minoría de ellos respectivamente (Díaz y Leyva, 2013).

Considerando que los ítems de esta prueba reportan índices de dificultad extremos (muy altos o bajos), se puede decir que estos resultados están estrechamente relacionados con los índices de discriminación ya que los ítems de esta prueba de admisión no discriminan y todos los ítems se encuentran en valores menores a 0.2 donde los ítems son considerados no adecuados y deben revisarse o eliminarse directamente.

Análisis Factorial Exploratorio AFP. En cuanto a el cumplimiento del cuarto objetivo enmarcado en la evaluación de la dimensionalidad de la prueba de admisión, podemos mencionar que las medidas de adecuación muestral KMO y prueba de esfericidad de Bartlett que se calcularon para este estudio de investigación mostraron valores meritorios. Para la prueba KMO de la Forma A fue de 0.62 y Forma B 0.66 lo cual es medianamente aceptable indicando que es un Modelo Regular; y la prueba de Esfericidad de Barlett contribuyó a

rechazar la hipótesis nula de no correlación entre las variables ya que ambas formas se obtuvieron valores inferiores al nivel de significancia ($,001$), con estos cálculos se concluyó que si es adecuado aplicar el Análisis Factorial Exploratorio.

Teoría de Respuesta al Ítem (Modelo de Rasch). El modelo de Rasch descansa sobre dos supuestos: 1) El atributo que se desea medir se puede representar en una única dimensión. 2) El nivel de la persona en el atributo (habilidad) y la dificultad del ítem determinan la probabilidad de obtener la respuesta correcta. (Jiménez y Montero, 2014). Primeramente, en este estudio se corroboraron estos supuestos para la aplicación del Modelo, y se afirmó que la prueba es unidimensional ya que al tomar en cuenta los índices de ajuste del Modelo de Rasch: INFINT y OUTFIT se obtuvieron valores igual a 1 que indican que la muestra es compatible con el modelo.

De un total de 80 ítems que forman parte de ambas formas (A y B) de la prueba de admisión, 45 ítems cumplen con los índices psicométricos mínimos que se requieren para ser considerados ítems aceptables; en cuanto a resultados obtenidos por estudiantes que contestaron la prueba solamente el 29.10% obtuvo una nota alta, considerando estos datos, es importante tomar en cuenta que los modelos matemáticos planteados en la TRI especifican que la probabilidad que tiene un evaluado de contestar correctamente un ítem depende de su nivel de aptitud y de las características de los ítems (Jiménez y Montero, 2014).

6.1. Limitaciones

Este estudio de investigación logró cumplir con los objetivos propuestos, sin embargo, se identificaron las siguientes limitaciones:

En primer lugar, la confusión generada por la inequidad entre el número de ítems que se encuentran en el documento de pauta de la prueba y el número de ítems codificados en la base de datos, ya que no se conocen cuales fueron incluidos en la prueba de admisión aplicada de acuerdo con la forma A y Forma B para estudiantes aspirantes a ingresar a las escuelas normales en el año 2011, y cuales fueron incluidos en la base de datos, sobre todo en el área de matemáticas y español, esta limitación fue superada al lograr identificar en la base de datos las formas que se aplicaron y los puntajes totales obtenidos haciendo una comparación con la cantidad de preguntas que se encontraban en las pautas de las pruebas en ambas formas.

En segundo lugar, no se cuenta con un informe técnico que indique procedimientos o elementos importantes para la creación de la base de datos con los resultados obtenidos por los estudiantes, tampoco se cuenta con un documento que explique generalidades del proceso de aplicación y resguardo de la prueba para su aplicación.

Finalmente, en la parte documental hay poco respaldo y acceso a estudios relacionados con la aplicación de exámenes de admisión en el nivel secundario a nivel regional y nacional.

6.2. Conclusiones

En conclusión, la prueba de Admisión a las Escuelas Normales aplicado a la muestra de estudiantes aspirantes en el año 2011 para optar a la carrera de magisterio en Honduras representa un instrumento con bondades significativas para evaluar aprendizajes de las áreas disciplinares básicas en el nivel de educación media, por lo cual lo que a continuación se resalta tiene con fin primordial fortalecer los procesos psicométricos en la elaboración de instrumentos que evalúen conocimientos básicos.

Uno de los elementos que se valoró para poder desarrollar este estudio de investigación fue el hecho de que a pesar de que las escuelas normales dejaron de existir en el país, el examen de admisión aplicado para ingresar a la carrera de magisterio es un instrumento que está construido con ítems basados en temáticas de conocimiento general del nivel de educación media, y al poder definir sus características psicométricas puede contribuir con aportes al área de la psicometría y la evaluación en el país.

La base de datos documentada en el caso de la aplicación de la prueba de admisión a las escuelas normales fue un excelente recurso que no se podía desestimar para poder ser utilizado y de esta manera considerar el análisis psicométrico que ha servido de base en este estudio de investigación.

Finalmente, el ámbito de la evaluación y la aplicación de instrumentos que valoren conocimientos a nivel estandarizado en el sistema educativo deben ser orientados y direccionados con lineamientos claros y coherentes, por tal motivo, es importante exhortar el impulso del área de la psicometría en el país, considerando que actualmente no es un área muy desarrollada y requiere un impulso grande para generar cambios oportunos en pro del mejoramiento de la calidad educativa.

6.3. Recomendaciones generales para estudios futuros.

En el marco del desarrollo del estudio de investigación sobre las propiedades psicométricas de la prueba de admisión a Escuelas Normales se recomienda lo siguiente:

Detallar los procedimientos logísticos, metodológicos y éticos utilizados para realizar la aplicación de un instrumento de gran importancia como lo es un examen de admisión.

En lo referente a la base de datos, es necesario incluir una nomenclatura y descripción mucho más detallada de las variables que se presentan para conocer con mayor claridad lo establecido de acuerdo con el instrumento aplicado.

Referencias

- Aiken, L. (2003). *Test psicológicos y evaluación* (Undécima edición). Pearson.
- Alas Solís, M. (2017). *Boletín Informativo: La Calidad de la Educación Básica en Honduras: Modestos avances en la última década*.
https://observatorio.upnfm.edu.hn/observatorio/wp-content/uploads/2018/09/Boletin_1_Observatorio_educativo_UPNFM.pdf
- Argibay, J. C. (2006). TECNICAS PSICOMETRICAS. CUESTIONES DE VALIDEZ Y CONFIABILIDAD. *Subjetividad Y Procesos Cognitivos* (8), 15–33.
- Arias, A. y Oviedo, H. (2008). Propiedades Psicométricas de una Escala: la Consistencia Interna. *Revista De Salud Pública*, 10(5), 831–839.
- Ato, M., López García, J. J. y Benavente, A. (2013). Un sistema de clasificación de los diseños de investigación en psicología. *Anales De Psicología*, 29(3), 1038–1059.
<https://doi.org/10.6018/analesps.29.3.178511>
- Attorresi, H. F., Lozzia, G. S., Abal, F. J. P., Galibert, M. S. y Aguerri, M. E. (2009). Teoría de Respuesta al Ítem. Conceptos básicos y aplicaciones para la medición de constructos psicológicos. *Revista Argentina De Clínica Psicológica*, XVIII(2), 179–188.
- Backhoff, E. (2018). Evaluación estandarizada de logro educativo: contribuciones y retos. *Revista Digital Universitaria*, 19(6).
- Barbero García, M. I. (2015). *Introducción a la Psicometría*. EDITORIAL SANZ Y TORRES, S. L.
- Bernal, C. A. (2010). *Metodología de la investigación: administración, economía, humanidades y ciencias sociales* (Tercera edición). Pearson.
- Bon, A. y Pini, M. (Eds.). (2019). *Cultura, Ciudadanía, Participación: Perspectivas de la Educación Inclusiva*. Pangea.
- Cardenoso Martínez, J., Cuesta, M. y Muñoz, J. (1996). Dimensionalidad y función de información de los test. *Psicothema*, 8(1), 215–220.
- Castillo, S. y Cabrerizo, J. (2010). *Evaluación educativa de aprendizajes y competencias*. Pearson.

- CEPAL. (2020). *COVID-19 y educación superior: de los efectos inmediatos al día después: Análisis de impactos, respuestas políticas y recomendaciones*. IESALC.
- Chiroleu, A. (2011). La educación superior en América Latina: ¿Problemas insolubles o recetas inadecuadas? *Campinas Sorocaba SP*, 16(3), 631–653.
<https://doi.org/10.1590/S1414-40772011000300008>
- Cortés, L. C. (2019). *Percepción de autoridades educativas y profesores del área de Letras de Educación media acerca de la PAA en cuatro colegios del municipio de Santa Rosa, Departamento de Copán* (24) [, Universidad Nacional Autónoma de Honduras]. CrossRef.
- Díaz, A. y Leyva, E. (2013). Metodología para determinar la calidad de los instrumentos de evaluación. *Educación Médica Superior*, 269–286.
- Doval, E. y Pinsach, J. (Eds.). (2019). *Cuadernos de Docencia Universitaria: Buenas prácticas en el uso de pruebas de alternativa múltiple* (Primera edición). OCTAEDRO, S.L.
- Espinoza, S. y Novoa, F. (2018). Ventajas del alfa ordinal respecto al alfa de Cronbach ilustradas con la encuesta AUDIT-OMS [Advantages of ordinal alpha versus Cronbach's alpha, illustrated using the WHO AUDIT test Vantagens do alfa ordinal em relação ao alfa de Cronbach verificadas na pesquisa AUDIT-OMS]. *Revista panamericana de salud publica = Pan American journal of public health*, 42, e65.
<https://doi.org/10.26633/RPSP.2018.65>
- FEREMA. (2017). *Informe de Progreso Educativo Honduras: Educacion: Una Deuda Pendiente*.
- Flores, J. (2017). La importancia de la Evaluación para la mejora de la educación y así obtener calidad educativa. *Atlante: Cuadernos De Educación Y Desarrollo*.
- Frías Navarro, D. (2021). *Recomendaciones para elaborar el informe de investigación: Discusión*. Open Science Framework. <https://doi.org/10.17605/OSF.IO/DFP94>
- Godino, J. D. (2010). Marcos Teóricos sobre el conocimiento y el aprendizaje matemático.
- González Llaneza, F. M. (2007). *Instrumentos de Evaluación Psicológica*. Ciencias Médicas.

- Hernández Rodríguez, R., Moncada Godoy, G. y Alas Solís, M. (2018a). *Boletín Informativo No. 7: ¿Cómo está la educación hondureña? Una visión desde los padres de familia* (núm. 7). UPNFM.
- Hernández Rodríguez, R., Moncada Godoy, G. y Alas Solís, M. (2018b). *Boletín No. 3: La deuda social en Honduras y los desafíos actuales y futuros que tiene el país con el financiamiento en educación*. UPNFM.
- Hernández Sampieri, R., Fernández Collado, C. y Baptista Lucio, P. (Eds.). (2010). *Metodología de la investigación* (Sexta edición). McGraw-Hill.
- Hervis, E. (2018). El desempeño del docente como factor asociado a la calidad educativa en América Latina. *Revista Educación*, 42(2), 717–739.
<https://doi.org/10.15517/revedu.v42i2.27033>
- Hidalgo Montesinos, M. y French, B. (2016). Una introducción didáctica a la Teoría de Respuesta al Ítem para comprender la construcción de escalas. *Revista De Psicología Clínica Con Niños Y Adolescentes*, 3(2), 13–21.
- IESALC. (2020). *Hacia el acceso universal a la educación superior: tendencias internacionales*.
- Jiménez, K. y Montero, E. (2014). Aplicación del modelo de Rasch, en el análisis psicométrico de una prueba de diagnóstico en matemática. *Revista Digital: Matemática, Educación E Internet*, 13(1). <https://doi.org/10.18845/rdmei.v13i1.1628>
- Jornet Meliá, J. M. (2017). Evaluación Estandarizada. *Revista Iberoamericana De Evaluación Educativa*, Artículo 10, 5–8.
- Lloret-Segura, S., Ferreras, A., Hernández-Baeza, A. y Marco, I. (2014). El análisis factorial exploratorio de los ítems: una guía práctica, revisada y actualizada. *Psicología*, 30(3). <https://doi.org/10.6018/analesps.30.3.199361>
- López Meza, A. (2019). Análisis Psicométrico de los Exámenes de Habilitación para el Ejercicio Profesional aplicados por el CEAACES. *Revista Ecuatoriana De Medicina EUGENIO ESPEJO*, 7(9), 37–45.
- Manzi, J. y San Martín, E. (2003). La necesaria complementariedad entre Teoría Clásica de la medición (TCM) y Teoría de Respuesta al Ítem (IRT): Aspectos conceptuales y aplicaciones. *Estudios Públicos*(90).

- Martínez, Rosa María, Pendás, Leonel, Martínez, M., Pérez, A., Cánovas y Ana María (2009). EL COEFICIENTE DE CORRELACION DE LOS RANGOS DE SPEARMAN CARACTERIZACION. *Revista Habanera De Ciencias Médicas*, 8(2).
- Martínez, C. y Sepúlveda, M. A. (2012). Introducción al análisis factorial exploratorio. *Revista Colombiana De Psiquiatría*, 41(1), 197–207. [https://doi.org/10.1016/S0034-7450\(14\)60077-9](https://doi.org/10.1016/S0034-7450(14)60077-9)
- Martínez Arias, M. R., Hernández Lloreda, M. J. y Hernández Lloreda, M. V. (Eds.). (2006). *Psicometría* (Edición en versión digital). Alianza Editorial.
- Martínez Rizo, F. (2008). La evaluación de aprendizajes en América Latina: Cuadernos de Investigación No. 32. *Resultados De Investigación*.
- Matas Terrón, A. (2010). Introducción al análisis de la Teoría de Respuesta al Ítem. *Aidesoc Formación Y Desarrollo*.
- Meneses, J., Barrios, M., Bonillo Albert, Cusculluela, A., Lozano, L. M., Turbany, J. y Valero, S. (Eds.). (2013). *Psicometría* (1º edición en lengua castellana). Editorial UOC.
- MIDEH. (2013). *Honduras MIDEH Project 2011-2016: Quarterly report for april through june*. USAID.
- Monzón Troncoso, M. Y. (2015). Evaluación del aprendizaje: un recorrido histórico y epistemológico. *ACADEMICUS*, 1(6).
- Mora Vargas, A. I. (2013). La evaluación educativa: concepto, períodos y modelos. *Actualidades Investigativas En Educación*, 4(2). <https://doi.org/10.15517/aie.v4i2.9084>
- Moreno Olivos, T. (2016). *Evaluación del aprendizaje y para el aprendizaje: Reinventar la evaluación en el aula*. Universidad Autónoma Metropolitana Unidad Cuajimalpa División de Ciencias de la Comunicación y Diseño. <https://doi.org/Moreno>
- Mullis, I., Martín, M., Ruddock, G., O'Sullivan, C. y Preuschoff, C. (2012). *TIMSS 2011 Marcos de la evaluación*. Madrid.
- Muñiz Fernández, J. (2018). *Introducción a la Psicometría: Teoría clásica y TRI* (1º edición electrónica). Ediciones Pirámide (Grupo Anaya, S. A.).
- Navas, M. J., Fidalgo, Á., Gabriel, C., Suárez, J. C., Brioso, Á., Gil, G., Martínez Arias, R. y Sarriá Encarnación. (2001). *Métodos, diseños y técnicas de investigación Psicológica: Psicopedagogía* (Primera edición). Fernández Ciudad, S. L.
- Navas Ara, M. J. (2001). *Métodos, diseños y técnicas de investigación psicológica*. UNED.

- OCDE. (2017). *Marco de Evaluación y de Análisis de PISA para el Desarrollo: Lectura, matemáticas y ciencias*. París.
- Odell, S. y Sucre, F. (2015). *El estado de las políticas públicas docentes en Centroamérica y República Dominicana*. PREAL.
- OECD. (2018). *PISA para el desarrollo: Resultados en Foco*. Programa de Evaluación Internacional de estudiantes.
- Oyanedel, J. C. (2017). Cálculo de Confiabilidad a través del uso del coeficiente Omega de McDonald. *Revista Medica De Chile*(145), 269–274.
- Padilla, M. R. (2019). *Boletín No. 7: Análisis de las posibilidades de lograr la universalización en la educación y el cumplimiento del Objetivo de Desarrollo Sostenible 4, en el contexto de la agenda 2030* (núm. 7).
- Pere, J. F. (1996). EVALUACIÓN DE LA UNIDIMENSIONALIDAD DE LOS ITEMS MEDIANTE ANÁLISIS FACTORIAL. *Psicothema*, 8, 397–410.
- Pere, J. F. y Anguiano, C. (2010). El análisis factorial como técnica de investigación en psicología: Centro de investigación para la evaluación y medida de la conducta. Universidad 'Rovira i Virgili', 3(1), 18–33.
- Pere, J. F. y Anguiano-Carrasco, C. (2010). El análisis factorial como técnica de investigación en psicología. *Papeles Del Psicólogo*, 31(1), 18–33.
- Pérez, E. y Medrano, L. (2010). Análisis Factorial Exploratorio: Bases Conceptuales y Metodológicas. *Revista Argentina De Ciencias Del Comportamiento RACC*, 2(1), 58–66.
- Pérez, J., Acuña, N. y Arratia, E. (2008). Nivel de dificultad y poder de discriminación del tercer y quinto examen parcial de la cátedra de cito-histología 2007 de la carrera de medicina de la UMSA. *Cuadernos*, 53(2).
- Pimienta Prieto, J. H. (2008). *Evaluación de los Aprendizajes: Un enfoque basado en competencias* (Primera edición). Pearson.
- Ramos Vega, Z. (2018). *Psicometría básica*. Fundación Universitaria del Área Andina. <https://doi.org/10.33132/9789585462793>
- Rápalo Castellanos, R. (2012). *Resultados Nacionales: Participación de Honduras en los estudios internacionales TIMSS y PIRLS 2011*. Honduras.

- Rojas, A. (2020). Desafíos en calidad y cobertura de la educación pública de Honduras 2014-2018. *Economía Y Administración (E&A)*, 11(2), 9–24.
<https://doi.org/10.5377/eya.v11i2.10517>
- Sánchez Sánchez, I. (2015). *Conceptos básicos de la metodología de la investigación*. Secretaría de Educación (Ed.). (2006). *Diseño curricular nacional para la educación básica 3*.
- Reglamentos Ley Fundamental de Educación, Diario Oficial la Gaceta (2014).
- Soler Cárdenas, S. (2008). Coeficientes de confiabilidad de instrumentos escritos en el marco de la teoría clásica de los tests. *Escuela Nacional De Salud Pública*,
- Sousa, V. D., Driessnack, M. y Mendes, I. A. C. (2007). Revisión de diseños de Investigación resaltantes para enfermería. Parte 1: Diseños de Investigación Cuantitativa. *Revista Latinoamericana De Enfermagem*, 15(3), 502–507.
<https://doi.org/10.1590/s0104-11692007000300022>
- UNESCO. (2015). *Informe de Resultados: Antecedentes iniciales: Relato TERCE*.
- UNESCO. (2016). *Recomendaciones de Políticas Educativas en América Latina en base al TERCE*. Santiago de Chile. (OREALC/UNESCO).
- Ventura-León, J. L. y Caycho-Rodríguez, T. (2014). El coeficiente Omega: un método alternativo para la estimación de la confiabilidad. *Revista Latinoamericana De Ciencias Sociales, Niñez Y Juventud*, 12(1), 291–308.
<https://doi.org/10.11600/1692715x.12117101613>
- Viquez, L. D. (2016). Diagnóstico Situacional de la Formación de Docentes en Honduras. *Revista Electrónica Calidad En La Educación Superior*, 7(1), 205–232.
<https://doi.org/10.22458/caes.v7i1.1439>
- Virla Quero, M. (2010). Confiabilidad y coeficiente Alpha de Cronbach. *Telos*, 12(2), 248–252.

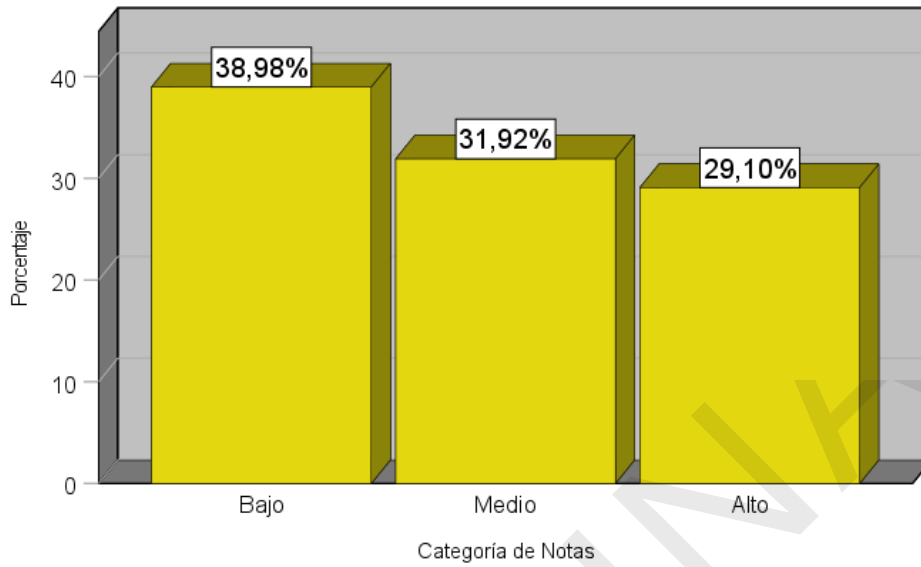
Declaración ética de la Investigación

Esta investigación ha sido realizada tomando en cuenta el código de ética de la investigación. De tal manera se afirma que no se reportan conflictos de interés.

DEGT-UNAH

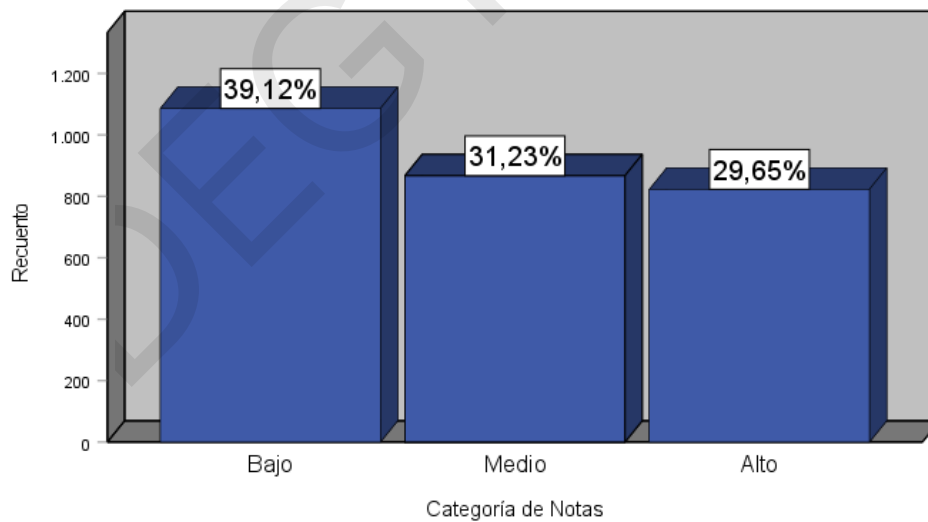
Apéndice

Figura 19. Porcentaje de resultados obtenidos por estudiantes en prueba de admisión



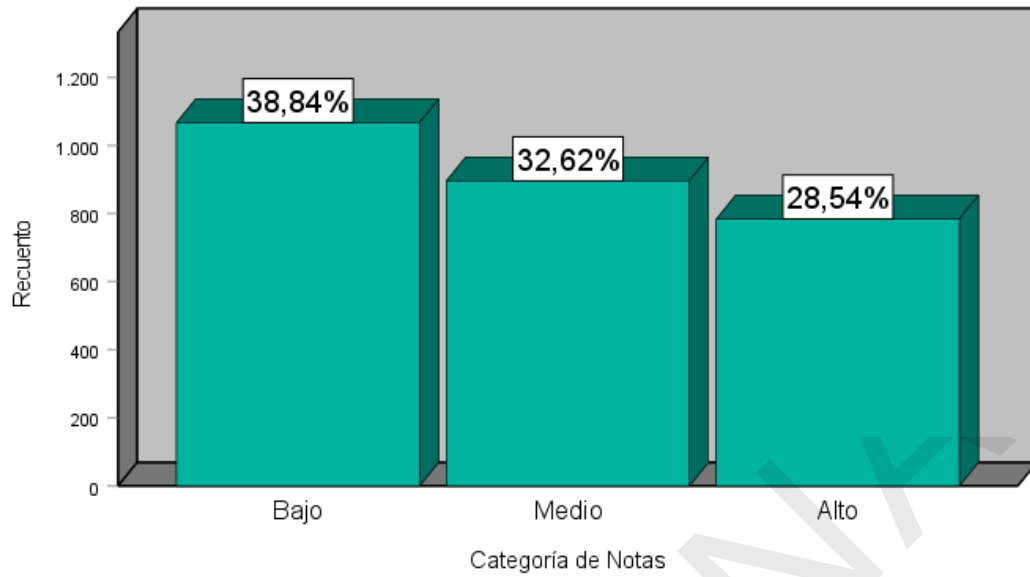
Nota. Fuente: Creación Propia utilizando RStudio

Figura 20. Porcentaje de resultados obtenidos en prueba de admisión Forma A



Nota. Fuente: Creación Propia utilizando RStudio

Figura 21. Porcentaje de resultados obtenidos en prueba de admisión Forma B



Nota. Fuente: Creación Propia utilizando RStudio

Tabla 14. Confiabilidad de los ítems utilizando Alpha de Cronbach y Omega de McDonald.

Forma / Componente Temático	Forma A		Forma B	
	Alpha de Cronbach (α)	Omega de McDonald (ω)	Alpha de Cronbach (α)	Omega de McDonald (ω)
Forma Total	0.866	0.860	0.835	0.829
Ciencias Sociales	0.612	0.608	0.841	0.844
CCSS_1	0.607	0.603	0.834	0.837
CCSS_2	0.575	0.574	0.833	0.837
CCSS_3	0.598	0.590	0.834	0.837
CCSS_4	0.585	0.581	0.835	0.838
CCSS_5	0.587	0.583	0.833	0.837
CCSS_6	0.598	0.593	0.836	0.839
CCSS_7	0.616	0.613	0.837	0.840
CCSS_8	0.606	0.601	0.837	0.840
CCSS_9	0.589	0.585	0.832	0.835
CCSS_10	0.605	0.602	0.832	0.835
CCSS_11	0.588	0.584	0.831	0.835
CCSS_12	0.576	0.575	0.832	0.835
CCSS_13	0.591	0.587	0.831	0.835
CCSS_14	0.618	0.615	0.835	0.839
CCSS_15	0.600	0.596	0.834	0.838
CCSS_16	0.609	0.604	0.833	0.836
CCSS_17	0.623	0.617	0.834	0.838
CCSS_18	0.601	0.595	0.834	0.837
CCSS_19	0.621	0.615	0.835	0.839
CCSS_20	0.598	0.594	0.834	0.838
Matemáticas	0.786	0.791	0.717	0.716
MM_21	0.781	0.787	0.708	0.707
MM_22	0.780	0.785	0.702	0.702
MM_23	0.772	0.778	0.704	0.703
MM_24	0.782	0.788	0.705	0.704
MM_25	0.777	0.782	0.699	0.698
MM_26	0.781	0.787	0.709	0.708
MM_27	0.778	0.783	0.703	0.702
MM_28	0.779	0.784	0.707	0.706
MM_29	0.775	0.780	0.703	0.701
MM_30	0.776	0.781	0.707	0.705
MM_31	0.776	0.781	0.708	0.707
MM_32	0.771	0.776	0.701	0.699
MM_33	0.776	0.777	0.711	0.707
MM_34	0.785	0.788	0.723	0.718
MM_35	0.770	0.776	0.696	0.696
MM_36	0.784	0.789	0.719	0.720
MM_37	0.780	0.784	0.712	0.710
MM_38	0.780	0.785	0.713	0.711
MM_39	0.763	0.771	0.685	0.687
MM_40	0.783	0.788	0.716	0.717
Español	0.429	0.432	0.406	0.403
ESP_41	0.406	0.411	0.388	0.387
ESP_42	0.416	0.422	0.387	0.385
ESP_43	0.413	0.417	0.398	0.392
ESP_44	0.412	0.415	0.390	0.386
ESP_45	0.388	0.398	0.381	0.381
ESP_46	0.431	0.434	0.409	0.404

ESP_47	0.418	0.424	0.388	0.385
ESP_48	0.406	0.414	0.363	0.369
ESP_49	0.420	0.420	0.401	0.397
ESP_50	0.423	0.428	0.415	0.413
ESP_51	0.435	0.434	0.417	0.408
ESP_52	0.419	0.423	0.400	0.399
ESP_53	0.416	0.414	0.404	0.401
ESP_54	0.431	0.435	0.405	0.401
ESP_55	0.412	0.416	0.382	0.482
ESP_56	0.393	0.400	0.368	0.372
ESP_57	0.418	0.420	0.399	0.395
ESP_58	0.417	0.417	0.397	0.390
ESP_59	0.418	0.421	0.385	0.379
ESP_60	0.424	0.427	0.397	0.389
Ciencias Naturales	0.853	0.855	0.614	0.611
CCNN_61	0.852	0.854	0.614	0.611
CCNN_62	0.850	0.852	0.599	0.595
CCNN_63	0.848	0.850	0.593	0.590
CCNN_64	0.848	0.850	0.592	0.590
CCNN_65	0.845	0.847	0.601	0.598
CCNN_66	0.848	0.850	0.594	0.592
CCNN_67	0.848	0.850	0.606	0.603
CCNN_68	0.847	0.849	0.606	0.602
CCNN_69	0.845	0.847	0.597	0.592
CCNN_70	0.846	0.848	0.598	0.593
CCNN_71	0.846	0.848	0.599	0.595
CCNN_72	0.842	0.844	0.598	0.594
CCNN_73	0.846	0.848	0.604	0.600
CCNN_74	0.842	0.844	0.588	0.584
CCNN_75	0.846	0.848	0.583	0.581
CCNN_76	0.844	0.846	0.620	0.619
CCNN_77	0.845	0.846	0.597	0.593
CCNN_78	0.849	0.851	0.609	0.606
CCNN_79	0.847	0.849	0.610	0.606
CCNN_80	0.853	0.854	0.623	0.618

Nota. Fuente: Creación Propia utilizando RStudio

Tabla 15. Dificultad y Discriminación de los ítems de Ciencias Sociales con TCT.

Ítem	Forma A		Forma B	
	Dificultad	Discriminación	Dificultad	Discriminación
CCSS1	0.05	0.02	0.08	0.06
CCSS2	0.59	0.16	0.60	0.16
CCSS3	0.82	0.13	0.80	0.13
CCSS4	0.26	0.13	0.29	0.14
CCSS5	0.07	0.02	0.08	0.04
CCSS6	0.18	0.09	0.19	0.06
CCSS7	0.07	0.02	0.10	0.06
CCSS8	0.20	0.10	0.22	0.16
CCSS9	0.37	0.17	0.39	0.15
CCSS10	0.08	0.04	0.09	0.04
CCSS11	0.30	0.13	0.28	0.14
CCSS12	0.20	0.09	0.17	0.08
CCSS13	0.26	0.13	0.30	0.11
CCSS14	0.09	0.03	0.10	0.06
CCSS15	0.51	0.18	0.51	0.18
CCSS16	0.28	0.11	0.30	0.12
CCSS17	0.12	0.05	0.15	0.07
CCSS18	0.32	0.13	0.30	0.10
CCSS19	0.61	0.16	0.56	0.16
CCSS20	0.22	0.12	0.21	0.09

Nota. Fuente: Creación Propia utilizando RStudio

Tabla 16. Dificultad y Discriminación de los ítems de Matemáticas con TCT.

Ítem	Forma A		Forma B	
	Dificultad	Discriminación	Dificultad	Discriminación
MM1	0.64	0.13	0.62	0.13
MM2	0.07	0.06	0.10	0.07
MM3	0.09	0.05	0.10	0.04
MM4	0.40	0.12	0.39	0.17
MM5	0.10	0.08	0.10	0.07
MM6	0.17	0.07	0.18	0.08
MM7	0.09	0.06	0.08	0.04
MM8	0.17	0.10	0.17	0.11
MM9	0.12	0.03	0.13	0.08
MM10	0.16	0.09	0.18	0.09
MM11	0.12	0.11	0.13	0.10
MM12	0.12	0.06	0.12	0.07
MM13	0.83	0.09	0.83	0.08
MM14	0.06	0.05	0.06	0.04
MM15	0.13	0.08	0.14	0.07
MM16	0.23	0.11	0.23	0.08
MM17	0.12	0.07	0.12	0.08
MM18	0.26	0.15	0.27	0.13
MM19	0.40	0.21	0.40	0.18
MM20	0.82	0.07	0.81	0.07

Nota. Fuente: Creación Propia utilizando RStudio

Tabla 17. Dificultad y Discriminación de los ítems de español con TCT.

Ítem	Forma A		Forma B	
	Dificultad	Discriminación	Dificultad	Discriminación
ESP1	0.17	0.08	0.18	0.09
ESP2	0.23	0.11	0.22	0.12
ESP3	0.24	0.12	0.28	0.12
ESP4	0.11	0.04	0.11	0.08
ESP5	0.29	0.11	0.22	0.10
ESP6	0.26	0.12	0.28	0.12
ESP7	0.70	0.11	0.73	0.12
ESP8	0.45	0.13	0.42	0.15
ESP9	0.56	0.15	0.58	0.13
ESP10	0.13	0.04	0.14	0.06
ESP11	0.27	0.07	0.26	0.04
ESP12	0.44	0.15	0.42	0.14
ESP53	0.05	0.02	0.06	0.01
ESP14	0.31	0.15	0.28	0.11
ESP15	0.29	0.14	0.29	0.10
ESP16	0.22	0.15	0.20	0.12
ESP17	0.04	0.04	0.03	0.01
ESP18	0.13	0.09	0.12	0.09
ESP19	0.29	0.10	0.25	0.12
ESP20	0.81	0.10	0.83	0.11

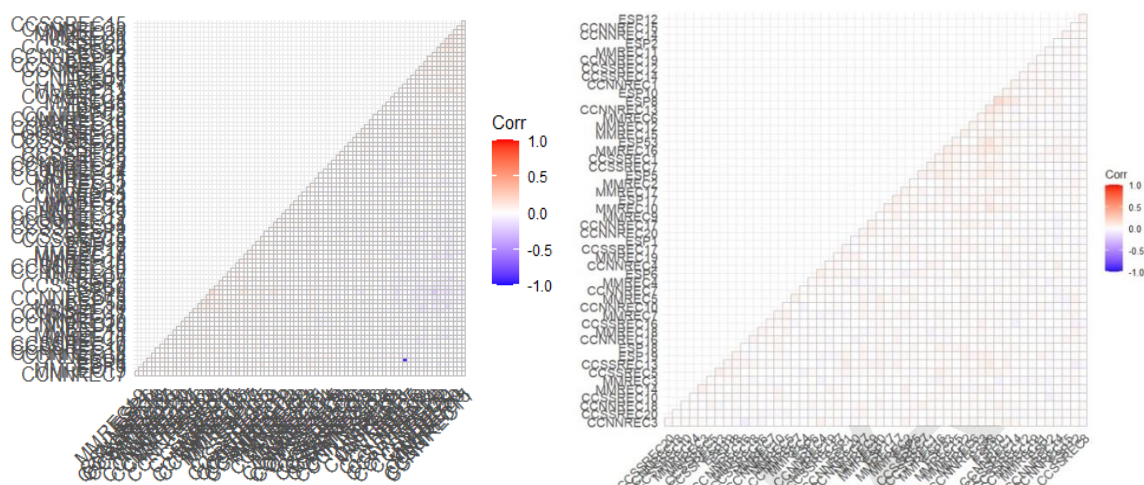
Nota. Fuente: Creación Propia utilizando RStudio

Tabla 18. Dificultad y Discriminación de los ítems de Ciencias Naturales con TCT.

Ítem	Forma A		Forma B	
	Dificultad	Discriminación	Dificultad	Discriminación
CCNN1	0.09	0.06	0.09	0.02
CCNN2	0.27	0.10	0.29	0.13
CCNN3	0.11	0.07	0.10	0.02
CCNN4	0.32	0.17	0.34	0.13
CCNN5	0.43	0.18	0.41	0.16
CCNN6	0.31	0.15	0.29	0.15
CCNN7	0.25	0.14	0.25	0.12
CCNN8	0.48	0.19	0.49	0.15
CCNN9	0.42	0.20	0.51	0.19
CCNN10	0.12	0.06	0.12	0.05
CCNN11	0.32	0.21	0.35	0.20
CCNN12	0.50	0.17	0.57	0.17
CCNN13	0.21	0.10	0.21	0.11
CCNN14	0.09	0.04	0.07	0.05
CCNN15	0.17	0.10	0.17	0.11
CCNN16	0.57	0.18	0.56	0.16
CCNN17	0.11	0.06	0.09	0.04
CCNN18	0.19	0.10	0.20	0.12
CCNN19	0.17	0.08	0.18	0.07
CCNN20	0.03	0.02	0.02	0.01

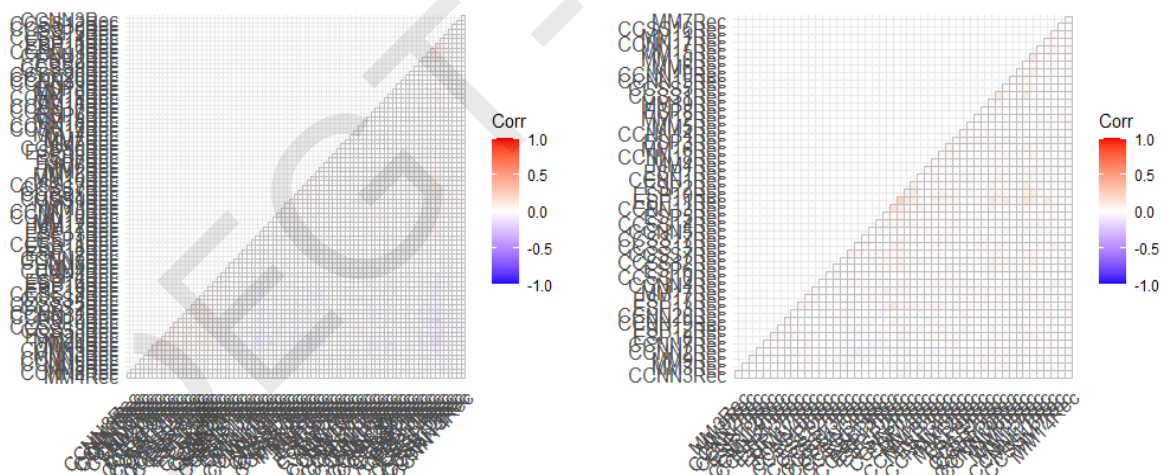
Nota. Fuente: Creación Propia utilizando RStudio

Figura 22. Matriz de Correlación inicial y final de la Forma A



Nota. Fuente: Creación Propia utilizando RStudio

Figura 23. Matriz de Correlación inicial y final de la Forma B



Nota. Fuente: Creación Propia utilizando RStudio

Tabla 19. Carga Factorial rotada de los ítems de la Forma A

Ítem	λ	h^2	u^2
ESP8	0.39	1.5e-01	0.85
MMREC1	-0.37	1.4e-01	0.86
MMREC20	-0.34	1.2e-01	0.88
ESP9	-0.32	1.0e-01	0.90
CCSSREC2	-0.28	7.6e-02	0.92
ESP10	0.27	7.2e-02	0.93
ESP20	-0.26	7.0e-02	0.93
CCSSREC18	-0.25	6.4e-02	0.94
CCNNREC11	-0.25	6.2e-02	0.94
MMREC10	0.24	5.9e-02	0.94
ESP14	-0.24	5.7e-02	0.94
MMREC13	-0.23	5.5e-02	0.94
MMREC16	0.23	5.3e-02	0.95
ESP53	0.23	5.2e-02	0.95
CCNNREC13	0.22	4.9e-02	0.95
ESP7	-0.21	4.4e-02	0.96
ESP5	0.20	3.9e-02	0.96
ESP6	0.19	3.8e-02	0.96
CCNNREC1	0.19	3.7e-02	0.96
CCSSREC7	0.19	3.5e-02	0.96
ESP16	-0.19	3.5e-02	0.96
CCSSREC15	-0.18	3.2e-02	0.97
MMREC17	0.18	3.1e-02	0.97
CCNNREC9	-0.18	3.1e-02	0.97
MMREC15	0.17	2.9e-02	0.97
CCNNREC8	-0.16	2.7e-02	0.97
CCSSREC5	0.16	2.4e-02	0.98
MMREC6	0.15	2.3e-02	0.98
ESP11	-0.15	2.2e-02	0.98
MMREC12	0.15	2.2e-02	0.98
CCSSREC1	0.15	2.1e-02	0.98
ESP19	0.14	2.0e-02	0.98
CCNNREC12	-0.14	2.0e-02	0.98
CCNNREC17	0.14	1.8e-02	0.98
CCSSREC13	0.13	1.6e-02	0.98
CCSSREC14 1	0.13	1.6e-02	0.98
MMREC7	0.12	1.4e-02	0.99
ESP18	0.12	1.3e-02	0.99
CCSSREC3	-0.11	1.2e-02	0.99
CCSSREC19	-0.11	1.2e-02	0.99
CCNNREC6	-0.11	1.2e-02	0.99
CCSSREC17	0.10	1.1e-02	0.99
CCNNREC5	-0.10	1.0e-02	0.99
CCSSREC11	-0.09	8.8e-03	0.99
CCNNREC10	0.09	8.1e-03	0.99
CCNNREC18	0.09	7.6e-03	0.99
ESP1	0.08	7.2e-03	0.99
CCSSREC12	0.08	6.2e-03	0.99
ESP17	0.08	5.8e-03	0.99
MMREC8	-0.07	5.5e-03	0.99
MMREC11	0.07	5.4e-03	0.99
MMREC5	0.07	4.8e-03	1.00

ESP15	-0.07	4.6e-03	1.00
CCNNREC3	0.07	4.6e-03	1.00
MMREC14	0.06	4.0e-03	1.00
CCNNREC14	0.06	3.6e-03	1.00
CCNNREC15	0.06	3.5e-03	1.00
MMREC9	0.06	3.5e-03	1.00
CCSSREC6	-0.05	2.9e-03	1.00
MMREC2	0.05	2.8e-03	1.00
CCSSREC10	0.05	2.8e-03	1.00
ESP12	0.05	2.2e-03	1.00
CCNNREC19	0.04	1.6e-03	1.00
CCNNREC2	-0.04	1.5e-03	1.00
CCSSREC4	-0.04	1.5e-03	1.00
CCNNREC7	0.04	1.4e-03	1.00
CCNNREC16	0.03	1.2e-03	1.00
CCSSREC16	0.03	1.0e-03	1.00
ESP2	0.03	1.0e-03	1.00
CCNNREC20	0.02	5.3e-04	1.00
ESP3	-0.02	5.0e-04	1.00
CCSSREC9	-0.02	3.4e-04	1.00
CCNNREC4	0.02	3.2e-04	1.00
MMREC18	0.01	1.8e-04	1.00
MMREC4	0.01	1.0e-04	1.00
ESP4	-0.01	7.2e-05	1.00
MMREC19	0.01	5.5e-05	1.00
CCSSREC8	0.01	4.2e-05	1.00
CCSSREC20	0.01	3.6e-05	1.00
MMREC3	0.00	1.6e-07	1.00

Nota. Fuente: Creación Propia utilizando RStudio

Tabla 20. Carga Factorial rotada de los ítems de la Forma A eliminando ítems.

No.	λ	h2	u2
ESP8	0.40	1.6e-01	0.84
ESP10	0.27	7.4e-02	0.93
MMREC10	0.25	6.3e-02	0.94
MMREC16	0.24	5.9e-02	0.94
ESP53	0.24	5.6e-02	0.94
CCNNREC13	0.24	5.5e-02	0.94
ESP5	0.23	5.1e-02	0.95
CCNNREC1	0.21	4.4e-02	0.96
MMREC15	0.19	3.6e-02	0.96
MMREC17	0.17	3.0e-02	0.97
CCSSREC7	0.16	2.6e-02	0.97
MMREC12	0.16	2.5e-02	0.97
MMREC6	0.15	2.4e-02	0.98
CCSSREC13	0.15	2.2e-02	0.98
CCNNREC17	0.15	2.1e-02	0.98
CCSSREC5	0.14	2.1e-02	0.98
ESP18	0.14	2.0e-02	0.98
CCSSREC1	0.14	1.9e-02	0.98
CCSSREC14	0.13	1.7e-02	0.98
ESP19	0.13	1.6e-02	0.98
MMREC7	0.11	1.3e-02	0.99
CCNNREC18	0.11	1.2e-02	0.99
ESP6	0.11	1.2e-02	0.99
CCSSREC17	0.11	1.1e-02	0.99
ESP17	0.10	1.1e-02	0.99
MMREC5	0.10	1.0e-02	0.99
MMREC2	0.09	7.4e-03	0.99
CCNNREC3	0.08	6.9e-03	0.99
CCNNREC10	0.08	6.6e-03	0.99
MMREC11	0.08	6.4e-03	0.99
MMREC14	0.08	6.2e-03	0.99
ESP1	0.07	5.6e-03	0.99
CCNNREC15	0.07	4.4e-03	1.00
MMREC19	0.07	4.3e-03	1.00
CCNNREC14	0.06	4.2e-03	1.00
CCSSREC12	0.06	3.8e-03	1.00
CCNNREC16	0.06	3.3e-03	1.00
CCNNREC19	0.06	3.2e-03	1.00
CCNNREC4	0.05	3.0e-03	1.00
CCNNREC7	0.05	2.2e-03	1.00
CCSSREC10	0.05	2.0e-03	1.00
ESP2	0.04	1.9e-03	1.00
MMREC9	0.04	1.8e-03	1.00
MMREC18	0.04	1.7e-03	1.00
CCNNREC20	0.04	1.6e-03	1.00
CCSSREC16	0.04	1.6e-03	1.00
ESP12	0.03	1.0e-03	1.00
CCSSREC20	0.02	2.6e-04	1.00
CCSSREC8	0.01	8.9e-05	1.00
MMREC4	0.00	1.3e-05	1.00
MMREC3	0.00	9.1e-07	1.00

Nota. Fuente: Creación Propia utilizando RStudio

Tabla 21. Carga Factorial rotada de los ítems de la Forma B

Ítem	λ	h2	u2
ESP11Rec	0.44	2.0e-01	0.80
MM1Rec	-0.41	1.7e-01	0.83
MM20Rec	-0.36	1.3e-01	0.87
ESP9Rec	-0.34	1.2e-01	0.88
ESP10Rec	0.34	1.1e-01	0.89
MM10Rec	0.26	6.5e-02	0.93
CCNN11Rec	-0.25	6.4e-02	0.94
ESP13Rec	0.25	6.3e-02	0.94
CCNN9Rec	-0.25	6.1e-02	0.94
CCSS2Rec	-0.24	5.9e-02	0.94
ESP20Rec	-0.23	5.3e-02	0.95
CCSS18Rec	-0.22	5.0e-02	0.95
CCNN13Rec	0.22	4.6e-02	0.95
MM13Rec	-0.21	4.6e-02	0.95
ESP14Rec	-0.21	4.5e-02	0.96
ESP2Rec	0.21	4.2e-02	0.96
MM16Rec	0.19	3.6e-02	0.96
MM15Rec	0.19	3.5e-02	0.96
CCSS3Rec	-0.18	3.4e-02	0.97
CCSS19Rec	-0.18	3.3e-02	0.97
CCNN1Rec	0.18	3.3e-02	0.97
CCNN6Rec	-0.18	3.3e-02	0.97
CCSS1Rec	0.18	3.1e-02	0.97
CCNN12Rec	-0.17	2.8e-02	0.97
CCSS7Rec	0.17	2.8e-02	0.97
CCSS5Rec	0.16	2.6e-02	0.97
CCSS15Rec	-0.16	2.6e-02	0.97
MM17Rec	0.16	2.5e-02	0.97
CCNN18Rec	0.16	2.4e-02	0.98
CCSS14Rec	0.15	2.1e-02	0.98
ESP19Rec	-0.14	1.9e-02	0.98
ESP3Rec	0.14	1.9e-02	0.98
MM6Rec	0.13	1.6e-02	0.98
ESP6Rec	0.13	1.6e-02	0.98
CCNN15Rec	0.12	1.5e-02	0.98
ESP16Rec	-0.11	1.3e-02	0.99
CCNN8Rec	-0.11	1.3e-02	0.99
CCSS13Rec	0.11	1.2e-02	0.99
ESP17Rec	0.11	1.2e-02	0.99
ESP1Rec	0.10	9.3e-03	0.99
ESP18Rec	0.10	9.1e-03	0.99
CCSS12Rec	0.09	8.9e-03	0.99
ESP5Rec	0.08	7.2e-03	0.99
CCNN17Rec	0.08	6.7e-03	0.99
MM12Rec	0.08	6.5e-03	0.99
ESP12Rec	0.08	5.8e-03	0.99
CCNN14Rec	0.08	5.8e-03	0.99
MM11Rec	0.08	5.7e-03	0.99
CCSS4Rec	-0.07	5.4e-03	0.99
CCNN10Rec	0.07	5.2e-03	0.99
ESP15Rec	-0.07	5.0e-03	0.99
CCNN3Rec	0.07	4.8e-03	1.00

CCNN4Rec	0.07	4.7e-03	1.00
CCSS16Rec	0.07	4.5e-03	1.00
CCSS10Rec	0.06	4.1e-03	1.00
MM7Rec	0.05	2.8e-03	1.00
ESP4Rec	-0.05	2.8e-03	1.00
CCSS11Rec	-0.05	2.3e-03	1.00
CCSS6Rec	-0.05	2.2e-03	1.00
CCSS17Rec	0.04	1.8e-03	1.00
CCSS11Rec	-0.05	2.3e-03	1.00
CCNN19Rec	0.04	1.8e-03	1.00
MM18Rec	-0.04	1.5e-03	1.00
ESP8Rec	0.04	1.4e-03	1.00
CCSS9Rec	-0.04	1.3e-03	1.00
MM9Rec	0.03	1.2e-03	1.00
CCNN5Rec	0.03	1.2e-03	1.00
CCNN2Rec	0.03	1.1e-03	1.00
MM2Rec	-0.03	1.0e-03	1.00
MM4Rec	0.03	9.9e-04	1.00
MM5Rec	0.03	9.8e-04	1.00
MM14Rec	0.03	6.9e-04	1.00
MM19Rec	0.03	6.7e-04	1.00
CCNN16Rec	-0.03	6.6e-04	1.00
CCSS20Rec	-0.03	6.4e-04	1.00
CCSS8Rec	-0.02	5.9e-04	1.00
CCNN7Rec	0.02	5.1e-04	1.00
ESP7Rec	-0.02	4.9e-04	1.00
MM8Rec	-0.02	3.1e-04	1.00
CCNN20Rec	0.01	1.4e-04	1.00
MM3Rec	0.00	2.5e-05	1.00

Nota. Fuente: Creación Propia utilizando RStudio

Tabla 22. Carga Factorial rotada de los ítems de la Forma B eliminando ítems.

Ítem	λ	h2	u2
ESP11Rec	0.45	0.20	0.80
SP10Rec	0.37	0.13	0.86
MM10Rec	0.27	0.07	0.93
CCNN13Rec	0.25	0.06	0.94
ESP13Rec	0.22	0.04	0.95
MM15Rec	0.21	0.04	0.96
ESP2Rec	0.20	0.04	0.96
CCSS1Rec	0.20	0.03	0.96
CCNN1Rec	0.19	0.03	0.97
MM16Rec	0.16	0.02	0.97
CCSS7Rec	0.16	0.02	0.97
CCSS5Rec	0.16	0.02	0.98
MM17Rec	0.15	0.02	0.98
CCSS14Rec	0.14	0.02	0.98
CCNN18Rec	0.14	0.01	0.98
ESP6Rec	0.12	0.01	0.98
MM11Rec	0.12	0.01	0.99
ESP3Rec	0.12	0.01	0.99
CCNN15Rec	0.11	0.01	0.99
CCSS12Rec	0.11	0.01	0.99
MM6Rec	0.11	0.01	0.99
ESP1Rec	0.10	0.01	0.99
CCSS13Rec	0.10	0.01	0.99
ESP18Rec	0.10	0.01	0.99
ESP5Rec	0.09	0.008	0.99
CCNN14Rec	0.09	0.008	0.99
MM12Rec	0.09	0.007	0.99
CCNN10Rec	0.08	0.006	0.99
CCSS16Rec	0.07	0.005	0.99
CCNN17Rec	0.07	0.005	0.99
CCNN4Rec	0.07	0.005	0.99
ESP12Rec	0.07	0.004	1.00
ESP17Rec	0.06	0.004	1.00
MM19Rec	0.06	0.003	1.00
CCNN3Rec	0.06	0.003	1.00
CCSS10Rec	0.06	0.003	1.00
CCNN2Rec	0.05	0.002	1.00
CCNN5Rec	0.05	0.002	1.00
MM7Rec	0.05	0.002	1.00
CCNN19Rec	0.05	0.002	1.00
CCSS17Rec	0.04	0.001	1.00
MM9Rec	0.04	0.001	1.00
MM4Rec	0.04	0.001	1.00
ESP8Rec	0.03	0.000	1.00
MM5Rec	0.02	0.000	1.00
MM14Rec	0.02	0.000	1.00
MM3Rec	-0.02	0.000	1.00
CCNN7Rec	-0.01	0.000	1.00
CCNN20Rec	0.01	0.000	1.00

Nota. Fuente: Creación Propia utilizando RStudio

Tabla 23. Índice de Dificultad basado en la TRI

Ítem	Forma A	Forma B
	I. Dificultad	I. Dificultad
CCSS1	2.88	2.40
CCSS2	-0.38	-0.44
CCSS3	-1.57	-1.43
CCSS4	1.02	0.86
CCSS5	2.48	2.35
CCSS6	1.50	1.40
CCSS7	2.55	2.12
CCSS8	1.33	1.23
CCSS9	0.49	0.44
CCSS10	2.32	2.23
CCSS11	0.84	0.90
CCSS12	1.34	1.54
CCSS13	1.00	0.83
CCSS14	2.29	2.14
CCSS15	-0.15	-0.06
CCSS16	0.94	0.81
CCSS17	1.90	1.71
CCSS18	0.72	0.81
CCSS19	-0.48	-0.25
CCSS20	1.25	1.32
MM1	-0.60	-0.49
MM2	2.53	2.14
MM3	2.24	2.17
MM4	0.38	0.41
MM5	2.12	2.13
MM6	1.53	1.46
MM7	2.22	2.35
MM8	1.53	1.56
MM9	1.96	1.84
MM10	1.62	1.47
MM11	1.92	1.86
MM12	1.92	1.96
MM13	-1.64	-1.60
MM14	2.59	2.59
MM15	1.86	1.79
MM16	1.16	1.18
MM17	1.91	1.97
MM18	0.99	0.95
MM19	0.40	0.40
MM20	-1.51	-1.49
ESP1	1.55	1.48
ESP2	2.24	2.28
ESP3	0.86	0.91
ESP4	2.02	2.00
ESP5	1.17	1.22
ESP6	1.01	0.91
ESP7	1.15	1.19
ESP8	0.35	0.30
ESP9	-0.27	-0.33
ESP10	1.86	1.78
ESP11	1.68	1.65

ESP12	0-22	0.31
ESP53	2.91	2.71
ESP14	0.79	0.92
ESP15	0.86	0.86
ESP16	1.24	1.36
ESP17	3.16	3.42
ESP18	1.81	1.96
ESP19	0.97	1.09
ESP20	-1.49	-1.62
CCNN1	2.25	2.24
CCNN2	0.97	0.85
CCNN3	2.04	2.14
CCNN4	0.72	0.62
CCNN5	0.25	2.20
CCNN6	0.77	0.85
CCNN7	1.05	1.08
CCNN8	0.06	0.00
CCNN9	0.28	-0.06
CCNN10	1.91	1.94
CCNN11	0.72	0.57
CCNN12	-0.00	-0.31
CCNN13	1.32	1.31
CCNN14	2.27	2.51
CCNN15	1.58	1.53
CCNN16	-0.29	-0.26
CCNN17	2.08	2.28
CCNN18	1.42	1.38
CCNN19	1.57	1.45
CCNN20	3.45	3.65

Nota. Fuente: Creación Propia utilizando RStudio