

DEGT-UNAH

UNIVERSIDAD NACIONAL AUTÓNOMA DE HONDURAS
FACULTAD DE CIENCIAS SOCIALES
ESCUELA DE CIENCIAS PSICOLÓGICAS
MAESTRÍA EN PSICOMETRÍA Y EVALUACIÓN EDUCATIVA



TESIS

**Análisis Psicométrico de la Prueba de Ciencias Naturales del TERCE en la
Muestra Aplicada a los Estudiantes Hondureños en el año 2013.**

Presentada Por

Sohany Alejandra Cáceres Perdomo

Para optar al grado de Máster en Psicometría y Evaluación Educativa

Asesor Temático: Mario Alas Solís, PhD

Asesor Metodológico: German Edgardo Moncada, PhD.

Tegucigalpa, M.D.C., Honduras, C. A.

Junio-2022

Tegucigalpa, M.D.C., 30 de junio de 2022

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 la Prueba de Ciencias Naturales del TERCE en la Muestra Aplicada a los Estudiantes Hondureños en el año 2013”, el cual he presentado como parte de los requisitos para optar por la Maestría en Psicometría y Evaluación Educativa.

Atentamente,

Sohany Alejandra Cáceres Perdomo

Cuenta No. PSE100207

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

Agradecimiento

Le agradezco a Dios por brindarme sabiduría, paciencia y fortaleza a lo largo de este camino, aún en los momentos de mayor dificultad.

Agradezco a la Maestría en Psicometría y Evaluación Educativa a cada uno de sus profesores, académicos y administrativos por el conocimiento brindado a lo largo de procesos formativos desarrollados en tiempos difíciles.

Agradezco a mis asesores PhD. Mario Alas, PhD. German Moncada por brindarme sus consejos, conocimientos y buen criterio académico.

A mis seres queridos y a mis compañeras Ana Ortiz y Wendoly Cruz por su apoyo y ánimo constante.

Resumen

En Honduras las pruebas estandarizadas para medir el desempeño académico han tenido mucho impacto para alcanzar las metas propuestas en educación, puesto que las mismas están alineadas con el currículo, los estándares y el sistema de evaluación nacional. Pese a que los estudios internacionales gozan de buena reputación, el contexto hondureño parece no estar familiarizado con los procedimientos que se utilizan en la elaboración de los instrumentos utilizados, por tanto el objetivo general de este estudio es validar la prueba de logros de aprendizajes en ciencias naturales para estudiantes de sexto grado del TERCE a través del análisis de procedimientos psicométricos que permitan la recopilación de evidencias que justifiquen su uso y las interpretaciones de los puntajes y brindar una experiencia que ejemplifique el cumplimiento de los estándares adecuados en el diseño de las pruebas. El análisis psicométrico consistió primeramente, en un análisis de confiabilidad mediante el cálculo del coeficiente alpha de Cronbach, seguidamente se hizo un análisis de dimensionalidad de la prueba por medio de análisis de varianza a través del análisis factorial exploratorio, posteriormente se estimaron los parámetros psicométricos de los ítems, dificultad y discriminación abordados en base a la teoría clásica de los test (TCT) y la teoría de respuesta al ítem (TRI) para cada uno de estos procedimientos se hizo uso de los programas Rstudio y Spss. Al estimar los parámetros psicométricos, se obtuvieron resultados satisfactorios para la mayoría los ítems que se incluyeron en la prueba, no obstante, los índices de discriminación mostraron una diferencia significativa entre las muestras nacional y regional. Respecto a los índices de dificultad se encontró que, aunque los ítems se encuentran dentro de los rangos aceptables estos resultan ser más difíciles para los estudiantes hondureños.

Palabras clave: Psicometría, Pruebas estandarizadas, Validez, Estándares

Índice de Contenido

Agradecimiento.....	I
Resumen	
Introducción.....	1
Capítulo 1. Planteamiento del problema.....	2
1.1 Situación problema.....	2
1.2 Objetivos de la investigación.....	4
1.2.1 Objetivo general.....	4
1.2.2 Objetivos específicos	4
1.3 Justificación	5
Capítulo 2. Marco teórico	6
2.1 Las pruebas estandarizadas y su uso en educación.....	6
2.1.1 Antecedentes.....	6
2.1.2 La evaluación estandarizada.	7
2.1.3 Pruebas de desempeño académico.	10
2.2 Tercer estudio regional comparativo y explicativo (TERCE).....	13
2.2.1 Antecedentes.....	13
2.2.2 Áreas evaluadas en ciencias naturales	14
2.2.3 Características de las pruebas	16
2.3 Teorías de medición en psicometría.....	17
2.3.1 Teoría clásica de los test TCT	17
2.3.1.1 Modelos de la teoría clásica de los test	18

2.3.1.2 Supuestos del modelo lineal clásico.....	19
2.3.1.3 Alpha de cronbach.....	22
2.3.2 Teoría de respuesta al ítem	23
2.3.2.1 Supuestos de TRI	27
2.3.2.2 Modelos de TRI.....	27
2.4 Análisis factorial.....	31
Capítulo 3. Marco contextual	33
3.1 El sistema educativo hondureño.....	33
3.1.1 Indicadores del sistema educativo hondureño.	34
3.1.2 Currículo nacional básico.	35
3.1.3 Evaluación estandarizada de los aprendizajes en Honduras.	38
3.1.4 Honduras en las evaluaciones internacionales.....	40
Capítulo 4. Método	42
4.1 Tipo de investigación.....	42
4.2 Población y muestra.....	42
4.3 Instrumento.....	44
4.4 Procedimientos para el análisis de datos.....	47
Capítulo 5. Resultados	49
5.1 Confiabilidad de la prueba de ciencias naturales del TERCE para Honduras.....	49
5.2 Análisis de dificultad de los ítems de la prueba de ciencias naturales del TERCE.....	51
5.3 Análisis de discriminación de los ítems de la prueba de ciencias naturales del TERCE.....	52

5.4 Análisis de dimensionalidad de la prueba de CCNN del TERCE.....	55
5.5 Análisis de los índices de ajuste infit y outfit de la prueba de CCNN del TERCE.....	57
5.6 Análisis de dificultad mediante tri para los ítems de la prueba de CCNN del TERCE.....	60
Capítulo 6. Discusión de los resultados.....	66
6.1 Conclusiones.....	69
6.2 Limitaciones.....	70
6.3 Recomendaciones.....	71
Referencias	72
Declaración de ética de la investigación.....	80
Apéndice	81

Índice de Tablas

Tabla 1. Pruebas estandarizadas aplicadas en América Latina.....	12
Tabla 2. Supuestos TCT	19
Tabla 3. Estándares educativos para sexto grado en el área de ciencias naturale.	36
Tabla 4. Número de estudiantes por cuadernillo	44
Tabla 5. Matriz de operacionalización de variables	45
Tabla 6. Especificaciones de la prueba de ciencias naturales de sexto grado	46
Tabla 7. Formas en la prueba de CCNN del TERCE	47
Tabla 8. Número de preguntas por bloque.....	47
Tabla 9. Coeficientes alpha de Cronbach por cuadernillo en la prueba CCNN honduras TERCE.....	50
Tabla 10. Resúmenes estadísticos KMO y prueba de esfericidad de Barlett	55
Tabla 11. Resumen varianza total explicada por el primer componente para cada cuadernillo	56
Tabla 12. Resumen parámetro dificultad TCT para el bloque 1.....	81
Tabla 13. Resumen parámetro dificultad TCT para el bloque 2.....	82
Tabla 14. Resumen parámetro dificultad TCT para el bloque 3.....	83
Tabla 15. Resumen parámetro dificultad TCT para el bloque 4.....	84
Tabla 16. Resumen parámetro dificultad TCT para el bloque 5.....	85
Tabla 17. Resumen parámetro dificultad TCT para el bloque 6.....	86

Tabla 18. Resumen parámetro discriminación tct para el bloque 1.....	87
Tabla 19. Resumen parámetro discriminación tct para el bloque 2.....	88
Tabla 20. Resumen parámetro discriminación tct para el bloque 3.....	89
Tabla 21. Resumen parámetro discriminación para el bloque 4.....	90
Tabla 22. Resumen parámetro discriminación para el bloque 5.....	91
Tabla 23. Resumen parámetro dificultad para bloque 6	92
Tabla 24. Varianza total explicada cuadernillo 1	93
Tabla 25. Varianza total explicada cuadernillo 2	94
Tabla 26. Varianza total explicada cuadernillo 3	95
Tabla 27. Varianza total explicada cuadernillo 4	96
Tabla 28. Varianza total explicada cuadernillo 5	97
Tabla 29. Varianza total explicada cuadernillo 6	98
Tabla 30. Resumen estadístico INFIT para el bloque 1.....	99
Tabla 31. Resumen estadístico INFIT para el bloque 2.....	100
Tabla 32. Resumen estadístico INFIT para el bloque 3.....	101
Tabla 33. Resumen estadístico INFIT para el bloque 4.....	102
Tabla 34. Resumen estadístico INFIT para el bloque 5.....	103
Tabla 35. Resumen estadístico INFIT para el bloque 6.....	104
Tabla 36. Resumen estadístico OUTIFIT el para bloque 1	105

Tabla 37. Resumen estadístico OUTIFIT para el bloque 2	106
Tabla 38. Resumen estadístico OUTIFIT para el bloque 3	107
Tabla 39. Resumen estadístico OUTIFIT para el bloque 4	108
Tabla 40. Resumen estadístico OUTIFIT para el bloque 5	109
Tabla 41. Resumen estadístico OUTIFIT para el bloque 6	110
Tabla 42 resumen dificultad TRI para el bloque 1	111
Tabla 43. Resumen dificultad TRI para el bloque 2	112
Tabla 44. Resumen dificultad TRI para el bloque 3.....	113
Tabla 45. Resumen dificultad TRI para el bloque 4.....	114
Tabla 46. Resumen dificultad TRI para el bloque 5.....	115
Tabla 47 Resumen dificultad TRI para el bloque 6.....	116

Índice de Figuras

Figura 1. Estudios desarrollados por OREALC/UNESCO.	14
Figura 2. Áreas evaluadas en la prueba de ciencias naturales del TERCE.....	16
Figura 3. Curva característica de un ítem	26
Figura 4. Procedimientos metodológicos para el análisis psicométrico de la prueba de CCNN	42
Figura 5. Plan de análisis de datos	48
Figura 6. Dificultad de los ítems según TCT	51
Figura 7. Dificultad de los ítems según TCT	52
Figura 8. Discriminación de los ítems según TCT	53
Figura 9. Discriminación de los ítems según TCT	54
Figura 10. Estadístico INFIT para los ítems de la prueba de ciencias naturales ..	57
Figura 11. Estadístico INFIT para los ítems de la prueba de ciencias naturales ..	58
Figura 12. Estadístico OUTFIT para los ítems de la prueba de ciencias naturales	58
Figura 13. Estadístico OUTFIT para los ítems de la prueba de ciencias naturales	59
Figura 14. Dificultad de los ítems según TRI.....	60
Figura 15. Dificultad de los ítem según TRI	61

Figura 16. Curva característica del ítem6.....	60
Figura 17. Curva característica del ítem 16	62
Figura 18. Curva de puntuación teórica ítem 24.....	61
Figura 19. Curva característica del ítem 24.....	63
Figura 20. Curva característica del ítem 23	62
Figura 21. Curva característica del ítem 34.....	64
Figura 22. Curva de puntuación teórica ítem 52.....	62
Figura 23. Curva característica del ítem 52.....	64

Introducción

La evaluación estandarizada se ha constituido en una necesidad en los sistemas educativos de la actualidad, hoy en día este tipo de evaluación es un requisito indispensable dentro de los sistemas educativos modernos ya que sirve para indagar sobre los logros de aprendizaje obtenidos por los estudiantes (G. D. Demarchi, 2020), actualmente esta ha mejorado gracias a los procedimientos psicométricos que garantizan la objetividad de sus instrumentos. (APA, AERA, NCME, 2018). Este estudio consistía en analizar una de las pruebas más importantes en el contexto Regional la Prueba sobre Logros de Aprendizaje en Ciencias Naturales del Tercer Estudio Regional Comparativo y Explicativo TERCE.

En el capítulo 1 se introduce el problema que da origen al estudio, así como el planteamiento del mismo y los objetivos que se persiguen. El capítulo 2 trata aspectos sobre las pruebas estandarizadas, sus antecedentes y uso educación, también se describen los antecedentes del TERCE y a partir de la información teórica recogida de diversas fuentes científico-académicas se abordan las dos grandes teorías de medición que sustentan el análisis psicométrico. El capítulo 3 brinda un panorama del contexto educativo nacional y su relación con la evaluación estandarizada. En el capítulo 4 se detallan los procedimientos metodológicos utilizados en el análisis, así como los aspectos más importantes de la muestra y del instrumento a validar. En el capítulo 5 se describen los resultados obtenidos una vez aplicados los análisis psicométricos correspondientes y finalmente en el capítulo 6 se presenta una discusión de resultados acerca de los principales hallazgos encontrados en este estudio.

CAPÍTULO 1.

Planteamiento del Problema

El presente apartado expone los argumentos que justifican la necesidad de realizar el análisis psicométrico de la prueba de Logros de Aprendizaje en Ciencias Naturales para estudiantes de sexto grado del Tercer Estudio Regional Comparativo y Explicativo TERCE. La estructura lógica en que se desarrollan las ideas es la siguiente: se describe la situación problemática donde se identifica la situación que da origen al estudio, en relación a las pruebas estandarizadas y en contraste con la importancia de conocer los parámetros de las mismas. Además, se subraya la necesidad de evaluar los logros de aprendizaje en ciencias naturales como parte fundamental del aprendizaje de los estudiantes.

1. 1 Situación Problema

En el sistema educativo nacional existe la necesidad de evaluar el desempeño académico de los estudiantes con el fin de obtener un punto de referencia que permita analizar el progreso de los estudiantes. Los estudios extranjeros a través de las evaluaciones estandarizadas permiten contrastar tanto el desempeño académico de los estudiantes en diferentes contextos, como los factores que inciden en dicho desempeño (Vaillant & Rodríguez, 2018).

La evaluación externa estandarizada es considerada en la actualidad como una “tarea necesaria y prioritaria de apoyo para el buen funcionamiento de los procesos educativos,

dado que aporta la información básica e indispensable para la toma de decisiones racionales y fundamentadas que orienten las políticas del sector” (Secretaría de Educación, 2015b).

Aunado a lo anterior dichas pruebas requieren un riguroso análisis de los ítems que se utilizan. Esto supone el requerimiento de procesos de validación que garanticen la calidad de los instrumentos (Secretaría de Educación, 2015b). Si bien es cierto a los estudios extranjeros les precede una buena reputación, en el contexto hondureño no son conocidos los parámetros que utilizan dichos estudios en la elaboración de sus instrumentos, por lo cual resulta oportuno realizar un estudio que de fe de que los mismos son confiables. Como señala

Muñiz, (2018) “es clave que los test utilizados cumplan unos requisitos de calidad psicométrica demostrables”.

Por otro lado, las evaluaciones sobre logros de aprendizaje en ciencias naturales son recientes en el contexto hondureño. Si bien es cierto se han aplicado varias evaluaciones la mayoría han sido en español y matemáticas. No obstante, en la actualidad, es necesario poseer comprensión de las ciencias naturales para que los estudiantes puedan entender el mundo que les rodea. Es fundamental asegurar que los estudiantes que dejan la escuela se vayan con un entendimiento global de las ciencias naturales (Mullis et al., 2012).

En los resultados del Estudio Internacional de Tendencias en Matemáticas y Ciencias 2011 (TIMSS) Honduras obtuvo resultados bajos en lo que respecta a ciencias (432 puntos en 4to básico y 396 puntos en 8vo básico) esto evidencia la necesidad de seguir realizando evaluaciones para valorar el avance de los estudiantes (Mullis et al., 2012).

Un dato alarmante en lo que respecta a las pruebas de Matemáticas y Ciencias aplicadas por TIMSS en el año 2011 fue la necesidad de aplicar a estudiantes hondureños de sexto grado, pruebas que han sido diseñadas para alumnos de cuarto. En estas pruebas inclusive, estos alumnos de sexto grado mostraron un desempeño sumamente bajo (Alas Solis, 2017).

En lo que respecta a los resultados obtenidos en las pruebas de ciencias naturales aplicadas por PISA-D, pese a que estos son superiores en comparación a español o matemáticas, siguen siendo bajos, Honduras obtuvo resultados en ciencias en el nivel 1a lo cual corresponde al nivel más básico establecido por PISA (OCDE, 2018).

También en el informe “Eficacia de los Aprendizajes Logrados por los Egresados de Educación Media” se encontró que los estudiantes egresados del sistema educativo nacional presentaban deficiencias muy marcadas en ciencias naturales. Los reactivos utilizados en esa prueba se diseñaron a fin de explorar y evaluar el razonamiento científico.

A causa de las consideraciones anteriores, se puede resaltar la relevancia del análisis psicométrico de la prueba de ciencias naturales para estudiantes de sexto grado. Puesto que los resultados de las pruebas son parte fundamental en la toma de decisiones que inciden en el sistema educativo nacional.(OCDE, 2018).

Resulta pertinente realizar un análisis psicométrico de un instrumento tan importante como lo es el Tercer Estudio Regional Comparativo y Explicativo TERCE de

OREALC - UNESCO para América Latina, puesto que su objetivo principal es evaluar la calidad de la educación en términos de logro de aprendizaje en los países participantes de América Latina y del Caribe, brindando un panorama sobre los factores asociados que deberían tomarse en cuenta para la formulación de políticas públicas y la revisión de prácticas educativas.

1.2 Objetivos de la Investigación

1.2.1 Objetivo General

Analizar los parámetros psicométricos de la prueba internacional del TERCE en el área de Ciencias Naturales para los estudiantes de sexto grado de Honduras aplicada en el año 2013.

1.2.2 Objetivos específicos

Determinar el nivel de confiabilidad de la prueba del TERCE en el área de Ciencias Naturales para los estudiantes de sexto grado de Honduras, por medio de la teoría clásica de los test.

Establecer los índices de dificultad de los ítems de la prueba del TERCE en el área de Ciencias Naturales para los estudiantes de sexto grado de Honduras por medio de la teoría clásica de los test.

Determinar los índices de discriminación de los ítems de la prueba del TERCE en el área de Ciencias Naturales para los estudiantes de sexto grado de Honduras a través de la teoría clásica de los test.

Evaluar la unidimensionalidad de la prueba del TERCE en el área de Ciencias Naturales para los estudiantes de sexto grado de Honduras por medio de análisis factorial exploratorio.

Establecer los niveles de dificultad de los ítems de la prueba internacional del TERCE en el área de Ciencias Naturales para los estudiantes de sexto grado de Honduras por medio de la teoría de respuesta al ítem.

1.3 Justificación

En este espacio se describe el aporte social, teórico y metodológico que brindará el análisis psicométrico de la prueba sobre logros de aprendizaje en ciencias naturales para estudiantes de sexto grado del TERCE. A su vez, aportando científicamente en la aplicación de procedimientos y metodologías de validación de instrumentos.

Este estudio es socialmente relevante debido a las implicaciones que tienen los resultados de las pruebas estandarizadas como parte fundamental a la hora de tomar decisiones en los diferentes sistemas educativos (Himelfarb, 2019).

Por otro lado, es oportuno señalar la relevancia teórica que la validación de este instrumento implica, pues la validez es el principio teórico más importante en psicometría (Martínez Arias et al., 2006). Muchos de los conflictos que se le atribuyen a las evaluaciones estandarizadas están directamente relacionados con la carencia de validez, por tal razón someter un instrumento a un proceso de validación garantiza su objetividad (Miguel & Meliá, 2017).

La relevancia metodológica del análisis psicométrico de la prueba de rendimiento académico en ciencias naturales para estudiantes de sexto grado radica en que permitirá reunir evidencias que justifiquen científicamente a favor o en contra del diseño, aplicación, uso e interpretación de los puntajes de la prueba. El estudio puede ser una experiencia práctica que ejemplifique el cumplimiento de los estándares adecuados en el diseño de las pruebas que se aplican en el país. Este tipo de análisis psicométrico contribuye a solventar falencias presentes en los instrumentos que ponen en duda la fiabilidad de la evaluación (Bakieva, 2019).

De igual manera aún prevalecen confusiones y vacíos importantes en cuanto a la concepción de la validez y también en lo que respecta a los procesos metodológicos de validación a que debe someterse un instrumento (Montero Rojas, 2013).

CAPÍTULO 2.

Marco Teórico

Esta sección describe el fundamento teórico que sustenta el análisis psicométrico de la prueba de desempeño de ciencias naturales. El cual reúne tres ejes principales, los cuales se irán desarrollando desde los planos más generales, hasta los más específicos. El primero de ellos son las pruebas estandarizadas, sus antecedentes y uso educación, también se estudiarán los antecedentes del TERCE y a partir de la información teórica recogida de diversas fuentes científico-académicas se abordarán las dos grandes teorías de medición que sustentan el análisis psicométrico.

2.1 Las Pruebas Estandarizadas y su Uso en Educación

2.1.1 Antecedentes

Los primeros registros de pruebas estandarizadas tienen su origen en China, en donde los ciudadanos para optar a un puesto de trabajo en el gobierno realizaban exámenes para determinar su conocimiento de la “filosofía y poesía confucianas”. Alfred Binet, que era un psicólogo francés, tal vez se considere el pionero del concepto moderno de pruebas estandarizadas. Comenzó a desarrollar una prueba de inteligencia que fue impulsada por Stanford y ahora se denomina Prueba de Inteligencia Stanford-Binet.

El campo de las pruebas se desarrolló rápidamente durante la Primera Guerra Mundial (1914-1918) cuando el problema de la selección profesional para las necesidades del ejército y la producción militar se convirtió en una prioridad. Durante ese tiempo, psicólogos destacados organizaron los exámenes Army Alpha y Army Beta para evaluar a los reclutas del ejército. Su éxito inspiró aún más a los psicólogos a abogar por las pruebas civiles (Himelfarb, 2019).

En el contexto estadounidense, durante el año 1845 se realizaron las primeras evaluaciones orales y junto con el Comité de Escuelas Públicas de Boston se aplicaron pruebas objetivas de forma escrita a los estudiantes, esto con el fin de brindar resultados relacionados con el nivel y la calidad de la enseñanza en las escuelas urbanas, así como comparar las escuelas y los maestros. En vista del éxito obtenido en las evaluaciones anteriores las pruebas escritas se introdujeron en todas las ciudades de Estados Unidos (Kaukab & Mehrunnisa, 2016).

En lo que respecta a América Latina, las evaluaciones a gran escala se extendieron a finales del siglo XX. El país que desarrolló inicialmente un verdadero sistema de evaluación fue Chile quien a partir de 1968 comenzó a utilizar pruebas a gran escala. En Cuba desde 1972 se puso en práctica el monitoreo y evaluación de los resultados por medio del programa Perfeccionamiento del Sistema Nacional de Educación. En Costa Rica en los setenta comenzó la aplicación de pruebas de educación básica a través de la Universidad de Costa Rica (Martínez, 2008).

En México se comenzaron a utilizar en los sistemas de admisión de educación superior, para 1990 el uso de instrumentos de evaluación estandarizada fue relevante pese a que no se daban a conocer los resultados. En Colombia se estableció el Instituto Colombiano de Fomento de la Educación Superior (ICFES) encargado de la elaboración de pruebas de ingreso a universidades y posteriormente aplicó pruebas de rendimiento para educación básica. Para finales de los noventa se comenzaron a implementar de forma acelerada a lo largo de toda la región sistemas nacionales de evaluación unos en mejor condición que otros (Martínez, 2008).

En el año 1994 se creó el Laboratorio Latinoamericano de la Calidad de la Educación (LLECE) coordinado por un equipo de la Oficina Regional de la Educación de la UNESCO para América Latina y el Caribe (OREALC/UNESCO) funcionando como una red de unidades de medición y evaluación de los sistemas educativos de los países de América Latina (UNESCO, 2016).

2.1.2 La evaluación estandarizada.

Hablar de evaluación estandarizada requiere de la comprensión del término prueba estandarizada y según Montero Rojas (2013) de (Nunnally & Bernstein, 1995) dice que las pruebas estandarizadas “son un medio empírico que permite generar puntuaciones en una escala numérica para representar una variable o constructo” (p.114). Estas puntuaciones facilitan la producción de inferencias relevantes sobre la conducta de los examinados (Muñiz & Fonseca-Pedrero, 2019).

Según Kaukab & Mehrunnisa (2016) una prueba estandarizada se define como “un tipo de prueba que es “estándar” o consistente en la calificación de todos los candidatos, quienes deben tomar la prueba con las mismas preguntas y con la misma cantidad de

tiempo”(p.1). El requisito previo es que se controlen todas las variables. Esto permite una comparación relativa de los rendimientos de los estudiantes. Si bien el formato de estas pruebas suele ser de opción múltiple, ya que se pueden calificar rápidamente, las preguntas se pueden incluir en forma de preguntas de verdadero o falso, preguntas de ensayo o preguntas de respuesta corta.

De acuerdo a APA, AERA, NCME (2018) el test es “Un dispositivo o procedimiento en el cual se obtiene y posteriormente se evalúa y califica una muestra del comportamiento de un individuo examinado en un dominio especificado, a través de un proceso estandarizado” (p.2).

Los instrumentos de evaluación, en este caso test o pruebas estandarizadas permiten, mediante un método específico, la recolección de información de muestras de comportamiento para medir un constructo (Yusoff, 2019).

Los instrumentos anteriormente descritos, se constituyen en el fundamento para la realización de evaluaciones estandarizadas, dichas evaluaciones, al menos a un nivel muestral, se están convirtiendo rápidamente en un elemento fundamental de los sistemas educativos modernos, dado que representan una parte necesaria del proceso de diseño, implementación y evaluación de programas para mejorar la calidad de la educación (Wolff, 2007).

La evaluación es un proceso clave en la educación. Solo a través de ella se puede averiguar si la instrucción ha tenido el efecto deseado, porque no se puede garantizar que incluso la instrucción mejor diseñada sea efectiva. Todos aquellos interesados en los resultados de la educación (alumnos, profesores, padres, otros contribuyentes, empleadores y la comunidad en general) quieren saber qué han aprendido los alumnos y parece plausible que esto pueda evaluarse fácilmente mediante el uso de instrumentos sencillos y familiares, como las pruebas de rendimiento (William, 2010).

Los resultados de las evaluaciones de desempeño académico son esperados anualmente, para valorar los logros en aprendizaje ya que por medio de las pruebas estandarizadas se puede adquirir información técnica, útil y específica que pueda orientar la toma decisiones en pro de la educación. Dichas pruebas tienen características especiales, como ítems equivalentes de dificultad comparables, que responden a una aplicación temporal dada en un espacio definido (Ravela, 2013).

Como se ha mencionado, las pruebas estandarizadas requieren de procedimientos rigurosos en su elaboración, es por ello que se demanda la intervención de la ciencia que se fundamenta en los modelos de medición, basados en la estadística y la probabilidad a fin de establecer las relaciones entre las puntuaciones de una prueba y el constructo medido, en este caso esa ciencia es la psicometría (Martínez Arias et al., 2006).

Las nuevas demandas de clasificación y selección han impulsado el desarrollo de toda una tradición en la medición psicológica, la tradición psicométrica basada en tres principios: estandarización, diferenciación y el uso de tareas generales. Los procedimientos de evaluación estandarizados se convirtieron en la norma para evaluar el logro de los estudiantes y en el único punto de referencia para la toma de decisiones tan importantes como el permitir o no a un alumno pasar de curso, ingresar a la universidad u obtener un diploma (Bravo Arteaga & Fernández Del Valle, 2000).

Por consiguiente, las pruebas estandarizadas han sido llamadas la mayor contribución social individual de la psicología y puede ser el método de evaluación más útil disponible para esfuerzos intensivos en recursos humanos. Sin embargo, durante la mayor parte de su historia, las pruebas estandarizadas se han desarrollado y administrado a gran escala y grandes organizaciones, típicamente políticamente sensibles, han controlado su uso (Phelps, 2008).

La evaluación estandarizada tiene varios beneficios para las instituciones educativas individuales:

Son útiles para identificar las fortalezas y debilidades de los estudiantes en relación con el promedio nacional e internacional de los estudiantes de similar edad y nivel de educación.

Permiten establecer la rendición de cuentas de los profesores y las escuelas porque generalmente el resultado de las pruebas pasa a ser un registro público.

Las pruebas estandarizadas han mejorado la gestión del tiempo educativo, al centrarse en los resultados del aprendizaje de un aspecto particular del plan de estudios y las áreas objetivo para los maestros.

Permiten comparar el nivel y el rendimiento de los estudiantes en las escuelas, distritos y provincias.

La naturaleza objetiva de las pruebas estandarizadas se debe a que las pruebas son calificadas por computadoras o por personas que no tienen ningún contacto con el estudiante. Además, las pruebas se desarrollan después de varias fases de revisión por expertos y se someten a escrutinio para eliminar el sesgo (Kaukab & Mehrunnisa, 2016, p.3).

Si bien es cierto, son diversos los beneficios de la evaluación estandarizada, así como su uso se expandió, también la crítica acerca de la cultura de la evaluación y los dilemas éticos en torno a su uso (Himelfarb, 2019). por tanto es importante entender que “Las pruebas estandarizadas no son herramientas de evaluación perfectas, sin embargo, se utilizan de forma válida y fiable, por lo tanto, una prueba estandarizada es más que un antídoto para el juicio sesgado” Hoy en día las pruebas estandarizadas han mejorado en calidad y conveniencia gracias a la psicometría (Phelps, 2008, p.3).

Lo expuesto anteriormente ha contribuido a que “no se conciba una mejora en los sistemas educativos si no es mediante la evaluación, especialmente a través de pruebas estandarizadas” (Reyes, 2020, p. 419) puesto que solo a través de esta se puede saber si las metodologías de enseñanza están teniendo la efectividad necesaria, además de permitir identificar otros factores que inciden en los procesos de aprendizaje.

También es importante añadir que debido al éxito de la evaluación estandarizada “Más allá de la evaluación de alumnado para procesos de selección o certificación, la estandarización también ha llegado a instaurarse en diversos procesos de evaluación, como por ejemplo la evaluación de docentes, instituciones y sistemas educativos” (Miguel & Meliá, 2017, p.7)

2.1.3 Pruebas de desempeño académico.

La complejidad del desempeño académico parte de su conceptualización. Lamas, (2015) de (Pizarro, 1985) se refiere al rendimiento académico como “una medida de las habilidades indicativas y de respuesta que expresan, de forma estimativa, lo que una persona ha aprendido como resultado de un proceso de educación o formación”.

Entre los principales fines de la evaluación del desempeño académico puede mencionarse que este se convierte en un indicador para saber en qué medida los estudiantes

logran su aprendizaje para lo cual dirigen su principal esfuerzo y por otro lado, proporciona conocimiento sobre la eficacia de la escolarización (Lamas, 2015).

Las pruebas estandarizadas de desempeño académico que realizan las agencias y organizaciones internacionales y que son aplicadas en varios países de manera conjunta, a distintos grupos o grados de estudiantes similares, son útiles a la hora de proporcionar información sobre los aprendizajes conceptuales y procedimentales en distintas áreas; también sirven para indagar acerca de los diferentes factores del contexto que influyen en el aprendizaje de los estudiantes (G. Demarchi, 2020).

Con las aplicaciones de diversas pruebas de desempeño académico a lo largo del tiempo, hoy en día resulta casi imposible negar la importancia de los organismos internacionales respecto a la evaluación educativa y su postura, la cual brinda claridad en cuanto a la búsqueda de mejorar la calidad de la educación a través de elementos que aporten en relevancia, pertinencia y equidad al proceso educativo (Reyes, 2020).

Es por ello que, desde las dos últimas décadas del siglo pasado, en América Latina se ha observado un interés más notable por la evaluación, estrechamente asociada a la calidad. Este fenómeno se ha traducido en un exagerado incremento en la aplicación de diversas evaluaciones (Sánchez & Sagrario, 2013). En la siguiente tabla se describen las pruebas estandarizadas más importantes aplicadas en el contexto latinoamericano.

Tabla 1. Pruebas estandarizadas aplicadas en américa latina

Prueba	Competencias que evalúa	Edades de los participantes o grado escolar
TIMMS		
Estudios sobre las tendencias en ciencias y matemáticas.	Ciencias Naturales y Matemáticas	Grados 3°, 4°, 7°, 8°, 11°
PERCE,	Matemáticas	Grados 3°, 4°
SERCE,	Lenguaje	Grados 3°, 6°
TERCE,	Ciencias Naturales.	Grados 6°
ERCE		
ICCS		
Estudio internacional de educación cívica y ciudadanía	Ciudadanía y Democracia, Identidad Nacional y Relaciones Internacionales	Grado 8 °
PISA		
Programa internacional para la evaluación de estudiantes.	Competencia Lectora, Competencia Matemática y Competencia Científica	Estudiantes entre 15 y 16 años de edad.
PIRLS		
Progreso internacional estudio de alfabetización en lectura.	Lectura	Grado 4 °

Fuente: Adaptado de (G. D. Demarchi, 2020)

2.2 Tercer Estudio Regional Comparativo y Explicativo (TERCE)

2.2.1 Antecedentes

La creación del LLECE en 1994, resultó ser un acontecimiento importante en el desarrollo de nuevas formas de cooperación entre los países de la región y en el papel de coordinación que asume la Oficina Regional de Educación de la UNESCO. Sus objetivos principales consisten en generar conocimientos que contribuyan a que los Ministerios de Educación de la región tomen decisiones informadas en materia de políticas educativas a partir del análisis de los resultados de las investigaciones evaluativas que el LLECE realiza (UNESCO, 2009, p.4).

Con el propósito de entregar información útil para formular y ejecutar políticas educativas en los países de la región, el Laboratorio Latinoamericano de Evaluación de Calidad de la Educación, LLECE, realizó en 1997 un Primer Estudio Internacional Comparativo en Lenguaje, Matemática y Factores Asociados, en tercer y cuarto grado de educación básica. En este estudio participaron: Argentina, Bolivia, Brasil, Chile, Colombia, Costa Rica, Cuba, Honduras, México, Paraguay, Perú, República Dominicana y Venezuela (UNESCO, 2000).

En vista del éxito obtenido en la aplicación del PERCE, entre los años 2004 y 2008 el LLECE realizó el Segundo Estudio Regional Comparativo y Explicativo (SERCE) que evaluó los mismos componentes que el PERCE. Para este estudio se contó con la participación de: 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 los 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 (UNESCO, 2009).

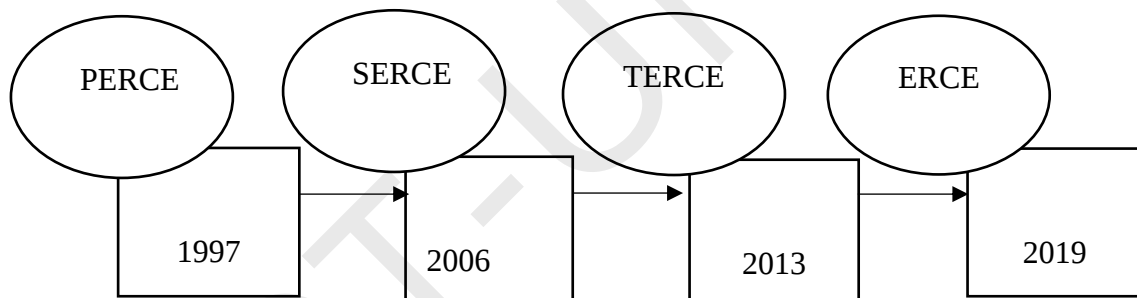
Los dos estudios previamente descritos son los antecesores del TERCE.

estudio de logro de aprendizaje a gran escala cuyo estudio piloto se llevó a cabo en 2012 y su aplicación definitiva en 2013, abarcó 15 países: Argentina, Brasil, Chile, Colombia, Costa Rica, Ecuador, Guatemala, Honduras, México, Nicaragua, Panamá, Paraguay, Perú, República Dominicana y Uruguay, más el Estado de Nuevo León (México). Su objetivo principal fue aportar

información para el debate sobre la calidad de la educación en la región, así como orientar la toma de decisiones en políticas públicas educativas. Las áreas evaluadas fueron lectura, matemáticas y ciencias naturales (UNESCO, 2016a, p.23).

Después de la aplicación del TERCE el siguiente estudio desarrollado por la UNESCO corresponde al Estudio Regional Comparativo y Explicativo ERCE aplicado en el año 2019, coordinado y elaborado por las mismas instituciones que sus antecesores, en esta ocasión participaron 16 países de la región y los resultados permiten a cada país mirar dónde están sus estudiantes en materia de logro escolar en un panorama comparativo regional, saber si lo que sus estudiantes están recibiendo en la sala de clases responde a lo que se espera que aprendan, valorar cuánto han progresado en estos últimos años e identificar cuáles son los desafíos hacia delante (UNESCO, 2019).

Figura 1. Estudios desarrollados por OREALC/UNESCO.



Fuente: Elaboración propia.

2.2.2 Áreas evaluadas en ciencias naturales

En lo que respecta a la prueba de Ciencias Naturales se evalúan cinco dominios

Salud: conocimiento de la estructura y funcionamiento del cuerpo humano, a partir de lo cual es posible comprender y valorar los aprendizajes acerca del cuidado general del cuerpo, los hábitos de higiene, la alimentación y la práctica del deporte, entre otros.

Seres vivos: reconocimiento de la diversidad de los seres vivos, las características de los organismos, la identificación de patrones comunes y la clasificación de los seres vivos basada en ciertos criterios.

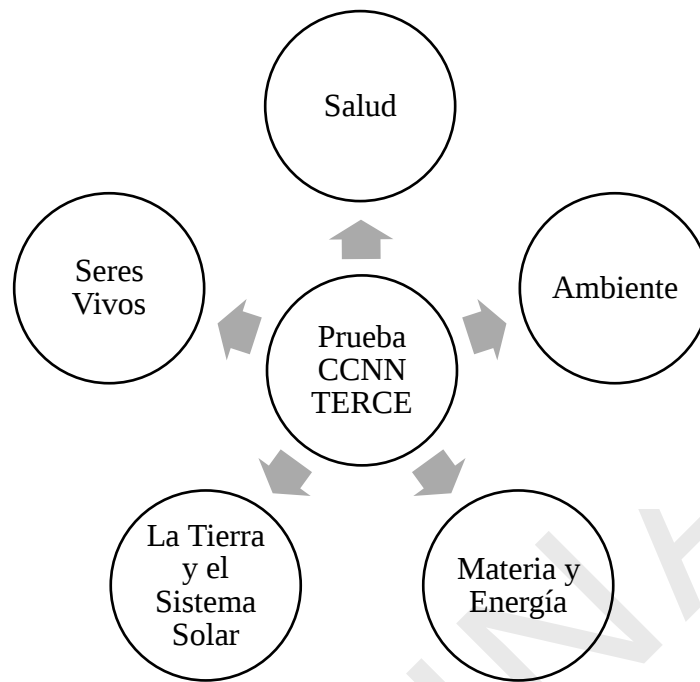
Ambiente: reconocimiento de la interacción entre los organismos y el ambiente; importancia del sol como la principal fuente de energía de todos

los seres vivos y del suelo y aire como las fuentes de materiales para la supervivencia de los seres vivos.

La Tierra y el sistema solar: temáticas orientadas a conocer y comprender las características físicas del planeta tierra, los movimientos de la tierra y la luna y su relación con fenómenos naturales observables.

Materia y energía: las temáticas de este dominio sirven para aprender que la energía toma diferentes formas; que la materia contiene energía y para que los seres vivos, los elementos naturales y los artefactos puedan moverse, funcionar o trabajar, se necesita energía (UNESCO, 2016b, p.36, 37).

La prueba de ciencias naturales también considera los siguientes procesos cognitivos: reconocimiento de información y conceptos, comprensión y aplicación de conceptos, pensamiento científico y resolución de problemas.

Figura 2. Áreas evaluadas en la prueba de ciencias naturales del TERCE

Fuente: Elaboración propia.

2.2.3 Características de las pruebas

Según el reporte técnico, en la prueba del TERCE se utilizan ítems de respuesta cerrada e ítems de producción de respuesta. Los ítems de respuesta cerrada utilizados se constituyen como ítems de selección múltiple. Las preguntas de producción son de respuesta corta en el caso de ciencias naturales y matemáticas. Se utilizan cuatro opciones de respuesta (la respuesta correcta y tres distractores). El estímulo o contexto corresponde a la descripción de una situación que sirve como base para que el estudiante tenga un punto de referencia concreto al momento de enfrentarse a la pregunta (UNESCO, 2016b).

Al momento de asignar puntuaciones a cada ítem se distinguen dos o tres categorías de respuesta, según su nivel de calidad. Si había tres niveles, correspondían a: correctas (crédito total), parcialmente correctas (crédito parcial) e incorrectas (sin crédito). Si había solamente dos categorías de respuestas, estas correspondieron a correctas/incorrectas. De acuerdo a UNESCO, (2016) “La prueba de ciencias está compuesta por ítems redactados adecuadamente en su mayoría. Asimismo, los ítems tienen un adecuado tratamiento del área. Es decir, “no se han detectado problemas atribuibles a errores científicos”.

Las pruebas, en su versión definitiva, están estructuradas en bloques de 11 a 16 preguntas; cada estudiante contesta un cuadernillo y cada cuadernillo está compuesto por dos bloques de preguntas, de modo que cada bloque ocupa la primera posición en un cuadernillo y la segunda en otro. Cada bloque es ensamblado para que constituya una “mini prueba” en términos de la representación de la tabla de especificaciones. En la prueba definitiva se ensamblaron seis cuadernillos de los cuales a dos de ellos se les considera bloques anclas de SERCE (UNESCO, 2016b).

2.3 Teorías de medición en psicometría

En psicometría existen las llamadas teorías de los test que guían y condicionan el proceso de diseño, análisis y utilización del test conforme a los avances en los procedimientos psicométricos y estadística, estas permiten una utilización rigurosa y científica. Dentro de estos modelos se encuentran la teoría clásica de los test (TCT) quizá la más famosa por su antigüedad, la cual se remonta a finales del siglo XIX y principios del siglo XX y también la teoría de respuesta al ítem (TRI) la cual vendría a ser una teoría más reciente pero muy usada por su eficacia (Muñiz, 2010).

Mediante el uso de ambas teorías se realiza la validación de los instrumentos, de manera general se pueden mencionar desde la teoría Clásica de los Test: El alfa de Cronbach y el índice de discriminación. Desde la teoría de respuesta al ítem: El modelo de Rasch, Teoría de la Generalizabilidad, Funcionamiento Diferencial del Ítem. Asimismo, el Análisis Factorial Exploratorio y Confirmatorio (Montero Rojas, 2013).

2.3.1 Teoría Clásica de los Test TCT

Erguven (2013) de (Marcoulides, 1999) dice que la TCT es un modelo simple que describe cómo los errores de medición pueden influir en los puntajes observados, cuyo objetivo principal es estimar la confiabilidad de los puntajes de una prueba. La teoría clásica de los test es una teoría sobre los puntajes de los exámenes que introduce tres conceptos: puntaje de prueba (a menudo llamado puntaje observado), puntaje verdadero y puntaje de error (Hambleton & Jones, 1993).

La teoría clásica de los test es considerada la primera teoría de la medición. Durante mucho tiempo, las características psicométricas de las medidas de personalidad han sido examinadas utilizando supuestos de TCT (Erguven, 2013).

A Spearman se le atribuye el desarrollo del anteproyecto de la teoría clásica de los test. Su teoría evolucionó hacia el llamado modelo de puntuación verdadera, otros contribuyentes importantes a este modelo lineal simple incluyen Guilford (1936) y Gulliksen (1950). TCT implica la estimación de un atributo como una combinación lineal de respuestas a elementos de prueba (Traub, 1997).

La teoría clásica de los test postula que cualquier observación es una combinación lineal de la puntuación real y el error. (Himelfarb, 2019). La principal ventaja de TCT es su simplicidad: la estimación de la medida. El rasgo se realiza simplemente agregando las respuestas puntuadas a los ítems. Esta teoría no involucra variables verdaderamente latentes; a pesar de que el puntaje real no es observable empíricamente, se puede definir operacionalmente como el puntaje promedio en el número infinito de repeticiones equivalentes del proceso de medición (Progar & Sočan, 2008).

A lo largo del tiempo la teoría clásica de los test ha sido considerada la teoría predominante tanto en la construcción como en el análisis de los test, se sabe que muchos de los test utilizados por los psicólogos fueron desarrollados mediante el uso de la teoría clásica. Es por ello que se considera un requisito para los profesionales de la psicometría la comprensión de la lógica clásica, sus posibilidades, así como sus limitaciones (Muñiz, 2010).

Pese a las consideraciones anteriores, cabe mencionar que esta teoría tiene sus limitantes entre las cuales se pueden mencionar:

- La puntuación del individuo depende del test que se aplique, puesto que la medición de las variables no es independiente del test (Guler et al., 2014).

- Una gran limitante radica en que las propiedades del test dificultad y discriminación dependen del tipo de personas utilizadas para estimarlas y del número de ítems que compone el test (Muñiz, 2018).

2.3.1.1 Modelos de la teoría clásica de los test

Spearman como predecesor de la TCT, a principios del siglo pasado propuso un modelo simple para las puntuaciones de los evaluados en los test, este modelo es llamado:

Modelo lineal clásico: El modelo asume que la puntuación que una persona obtiene en un test que a menudo es llamada (puntuación empírica u observada) está formada por dos componentes: la puntuación verdadera o esperada, es la que la persona obtiene en el

test y un error, que puede ser originado por distintas causas. En vista de lo anterior el modelo puede expresarse de la siguiente forma:

$$X = V + E$$

Donde X representa la puntuación empírica u observada, V representa la puntuación verdadera y E el error (Muñiz, 2010).

2.3.1.2 Supuestos del modelo lineal clásico

Ellis & Mead, (2008) describen los tres supuestos bajo los cuales funciona el modelo lineal clásico:

El valor esperado o la media de las puntuaciones de error para una población de examinados es cero. $\mu_E = 0$

La correlación entre la puntuación verdadera y el error para una población de examinados es cero o. $\rho_{VE} = 0$

La correlación entre el error en la prueba 1 (E_1) y el error en la prueba 2 (E_2) (cuando los errores son de dos pruebas o dos administraciones de una sola prueba) es cero o $\rho_{E_1E_2} = 0$.

A continuación, se resume el modelo lineal clásico y sus supuestos:

Tabla 2. Supuestos TCT

Modelo	$X = V + e$
Supuestos	1. $\mu_E = 0$
	2. $\rho_{(v, e)} = 0$
	3. $\rho_{E_1E_2} = 0$

Fuente: Elaboración propia

Ellis & Mead, (2008) añaden que el objeto de interés es la puntuación real del examinado, pero solo está disponible la puntuación observada, tiene sentido indagar la relación entre puntuaciones observadas y verdaderas. La correlación entre los puntajes observados y verdaderos para una población de examinados se conoce como índice de confiabilidad (ρ_{XT}) expresada como la relación entre la desviación estándar de puntajes verdaderos y puntajes observados:

$$\rho_{XT} = \sigma_T / \sigma_X$$

No obstante, una de las limitantes se debe a que se desconoce la desviación estándar de la distribución de puntajes verdaderos. Sin embargo, es posible evaluar a los examinados en la misma prueba dos veces o en dos pruebas paralelas. Las medias y las variaciones para las pruebas paralelas serán iguales. La correlación entre pruebas paralelas (o la misma prueba administrada dos veces: ρ_{X1X2}) o la confiabilidad coeficiente, es conceptualmente similar al índice de confiabilidad incognoscible. Matemáticamente, el coeficiente de confiabilidad es la relación entre la varianza de la puntuación real y la varianza de la puntuación observada (Ellis & Mead, 2008):

$$\rho_{X1X2} = \sigma_T^2 / \sigma_x^2$$

Los principales índices utilizados en el enfoque TCT para el análisis de ítems son la dificultad y la discriminación. El índice de dificultad de un ítem está definido como el cociente entre el número de examinados que han acertado el ítem y el número de examinados que lo han respondido. La forma más frecuente a la hora de representar la dificultad de un ítem en particular, es como porcentaje de respuestas correctas (PRC) el que toma valores entre 0% cuando ningún examinado responde correctamente el ítem y 100% cuando todos los examinados lo hacen. Por lo tanto, a mayor índice menor dificultad (Wood, 1960).

Gulliksen, (1950) sugiere que es válido utilizar la proporción de respuesta correcta para pruebas de potencia, pues se espera que los reactivos no contestados indiquen desconocimiento por parte del examinado, por otro lado, si la prueba es de velocidad en este sentido la proporción no tiene que ver con la aptitud, es más bien una característica de la posición del reactivo en la prueba. A causa de lo anterior la proporción de respuesta correcta no es recomendable en pruebas de velocidad. (p.367). La proporción de respuesta correcta puede calcularse de acuerdo a la siguiente fórmula:

$$PCRi = \frac{A_i}{N_i} 100$$

i es el indicador del ítem.

A_i representa el número de examinados que contestan correctamente el ítem.

N_i representa el número de examinados que contestan el ítem.

Por otra parte, la discriminación de un ítem se refiere a la capacidad que tiene el ítem para diferenciar adecuadamente a examinados que poseen diferentes niveles de habilidad. Las formas más usadas para determinar la discriminación son aquellas basadas en la correlación entre el puntaje de cada ítem y el puntaje total de la prueba. Para el análisis de los ítems de las pruebas, se utiliza la correlación punto biserial, que corresponde a un caso particular de la correlación de Pearson, que se utiliza cuando una de las variables es dicotómica y la otra continua. (Chávez & Saade 2009). Esta puede calcularse haciendo uso de la siguiente ecuación.

$$P_{pbis_i} = \frac{(\hat{\mu}_{+i} - \hat{\mu}_{xi})}{\hat{\sigma}_{xi}} \sqrt{\frac{p_i}{(1-p_i)}}$$

Donde:

$\hat{\mu}_{+i}$ corresponde al puntaje total de los examinados que contestaron correctamente el ítem i

$\hat{\mu}_{xi}$ corresponde al puntaje promedio de los examinados que contestaron el ítem i

$\hat{\sigma}_{xi}$ corresponde a la desviación estándar de los puntajes de todos los examinados que respondieron el ítem i

P_i es la proporción de respuesta correcta del ítem i

La correlación biserial puntual es conveniente puesto que sirve para identificar si las personas “adecuadas” son aquellas que responden de forma correcta, dicho de otro modo, sugiere que tanto poder predictivo tiene el ítem y como contribuye a las predicciones. La correlación biserial puntual está implícita en lo que se refiere a la validez predictiva del test, una propiedad relevante (Henrysson ,1971).

Es oportuno señalar que la confiabilidad o fiabilidad, se refiere a la consistencia o estabilidad de una medida. Una definición técnica de confiabilidad que ayuda a resolver tanto problemas teóricos como prácticos es aquella que parte de la investigación de qué tanto error

de medición existe en un instrumento, considerando tanto la varianza sistemática como la varianza por el azar (Kerlinger & Lee, 2002).

2.3.1.3 Alpha de Cronbach

Como describe Miller, (1995) el coeficiente alfa (α) “es un índice de fiabilidad de las pruebas que se utiliza habitualmente. Se puede utilizar para cualquier prueba en la que se obtengan puntuaciones sumando las puntuaciones de dos o más elementos de la prueba”. Fue desarrollado por varios investigadores durante la primera mitad del siglo pasado y fue nombrado por Cronbach, (1951) el primer escritor en reconocer su utilidad general. A menudo se le llama alfa de Cronbach en honor a su contribución. (p. 255)

Cronbach (1951), propuso que el alfa cuantifica el factor dominante entre los ítems, lo que lo llevó a relacionar el coeficiente alfa con lo que llamó la consistencia interna de una prueba. Cronbach sugirió entonces cuantificar la consistencia interna mediante la correlación media entre ítems y relacionarla con interpretabilidad psicológica. Relacionar el alfa con la consistencia interna se convirtió en una de las mayores contribuciones en la psicometría.

Según Rosner & Cronbach (1960), el coeficiente alfa es un modelo de consistencia interna basado en el promedio de las correlaciones entre los ítems y se refiere a que tan bien los ítems miden un constructo unidimensional. La forma más estable de medir la confiabilidad de un instrumento es mediante la consistencia interna.

Las principales ideas sobre la validez fueron propuestas a comienzos del siglo pasado, con la definición del coeficiente de validez, el cual se entendía como la correlación entre un predictor y un criterio (Cronbach & Meehl, 1955).

Montero Rojas, (2013) se refiere al alfa de Cronbach como el resultado de mayor importancia práctica desde la teoría clásica de los test puesto que “mide la precisión de la prueba en términos del grado de consistencia interna del instrumento y apunta hacia el grado de estabilidad de los puntajes (p. 121).

El Alpha de Cronbach es un coeficiente que sirve para estimar la confiabilidad de una escala de medida utilizando Teoría Clásica de los Test. Se utiliza en situaciones en que la escala está conformada por diversos ítems que pretenden evaluar una misma habilidad. (Muñiz, 2018) Puede calcularse de acuerdo a la siguiente ecuación:

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum \hat{\sigma}_i^2}{\hat{\sigma}_x^2} \right)$$

K corresponde al número de ítems en la prueba

$\hat{\sigma}_i^2$ corresponde a la varianza del puntaje en el ítem i

$\hat{\sigma}_x^2$ corresponde a la varianza del puntaje total

Este coeficiente puede tomar valores entre 0 y 1, considerándose convencionalmente que confiabilidades superiores a 0,8 son aceptables. Nunnally & Bernstein, (1995) dicen que en lo que respecta a pruebas de alto impacto en donde los resultados implican decisiones directas sobre los examinados, sugieren que dichas pruebas deberían 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.

Frías-navarro & Valencia, (2021) indican que valores de consistencia interna menores a .70 muestran una baja correlación entre los ítems y valores por encima de .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.

2.3.2 Teoría de Respuesta al Ítem

La teoría de la respuesta al ítem es, en última instancia, una teoría sobre los procesos que subyacen a la respuesta de una persona a una pregunta (un ítem). La teoría se basa en los primeros trabajos de Thurstone en la década de 1920 sobre el concepto del proceso de respuesta y de Lazarsfeld y Lord en la década de 1950 sobre la relación de las respuestas a los ítems y los puntajes de las pruebas con las variables latentes (no observadas). (Thissen & Steinberg, 1988)

No obstante la mayor relevancia o auge de la Teoría de Respuesta al ítem se remonta a los años 1968 y 1980 con la obra de Tatsuoaka et al., (1971) y Lord, (1980) quienes impulsaron los primeros modelos a pesar de su complejidad matemática y la carencia de programas que facilitaran su uso en esa época. Sus aportes vendrían a establecer cambios

importantes en lo que se refiere a los análisis psicométricos y por supuesto se constituyeron como una de las mayores contribuciones de la psicometría.

En la década de 1970 la teoría de respuesta al ítem (TRI) ganó mucha popularidad debido a su aplicación en pruebas estandarizadas importantes como la prueba de aptitud académica (SAT). Años después TRI se convirtió en el más importante método psicométrico de validación de escalas porque proporciona un método para resolver muchos de los desafíos de medición que deben abordarse al construir una prueba o escala (Yang & Kao, 2014).

Es entre las décadas de 1980-1990 cuando TRI se afianza y tiene su mayor expansión en la psicometría, debido a que se tenía mayor aplicabilidad gracias a los trabajos de diversos autores, en este tiempo la comunidad académica se muestra menos escéptica y se da apertura en congresos, revistas y numerosos trabajos relacionados con TRI y diversidad de textos comprensivos (Muñiz Fernández & Hambleton, 1992).

En 1990, el uso de TRI en programas de pruebas operativas se expandió considerablemente. La metodología de la TRI para los datos de respuesta a ítems dicotómicos estaba bien desarrollada y se utilizó ampliamente a fines de la década de 1980. En los primeros años de la década de 1990, se desarrollaron modelos para datos de respuesta a elementos politómicos y se comenzaron a utilizar en programas operativos (Carlson & Davier, 2013).

La teoría de respuesta al ítem es considerada como sinónimo de la teoría de los rasgos latentes. A veces se le conoce como la teoría de puntaje verdadero fuerte o la teoría de prueba mental moderna porque, la TRI es un cuerpo de teoría más reciente y hace suposiciones más fuertes en comparación con la teoría clásica de los test (Magno, 2009).

La TRI brinda información específica respecto al grado de exactitud con que se mide la variable en función de sus diferentes niveles. Estas medidas de precisión locales se hacen operativas mediante las funciones de información de los ítems y del test. (Birnbbaum, 1968)

El objetivo fundamental de la TRI es la construcción de instrumentos de medición con propiedades invariantes entre poblaciones. Si dos individuos presentan el mismo nivel de rasgo medido ambos tendrán igual probabilidad de dar la misma respuesta, independientemente de la población a la que pertenecen (Attorresi et al., 2009).

A pesar de las ventajas antes mencionadas, en vista de que los modelos de la TRI implicaban complejos procesos matemáticos, fue hasta el desarrollo de software apropiados, en los años 80 (como BICAL, LOGISTIC, BILOG, MULTILOG, MICROCAT) que comenzó su aplicabilidad (Martínez Arias., et al 2006).

La TRI resuelve importantes problemas en la medición psicológica que no encontraban solución en la Teoría Clásica cuyo objetivo principal es obtener mediciones invariantes respecto a las personas implicadas y más allá de los ítems o instrumentos utilizados (Muñiz, 2018).

Entre los principios básicos de esta teoría pueden mencionarse los siguientes:

- El rasgo latente (θ) explica el patrón de respuestas a un conjunto de ítems en un test. (Attorresi et al., 2009)

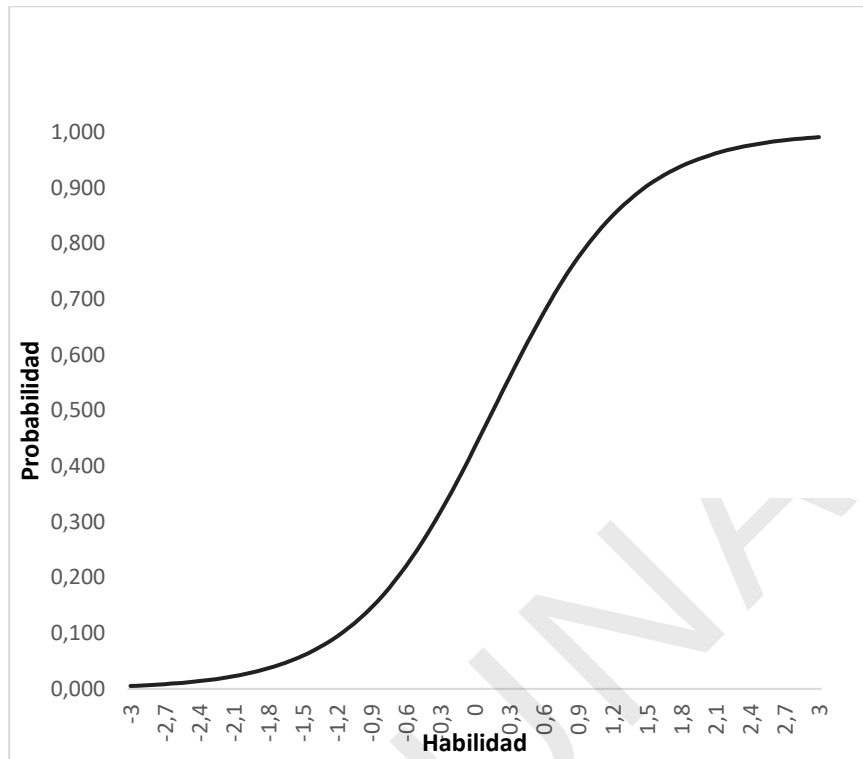
- La curva característica del ítem (CCI) representa una relación funcional expresada entre los valores de la variable que miden los ítems y la probabilidad de acertar estos (Hidalgo Montesinos & French, 2016).

- La puntuación verdadera es independiente del ítem y del test utilizado para estimarlo. Por lo tanto, la respuesta del sujeto depende únicamente del rasgo latente. (Matas, 2010)

- Invarianza de los parámetros, se supone que la estimación de los parámetros del ítem y el rasgo latente son independientes de las características de la muestra dentro de una población (Matas, 2010).

Aunado a lo anterior, existen diversos modelos matemáticos que pueden demostrar la relación entre la respuesta de un sujeto a un ítem y el rasgo latente, frecuentemente esta puede describirse como una función monótona creciente, expresada en términos probabilísticos, en específico se basa en análisis de regresión de datos multivariante que enlazan una variable respuesta con otra no observada. Normalmente, este análisis se obtiene por medio de funciones de probabilidad normal acumulada y logística y es representado gráficamente mediante la Curva Característica del Ítem (CCI). (Guler et al., 2014).

Gráficamente suele utilizarse la ojiva logística creciente, dicha curva se compone de dos asíntotas una superior y otra inferior, con una inflexión en el centro de la curva, que es el punto de máxima pendiente. La CCI se suele representar en una escala situada entre -3 y +3 puntos de habilidad (Navas Ara, 2002).

Figura 3. Curva característica de un ítem

Nota: La figura 2 muestra la probabilidad de acertar un ítem de acuerdo al nivel de habilidad.

El comportamiento de un individuo en un ítem es en función no solo de su nivel de aptitud, sino también de las características específicas del ítem (Martínez Arias et al., 2006). Por tanto, la forma de la CCI puede ser determinada por los siguientes parámetros del ítem:

Parámetro a: Mide el poder discriminatorio para identificar los sujetos en función del nivel en el rasgo latente que poseen. Este es proporcional a la pendiente de la recta tangente a la CCI en el punto de máxima pendiente de esta. Tiene un rango teórico entre $-\infty < a < +\infty$, aunque normalmente es comprendida entre -2.8 y 2.8. Cuanto mayor sea la pendiente, mayor será el poder discriminatorio del ítem (Muñiz, 2018).

Parámetro b: Representa el nivel de dificultad del ítem y se define como el nivel de habilidad en la que un individuo tiene una probabilidad del 50% de responder correctamente al ítem. Este describe dónde está situado el ítem en la escala de aptitud, por lo que se expresa en la misma escala que el rasgo latente, correspondiente al punto en el eje x de máxima pendiente de la CCI. La escala adoptada para medir θ tiene un rango teórico de $-\infty$ a $+\infty$, con media 0 y desvío típico 1. Cuanto mayor es el valor de b más difícil es el ítem (Muñiz, 2018; Navas Ara, 2001).

Parámetro c: La probabilidad de que un sujeto responda correctamente, poseyendo un nivel muy bajo de rasgo o acertándolo por azar. Es definido como el valor de la asíntota inferior de la curva característica del ítem. (Attorresi et al., 2009; Hidalgo-Montesinos y French, 2016).

2.3.2.1 Supuestos de TRI

Según Himelfarb, (2019) la teoría de respuesta al ítem es una colección de estadísticas y métodos psicométricos utilizados para modelar las respuestas a los ítems de los examinados. Los supuestos básicos de TRI descritos por Yang & Kao, (2014) son los siguientes :

-Una suposición de la TRI es la monotonidad, esto se representa mejor mediante la curva característica del ítem, en donde se ve reflejada gráficamente la verdadera relación entre el rasgo y las respuestas al elemento. Esta función especifica que, a medida que el nivel de habilidad aumenta, la probabilidad de responder correctamente también lo hace.

-La independencia local es otro supuesto de TRI. Esto asume que las respuestas del examinado a las preguntas no están estadísticamente relacionadas entre sí, incluso después de que el rasgo latente se toma en consideración o se mantiene estadísticamente constante. Hay dos componentes en la independencia local: el primero es que solo se considera un rasgo latente; el segundo es que la respuesta a una pregunta no depende de la respuesta a otra pregunta.

-Complementario al supuesto de independencia local es la unidimensionalidad: solo un rasgo latente se mide por el conjunto de ítems en la escala o prueba.

2.3.2.2 Modelos de TRI

Según Muñiz, (2010) existen diversos modelos que podrían plantearse dentro de la teoría de respuesta al ítem, no obstante sobresalen tres :

Modelo de 1 parámetro (1PL): También conocido como modelo de Rasch, describe la probabilidad de que un examinado responda correctamente considerando solamente la dificultad del ítem. En este modelo se define un parámetro de discriminación, cuyo valor es constante (1) (Rash, 1961).

En lo que respecta a este estudio, como se plantea en (UNESCO, 2016b) “Debido a las características y objetivos del TERCE, el modelo de Rasch presenta algunas ventajas por sobre los otros modelos de la TRI, como permitir la generación de una escala lineal de medición de la habilidad de los examinados. Además, es el que permite realizar estimaciones más robustas de sus parámetros debido a la existencia de estadísticos suficientes que permiten relacionar el puntaje bruto con la estimación de la habilidad. Por estos motivos, en el TERCE se utiliza el modelo Rasch”

Por consiguiente, resulta oportuno describir matemáticamente este modelo. El modelo de Rasch, describe que la probabilidad de responder correctamente es la misma para un examinado con un nivel de habilidad 1 al enfrentarse a un ítem con un grado de dificultad 1, que la de un examinado con el doble de habilidad enfrentándose a un ítem doblemente difícil (Thissen, D., & Wainer, 2001).

Matemáticamente, el modelo de Rasch se describe de la siguiente forma:

$$P(Y_{ij} = 1 | \theta_i, \beta_j) = \frac{\exp(\theta_i - \beta_j)}{1 + \exp(\theta_i - \beta_j)} \quad \text{con } \sum_i \theta_i = 0$$

Y_{ij} = Respuesta del individuo i al ítem j

θ_i = Habilidad del examinado i

β_j = Dificultad del ítem j

El parámetro que caracteriza al ítem en este de modelo es la dificultad (cuyos valores pueden fluctuar entre $(+\infty$ y $-\infty)$ y esta, al igual que en la teoría clásica de los test, no debe llegar a los extremos. Ítems de buena calidad no son demasiado fáciles ni demasiado difíciles.

Es preciso añadir que para hacer uso del modelo de Rasch es necesario un ajuste de los datos empíricos al modelo. (Karabatsos, 2000). En el TERCE el ajuste de datos se hace utilizando los estadísticos INFIT y OUTFIT (UNESCO, 2016b).

El INFIT es un estadístico de ajuste que brinda información ponderada cuyo enfoque es el comportamiento general de un ítem o de una persona. Se calcula con el promedio ponderado de las desviaciones cuadráticas estandarizadas entre el desempeño observado y el esperado. Este valor es sensible al comportamiento inesperado que afecta a

ítems cuya dificultad se encuentra cerca del nivel de habilidad del examinado. El OUTFIT permite determinar eventos poco usuales que ocurren de modo inesperado. Se determina por el promedio de las desviaciones cuadráticas estandarizadas entre el desempeño observado y el esperado. Detecta desajustes en las desviaciones lejos de la zona de medición del ítem. Es sensible al comportamiento inesperado que afecta a ítems cuya dificultad se encuentra lejos del nivel de habilidad del examinado (De Ayala, 2009). Matemáticamente estos estadísticos se representan de la siguiente manera:

$$INFIT_j = \sum_{i=1}^N \frac{V_{ij} Z_{ij}}{\sum_{k=1}^N V_{kj}} \quad \text{y} \quad OUTFIT_j = \frac{1}{N} \sum_{i=1}^N z_{ij}$$

Donde:

$Z_{ij} = \frac{Y_{ij} - E(Y_{ij})}{\sqrt{v_{ij}}}$ corresponde a la respuesta estandarizada del individuo i al ítem j

Y_{ij} corresponde a la respuesta del individuo i al ítem j

$E(Y_{ij})$ corresponde al puntaje esperado del individuo i al ítem j , es decir,

$$E(Y_{ij}) = \begin{cases} p_{ij}, & \text{si } Y_{ij} \in \{0,1\} \\ \sum_{k=1}^{k_j} k p_{ijk}, & \text{si } Y_{ijk} \in \{0,1, \dots, k_j\} \end{cases}$$

Donde:

$v_{ij} = \text{Var}(Y_{ij})$ es la varianza del puntaje del estudiante i en el ítem j .

Modelo de dos parámetros (2PL): Este modelo fue desarrollado por Binbaum a mitad del siglo pasado, aquí se asume que además de la dificultad se tiene en cuenta el índice de discriminación de los ítems. Existe el planteamiento de que la probabilidad de responder correctamente un ítem depende, además de la habilidad del examinado, de la dificultad de los ítems y de su grado de discriminación (Muñiz, 2018). Matemáticamente se describe de la siguiente manera:

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

Donde:

$P_i(\theta)$: Probabilidad de acertar el ítem i a determinado nivel θ

θ : Valores de la variable medida

b_i : Índice de dificultad del ítem i

a_i : Índice de discriminación del ítem i

e : Base de los logaritmos neperianos (2.72)

D : Constante ($D = 1.7$).

Modelo de tres parámetros (3PL): En este modelo, aunado a los parámetros anteriores dificultad y discriminación se añade la probabilidad de acertar el ítem al azar (Muñiz, 2018).

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

Donde:

$P_i(\theta)$: Probabilidad de acertar el ítem i a determinado nivel θ

θ : Valores de la variable medida

b_i : Índice de dificultad del ítem i

a_i : Índice de discriminación del ítem i

c_i : índice de pseudoazar del ítem i

e : Base de los logaritmos neperianos (2.72)

D : Constante ($D = 1.7$).

2.4 Análisis Factorial

Si bien es cierto son distintos los procedimientos para evidenciar la validez de un instrumento, un análisis sobre la validez de estructura interna es indispensable. La evidencia de validez de estructura interna implica un estudio para evaluar si la relación entre los ítems y componentes de la prueba se ajustan al marco conceptual del constructo y sus interpretaciones. (APA, AERA, NCME, 2018).

Según, Pérez-Gil et al., (2000) “ La validez de constructo es la principal de los tipos de validez, en tanto que la validez de constructo es el concepto unificador que integra las consideraciones de validez de contenido y de criterio en un marco común para probar hipótesis acerca de relaciones teóricamente relevantes”

Por consiguiente, la evidencia de validez de estructura interna implica un análisis respecto a la dimensionalidad de la prueba, el cual consiste en determinar si las interrelaciones de los ítems apoyan las puntuaciones propuestas que serán usadas para hacer inferencias (Rios & Wells, 2014).

El proceso de recolección de evidencias de validez de estructura interna implica una verificación empírica del ajuste de los ítems a la dimensionalidad propuesta del constructo (Prieto & Delgado, 2010).

Para evaluar la dimensionalidad del instrumento uno de los métodos más utilizados desde la teoría de respuesta al ítem (TRI) es el análisis factorial, que podría ser de carácter exploratorio o confirmatorio (Martínez Arias et al. 2006).

De acuerdo a Pere Joan Ferrando, (2006) el análisis factorial plantea que las relaciones entre un conjunto de variables, podrían ser explicadas a partir de un conjunto de variables no observadas llamadas “Latentes” a menudo denominadas factores, siendo el número de estas inferior al de las variables.

Por lo tanto, se sugiere tomar en consideración las hipótesis rivales, es decir, modelos teóricos alternativos para unidimensional y multidimensionales, esto con el propósito de identificar el modelo que se ajusta mejor a los datos. De acuerdo a diversas experiencias se recomienda el modelo unidimensional (Rios & Wells, 2014).

El estudio de validez de estructura interna de un test también implica el análisis de la invarianza y confiabilidad. La invarianza se refiere a la comparación de características de los ítems como discriminación y dificultad son afectados según diferentes grupos (sexo,

raza). El Análisis Factorial Confirmatorio Multigrupo permite determinar el nivel de invarianza entre los distintos grupos (Rios & Wells, 2014).

DEGT-UNAH

CAPÍTULO 3.

Marco Contextual

El siguiente apartado brinda una visión panorámica del contexto educativo nacional. Inicialmente, se presentan de manera cronológica la estructura del sistema educativo hondureño, se detallan los aspectos más relevantes del currículo nacional básico, finalmente la evaluación estandarizada en Honduras y su impacto en el sistema educativo.

3.1 El Sistema Educativo Hondureño

La Constitución de la República de Honduras, Art. 151 (1982) señala que "La educación es función esencial del Estado para la conservación, el fomento y difusión de la cultura, la cual deberá proyectar sus beneficios a la sociedad sin discriminación de ninguna naturaleza". Asimismo, establece en el (Art. 156), que la educación en todos los niveles del sistema educativo formal, excepto el nivel superior, es autorizada, organizada, dirigida y supervisada por el Poder Ejecutivo a través de la Secretaría de Educación.

La Ley Fundamental de educación (D.S. No 262-2011-PCM, 2011) (Art. 15) define la estructura del Sistema Nacional de Educación. En dicho artículo señala que está "conformado por los componentes siguientes: 1) Educación Formal; 2) Educación No Formal; y 3) Educación Informal". El (Art. 20) a su vez detalla que "la Educación Formal se organiza en los niveles: 1) Educación Pre-básica; 2) Educación Básica ;3) Educación Media; y, 4) Educación Superior. El desarrollo de la educación pre-básica, básica y media corresponde a la Secretaría de Estado en el Despacho de Educación, a través de sus correspondientes estructuras del nivel central. Le corresponde a la UNAH, a través del Consejo de Educación Superior, organizar, dirigir y desarrollar la Educación Superior y Profesional"

Una vez descrita la estructura del sistema educativo hondureño, es preciso mencionar que en el año 2000 la Secretaría de Educación comenzó a realizar esfuerzos encaminados a mejorar el sistema educativo nacional. Parte de estos esfuerzos fueron la creación de un nuevo currículo nacional básico, estándares educativos y un nuevo sistema de evaluación nacional alineados entre sí. Esta transformación surgió con el propósito y la necesidad de asegurar la equidad y mejorar la calidad educativa del país (Secretaría de Educación, 2009).

El currículo Nacional Básico, constituye una referencia que guía a los docentes, alumnos, padres de familia y autoridades del sistema educativo nacional. Los estándares conforman los objetivos educativos, coordinan los elementos del sistema educativo, son una base en el establecimiento de criterios para el diseño de pruebas, por lo que son útiles para contrastar el rendimiento entre instituciones o regiones del país, al establecer criterios de evaluación precisos para realizar evaluaciones internas y externas (Secretaría de Educación, 2009).

La Secretaría de Educación a través de la Dirección General de Currículo y Evaluación, es la responsable de reglamentar la evaluación que se aplica en todos los ciclos y modalidades del Sistema Educativo Nacional, el Sistema de Evaluación Nacional considera la evaluación diagnóstica y la evaluación formativa, también se definen las dimensiones que se evalúan en el proceso educativo. Los procesos de evaluación como se mencionó en los párrafos precedentes consideran los estándares definidos en el Currículo Nacional Básico (Secretaría de Educación, 2005b).

3.1.1 Indicadores del sistema educativo Hondureño.

Según datos proporcionados por la secretaría de educación para el año 2018, existen 25,878 centros educativos a nivel nacional, los mismos atienden educación prebásica, básica y media, de estos, 23,799 pertenecen a la administración gubernamental y 2,079 a la no gubernamental, 12,390 atienden la educación prebásica, 13,409 educación básica y 2411 centros atienden la educación media. (Sistema de Administración de Centros Educativos (SACE). 2018)

Un indicador relevante en el contexto educativo hondureño es el nivel de cobertura, decisivo a la hora de medir el desempeño académico de los estudiantes, para el año 2019 se estimaba que la tasa de cobertura total para la población en un rango de edad entre 5 y 18 años era de 54.2%. (Instituto Nacional de Estadísticas (INE), 2019).

En lo que respecta a la tasa de matrícula en el nivel de Educación Prebásica, se refleja un aumento desde el año 2014, 37.60 % al año 2018 con un 42.38% , en cambio en el nivel de Educación Básica (I y II ciclo) muestra un descenso, en el año 2014 un 99.42% hasta el año 2018 con un porcentaje de 95.40%, igualmente sucede con el III ciclo, en el año 2014; 69.16% hasta el año 2018 con un 67.61% y en el nivel de Educación Media ha habido

un leve aumento del año 2014; 42.2% a un 44.3% en el año 2018, lo que significa que ha habido una leve disminución en la matrícula bruta en el período 2014-2018 (Secretaría de Educación, 2019).

Sin embargo, una recomendación importante por parte de los organismos internacionales a partir de los resultados en pruebas internacionales de evaluación, es que los países de América Latina tienen la necesidad de establecer reformas vinculadas a la calidad educativa, bajo la premisa de que la escolarización no es garantía de aprendizaje (Rojas, 2020).

3.1.2 Currículo nacional básico.

El Currículo Nacional Básico (CNB) "es el instrumento normativo que establece las habilidades, competencias, destrezas y actitudes que debe lograr todo sujeto del Sistema Educativo Nacional en los niveles, ciclos y/o modalidades que rectora la Secretaría de Educación". Los fundamentos legales del Currículo Nacional Básico están determinados por las normativas legales vigentes, establecidas en la Constitución de la República, Ley Orgánica de Educación, reglamentos educativos, y otras leyes nacionales e internacionales (Secretaría de Educación, 2005).

El Currículo Nacional Básico del Sistema Educativo Hondureño define tres niveles: Educación prebásica que incluye niños entre 3 y 6 años de edad y tiene una duración de tres años, educación básica con estudiantes entre 7 y 15 años de edad, la misma tiene una organización en tres ciclos de tres años cada uno e incluye los grados de 1º a 9º y educación media con estudiantes entre 16 y 18 años que incluye los bachilleratos. El instrumento a validar corresponde al segundo nivel del sistema educativo hondureño (Secretaría de Educación, 2005b).

Con el propósito de mejorar la calidad educativa del país, aunado al currículo nacional básico también se han puesto en práctica los Estándares Educativos Nacionales, con ello la secretaría de educación pretende alcanzar la equidad en los aprendizajes en el contexto educativo hondureño, ya que los mismos señalan lo que los alumnos tienen que saber (conocimientos) y saber hacer (destrezas), independientemente de su contexto geográfico, cultural o social (Secretaría de Educación, 2009).

En el año 2009 se presentaron los estándares de 1° a 9° grado para las áreas de ciencias naturales y ciencias sociales. Estos estándares fueron elaborados por personal técnico y consultados con distintos grupos de actores del proceso educativo. Los mismos tienen una estructura coherente de acuerdo a lo establecido en el CNB (Secretaría de Educación, 2009).

Como el análisis psicométrico corresponde a la prueba de ciencias naturales para estudiantes de sexto grado de TERCE, es conveniente presentar los estándares educativos de Honduras en el área de ciencias naturales para estudiantes de sexto grado de acuerdo a los ejes temáticos evaluados en la prueba de TERCE.

Tabla 3. Estándares educativos para sexto grado en el área de ciencias naturales

Bloque	Estándares
Seres Vivos	Describen la estructura de la célula, su funcionamiento y principios de la teoría celular. Describen las características de los seres vivos según el reino al que pertenecen. Reconocen las semejanzas y diferencias entre individuos de una misma especie a través de la herencia de genes. Determinan las características e importancia de los diferentes ecosistemas. Identifican las características de los recursos naturales y peligros a los que están expuestos por las acciones del ser humano. Describen los ecosistemas naturales más importantes de Honduras mencionando su ubicación.

El cuerpo Humano	Identifican las características,
Los Sentidos	funciones y cuidados de los sistemas
Higiene, Salud y Nutrición	reproductores femenino y masculino.
	Identifican las principales
	características, funciones y cuidados de
	los sistemas nervioso y endocrino.
	Describen las características y
	cuidados de los órganos de los sentidos y
	su relación con el sistema nervioso.
	Identifican las características,
	formas de prevención, cuidado y
	tratamiento de las enfermedades
	frecuentes y adicciones. Identifican los
	órganos y funciones del sistema
	inmunológico del ser humano.
	Reconocen los riesgos, medidas
	de prevención y la forma de actuar en caso
	de que se produzcan accidentes y
	fenómenos meteorológicos
El Universo	Describen el origen y
El Sistema Solar	organización del universo, mencionando
Composición de la Tierra	los avances tecnológicos para su
	exploración.
	Describen las características de
	los planetas, mencionando la importancia
	que ejerce el sol sobre todos los elementos
	del sistema.
	Identifican los elementos que
	intervienen en los cambios del clima,
	mencionando los instrumentos de
	medición.

	Reconocen las características de las capas y subcapas de la tierra y su utilidad a los seres vivos.
	Mencionan las características de los procesos que contribuyeron a la formación de la corteza terrestre.
La Materia	Describen las características de los cambios físicos y químicos de la materia.
La Energía	Reconocen las características de las formas, fuentes y transformaciones de la energía.

Nota : Adaptado de (Secretaría de Educación, 2009)

3.1.3 Evaluación estandarizada de los aprendizajes en Honduras.

En el contexto educativo hondureño a partir de 1990 a 1994 tuvieron lugar los primeros intentos de llevar a cabo evaluaciones a gran escala, como parte del Programa de Mejora de la Educación Primaria, apoyado por la Agencia de Estados Unidos para el Desarrollo Internacional (USAID). En 1995 se creó la Unidad de Medición de la Calidad de la Educación (UMCE), de la Universidad Pedagógica Nacional Francisco Morazán, que asumió la responsabilidad de las actividades de evaluación en el país, en cooperación con la Secretaría de Educación Pública (Martínez, 2008).

El primer levantamiento de datos tuvo lugar en 1997, cubriendo los grados 3° y 6° de primaria y las áreas de Español y Matemáticas. 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. En 1999 se evaluaron los grados 2°, 3°, 4° y 5°; en 2000, 2002 y 2004, sólo 3° y 6°, incorporándose en la última ocasión el área de Ciencias Naturales (Martínez, 2008).

A partir de la implementación del currículo nacional básico, en el año 2003 también se puso en práctica un sistema de evaluación centrado en los estándares educativos, este

sistema de evaluación nacional incluye pruebas diagnósticas, aplicadas al inicio del año escolar, pruebas formativas mensuales y pruebas de fin de grado. Es para el año 2012 cuando se destacan dos acontecimientos importantes referente a la evaluación nacional: 1) Evaluación de rendimiento académico a nivel censal y a nivel muestral de 1° a 9° grado. 2) Evaluación de escritura (Español) de 1° a 9° grado (Secretaría de Educación, 2012).

La responsabilidad técnica de la elaboración y aplicación de las pruebas externas y estandarizadas del 2012, fue coordinado por el Proyecto Mejorando el Impacto al Desempeño Estudiantil de Honduras (MIDEH) bajo un convenio realizado desde finales del año 2004 entre la Secretaría de Educación en cooperación con (USAID) y con la participación de la Dirección General de Evaluación de la Calidad Educativa (DIGECE) (FEREMA, 2018).

En noviembre del 2015, por cuarto año consecutivo en Honduras, la Secretaría de Educación por medio de DIGECE, coordinó la evaluación estandarizada de fin de grado en los espacios curriculares de Matemáticas y Español de 1° a 9° grado. Es oportuno señalar que, para este año por primera vez en la historia de la educación hondureña, la evaluación se realizó considerando una muestra con representatividad hasta el nivel municipal, lo que elevó considerablemente la inversión para el desarrollo de la evaluación (Secretaría de Educación, 2015b).

Otra evaluación realizada en el 2015 fue la que se aplicó a la población conformada por estudiantes que en el 2015 egresaron de Educación Media, de esta población se seleccionó una muestra representativa, en total se planificó la aplicación a 11,000 egresados distribuida en los 18 departamentos del país. Participaron en total 160 Centros Educativos. Esta evaluación incluyó español, matemáticas, ciencias naturales y ciencias sociales. En esta evaluación los resultados fueron bajos en todas las áreas resaltando la importancia de implementar mejoras en el sistema educativo (Secretaría de Educación, 2015a).

En noviembre del año 2018 se aplicaron las pruebas de evaluación del desempeño estudiantil a una muestra conformada por 648 centros educativos y a un total de 23,786 estudiantes evaluados en Español (solamente el área de la Lectura) y Matemáticas del 3er, 6to y 9no grado de educación básica. De manera general los resultados señalan que los estudiantes evaluados en el año 2018 siguen mostrando que tienen mejores niveles de

desempeño en Español (Lectura) que en el área de Matemáticas, aunque no son los resultados esperados en el sistema educativo (Secretaría de Educación, 2018).

3.1.4 Honduras en las evaluaciones internacionales.

En lo que respecta a las pruebas estandarizadas a nivel regional, Honduras participó en el año 1997 en el primer estudio regional comparativo y explicativo PERCE, también participó en el año 2013 en el tercer estudio regional comparativo y explicativo TERCE. Finalmente, para el año 2019, Honduras tuvo participación en el estudio regional comparativo y explicativo ERCE (UNESCO, 2019).

Otros estudios internacionales importantes que implican pruebas estandarizadas son TIMSS (Estudio Internacional de Tendencias en Matemáticas y Ciencias) y PIRLS (Estudio Internacional de Progreso en Comprensión Lectora) ambos evalúan el rendimiento del alumnado en 4º curso de Educación Primaria (10 años) y TIMSS también en 2º curso (14 años). En ambos estudios Honduras fue participe (Mullis et al., 2012).

La participación de Honduras en estudios internacionales ha sido relevante puesto que ha permitido obtener una visión panorámica acerca del estado actual de la educación en el país. Refiriéndose en específico a los resultados obtenidos en las pruebas de la UNESCO en el estudio TERCE, Honduras se encuentra entre los países cuyos promedios en ciencias naturales están por debajo del promedio (668) de todos los países evaluados en el TERCE. (UNESCO, 2016a) y en los resultados del ERCE recientemente publicados, el país se encuentra en la categoría de países que mantienen su promedio, en el área de ciencias en la prueba ERCE Honduras obtuvo (678) puntos (UNESCO, 2021).

En lo que respecta a las pruebas de desempeño PISA, es oportuno señalar que después de las experiencias previas en el año 2000 con países de ingresos medios, en un esfuerzo por responder a la demanda emergente para atender a una gama más amplia de países, la OCDE lanzó la iniciativa PISA para el Desarrollo (PISA-D) en 2014. Este proyecto piloto único, que abarca seis años, tiene como objetivo hacer que la evaluación sea más accesible y relevante para los países de ingresos bajos a medios, a este grupo de países pertenece Honduras (OCDE, 2018).

Respecto a los resultados obtenidos en la evaluación de las pruebas PISA-D (PISA para el desarrollo) para estudiantes de 15 años de edad y en específico en el área de ciencias naturales, para los estudiantes hondureños el nivel (1a) fue el nivel promedio. La evaluación

mide la capacidad de los estudiantes para: explicar fenómenos científicamente; evaluar y diseñar la investigación científica; e interpretar datos y evidencia científicamente (OECD, 2018).

En TIMMS, Honduras obtuvo 396 puntos en Matemáticas y 432 en Ciencias. Esta evaluación como tal proporciona a los países una oportunidad sin precedentes de medir el progreso del rendimiento educativo en matemáticas y ciencias. Como se mencionó antes, una particularidad de estas pruebas es que las pruebas de 4 ° grado fueron aplicadas a los estudiantes de sexto grado del sistema de educación hondureño (Mullis et al., 2012).

En términos generales, pese a los bajos resultados es preciso añadir que la información generada a partir de las evaluaciones estandarizadas internacionales es detallada, amplia y valiosa, puesto que se genera un sentido de responsabilidad por parte de los diferentes actores del sector educativo a partir de los resultados obtenidos, en tanto que se ha visto reflejado que las evaluaciones nacionales producen un menor impacto en los mismos (FEREMA, 2018).

Los avances realizados han sido considerables en las técnicas de evaluación de resultados de rendimiento académico a nivel nacional, desarrolladas gracias al apoyo de organismos internacionales que han brindado ayuda técnica y económica, no obstante, es importante hacer hincapié en que, si bien la evaluación de rendimiento académico no es la solución absoluta, es una condición necesaria para la mejora de la calidad de la educación (Secretaría de Educación, 2017).

CAPÍTULO 4.

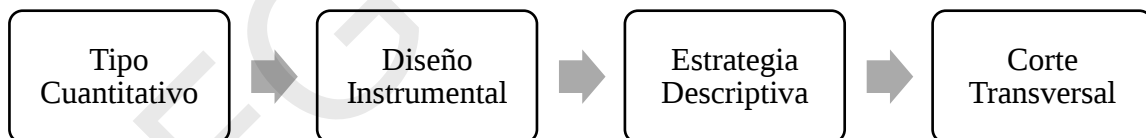
Método

En esta sección se describen los procedimientos metodológicos a utilizar en el análisis psicométrico de la prueba de desempeño académico en ciencias naturales para estudiantes de sexto grado. Este incluye el tipo de investigación, operacionalización de variables, caracterización de la muestra y una descripción del instrumento a validar.

4.1 Tipo de investigación

El estudio de evidencias de validez de la prueba sobre logros de aprendizaje en ciencias naturales es de tipo cuantitativo, ya que se analiza la base de datos de los puntajes obtenidos en la aplicación de la prueba. (Hernández Sampieri et al., 2010). Consiste en un diseño Instrumental, ya que se realiza un análisis psicométrico para reunir evidencias de validez del test. La estrategia utilizada es descriptiva, puesto que se pretenden describir los parámetros psicométricos que determinan el comportamiento de los ítems en la prueba, es de corte transversal, puesto que se tomarán en cuenta los datos del Test, aplicado en el año 2013 (Ato et al., 2013).

Figura 4. Procedimientos metodológicos para el análisis psicométrico de la prueba de CCNN



Fuente: Elaboración propia

4.2 Población y muestra.

La población objeto de estudio está compuesta por estudiantes que cursaron el sexto grado de educación básica del sistema educativo nacional en el año 2013. El tamaño de la muestra es de 3,886 estudiantes. No se conocen las características sociodemográficas.

Se optó por un diseño muestral estratificado, por conglomerados, un sistema de selección de escuelas aleatorio. Las variables de estratificación fueron las siguientes:

- Dependencia administrativa de la escuela: pública y privada
- Área: rural y urbana

-Grados en la escuela: solo tercer grado, solo sexto grado, tercer y sexto grado

Los criterios de inclusión para seleccionar las escuelas participantes fueron los siguientes:

- Escuelas con una ubicación física única.
- Escuelas con un número definido e identificable de alumnos y profesores.
- Escuelas que contaban con un sistema de administración educativo único.
- Escuelas con un espacio social identificable.

Se consideró población no apta:

- Escuelas en donde se imparte educación no formal.
- Escuelas para alumnos con necesidades especiales.
- Alumnos en escuelas regulares pero que tienen necesidades especiales.
- Escuelas en donde se imparte educación para adultos.

Como parte de las normas éticas para la recolección de datos los alumnos con necesidades especiales en escuelas regulares, participaron de las sesiones de aplicación, pero luego sus registros no fueron considerados en los análisis posteriores de los datos. Esto, para cumplir con los criterios de no discriminación imperantes en la mayoría de los sistemas educativos de la región.

La única categoría de exclusión dentro de la población objetivo que se considera válida en el TERCE son las escuelas pequeñas. Se considera este tipo de escuelas como sujetas de exclusión por la relación que existe entre los costos que implica aplicar los test, versus los beneficios que aporta a la validez de los resultados, un grupo pequeño de alumnos. Sin embargo, en el TERCE, se decide excluir estas escuelas hasta un cierto porcentaje,

Dado esto, se ha fijado un criterio de exclusión de un 2% de escuelas pequeñas en los países del TERCE.

En la siguiente tabla se detalla únicamente el número de estudiantes que contestó cada una de las formas de la prueba.

Tabla 4. Número de estudiantes por cuadernillo

Cuadernillo	Número de Estudiantes
Cuadernillo No 1	655
Cuadernillo No 2	643
Cuadernillo No 3	654
Cuadernillo No 4	651
Cuadernillo No 5	645
Cuadernillo No 6	638
Número Total de Estudiantes	3886

Como parte del análisis, primero se realizó una revisión de la base de datos, se identificó que faltaba una variable que identificara el cuadernillo que contestó cada estudiante por lo que en Excel se creó la misma con el fin de identificar la forma de la prueba resuelta por cada examinado, esto era necesario para poder seccionar la base tanto a nivel de cuadernillo como por bloque para realizar los análisis correspondientes.

4.3 Instrumento

El instrumento a validar es la prueba sobre logros en aprendizajes en ciencias naturales para estudiantes de sexto grado. Este instrumento fue elaborado por el Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (LLECE), coordinado por la OREALC/UNESCO. El instrumento evalúa conocimiento en ciencias naturales, considera el reconocimiento de información y conceptos, la comprensión y aplicación de los mismos, el pensamiento científico y la resolución de problemas. La prueba considera cinco ejes temáticos o dominios (salud, seres vivos, ambiente, la tierra y el sistema solar, materia y energía) cada uno de estos dominios está incluido en los bloques que conforman cada cuadernillo. Se pilotearon 19 ítems sobre salud, 22 sobre seres vivos, 22 sobre ambiente, 13 sobre la tierra y el sistema solar y 12 sobre materia y energía para un total de 86, de los cuales finalmente solo se consideraron 82. El tiempo máximo permitido para resolver la prueba eran 60 minutos. La construcción de la prueba se basó en ítems elaborados en talleres en los que

se contó con la participación de todos los países que forman parte del estudio. Para ello lo primero que se hizo fue la revisión de los marcos curriculares de los países participantes, esto lo hizo el Instituto Colombiano para la Evaluación de la Educación (ICFES) lo que permitió definir la estructura de una prueba apta para medir la calidad de la educación a nivel regional. Los ítems utilizados en la prueba son de selección única, con cuatro opciones de respuesta, también se utilizaron ítems de respuesta corta

Tabla 5. Matriz de operacionalización de variables

Variable	Definición conceptual	Definición operacional
Logros de aprendizaje	se constituye en un indicador del aprendizaje alcanzado por el estudiante, representa el nivel de eficacia en la consecución de los objetivos (UNESCO, 2016b).	en ciencias naturales se mide a través del dominio de las habilidades y destrezas que demuestran los estudiantes en las diferentes ramas de las ciencias naturales (Jaramillo, 2019).

Fuente: Elaboración propia.

La tabla de especificaciones de la prueba de ciencias de sexto grado de acuerdo al Reporte Técnico del TERCE 2013 es la siguiente:

Tabla 6. Especificaciones de la prueba de ciencias naturales de sexto grado

Dominio	Proceso			Total	%
	Reconocimiento de información y conceptos	Comprensión y aplicación de conceptos	Pensamiento científico y resolución de problemas		
Salud	5	7	7	19	21%
Seres Vivos	9	10	6	25	27%
Ambiente	4	15	3	22	24%
La Tierra y el Sistema Solar	3	6	4	13	14%
Materia y Energía	3	6	4	13	14%
Total	24	44	24	92	100%
%	26%	48%	26%	100%	---

Nota : Tomado de (UNESCO, 2016b)

Tabla 7. Formas en la prueba de CCNN del TERCE

Cuadernillo	Bloques en cada cuadernillo
Cuadernillo No 1	B1-B2
Cuadernillo No 2	B3-B4
Cuadernillo No 3	B5-B1
Cuadernillo No 4	B6-B3
Cuadernillo No 5	B2-B6
Cuadernillo No 6	B4-B5

Nota : Tomado de (UNESCO, 2016b)

Tabla 8. Número de preguntas por bloque

Bloque	Número de Preguntas
B1	15
B2	15
B3	13
B4	14
B5	12
B6	13

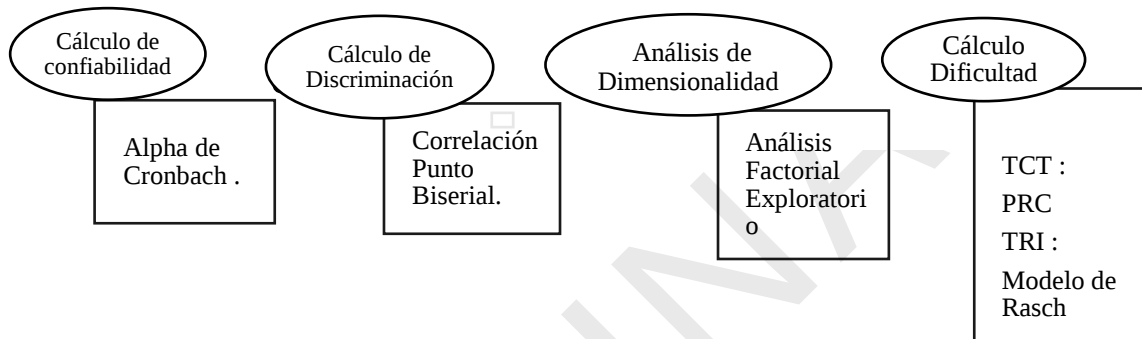
4.4 Procedimientos para el análisis de datos.

1. El nivel de confiabilidad de la prueba por medio de la teoría clásica de los test, se determinó haciendo uso del coeficiente alpha de Cronbach.
2. Para determinar el grado de dificultad de los ítems mediante la teoría clásica de los test se hizo el cálculo del porcentaje de respuesta correcta PRC.
3. Para determinar el poder de discriminación de los ítems mediante TCT, se hizo uso de la correlación biserial puntual.
4. Se evaluó la dimensionalidad del test, para ello se realizó un análisis factorial exploratorio, en el cual se tomaron en cuenta la prueba KMO y prueba

de esfericidad de Barlett y el porcentaje de varianza explicado por el primer componente.

5. Se determinó el ajuste de datos mediante los estadísticos INFIT y OUTFIT y para determinar el grado de dificultad por medio de TRI se aplicó el modelo de Rash, en donde se hizo un análisis del parámetro b (dificultad).

Figura 5. Plan de análisis de datos



Nota: Para los análisis se utilizarán los programas R Studio, SPSS, y Excel.

CAPÍTULO 5.

Resultados

En el siguiente apartado se detallan los principales hallazgos obtenidos al aplicar los procedimientos psicométricos pertinentes a este estudio de validación. Primero se exponen los resultados obtenidos respecto al análisis de confiabilidad, luego se presentan los resultados obtenidos sobre los índices de dificultad y discriminación por medio de la TCT. Se presentan un análisis factorial utilizando un análisis de varianza, así como los resultados obtenidos referidos a los índices de ajuste y finalmente se detallan los índices de dificultad por medio de la TRI para cada uno de los ítems de la prueba de ciencias naturales del TERCE.

5.1 Confiabilidad de la Prueba de Ciencias Naturales del TERCE para Honduras

El propósito de este apartado era analizar la fiabilidad de cada una de las formas (cuadernillos) de la prueba de ciencias naturales del tercer estudio regional comparativo y explicativo TERCE, a fin de establecer que la prueba es consistente y que se permiten realizar los procedimientos psicométricos necesarios como parte del estudio de validación de este instrumento. El análisis de confiabilidad por medio del alpha de cronbach permite evaluar la configuración de estructura interna del test.

Tabla 9. Coeficientes Alpha de Cronbach por cuadernillo en la prueba CCNN Honduras TERCE

Número de Cuadernillo			Alpha de Cronbach TERCE Muestra Honduras	Alpha de Cronbach TERCE Muestra Regional
1	Cuadernillo	No	0.70	0.74
2	Cuadernillo	No	0.70	0.75
3	Cuadernillo	No	0.70	0.73
4	Cuadernillo	No	0.66	0.69
5	Cuadernillo	No	0.68	0.71
6	Cuadernillo	No	0.70	0.74
Grado de confiabilidad promedio			0.69	0.72

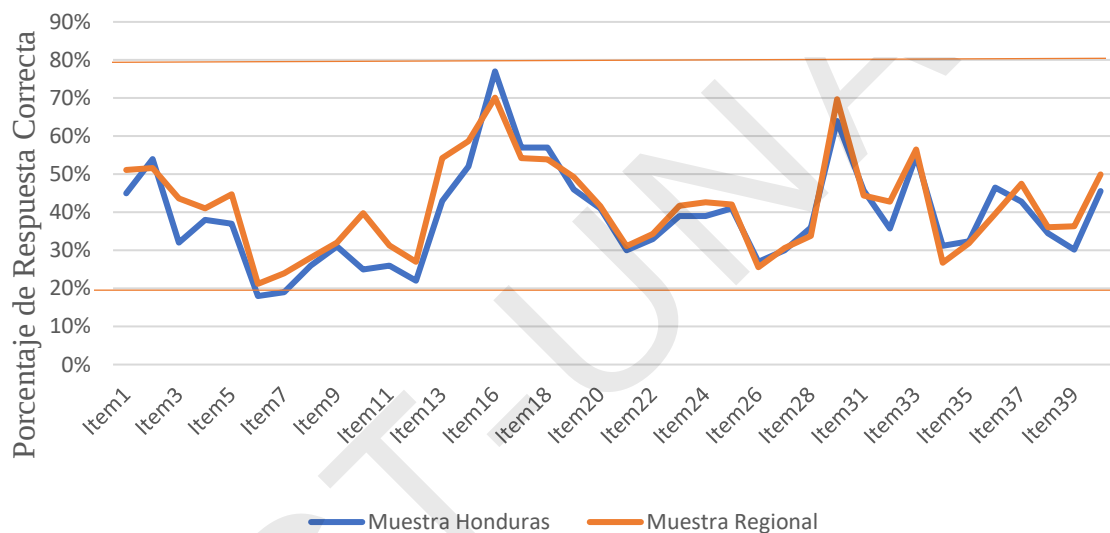
Para cumplir con el primer objetivo, al realizar el análisis de confiabilidad de la prueba, se hizo el cálculo del coeficiente Alpha de Cronbach para cada uno de los seis cuadernillos, la tabla 9 muestra de manera general coeficientes con valores cercanos al mínimo aceptable (Bojórquez et al., 2013) para cada una de las formas de la prueba, así como el grado de confiabilidad promedio, esto también con relación a los rangos establecidos por el TERCE siendo los valores iguales o superiores a 0,7 más convenientes.

En comparación con los resultados de confiabilidad obtenidos en la aplicación a la muestra regional en la prueba TERCE, puede verse que los resultados tienen ligeras variaciones, pero en general el grado de confiabilidad es similar para ambas muestras.

5.2 Análisis de dificultad de los ítems de la prueba de ciencias naturales del TERCE.

El objetivo de este apartado consistía en realizar un análisis de la dificultad de cada uno de los 82 ítems de la prueba de ciencias naturales, haciendo uso de la teoría clásica de los test, puesto que la dificultad se constituye en uno de los parámetros más determinantes a la hora de determinar la calidad métrica de un ítem.

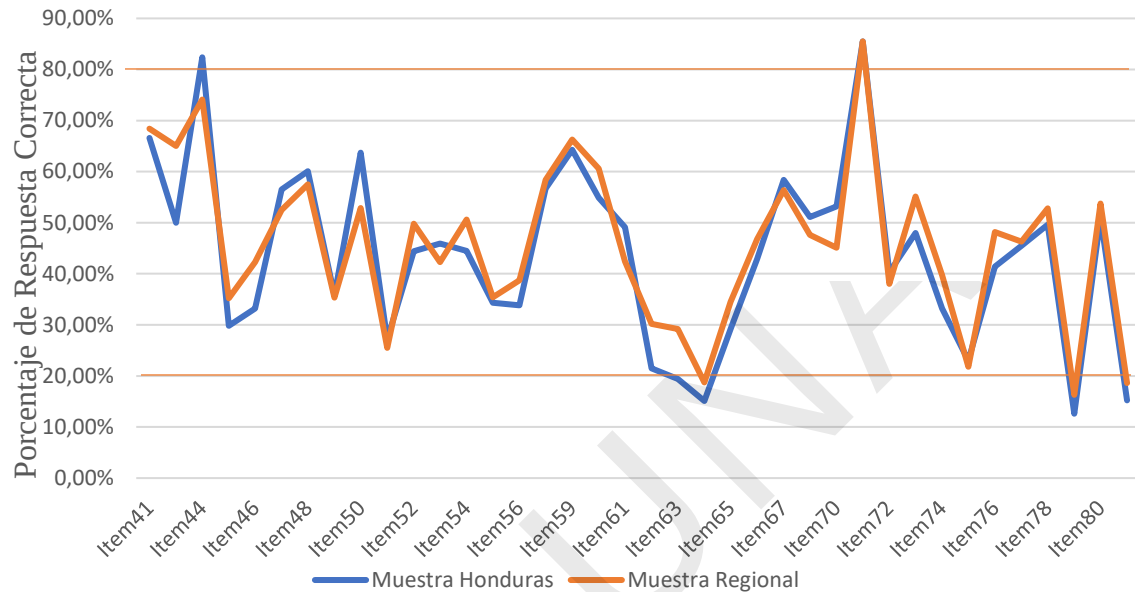
Figura 6. Dificultad de los ítems 1- 40 según TCT para las muestras nacional y regional



Como se mencionó previamente el siguiente objetivo de este estudio consistía en determinar el nivel de dificultad de los ítems; para ello se hizo el cálculo del porcentaje de respuesta correcta y de acuerdo con los criterios establecidos en el TERCE se consideran ítems con dificultades aceptables aquellos localizados entre un rango de 20 % y 80%. De manera general pueden observarse valores de dificultad similares y aceptables para ambas muestras. No obstante, puede notarse en la figura 4, que para los estudiantes hondureños los ítems parecen ser más difíciles sobre todo en el conjunto de ítems del 1 al 14 puede observarse una diferencia más visible. Sin embargo, en los ítems 16, 17, 18, el PRC es mayor en el caso de los estudiantes hondureños. El ítem 16 parece ser el más fácil para ambos grupos. También cabe señalar que los ítems 6 y 7 tendrían que ser eliminados de acuerdo con los criterios de

exclusión previamente descritos, puesto que su porcentaje de respuesta correcta es inferior al 20%. El detalle de estos valores puede verse en el apéndice.

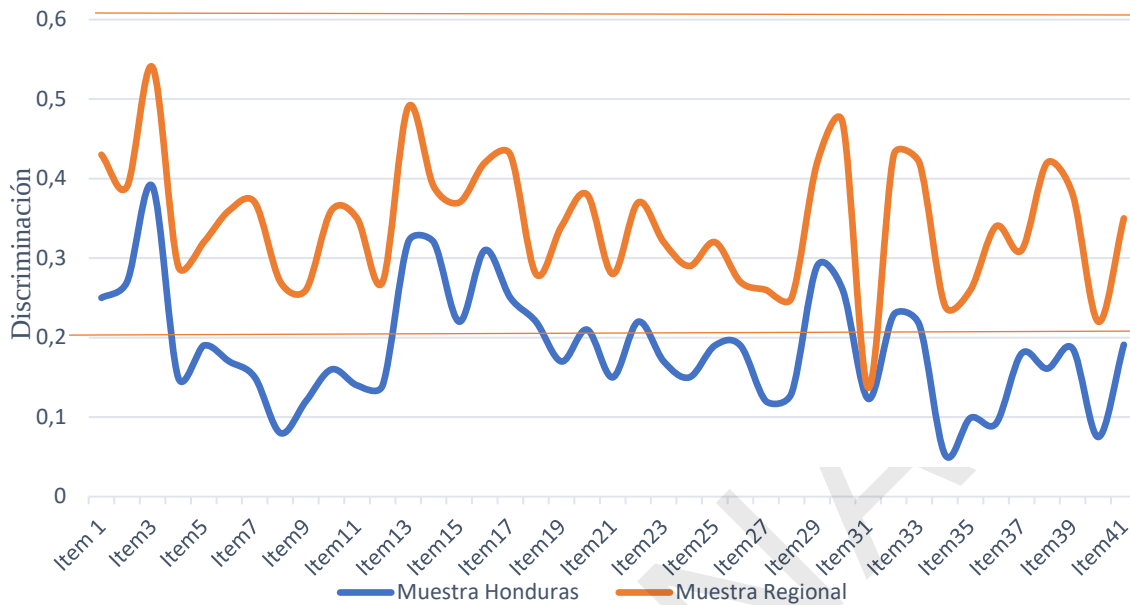
Figura 7. Dificultad de los ítems 41- 81 según TCT para las muestras nacional y regional



Para el conjunto de ítems restantes del 41 al 81, puede observarse una tendencia similar, mayoría de dificultad para la muestra de estudiantes hondureños, los ítems 64, 79 y 81 tendrían que ser eliminados de acuerdo con los criterios de exclusión previamente descritos, tanto en la muestra de estudiantes hondureños como en la muestra regional los niveles de dificultad se encuentran fuera del rango de tolerancia. Es importante añadir que de manera general en lo que respecta al parámetro dificultad mediante TCT, los ítems muestran una variabilidad importante.

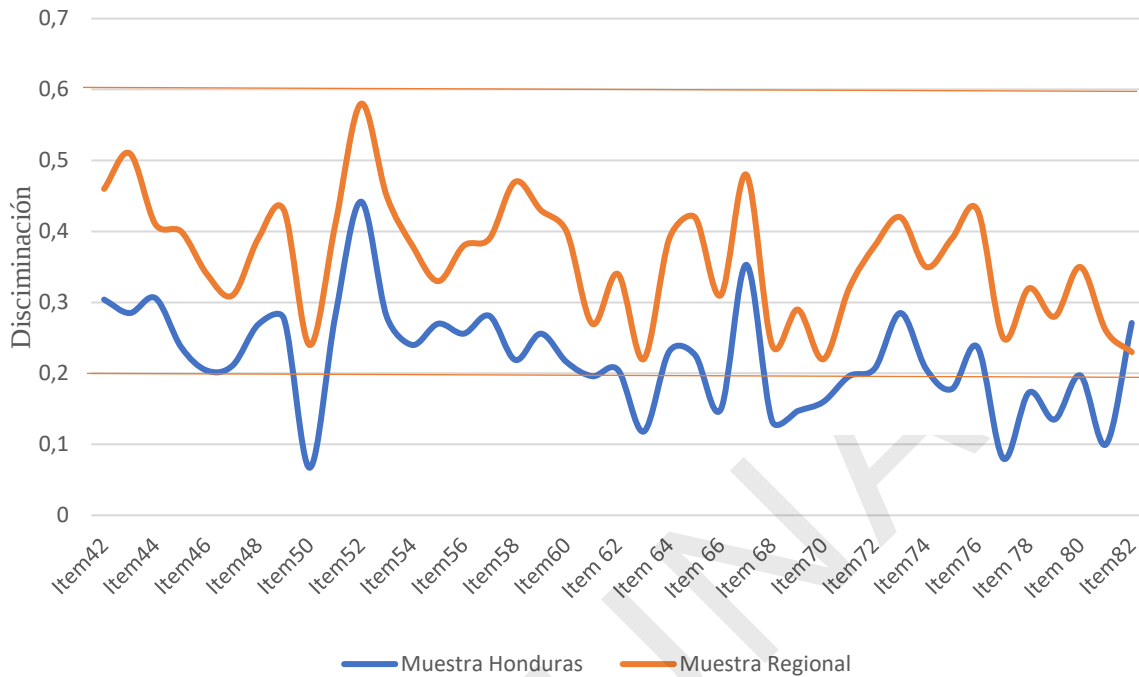
5.3 Análisis de discriminación de los ítems de la prueba de ciencias naturales del TERCE

La finalidad del siguiente apartado consistía en realizar un análisis sobre el parámetro discriminación mediante el uso de la teoría clásica de los test para cada ítem, a fin de conocer si los ítems permiten diferenciar adecuadamente a los sujetos con diferentes niveles de habilidad o rasgo que evalúa el test.

Figura 8. Discriminación de los ítems 1- 41 según TCT para las muestras nacional y regional

En cumplimiento con el tercer objetivo de este estudio, al analizar el poder discriminativo de los ítems y de acuerdo con los criterios de exclusión establecidos por el TERCE ($\rho > 0,3$ el ítem discrimina en forma aceptable, $0,2 \leq \rho \leq 0,3$ el ítem discrimina regularmente y debe ser revisado, $\rho < 0,2$ el ítem debiera ser eliminado de la prueba). La figura 8 muestra diferencias muy marcadas entre ambas muestras, en lo que respecta a los ítems desde 1 al 41, el patrón más notable para la muestra de estudiantes hondureños es el de ítems con una correlación biserial puntual inferior a 0.2, tan solo 3 de estos 41 ítems (3, 14, 16) tienen correlaciones superiores o iguales a 0.3, es decir que tan solo tres discriminan de forma aceptable. No obstante, en lo que respecta a la muestra regional se observa un patrón diferente, pues la mayoría de los ítems discrimina de forma aceptable con excepción del ítem 31 en el cual la correlación es de 0.137 aún en la muestra regional, por lo cual debió ser eliminado (pueden verse más detalles en las tablas que se adjuntan en el apéndice)

Figura 9. Discriminación de los ítems 42- 82 según TCT para las muestras nacional y regional



Continuando con el análisis de discriminación del conjunto de ítems del 42 al 82, se observa un incremento en los valores de correlación biserial puntual para la muestra de Honduras, en comparación con los ítems precedentes, aquí se reflejan valores superiores a 0.2 incluso valores entre 0.3 y 0.4 para algunos ítems, aunque al menos 15 de estos ítems presentan correlaciones inferiores a 0.2 lo cual sugiere una revisión, el ítem 50 en la muestra de Honduras presenta una correlación de 0.06 sumamente bajo y en la muestra regional 0.24 no satisfactorio, por tanto este ítem no debería incluirse en la prueba. En el caso de la muestra regional los valores son satisfactorios para la mayoría del conjunto de ítems, por lo que se marca una brecha significativa entre ambas muestras. El detalle de estos valores puede observarse en el apéndice.

5.4 Análisis de Dimensionalidad de la Prueba de CCNN del TERCE

El siguiente objetivo de este estudio consistía en realizar un análisis de dimensionalidad del instrumento con el fin de evaluar la configuración interna del test, ya que, para proceder con la aplicación de la teoría de respuesta al ítem, es necesario el cumplimiento del supuesto de unidimensionalidad, este supuesto es muy relevante para obtener buenas mediciones. Para ello se realizaron las pruebas estadísticas correspondientes y un análisis factorial en donde se evaluó la varianza explicada por los componentes de cada una de las formas de la prueba.

Tabla 10. Resúmenes estadísticos KMO y prueba de esfericidad de Barlett

	Cuadernillo	Prueba KMO	Prueba de Esfericidad de Barlett
1	Cuadernillo	0,767	,000
2	Cuadernillo	0,780	,000
3	Cuadernillo	0,777	,000
4	Cuadernillo	0,742	,000
5	Cuadernillo	0,733	,000
6	Cuadernillo	0,786	,000

Para evaluar el supuesto de unidimensionalidad se requiere de los índices de medida de adecuación muestral Kaiser-Meyer-Olkin (KMO) en este estudio los valores resultaron satisfactorios para cada cuadernillo (formas de la prueba) con valores próximos a 1, lo cual indica que las correlaciones entre pares de variables pueden explicarse por otras

variables, por lo que en este caso sería una evidencia más de que se puede realizar el análisis factorial. En cuanto a la prueba de esfericidad de Barlett pueden observarse valores inferiores (,000) al nivel de significancia por tanto los datos son adecuados para realizar una reducción de datos.

Tabla 11. Resumen varianza total explicada por el primer componente para cada cuadernillo

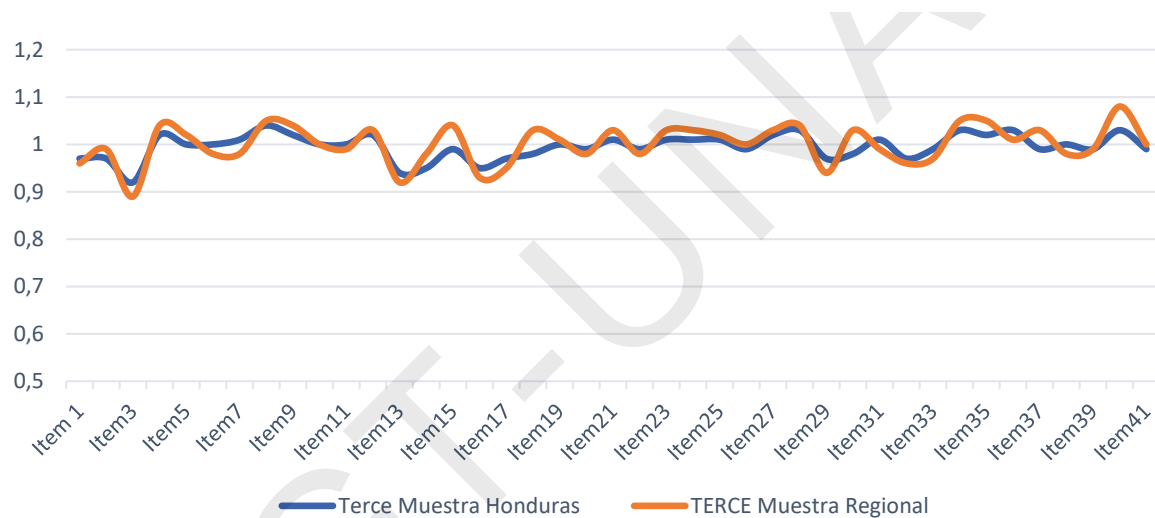
Autovalores Iniciales					Sumas de Cargas al Cuadrado de la Extracción		
Cuadernillo	Componente	Total	% de Varianza	% Acumulado	Total	% de Varianza	% Acumulado
1	1	,949	13,281	13,281	,949	13,281	13,281
2	1	,778	12,767	12,767	,778	12,767	12,767
3	1	,788	13,385	13,385	,778	13,385	13,385
4	1	,857	14,134	14,134	,857	14,134	14,134
5	1	1,055	15,024	15,024	1,055	15,024	15,024
6	1	,759	13,240	13,240	,759	13,240	13,240

Con el objetivo de evaluar la unidimensionalidad de la prueba se procedió con el análisis factorial exploratorio, es oportuno señalar que, para este estudio, el objetivo del análisis factorial exploratorio no era realizar interpretaciones sustantivas de lo que significan los componentes, sólo se aplica el análisis factorial exploratorio en busca de evidencias de unidimensionalidad. En la Tabla 11 se muestran los porcentajes de varianza para los primeros componentes de cada cuadernillo (a detalle se pueden ver los valores de los siguientes componentes en las tablas 24, 25, 26, 27 y 28 que se adjuntan en el apéndice). De manera general para cada uno de los 6 cuadernillos el primer factor es el que explica el mayor porcentaje de la varianza, lo cual constituye un indicador de unidimensionalidad, puesto que cuanto más varianza explique el primer factor, más unidimensionalidad existe (Muñiz, 2018) no obstante este porcentaje es bajo, ninguno logra explicar más del 20% de la varianza.

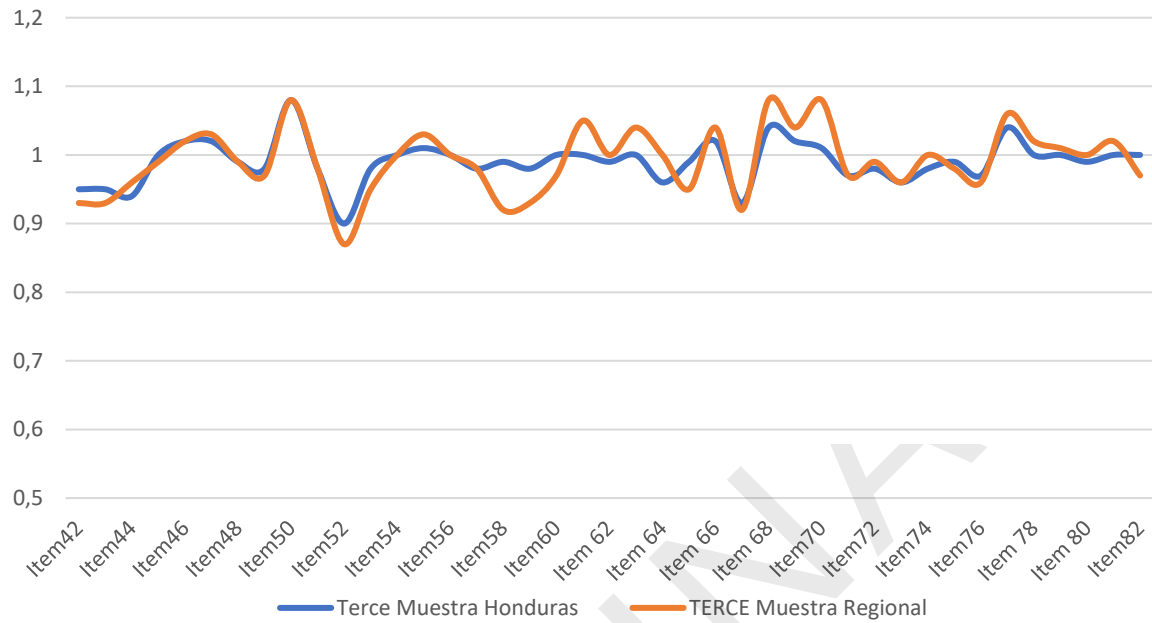
5.5 Análisis de los índices de ajuste INFIT y OUTFIT de la prueba de CCNN del TERCE.

En vista de que uno de los principales objetivos de este estudio, era obtener los valores de dificultad haciendo uso de la teoría de respuesta al ítem y específicamente haciendo uso del modelo de Rasch, resulta imprescindible que los datos empíricos se ajusten al modelo. En lo que respecta a los análisis del TERCE se ocupan aquellos basados en los residuos (las diferencias entre las respuestas observadas y esperadas): INFIT y OUTFIT

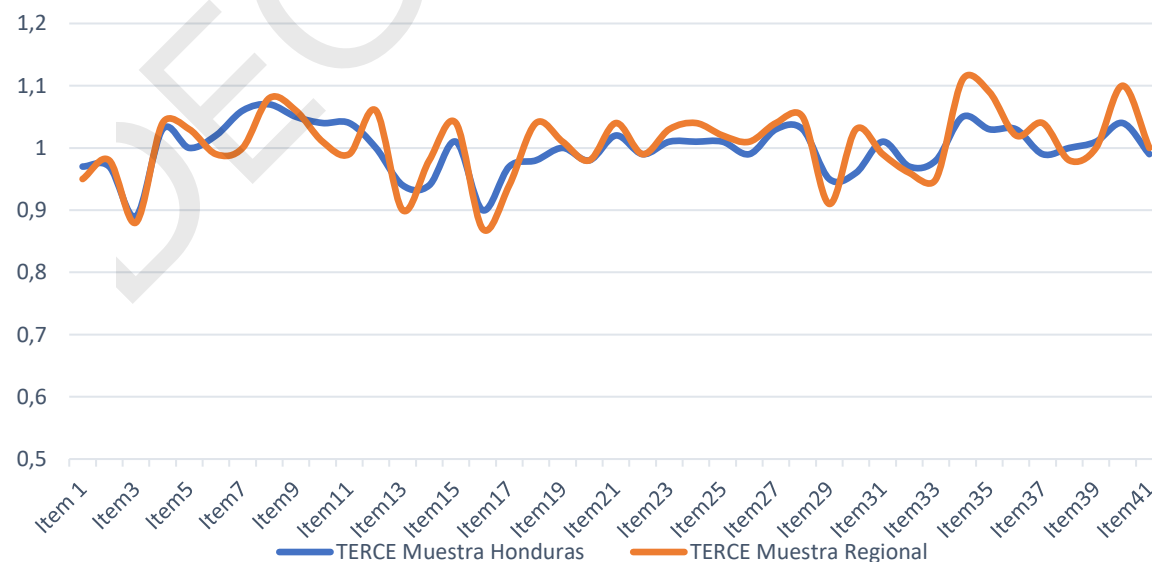
Figura 10. Estadístico INFIT para los ítems 1- 41 para las muestras nacional y regional



Para el análisis del TERCE, el intervalo de tolerancia según (UNESCO, 2016b) es de (0,5 a 1,5) para los INFIT y OUTFIT este rango es utilizado dentro de los criterios de exclusión de ítems. En las figuras 10 puede verse que para el parámetro INFIT los ítems 1- 41 se encuentran dentro del rango de tolerancia establecido y ninguno de los ítems se encuentra fuera del mismo, de igual forma, los valores obtenidos entre ambas muestras son bastante cercanos (Revisar el apéndice para visualizar más detalles) con lo cual la diferencia resulta ser mínima y los datos resultan ser apropiados para hacer uso del modelo de Rasch y obtener las ventajas que dicho modelo brinda.

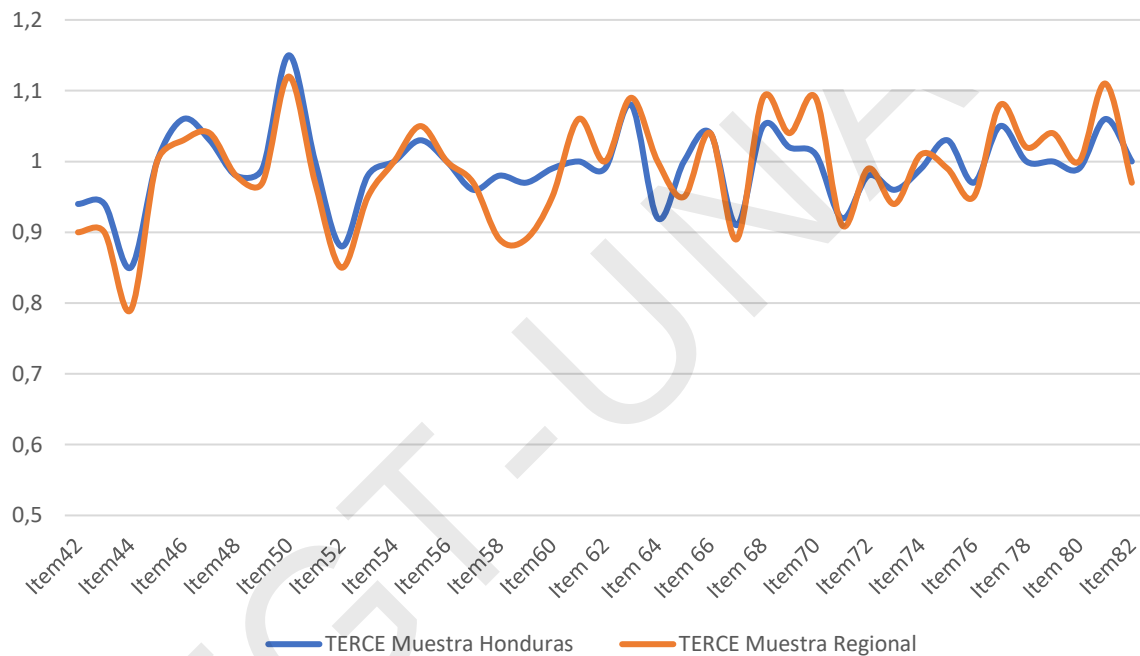
Figura 11. Estadístico INFIT para los ítems 42- 82 para las muestras nacional y regional

Para los ítems 42- 82 que se observan en la figura 11 se observa que también se encuentran dentro del rango de tolerancia establecido, de igual forma, los valores obtenidos entre ambas muestras son bastante cercanos (Revisar el apéndice para visualizar más detalles)

Figura 12. Estadístico OUTFIT para los ítems 1-41 para las muestras nacional y regional

De manera similar la figura 12, muestra valores de OUTFIT satisfactorios para el conjunto de ítems 1- 41 para ambas muestras (Honduras y Muestra Regional) con diferencias más marcadas para los ítems 34, 35 ,36, 40 y 41 entre ambas muestras. Las diferencias entre los valores obtenidos para ambos grupos son bajas, estos datos reflejan que los datos son apropiados para poder proceder con la aplicación del modelo de 1 parámetro.

Figura 13. Estadístico OUTFIT para los ítems 42-82 para las muestras nacional y regional

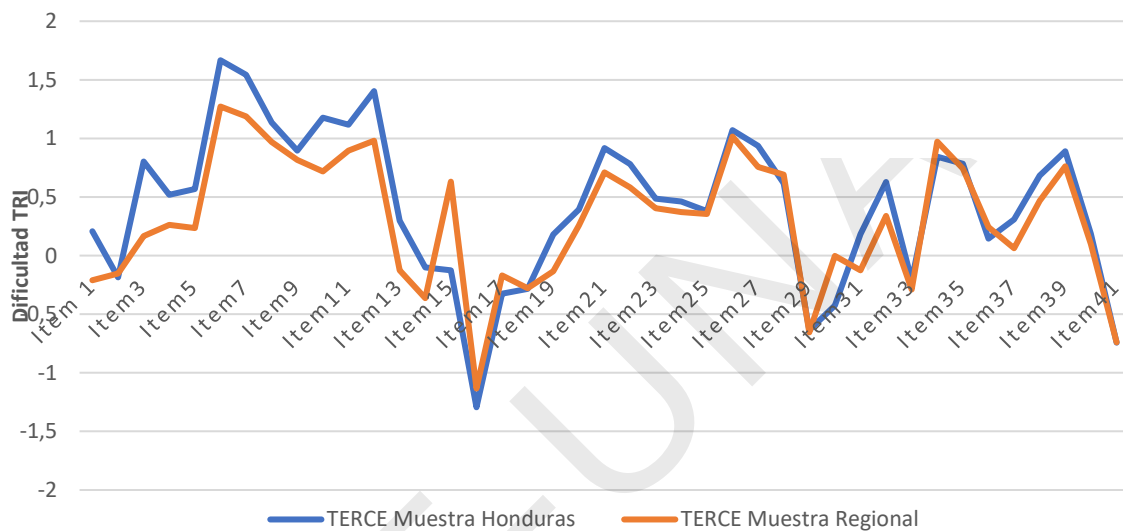


En lo que respecta a la figura 13 se ve que existen diferencias más notorias en los ítems 57, 58, y 59 entre ambas muestras. Los resultados obtenidos en este parámetro son favorables puesto que el OUTFIT es un estadístico de ajuste sensible a los casos atípicos, que permite determinar eventos poco usuales que ocurren de modo inesperado. De manera general los resultados obtenidos para los estadísticos INFIT y OUTFIT en la prueba de ciencias naturales del TERCE muestran un ajuste adecuado al modelo de Rasch.

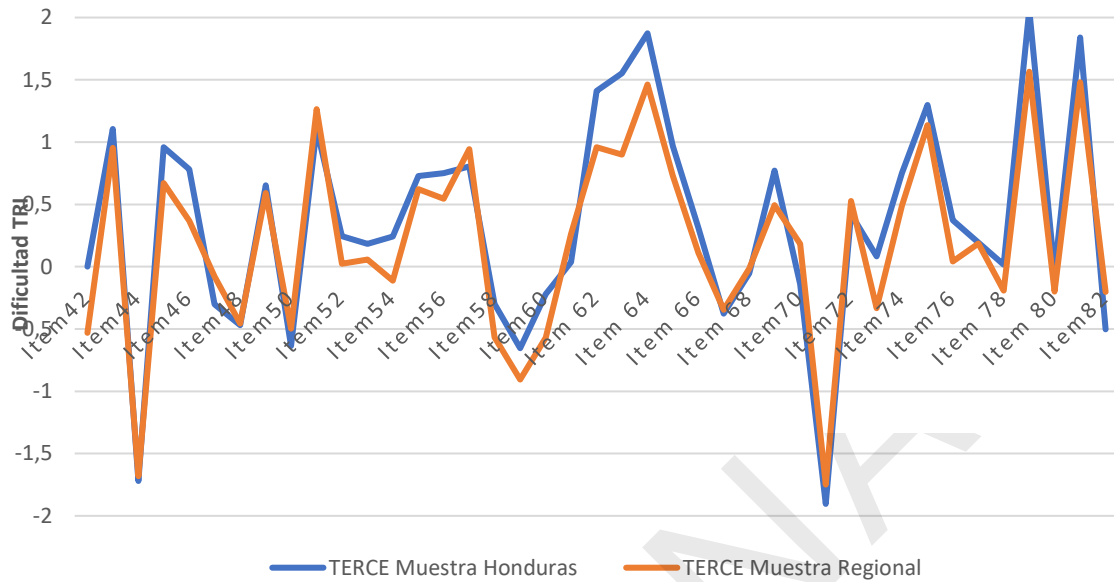
5.6 Análisis de dificultad mediante TRI para los ítems de la prueba de CCNN del TERCE.

El objetivo de este apartado se basaba en establecer la dificultad de cada uno de los ítems de la prueba por medio de la teoría de respuesta al ítem a fin de determinar la calidad métrica de los mismos, para ellos se hizo uso del modelo de 1 parámetro o modelo de Rasch.

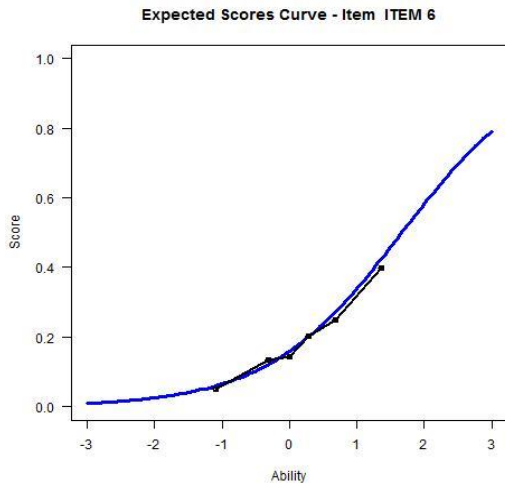
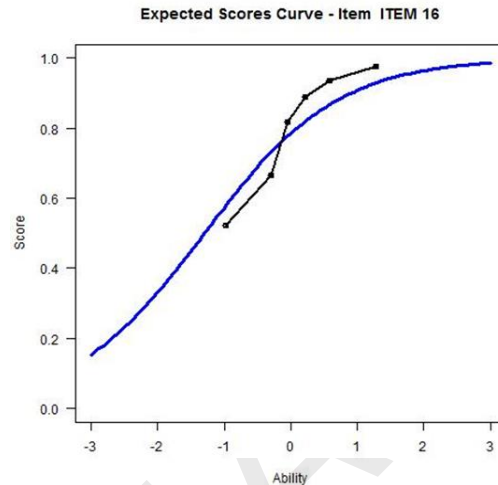
Figura 14. Dificultad de los Ítems 1- 41 según TRI para las muestras nacional y regional



En la figura 14 se presentan los valores de dificultad TRI obtenidos a partir de la aplicación del modelo de Rasch para el conjunto de ítems del 1 al 41, tomando en cuenta que para este parámetro los valores pueden fluctuar entre $+\infty$ y $-\infty$ y que en el TERCE se aceptan ítems con valores de dificultad entre -3 y 3.(UNESCO, 2016b) Pueden observarse valores de dificultad bastante favorables tomando en consideración que ítems de buena calidad no son demasiado fáciles ni demasiado difíciles, además puede percibirse una diferencia más notoria en el conjunto de ítems entre 1 y 16 para ambas muestras, el ítem 15 se comporta de manera singular al ser notoriamente más difícil para los estudiantes de la muestra regional, para los ítems desde el 17 al 41 la diferencia parece ser menor (esta puede verse a detalle en las tablas que se adjuntan en el apéndice) también puede percibirse que el ítem 16 es el que resulta ser más fácil para ambas muestras de estudiantes y por el contrario los ítems 6 y 12 los más difíciles, también para los ítems 37, 38 y 39 se nota mayor dificultad en el caso de los estudiantes hondureños.

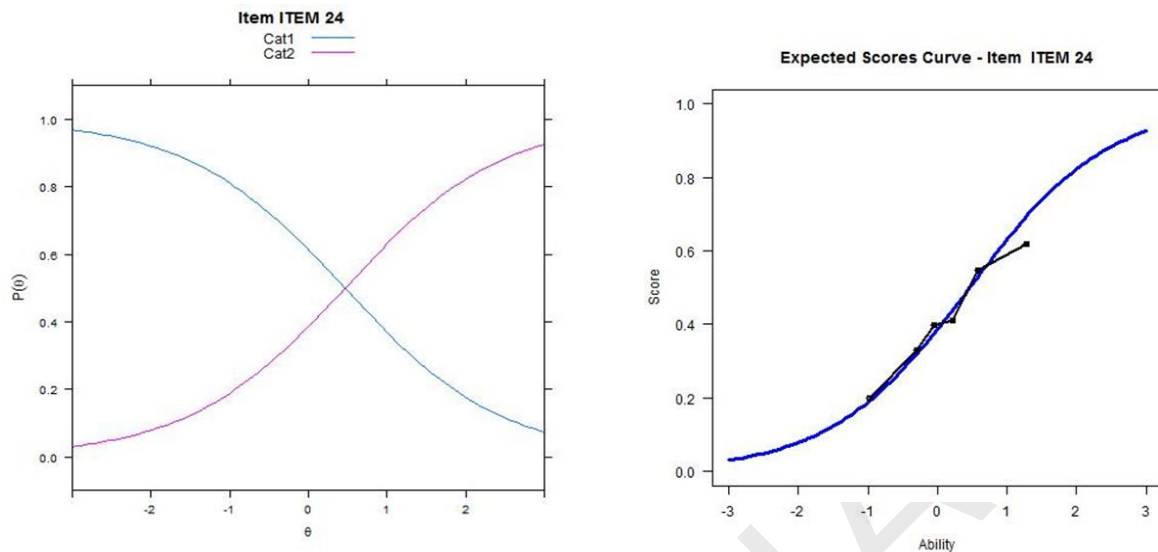
Figura 15. Dificultad de los Ítems 42- 82 según TRI para las muestras nacional y regional

Para el siguiente conjunto de ítems (42 al 82), la tendencia continúa con valores de dificultad satisfactorios con base a los criterios previamente establecidos. De manera general los ítems 44 y 71 presentan menor dificultad entre ambos grupos, y por el contrario los ítems 64, 79 y 81 se perciben como los más difíciles lo cual refleja un comportamiento similar de los ítems en ambas muestras, los ítems 51, 57, 61, 70, 72 se comportan de manera distinta al presentar mayor dificultad en la muestra regional, en comparación al resto de los ítems que resultan ser más difíciles para el grupo de estudiantes hondureños, con énfasis en los ítems 62, 63, 64 y 65 que es en donde se observa una tendencia más clara en cuanto al incremento de dificultad. Cabe añadir que del total de ítems que conforman la prueba del TERCE en ciencias naturales, son pocos los ítems con dificultades bajas y mayor la cantidad de ítems con dificultades altas.

Figura 16. Curva característica del ítem 6**Figura 17.** Curva característica del ítem 16

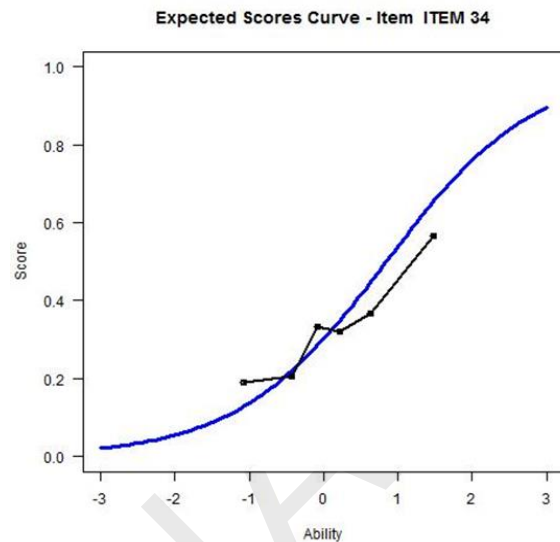
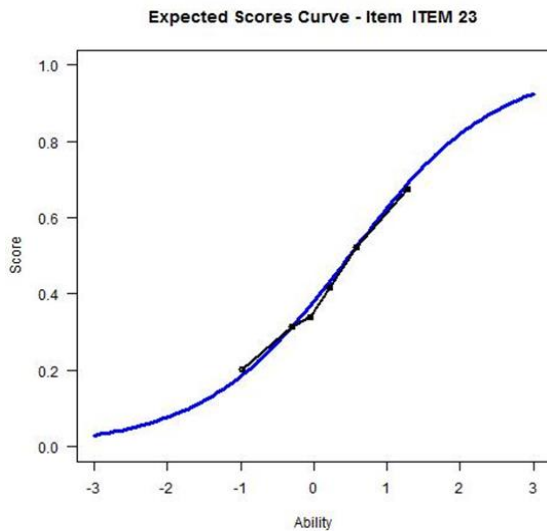
Nota: Las curvas azules representan el valor esperado y las negras representan el valor observado.

En las figuras 16 y 17 se puede observar la diferencia entre los ítems 6 y 16 respectivamente, se aprecia una comparativa entre dos ítems que se comportan de manera diferente en la muestra de estudiantes hondureños. En la figura correspondiente al ítem 6, la curva muestra como aún un estudiante con un nivel de habilidad 3 (alta) no alcanzaría una probabilidad de 1 de acertar este ítem, con lo cual el mismo denota una dificultad muy alta (1.66) en las figuras adjuntadas en el apéndice pueden observarse como muchos de los ítems de la prueba siguen la misma tendencia. Por otro lado, para el ítem 16 que resulta ser el más fácil para los estudiantes hondureños se puede apreciar en la curva como un estudiante con un nivel de habilidad incluso inferior a 3 tendría una probabilidad más elevada para responder correctamente este ítem, exhibiendo así, una dificultad inferior (-1.29).

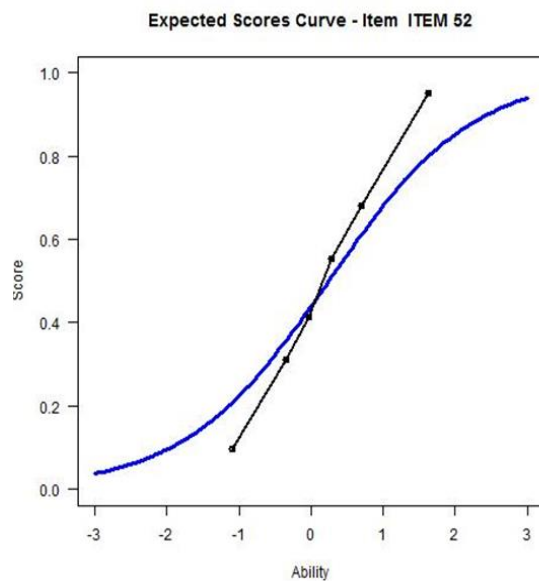
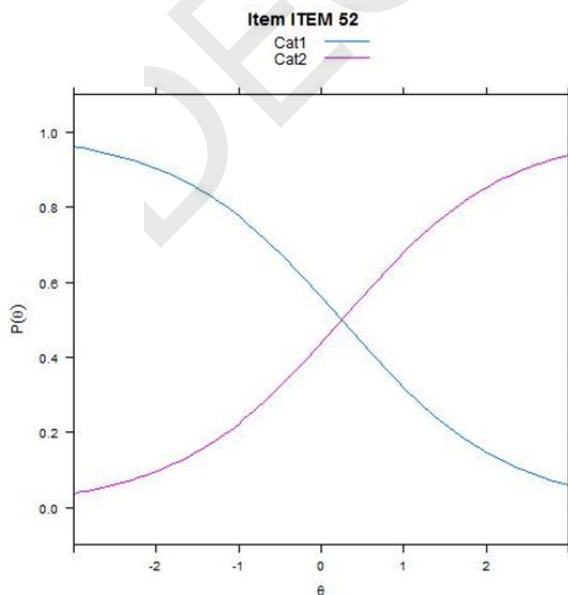
Figura 18. Curva de puntuación teórica del ítem 24 **Figura 19.** Curva característica del ítem 24

Nota: Las curvas representan la información del ítem respecto al nivel de habilidad.

En la figura 18 la curva de color rojo denota como para el ítem número 24 se cumplen los supuestos planteados desde la teoría de respuesta al ítem, en donde al incrementarse el nivel de habilidad del examinado se observa también un incremento en la probabilidad de acertar este ítem. Es decir que un estudiante con un nivel de habilidad superior a 2 alcanzaría una probabilidad muy cercana a 1 de responder correctamente a este ítem. La curva azul representa el caso contrario, un estudiante con un nivel de habilidad inferior, tendrá menos probabilidad de acertar este ítem. De igual forma en la figura 19 se observa en el gráfico de color negro (a partir de los datos obtenidos con la aplicación a la muestra de estudiantes hondureños) que estos se ajustan a la curva teórica y que la mayoría siguen la misma tendencia. No obstante, se observa que, para una parte de la muestra, específicamente aquellos estudiantes con un nivel de habilidad entre 1 y 2 la probabilidad de acertar se mantiene constante y no en aumento.

Figura 20. Curva característica del ítem 23 **Figura 21.** Curva característica del ítem 34

Para el ítem 23 puede observarse en la figura 20, como los datos a partir de la muestra se ajustan a la curva teórica, este ítem en particular muestra una dificultad aceptable (0.4 para ambas muestras, regional y Honduras) también se muestra que los estudiantes hondureños presentan un nivel de habilidad cercano a 2 este es uno de los pocos ítems que reflejan este comportamiento, pues para la mayoría de los ítems los estudiantes hondureños muestran un nivel de habilidad muy bajo. El ítem 34 presenta un índice de dificultad de 0.8 puede verse como aún un estudiante hondureño con un nivel de habilidad muy cercano a 2 tan solo tendría una probabilidad entre 0.4 y 0.5 de acertarlo, es decir muy baja.

Figura 22. Curva de puntuación del ítem 52 **Figura 23.** Curva característica del ítem 52

El ítem 52 parece comportarse de forma singular. En la figura 22 por medio de la curva de puntuación teórica, se observa el comportamiento o tendencia que debería seguir el mismo. Este ítem en particular es uno de los más fáciles dentro de la prueba con una dificultad de 0.2 tanto en la muestra regional como en la de estudiantes hondureños. La curva característica de este ítem muestra como los examinados con un nivel de habilidad muy inferior a 3 tendrían una probabilidad muy alta, casi de 1 de acertarlo. Es preciso añadir que no se sugiere incluir ítems demasiado fáciles como en este caso ni demasiado difíciles en la prueba, pero si ítems de dificultad intermedia.

DEGT-UNAH

CAPÍTULO 6.

Discusión de los Resultados

El objetivo general de este estudio consistía en analizar los parámetros psicométricos de la prueba del TERCE en el área de ciencias naturales, para ello se utilizaron los análisis psicométricos apropiados según (APA, AERA, NCME, 2018) desde las dos grandes teorías de la medición.

En cumplimiento con el objetivo relacionado con el análisis de confiabilidad de la prueba mediante el coeficiente alpha de Cronbach tan solo dos de los cuadernillos 4 y 5 (formas de la prueba) presentaron valores de 0.66 y 0.68 en el reporte técnico solamente el cuadernillo 4 presenta un alpha de 0.69 inferior pero próximo a 0.7 que es el valor mínimo aceptado por el TERCE y también como sugieren (Oviedo Celina & Campo-Arias, 2005) 0.7 es el valor mínimo aceptable de alpha de Cronbach. Es importante añadir que este estudio tiene una finalidad investigativa, sin embargo, no se puede obviar el hecho de que los resultados obtenidos en pruebas estandarizadas a gran escala tienen implicaciones importantes en la toma de decisiones de los sistemas educativos, por lo tanto, serían recomendables resultados superiores.

A pesar de que en el contexto hondureño no se documentan estudios psicométricos sobre las pruebas del Tercer Estudio Regional Comparativo y Explicativo TERCE, de las cuales únicamente se conocen los resultados, en el contexto internacional puede mencionarse: el estudio de (Atorresi et al., 2019) sobre la “Evaluación de la escritura a gran escala: análisis de la prueba de escritura del Tercer Estudio Regional Comparativo y Explicativo (TERCE)”. Este estudio presenta objetivos y metodología similares entre los cuales se pueden mencionar un análisis de validez y confiabilidad, un análisis de unidimensionalidad, al igual que un análisis sobre parámetros psicométricos como la dificultad aplicando el modelo de Rasch.

En el estudio mencionado anteriormente, en lo que respecta al análisis de confiabilidad aplicando el Alpha de Cronbach, se obtuvieron valores similares a los obtenidos en este estudio, generalmente satisfactorios; al igual que al analizar la dificultad de los ítems mediante TRI. En contraste el análisis factorial proporciona evidencias en contra del supuesto de unidimensionalidad. Como se detalló en el capítulo de resultados, el porcentaje de varianza explicada por el primer factor en cada una de las formas de la prueba

de ciencias naturales, aunque es superior al de los otros componentes, es bajo, lo cual sugiere un comportamiento similar. Es preciso añadir que el reporte técnico del TERCE (UNESCO, 2016b) no proporciona información detallada respecto al análisis de unidimensionalidad mediante análisis factorial para ninguna de las pruebas, únicamente para los factores asociados.

Con relación a los niveles de dificultad de la prueba objeto de estudio en el capítulo de resultados se concluyó que, aunque estos en su mayoría si se encuentran dentro de los criterios establecidos por las UNESCO, son más difíciles para los estudiantes hondureños, tanto al utilizar TCT como TRI, lo cual se refleja en los resultados obtenidos (inferiores al promedio). Los estudios regionales e internacionales confirman esta tendencia (PERCE, 1997; TIMMS, 2011; PISA, 2017 ; ERCE, 2019). Una alternativa interesante a esta tendencia, sería el desarrollo de una prueba desarrollada por la UNESCO para los países que han mostrado resultados inferiores en sus tres estudios, siguiendo la pauta de PISA ya que, en vista de los resultados distantes con respecto al resto de los países participantes, se sugirió el desarrollo de PISA-D, de esta manera PISA se volvió más accesible y relevante para países de bajo y mediano ingreso, como el nuestro. Además, una mejora importante en esta iniciativa son los instrumentos de recolección de datos, los cuales fueron rediseñados para capturar una gama más amplia de niveles de desempeño y contextos sociales, pero siempre dentro de las escalas utilizadas en la evaluación PISA regular (OCDE, 2018)

De los resultados obtenidos también se concluye la necesidad de revisar los ítems que no cumplieron estos criterios psicométricos en la prueba de ciencias naturales del TERCE y de ampliar la muestra de los ítems ya que muchos de ellos presentan niveles de dificultad muy elevados. Es importante añadir que esta tendencia en la dificultad de los ítems podría ser evidencia de penalización en el caso de la muestra de estudiantes hondureños, por lo que además de la revisión de los mismos, sería recomendable ampliar los niveles de desempeño e incluir expertos en el contexto educativo y de la psicometría en Honduras, a fin de establecer las escalas de evaluación de este tipo de estudios.

Respecto a la discriminación de los ítems, este estudio encontró diferencias notables entre la muestra de estudiantes hondureños y la muestra regional. En lo que respecta a Honduras, la mayoría de los ítems muestran correlaciones bajas cercanas o inferiores a 0.2 en contraste con lo que propone Escudero (2000), al mencionar que las correlaciones entre

0.00 - 0.20 son pobres por lo que se sugiere una revisión más detallada, ya que entre más alto es el índice de discriminación, el reactivo diferenciará mejor a los examinados que poseen más conocimientos de los que no lo poseen.

El análisis de unidimensionalidad es importante puesto que es necesario para poder aplicar el modelo de Rasch. (Chang et al., 2014). Estudios como el de (Jiménez Alfaro & Montero Rojas, 2013) sobre la “Aplicación del modelo de Rasch en el análisis psicométrico de una prueba de diagnóstico en matemática” sugiere que resulta muy difícil encontrar unidimensionalidad absoluta y que si el primer factor explica el mayor porcentaje de la varianza (aunque este porcentaje sea bajo) este brindaría evidencias de unidimensionalidad.

No obstante el análisis de varianza no es el único criterio suficiente a la hora de determinar la unidimensionalidad, por lo que en este estudio se tomaron en cuenta los índices de ajuste al modelo de Rasch, INFIT y OUTFIT en donde se obtuvieron valores cercanos a 1 que indican que la muestra es compatible con el modelo de Rasch como lo sugieren (Uzun & Öğretmen, 2021).

Respecto a las medidas de adecuación muestral KMO y esfericidad de Bartlett, en este estudio se obtuvieron valores meritorios según Romero & Mora, (2020) para la prueba KMO, los valores son cercanos a 0.8 lo que permite valorar el grado en que cada una de las variables es predecible a partir de las demás, cuanto mayor es el valor, más relacionadas estarán las variables entre sí, para la siguiente prueba este valor es confirmado por la significatividad asociada al test de esfericidad de Bartlett, que es 0,000, por lo que se puede rechazar la hipótesis nula de incorrelación entre variables, por tanto los datos si resultaron adecuados para realizar un análisis factorial.(Guti, 2019)

Finalmente es preciso añadir las implicaciones a nivel internacional al penalizar a los países con promedios inferiores.

6.1 Conclusiones

En conclusión, la prueba del TERCE sobre logros de aprendizaje en el área de ciencias naturales representa un logro incuestionable y un buen ejemplo para las diferentes agencias internacionales que se dedican a evaluar logros de aprendizaje. Las sugerencias que se plantean en este apartado solo tienen la finalidad de fortalecer la evaluación de logros de aprendizaje.

Un aspecto que motivó el desarrollo de este estudio fueron los resultados publicados por la UNESCO en donde Honduras se ubica en las últimas posiciones, a los que de una u otra manera se les brinda mayor relevancia, sin embargo, es importante conocer los parámetros y procedimientos utilizados por estas organizaciones a fin de obtener mayor garantía y valorar mejor la información brindada.

Se destaca, además, que la metodología utilizada en este estudio resulta importante al ejemplificar los estándares que deben cumplir las pruebas que evalúan rendimiento académico y que al cumplirlos brindan más credibilidad a la hora de presentarse los resultados y ser discutidos por los actores interesados en la sociedad.

Las bases de datos de las pruebas de la UNESCO son un excelente recurso a la hora de realizar estudios como este debido a la complejidad de sus datos, a la amplitud de la muestra y por su puesto a los criterios psicométricos utilizados.

Finalmente, las pruebas estandarizadas son de vital importancia en los contextos educativos modernos y no se puede hacer caso omiso de sus consecuencias e implicaciones. Si bien es cierto el énfasis de las mismas radica en los resultados presentados por las agencias internacionales dedicadas a este tipo de evaluación, es importante ahondar en el análisis del cumplimiento de los estándares que las mismas deben cumplir.

6.2 Limitaciones

Una de las principales limitantes de este estudio es que no se conocen los ítems incluidos en la prueba sobre logros de aprendizaje en ciencias naturales para estudiantes de sexto grado.

Si bien es cierto en el reporte técnico se detallan los procedimientos para seleccionar la muestra de estudiantes, la base de datos no cuenta con las características sociodemográficas de la misma.

No se cuenta con niveles de desempeño específicos para los índices de dificultad, tan solo los rangos de aceptación o exclusión de los mismos.

6.3 Recomendaciones

Para futuros estudios sería recomendable realizar un análisis de las puntuaciones o niveles de desempeño, debido a que el rendimiento académico depende de diversos factores, por lo cual realizar este análisis en donde se incluyan estos factores asociados en las diferentes escalas de puntuación resultaría en un buen ejercicio, pertinente y relevante.

Se recomienda detallar los procedimientos éticos utilizados a la hora de realizar la selección de la muestra de estudiantes participantes.

En lo que respecta a la base de datos, resultaría oportuno añadir la variable que identifica la forma de la prueba que contestó cada estudiante, puesto que la prueba como tal contiene 6 formas, incluir esta variable resulta necesario para poder realizar los distintos procedimientos psicométricos.

Referencias

Alas Solis, M. (2017). La calidad de la educación básica en Honduras: Modestos avances en la última década. Honduras.

APA, AERA, NCME. (2018). *Estándares para pruebas educativas y psicológicas (Spanish Edition)*. American Educational Research Association. Original work published 2014

Ato, M., López, J. J., & Benavente, A. (2013). A Classification System for Research Designs in Psychology. *Anales de Psicología / Annals of Psychology*, 29(3), 1038–1059. <https://doi.org/10.6018/analesps.29.3.178511>

Atorresi, A., Villegas, F., & Pardo, C. (2019). *Evaluación de la escritura a gran escala : Análisis de la prueba de escritura del tercer estudio regional comparativo y explicativo (TERCE)*. 55, 7–26.

Attorresi, H. F., Lozzia, G. S., Abal, F. J. P., Galibert, M. S., & Aguerri, M. E. (2009). Teoría de respuesta al ítem: conceptos básicos y aplicaciones para la medición de constructos psicológicos. *Revista Argentina de Clinica Psicologica*, 18(2), 179–188.

Birnbaum, A. (1968). *Some latent trait models and their use in inferring an examinee's ability. Statistical Theories of Mental Test Scores*, Addison-Wesley,.

Bojórquez, J., López, L., Hernández, M., & Jiménez, E. (2013). Utilización del alfa de Cronbach para validar la confiabilidad de un instrumento de medición de satisfacción del estudiante en el uso del software Minitab. *Eleventh LACCEI Latin American and Caribbean Conference for Engineering and Technology (LACCEI'2013)*, 1–9. <http://www.laccei.org/LACCEI2013-Cancun/RefereedPapers/RP065.pdf>

Bravo Arteaga, A., & Fernández Del Valle, J. (2000). La Evaluación Convencional frente a los Nuevos Modelos de Evaluación Auténtica. *Psicothema*, 12(SUPPL. 2), 95–99.

Carlson, J. E., & Davier, M. von. (2013). Contributions Series Item Response Theory. *ETS R&D Scientific and Policy Contributions Series*, RR-13-28(December), 79. <https://www.ets.org/Media/Research/pdf/RR-13-28.pdf>

Chang, K., Wang, J., Tang, H., Cheng, C., & Lin, C. (2014). *Psychometric evaluation , using Rasch analysis , of the WHOQOL-BREF in heroin-dependent people undergoing methadone maintenance treatment : Further item validation*. 1–9.

Chávez, C. & Saade, H. 2009. (n.d.). Procedimientos básicos para el análisis de reactivos. Cuaderno Técnico 8. *Centro Nacional de Evaluación Para La Educación Superior, A.C. (CENEVAL)-México*.

Cronbach, L. J., & Meehl, P. E. (1955). Construct Validity in Psychological Tests. *Psychological Bulletin*, 52(4), 281(July).
<https://doi.org/https://doi.org/10.1037/h0040957>

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334. <https://doi.org/10.1007/BF02310555>

Demarchi, G. (2020). La Evaluación Desde Las Pruebas Estandarizadas En La Educación. *En Contexto*, 8(13), 107–133.
[https://www.icfes.gov.co/documents/20143/193784/Alineacion examen Saber 11.pdf](https://www.icfes.gov.co/documents/20143/193784/Alineacion+examen+Saber+11.pdf)

Demarchi, G. D. (2020). *La evaluación desde las pruebas estandarizadas en la educación en Latinoamérica*. 8(13), 107–133.

Ellis, B. B., & Mead, A. D. (2008). Item analysis: theory and practice using classical and modern test theory. *Handbook of Research Methods in Industrial and Organizational Psychology*, 324–343. <https://doi.org/10.1002/9780470756669.ch16>

Erguven, M. (2013). Two approaches in psychometric process: classical test theory & item response theory. *Journal of Education*, 2(2), 23–30.

FEREMA. (2018). *Informe de progreso educativo Honduras*. 53(9), 51.

Frías-navarro, D., & Valencia, U. De. (2021). *Apuntes de consistencia interna de las puntuaciones de un instrumento de medida*. 1–22.

George Reyes, C. E. (2020). Pruebas estandarizadas y calidad de la educación en México, Sexenio 2012-2018. *Universidad y Sociedad*, 9(2), 313–318.

Guler, N., Uyanik, G. K., & Teker, G. T. (2014). Terms of item parameters 1. *Internatinal Association of Social Science Research*, 2013(June 2013), 1–6.

Gulliksen, H. (1950). *Theory of mental tests*. Wiley.

Guti, L. (2019). *Cómo realizar e interpretar un análisis factorial exploratorio utilizando SPSS*. 12, 1–14.

Hambleton, R. K., & Jones, R. W. (1993). Comparison of classical test theory and item tesponse theory and their applications to test development. *Instructional Topics in Educational Measurement*, 38–47.

Hidalgo Montesinos, M., & 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.

Himelfarb, I. (2019). A primer on standardized testing: history, measurement, classical test theory, item response theory, and equating. *Journal of Chiropractic Education*, 33(2), 151–163. <https://doi.org/10.7899/JCE-18-22>

Jaramillo, L. (2019). Natural Sciences as an integrating knowledge. *Sophia-Coleccion De Filosofia De La Educacion*, 1(26), 199–221. <https://www.redalyc.org/jatsRepo/4418/441857903006/441857903006.pdf>

Jiménez Alfaro, K., & Montero Rojas, E. (2013). *Aplicación del modelo de Rasch, en el análisis psicométrico de una prueba de diagnóstico en matemática*. 13(1), 1–24.

Kaukab, R., & Mehrunnisa, S. (2016). History and evolution of standarized testing– A. *International Journal of Research -Granthaalayah*, 4(5), 126–132. <https://doi.org/10.29121/granthaalayah.v4.i5.2016.2688>

Kerlinger, F.N. & Lee, H. B. (2002). *Investigación del Comportamiento*. McGraw-Hill.

Lamas, H. (2015). *School performance. Propósitos y representaciones*. 3.

Lord, F. M. (1980). Applications of item response theory to practical testing problems. In *Applications of Item Response Theory To Practical Testing Problems*. <https://doi.org/10.4324/9780203056615>

Magno, C. (2009). Demonstrating the Difference between Classical Test Theory and Item Response Theory Using Derived Test Data. *The International Journal of Educational and Psychological Assessment*, 1(1), 1–11.

Martínez Arias, M. Rosario; Hernández Lloreda, M. José; Hernández Lloreda, M. V. (2006). *Psicometría* (A. Editorial (ed.); Edición El).

Martínez, F. (2008). *La evaluación de aprendizajes en América Latina*. 1–28. <http://publicaciones.inee.edu.mx/buscadorPub/P1/C/140/P1C140.pdf>

Matas, A. (2010). *Introducción al análisis de la teoría de respuesta al Ítem*.

Miguel, J., & Meliá, J. (2017). *Editorial: Evaluación Estandarizada Standardized Assessment*. 10(1), 5–8.

Miller, M. B. (1995). Coefficient Alpha: A Basic Introduction from the Perspectives of Classical Test Theory and Structural Equation Modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 2(3), 255–273. <https://doi.org/10.1080/10705519509540013>

Montero Rojas, E. (2013). Referentes conceptuales y metodológicos sobre la noción moderna de validez de instrumentos de medición: Implicaciones para el caso de personas con necesidades educativas especiales. *Actualidades En Psicología*, 27(114), 113–128. <https://doi.org/10.15517/ap.v27i114.7900>

Mullis, I. V. S., Martin, M. O., Ruddock, G. J., O’Sullivan, C. Y., & Preuschoff, C. (2012). *TIMSS 2011. Marcos de la Evaluación* (Issue Book, Whole). <http://www.mecd.gob.es/dctm/ievaluacion/internacional/inee-timss-2011.-marcos-de-la-evaluacion.pdf?documentId=0901e72b8127e807>

Muñiz Fernández, J., & Hambleton, R. (1992). Medio Siglo de Teoría de Respuesta a los Items. *Anuario de Psicología*, 52, 41–66. <https://doi.org/10.1344/anuario.any.volum.numero>

Muñiz, J. (2018). *Introducción a la psicometría: teoría clásica y TRI. Pirámide*.

Muñiz, José. (2010). Las teoría de los tests: TCT y TRI. *Papeles Del Psicólogo*, 31(1), 57–66. <http://www.papelesdelpsicologo.es/pdf/1796.pdf>

Muñiz, José, & Fonseca-Pedrero, E. (2019). Ten steps for test development. *Psicothema*, 31(1), 7–16. <https://doi.org/10.7334/psicothema2018.291>

Navas Ara, M. -a. J. (2002). *Métodos , diseños y técnicas de investigación psicológica*.

Nunnally, J.C. & Bernstein, I. J. (1995). *Teoría psicométrica* (3ª ed). McGrawHill Latinoamericana.

OECD. (2018). PISA for development assessment and analytical framework: Reading, Mathematics and Science. In *Pro*.

Organización para la Cooperación y el Desarrollo Económicos (OCDE). (2018). *PISA para el desarrollo. Resultados en foco*. 20. https://www.oecd.org/pisa/pisa-for-development/PISA_D_Resultados_en_Foco.pdf

Oviedo Celina, H., & Campo-Arias, a. (2005). Aproximación al uso

Coeficiente Alfa de Cronbach. *Revista Colombiana de Psiquiatría*, 34(4), 572–580.
<https://doi.org/10.1590/S1135-57272002000200001>

Pere Joan Ferrando, C. A.-C. (2006). El Análisis Factorial Como Técnica De Investigación En Psicología. *Papeles Del Psicólogo*, 31(1), 18–33.
<http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Sección+Monografía+ca#2>

Pérez-Gil, J. A., Chacón Moscoso, S., & Moreno Rodríguez, R. (2000). Validez de Constructo: El uso de análisis factorial exploratorio-confirmatorio para obtener evidencias de validez. *Psicothema*, 12(SUPPL. 2), 442–446.

Phelps, R. P. (2008). The role and importance of standardized testing in the world of teaching and training. *Nonpartisan Education Review*, 4(3), 1–9.

Progar, Š., & Sočan, G. (2008). An empirical comparison of item response theory and classical test theory. *Horizons of Psychology*, 17, 5–24.

Rash, G. (1961). On General Laws and the Meaning of Measurement in Psychology. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 4: Contributions to Biology and Problems of Medicine*, v, 321–334.

Ravela, P. (2013). La evaluación de y en la educación (I). *Arista - INEE*, I.
https://www.uv.mx/personal/jomartinez/files/2011/08/La_Evaluación_Educativa.pdf

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>

Romero, K. P., & Mora, O. M. (2020). Análisis factorial exploratorio mediante el uso de las medidas de adecuación muestral kmo y esfericidad de bartlett para determinar factores principales. 5, 903–924.

Rosner, B., & Cronbach, L. J. (1960). Essentials of Psychological Testing. *The American Journal of Psychology*, 73(2), 323. <https://doi.org/10.2307/1419921>

Sánchez, M., & Sagrario, F. M. del. (2013). Las evaluaciones estandarizadas : sus efectos en tres países latinoamericanos Standardized assessments : their effects on three latin american countries. *Revista Latinoamericana de Estudios Educativos*, XLIII(1), 97–124.

Secretaría de Educación. (2005a). *Currículo Nacional Básico*.

Secretaría de Educación. (2005b). *Estándares Educativos Nacionales Español y Matemáticas*.

Secretaría de Educación. (2009). *Estándares Educativos Nacionales Ciencias Sociales y Ciencias Naturales*.

Secretaría de Educación. (2015a). *Eficacia de los aprendizajes logrados por los egresados de educación media*. 1–4.

Secretaría de Educación. (2015b). *Informe nacional de rendimiento académico 2015*.

Secretaría de Educación. (2017). *Factores asociados al rendimiento académico*. 1–20. <https://doi.org/10.35362/rie3312872>

Secretaría de Educación. (2018). *Informe Nacional de Desempeño Académico Español y Matemáticas 3ro, 6to y 9no Grado*. <http://www2.icfesinteractivo.gov.co/investigacionFormulario/docman/instituciones-educativas-y-secretarias/saber-11/informes-de-saber-11/4863-informe-nacional-de-resultados-del-examen-saber-11-de-2014-2-2016-2/file?force-download=1>

Secretaría de Educación. (2019). *Planificación estratégica institucional 2018-2022*.

Tatsuoka, M. M., Lord, F. M., Novick, M. R., & Birnbaum, A. (1971). Statistical theories of mental test scores. *Journal of the American Statistical Association*, 66(335), 651. <https://doi.org/10.2307/2283550>

Thissen, D., & Steinberg, L. (1988). Data Analysis Using Item Response Theory. *Psychological Bulletin*, 104(3), 385–395. <https://doi.org/10.1037//0033-2909.104.3.385>

Traub, R. E. (1997). Classical Test Theory in Historical Perspective. *Educational Measurement: Issues and Practice*, 16(4), 8–14. <https://doi.org/10.1111/j.1745-3992.1997.tb00603.x>

UNESCO. (2009). *SERCE: segundo estudio regional comparativo y explicativo: Los aprendizajes de los estudiantes de América Latina y el Caribe; Reporte técnico*.

UNESCO. (2016a). Informe de resultados TERCE. Logros de aprendizaje. In

Organización de las Naciones Unidas para la Educación (Vol. 3, Issue 2).

UNESCO. (2016b). *Reporte Técnico Tercer Estudio Regional Comparativo y Explicativo. TERCE*.

UNESCO. (2019). *Los Aprendizajes Fundamentales en América Latina y el Caribe* (Issue Erce).

UNESCO. (2021). *Presentación de Resultados de Logros de Aprendizaje ERCE 2019*.

UNESCO Regional Office for Education in Latin America and the Caribbean. (2000). *Primer estudio internacional comparativo sobre lenguaje, matemática y factores asociados en tercero y cuarto grado. segundo informe*. 48223.

Uzun, Z., & Öğretmen, T. (2021). *Test equating with the Rasch model to compare pre-test and post-test measurements*. 12(4), 336–347. <https://doi.org/10.21031/epod.957614>

Vaillant, D., & Rodríguez, E. (2018). *Calidad de la Educación en Iberoamérica: Discursos, Políticas y Prácticas*. Dykinson, 223. <https://www.dykinson.com/cart/download/ebooks/8765/>

William, D. (2010). *Standardized Testing and School Accountability*. *Educational Psychologist*, 45(2), 107–122. <https://doi.org/10.1080/00461521003703060>

Wolff, L. (2007). *Los Costos de la Evaluación del Aprendizaje en América Latina*. https://www.mineduacion.gov.co/cvn/1665/articles-137083_archivo_pdf.pdf%0Ahttp://www.mineduacion.gov.co/cvn/1665/articles-137083_archivo_pdf.pdf

Wood, D. A. (1960). *Test Construction: Development and Interpretation of Achievement Tests*. Columbus: OH : Charles E. Merrill Books, Inc. *Educational and Psychological Measurement*, 21(1), 237–238. <https://doi.org/10.1177/001316446102100125>

Yang, F. M., & Kao, S. T. (2014). *Item response theory for measurement validity*. *Shanghai Archives of Psychiatry*, 26(3), 171–177. <https://doi.org/10.3969/j.issn.1002-0829.2014.03>

Yusoff, M. S. B. (2019). *ABC of content validation and content validity index*

calculation. *Education in Medicine Journal*, 11(2), 49–54.
<https://doi.org/10.21315/eimj2019.11.2.6>

DEGT-UNAH

Declaración de ética de la investigación

Esta investigación ha sido realizada tomando en consideración el código de ética de investigación. No se reportan conflictos de interés.

DEGT-UNAH

Apéndice

Tabla 12. Resumen parámetro dificultad TCT para el bloque 1

Bloque	Ítem	TERCE Honduras	Muestra TERCE Regional	Muestra Diferencia
1	1	45%	51,1%	6,10%
1	2	54%	51,6%	2,40%
1	3	32%	43,6%	11,60%
1	4	38%	41,0%	3,00%
1	5	37%	44,7%	7,70%
1	6	18%	21,2%	3,20%
1	7	19%	24,0%	5,00%
1	8	26%	28,1%	2,10%
1	9	31%	32,0%	1,00%
1	10	25%	39,8%	14,80%
1	11	26%	31,3%	5,30%
1	12	22%	27,0%	5,00%
1	13	43%	54.2%	11.2 %
1	14	52%	58.7%	6.7%
1	15	Pregunta Abierta	Pregunta Abierta	Pregunta Abierta

Tabla 13. Resumen parámetro dificultad TCT para el bloque 2

Bloque	Ítem	TERCE Muestra Honduras	TERCE Muestra Regional	Diferencia
2	16	77 %	70,1%	6,9%
2	17	57%	54,2%	2,8%
2	18	57%	53,9%	3,1%
2	19	46%	49,3%	3,3%
2	20	41%	41,6%	0,6%
2	21	30%	31,1%	1,1%
2	22	33%	34,3%	1,3%
2	23	39%	41,7%	2,7%
2	24	39%	42,6%	3,6%
2	25	41%	42,0%	1%
2	26	27%	25,6%	1,4%
2	27	30%	30,6%	0,6%
2	28	36%	33,8%	2,2%
2	29	64%	69,7%	5,7%
2	30	Pregunta Abierta	Pregunta Abierta	Pregunta Abierta

Tabla 14. Resumen parámetro dificultad TCT para el bloque 3

Bloque	Ítem	TERCE Muestra Honduras	TERCE Muestra Regional	Diferencia
3	31	45.7%	44,4%	1.3%
3	32	35.7%	42,8%	7.1%
3	33	54.8%	56.5%	1,7%
3	34	31.2%	26.7%	4,5%
3	35	32.4%	31,9%	0,5%
3	36	46.5%	39,6%	6,9%
3	37	42.8%	47,5%	4,7 %
3	38	34.5%	36,1%	1,6%
3	39	30.2%	36,3%	6,1 %
3	40	45.6%	49,9%	4,3%
3	41	66.6%	68,4%	1,8 %
3	42	5%	65,0%	60%
3	43	Pregunta Abierta	Pregunta Abierta	Pregunta Abierta

Tabla 15. Resumen parámetro dificultad TCT para el bloque 4

Bloque	Ítem	TERCE Muestra Honduras	TERCE Muestra Regional	Diferencia
4	44	82.4%	74,1%	8.3%
4	45	29.8%	35,2%	5,4%
4	46	33.2%	42,3%	9,1%
4	47	56.5%	52,5%	4%
4	48	60.1%	57,5%	2,6%
4	49	35.8%	35,3%	0,5%
4	50	63.7%	52,9%	10,8%
4	51	27.1%	25,5%	1,6%
4	52	44.4%	49,8%	5,4%
4	53	45.9%	42,3%	3,6%
4	54	44.5%	50,6%	6,1%
4	55	34.3%	35,4%	1,1%
4	56	33.8%	38,7%	4,9%
4	57	Pregunta Abierta	Pregunta Abierta	Pregunta Abierta

Tabla 16. Resumen parámetro dificultad TCT para el bloque 5

Bloque	Ítem	TERCE Muestra Honduras	TERCE Muestra Regional	Diferencia
5	58	56.7%	58,4%	1,7%
5	59	64.3%	66,3%	2%
5	60	54.9%	60,6%	5,7%
5	61	49.1%	42,4%	6,7%
5	62	21.5%	30,2%	8,7%
5	63	19.4%	29,2%	9,8%
5	64	15.1%	18,7%	3,6%
5	65	29.2%	34,5%	5,3%
5	66	42.9%	46,8%	3,9%
5	67	58.4%	56,4%	2%
5	68	51.1%	47,6%	3,5%
5	69	Pregunta Abierta	Pregunta Abierta	Pregunta Abierta

Tabla 17. Resumen parámetro dificultad TCT para el bloque 6

Bloque	Ítem	TERCE Muestra Honduras	TERCE Muestra Regional	Diferencia
6	70	53.2%	45,1%	8,1%
6	71	85.5%	85,5%	0%
6	72	40.3%	38,0%	2,3%
6	73	48%	55,1%	7,1%
6	74	33.2%	39,7%	6,5%
6	75	23%	21,8%	1,2%
6	76	41.4%	48,2%	6,8%
6	77	45.5%	46,3%	0,8%
6	78	49.6%	52,8%	3,2%
6	79	12.6%	16,3%	3,7%
6	80	51%	53,73%	2,73%
6	81	15.2%	18,6%	3,4%
6	82	Pregunta Abierta	Pregunta Abierta	Pregunta Abierta

Tabla 18. Resumen parámetro discriminación TCT para el bloque 1

Bloque	Ítem	TERCE Muestra Honduras	TERCE Muestra Regional	Diferencia
1	1	0,25	0,43	0,18
1	2	0,27	0,39	0,12
1	3	0,39	0,54	0,15
1	4	0,15	0,29	0,14
1	5	0,19	0,32	0,13
1	6	0,17	0,36	0,19
1	7	0,15	0,37	0,22
1	8	0,08	0,27	0,19
1	9	0,12	0,26	0,14
1	10	0,16	0,36	0,2
1	11	0,14	0,35	0,21
1	12	0,14	0,27	0,13
1	13	0,32	0,49	0,17
1	14	0,32	0,39	0,07
1	15	0,22	0,37	0,15

Tabla 19. Resumen parámetro discriminación TCT para el bloque 2

Bloque	Ítem	TERCE Muestra Honduras	TERCE Muestra Regional	Diferencia
2	16	0,31	0,42	0,11
2	17	0,25	0,43	0,18
2	18	0,22	0,28	0,06
2	19	0,17	0,34	0,17
2	20	0,21	0,38	0,17
2	21	0,15	0,28	0,13
2	22	0,22	0,37	0,15
2	23	0,17	0,32	0,15
2	24	0,15	0,29	0,14
2	25	0,19	0,32	0,13
2	26	0,19	0,27	0,08
2	27	0,12	0,26	0,14
2	28	0,13	0,25	0,12
2	29	0,29	0,42	0,13
2	30	0,26	0,47	0,21

Tabla 20. Resumen parámetro discriminación TCT para el bloque 3

Bloque	Ítem	TERCE Muestra Honduras	TERCE Muestra Regional	Diferencia
3	31	0,123	0,137	0,014
3	32	0,229	0,43	0,201
3	33	0,216	0,42	0,204
3	34	0,053	0,24	0,187
3	35	0,099	0,26	0,161
3	36	0,092	0,34	0,248
3	37	0,180	0,31	0,13
3	38	0,161	0,42	0,259
3	39	0,186	0,38	0,194
3	40	0,075	0,22	0,145
3	41	0,191	0,35	0,159
3	42	0,304	0,46	0,156
3	43	0,285	0,51	0,225

Tabla 21. Resumen parámetro discriminación para el bloque 4

Bloque	Ítem	TERCE Muestra Honduras	TERCE Muestra Regional	Diferencia
4	44	0,306	0,41	0,104
4	45	0,237	0,40	0,163
4	46	0,204	0,34	0,136
4	47	0,211	0,31	0,099
4	48	0,269	0,39	0,121
4	49	0,277	0,43	0,153
4	50	0,067	0,24	0,173
4	51	0,281	0,41	0,129
4	52	0,442	0,58	0,138
4	53	0,280	0,45	0,17
4	54	0,240	0,38	0,14
4	55	0,270	0,33	0,06
4	56	0,256	0,38	0,124
4	57	0,281	0,39	0,109

Tabla 22. Resumen parámetro discriminación para el bloque 5

Bloque	Ítem	TERCE Muestra Honduras	TERCE Muestra Regional	Diferencia
5	58	0,219	0,47	0,251
5	59	0,256	0,43	0,174
5	60	0,216	0,40	0,184
5	61	0,196	0,27	0,074
5	62	0,205	0,34	0,135
5	63	0,118	0,22	0,102
5	64	0,231	0,39	0,159
5	65	0,226	0,42	0,194
5	66	0,149	0,31	0,161
5	67	0,353	0,48	0,127
5	68	0,133	0,24	0,107
5	69	0,147	0,29	0,143

Tabla 23. Resumen parámetro dificultad para bloque 6

Bloque	Ítem	TERCE Muestra Honduras	Muestra Regional	Diferencia
6	70	0,160	0,22	0,06
6	71	0,196	0,32	0,124
6	72	0,207	0,38	0,173
6	73	0,285	0,42	0,135
6	74	0,206	0,35	0,144
6	75	0,178	0,39	0,212
6	76	0,237	0,43	0,193
6	77	0,080	0,25	0,17
6	78	0,173	0,32	0,147
6	79	0,135	0,28	0,145
6	80	0,197	0,35	0,153
6	81	0,100	0,26	0,16
6	82	0,271	0,23	0,041

Tabla 24. Varianza total explicada cuadernillo 1

Componente	Autovalores Iniciales			Sumas de Cargas al Cuadrado de la Extracción		
	Total	% de Varianza	% Acumulado	Total	% de Varianza	% Acumulado
1	,949	13,281	13,281	,949	13,281	13,281
2	,457	6,401	19,682	,457	6,401	19,682
3	,385	5,383	25,065	,385	5,383	25,065
4	,295	4,135	29,199	,295	4,135	29,199
5	,280	3,919	33,118	,280	3,919	33,118
6	,268	3,751	36,869	,268	3,751	36,869
7	,257	3,603	40,473	,257	3,603	40,473
8	,254	3,550	44,023	,254	3,550	44,023
9	,248	3,468	47,490	,248	3,468	47,490
10	,244	3,419	50,909	,244	3,419	50,909

Tabla 25. Varianza total explicada cuadernillo 2

Componente	Autovalores Iniciales			Sumas de Cargas al Cuadrado de la Extracción		
	Total	% de Varianza	% Acumulado	Total	% de Varianza	% Acumulado
1	,778	12,767	12,767	,778	12,767	12,767
2	,334	5,479	18,246	,334	5,479	18,246
3	,304	4,992	23,238	,304	4,992	23,238
4	,291	4,773	28,011	,291	4,773	28,011
5	,274	4,493	32,504	,274	4,493	32,504
6	,262	4,297	36,801	,262	4,297	36,801
7	,257	4,220	41,021	,257	4,220	41,021
8	,251	4,124	45,145	,251	4,124	45,145
9	,243	3,977	49,122	,243	3,977	49,122
10	,230	3,770	52,892	,230	3,770	52,892

Tabla 26. Varianza total explicada cuadernillo 3

Componente	Autovalores Iniciales			Sumas de Cargas al Cuadrado de la Extracción		
	Total	% de	%	Total	% de	%
		Varianza	Acumulado		Varianza	Acumulado
1	,788	13,385	13,385	,788	13,385	13,385
2	,341	5,793	19,178	,341	5,793	19,178
3	,300	5,088	24,266	,300	5,088	24,266
4	,281	4,772	29,038	,281	4,772	29,038
5	,260	4,410	33,448	,260	4,410	33,448
6	,252	4,279	37,727	,252	4,279	37,727
7	,240	4,077	41,803	,240	4,077	41,803
8	,231	3,920	45,723	,231	3,920	45,723
9	,222	3,768	49,492	,222	3,768	49,492

Tabla 27. Varianza total explicada cuadernillo 4

Componente	Autovalores Iniciales			Sumas de Cargas al Cuadrado de la Extracción		
	Total	% de Varianza	% Acumulado	Total	% de Varianza	% Acumulado
1	,857	14,134	14,134	,857	14,134	14,134
2	,423	6,987	21,121	,423	6,987	21,121
3	,288	4,755	25,876	,288	4,755	25,876
4	,280	4,625	30,501	,280	4,625	30,501
5	,268	4,423	34,924	,268	4,423	34,924
6	,266	4,394	39,319	,266	4,394	39,319
7	,259	4,277	43,595	,259	4,277	43,595
8	,257	4,243	47,838	,257	4,243	47,838
9	,247	4,070	51,908	,247	4,070	51,908
10	,243	4,009	55,917	,243	4,009	55,917

Tabla 28. Varianza total explicada cuadernillo 5

Autovalores Iniciales				Suma de Cargas al Cuadrado		
De Extracción						
Componente	Total	% de	%	Total	% de	%
		Varianza	Acumulado		Varianza	Acumulado
1	1,055	15,024	15,024	1,055	15,024	15,024
2	,481	6,843	21,867	,481	6,843	21,867
3	,442	6,287	28,155	,442	6,287	28,155
4	,321	4,563	32,718	,321	4,563	32,718
5	,293	4,169	36,887	,293	4,169	36,887
6	,280	3,989	40,876	,280	3,989	40,876
7	,270	3,838	44,714	,270	3,838	44,714
8	,261	3,711	48,425	,261	3,711	48,425

Tabla 29. Varianza total explicada cuadernillo 6

Autovalores Iniciales				Suma de Cargas al Cuadrado		
De Extracción						
Componente	Total	% de	%	Total	% de	%
		Varianza	Acumulado		Varianza	Acumulado
1	,759	13,240	13,240	,759	13,240	13,240
2	,317	5,537	18,777	,317	5,537	18,777
3	,288	5,031	23,807	,288	5,031	23,807
4	,281	4,907	28,715	,281	4,907	28,715
5	,268	4,685	33,399	,268	4,685	33,399
6	,253	4,413	37,812	,253	4,413	37,812
7	,251	4,372	42,183	,251	4,372	42,183
8	,243	4,236	46,419	,243	4,236	46,419
9	,235	4,093	50,513	,235	4,093	50,513
10	,231	4,034	54,546	,231	4,034	54,546

Tabla 30. Resumen estadístico INFIT para el bloque 1

Bloque	Ítem	TERCE Muestra Honduras	TERCE Muestra Regional	Diferencia
1	1	0,97	0,96	0,01
1	2	0,97	0,99	0,02
1	3	0,92	0,89	0,03
1	4	1,02	1,04	0,02
1	5	1,00	1,02	0,02
1	6	1,00	0,98	0,02
1	7	1,01	0,98	0,03
1	8	1,04	1,05	0,01
1	9	1,02	1,04	0,02
1	10	1,0	1	0
1	11	1,0	0,99	0,01
1	12	1,02	1,03	0,01
1	13	0,94	0,92	0,02
1	14	0,95	0,98	0,03
1	15	0,99	1,04	0,05

Tabla 31. Resumen estadístico INFIT para el bloque 2

Bloque	Ítem	TERCE Muestra Honduras	Muestra Regional	Diferencia
2	16	0,95	0,93	0,02
2	17	0,97	0,95	0,02
2	18	0,98	1,03	0,05
2	19	1,00	1,01	0,01
2	20	0,99	0,98	0,01
2	21	1,01	1,03	0,02
2	22	0,99	0,98	0,01
2	23	1,01	1,03	0,02
2	24	1,01	1,03	0,02
2	25	1,01	1,02	0,01
2	26	0,99	1	0,01
2	27	1,02	1,03	0,01
2	28	1,03	1,04	0,02
2	29	0,97	0,94	0,03
2	30	0,98	1,03	0,04

Tabla 32. Resumen estadístico INFIT para el bloque 3

Bloque	Ítem	TERCE Muestra Honduras	TERCE Muestra Regional	Diferencia
3	31	1,01	0,99	0,02
3	32	0,97	0,96	0,02
3	33	0,99	0,97	0,02
3	34	1,03	1,05	0,02
3	35	1,02	1,05	0,03
3	36	1,03	1,01	0,02
3	37	0,99	1,03	0,03
3	38	1,00	0,98	0,02
3	39	0,99	0,99	0,00
3	40	1,03	1,08	0,05
3	41	0,99	1	0,01
3	42	0,95	0,93	0,02
3	43	0,95	0,93	0,03

Tabla 33. Resumen estadístico INFIT para el bloque 4

Bloque	Ítem	TERCE Muestra Honduras	TERCE Muestra Regional	Diferencia
4	44	0,94	0,96	0,02
4	45	1,00	0,99	0,01
4	46	1,02	1,02	0,00
4	47	1,02	1,03	0,01
4	48	0,99	0,99	0,00
4	49	0,98	0,97	0,01
4	50	1,08	1,08	0,00
4	51	0,98	0,98	0,01
4	52	0,90	0,87	0,04
4	53	0,98	0,95	0,04
4	54	1,00	1	0.00
4	55	1,01	1,03	0,02
4	56	1,00	1	0,00
4	57	0,98	0,98	0,00

Tabla 34. Resumen estadístico INFIT para el bloque 5

Bloque	Ítem	TERCE Muestra Honduras	TERCE Muestra Regional	Diferencia
5	58	0,99	0,92	0,07
5	59	0,98	0,93	0,05
5	60	1,00	0,97	0,03
5	61	1,00	1,05	0,005
5	62	0,99	1	0,01
5	63	1.0	1,04	0,04
5	64	0,96	1	0,04
5	65	0,99	0,95	0,04
5	66	1,02	1,04	0,02
5	67	0,93	0,92	0,01
5	68	1,04	1,08	0,04
5	69	1,02	1,04	0,02

Tabla 35. Resumen estadístico INFIT para el bloque 6

Bloque	Ítem	TERCE Muestra Honduras	TERCE Muestra Regional	Diferencia
6	70	1,01	1,08	0,07
6	71	0,97	0,97	0,00
6	72	0,98	0,99	0,01
6	73	0,96	0,96	0,00
6	74	0,98	1	0,02
6	75	0,99	0,98	0,01
6	76	0,97	0,96	0,01
6	77	1,04	1,06	0,02
6	78	1,00	1,02	0,02
6	79	1,00	1,01	0,01
6	80	0,99	1	0,01
6	81	1,00	1,02	0,02
6	82	1,00	0,97	0,03

Tabla 36. Resumen estadístico OUTIFIT el para bloque 1

Bloque	Ítem	TERCE Muestra Honduras	TERCE Muestra Regional	Diferencia
1	1	0,97	0,95	0,02
1	2	0,97	0,98	0,01
1	3	0,89	0,88	0,01
1	4	1,03	1,04	0,01
1	5	1,0	1,03	0,03
1	6	1,02	0,99	0,03
1	7	1,06	1	0,06
1	8	1,07	1,08	0,01
1	9	1,05	1,06	0,01
1	10	1,04	1,01	0,03
1	11	1,04	0,99	0,05
1	12	1,0	1,06	0,06
1	13	0,94	0,9	0,04
1	14	0,94	0,98	0,04
1	15	1,01	1,04	0,03

Tabla 37. Resumen estadístico OUTIFIT para el bloque 2

Bloque	Ítem	TERCE Muestra Honduras	TERCE Muestra Regional	Diferencia
2	16	0,90	0,87	0,03
2	17	0,97	0,94	0,03
2	18	0,98	1,04	0,06
2	19	1,00	1,01	0,01
2	20	0,98	0,98	0,00
2	21	1,02	1,04	0,02
2	22	0,99	0,99	0,00
2	23	1,01	1,03	0,02
2	24	1,01	1,04	0,03
2	25	1,01	1,02	0,01
2	26	0,99	1,01	0,02
2	27	1,03	1,04	0,01
2	28	1,03	1,05	0,02
2	29	0,95	0,91	0,04
2	30	0,96	1,03	0,07

Tabla 38. Resumen estadístico OUTIFIT para el bloque 3

Bloque	Ítem	TERCE Muestra Honduras	TERCE Muestra Regional	Diferencia
3	31	1,01	0,99	0,02
3	32	0,97	0,96	0,01
3	33	0,98	0,95	0,03
3	34	1,05	1,11	0,06
3	35	1,03	1,09	0,06
3	36	1,03	1,02	0,01
3	37	0,99	1,04	0,05
3	38	1,00	0,98	0,02
3	39	1,01	1	0,01
3	40	1,04	1.1	0,06
3	41	0,99	1	0,01
3	42	0,94	0,9	0,04
3	43	0,94	0,9	0,04

Tabla 39. Resumen estadístico OUTIFIT para el bloque 4

Bloque	Ítem	TERCE Muestra Honduras	TERCE Muestra Regional	Diferencia
4	44	0,85	0,79	0,06
4	45	1,00	1	0,00
4	46	1,06	1,03	0,03
4	47	1,03	1,04	0,01
4	48	0,98	0,98	0,00
4	49	0,99	0,97	0,02
4	50	1,15	1,12	0,03
4	51	1,00	0,97	0,03
4	52	0,88	0,85	0,03
4	53	0,98	0,95	0,03
4	54	1,00	1	0,00
4	55	1,03	1,05	0,02
4	56	1,00	1	0,00
4	57	0,96	0,97	0,01

Tabla 40. Resumen estadístico OUTIFIT para el bloque 5

Bloque	Ítem	TERCE Muestra Honduras	TERCE Muestra Regional	Diferencia
5	58	0,98	0,89	0,09
5	59	0,97	0,89	0,08
5	60	0,99	0,95	0,04
5	61	1,00	1,06	0,06
5	62	0,99	1	0,01
5	63	1,08	1,09	0,01
5	64	0,92	1	0,08
5	65	1,00	0,95	0,05
5	66	1,04	1,04	0,00
5	67	0,91	0,89	0,02
5	68	1,05	1,09	0,04
5	69	1,02	1,04	0,02

Tabla 41. Resumen estadístico OUTIFIT para el bloque 6

Bloque	Ítem	TERCE Muestra Honduras	TERCE Muestra Regional	Diferencia
6	70	1,01	1,09	0,08
6	71	0,92	0,91	0,01
6	72	0,98	0,99	0,01
6	73	0,96	0,94	0,02
6	74	0,99	1,01	0,02
6	75	1,03	0,99	0,04
6	76	0,97	0,95	0,02
6	77	1,05	1,08	0,03
6	78	1,00	1,02	0,02
6	79	1,00	1,04	0,04
6	80	0,99	1	0,01
6	81	1,06	1,11	0,05
6	82	1,00	0,97	0,03

Tabla 42 Resumen dificultad TRI para el bloque 1

Bloque	Ítem	TERCE Muestra Honduras	TERCE Muestra Regional	Diferencia
1	1	0,209	-0,211	0,002
1	2	-0,186	-0,154	0,032
1	3	0,803	0,168	0,635
1	4	0,519	0,262	0,257
1	5	0,568	0,236	0,332
1	6	1,667	1,272	0,395
1	7	1,543	1,189	0,354
1	8	1,135	0,97	0,165
1	9	0,895	0,817	0,078
1	10	1,178	0,717	0,461
1	11	1,118	0,896	0,222
1	12	1,403	0,98	0,423
1	13	0,297	-0,125	0,172
1	14	-0,102	-0,364	0,262
1	15	-0,125	0,63	0,505

Tabla 43. Resumen dificultad TRI para el bloque 2

Bloque	Ítem	TERCE Muestra Honduras	TERCE Muestra Regional	Diferencia
2	16	-1,294	-1,134	0,16
2	17	-0,325	-0,17	0,155
2	18	-0,284	-0,277	0,007
2	19	0,180	-0,134	0,046
2	20	0,393	0,258	0,135
2	21	0,918	0,71	0,208
2	22	0,780	0,582	0,198
2	23	0,487	0,403	0,084
2	24	0,463	0,372	0,091
2	25	0,380	0,355	0,025
2	26	1,071	1,016	0,055
2	27	0,937	0,756	0,181
2	28	0,618	0,692	0,074
2	29	-0,636	-0,655	0,019
2	30	-0,426	-0,003	0,423

Tabla 44. Resumen dificultad TRI para el bloque 3

Bloque	Ítem	TERCE Muestra Honduras	TERCE Muestra Regional	Diferencia
3	31	0,182	-0,126	0,056
3	32	0,629	0,339	0,29
3	33	-0,210	-0,289	0,079
3	34	0,843	0,971	0,128
3	35	0,783	0,744	0,039
3	36	0,145	0,243	0,098
3	37	0,309	0,062	0,247
3	38	0,683	0,468	0,215
3	39	0,890	0,762	0,128
3	40	0,185	0,099	0,086
3	41	-0,743	-0,74	0,003
3	42	-0,000	-0,531	0,531
3	43	1,104	0,954	0,15

Tabla 45. Resumen dificultad TRI para el bloque 4

Bloque	Ítem	TERCE Muestra Honduras	TERCE Muestra Regional	Diferencia
4	44	-1,720	-1,682	0,038
4	45	0,960	0,67	0,29
4	46	0,782	0,368	0,414
4	47	-0,302	-0,08	0,222
4	48	-0,469	-0,456	0,013
4	49	0,654	0,592	0,062
4	50	-0,636	-0,498	0,138
4	51	1,105	1,264	0,159
4	52	0,246	0,024	0,222
4	53	0,182	0,057	0,125
4	54	0,242	-0,112	0,13
4	55	0,727	0,623	0,104
4	56	0,750	0,546	0,204
4	57	0,805	0,943	0,138

Tabla 46. Resumen dificultad TRI para el bloque 5

Bloque	Ítem	TERCE Muestra Honduras	TERCE Muestra Regional	Diferencia
5	58	-0,300	-0,566	0,536
5	59	-0,653	-0,906	0,253
5	60	-0,220	-0,569	0,349
5	61	0,036	0,264	0,228
5	62	1,410	0,958	0,452
5	63	1,553	0,899	0,654
5	64	1,872	1,463	0,409
5	65	0,969	0,729	0,24
5	66	0,311	0,11	0,201
5	67	-0,377	-0,347	0,030
5	68	-0,052	-0,014	0,038
5	69	0,772	0,495	0,277

Tabla 47 Resumen dificultad TRI para el bloque 6

Bloque	Ítem	TERCE	TERCE	Diferencia
		Muestra	Muestra Regional	
		Honduras		
6	70	-0,141	0,182	0,041
6	71	-1,904	-1,749	0,155
6	72	0,422	0,528	0,106
6	73	0,082	-0,334	0,252
6	74	0,752	0,488	0,264
6	75	1,297	1,136	0,161
6	76	0,373	0,043	0,33
6	77	0,193	0,185	0,008
6	78	0,015	-0,193	0,178
6	79	2,064	1,565	0,469
6	80	-0,044	-0,2	0,156
6	81	1,841	1,48	0,361
6	82	-0,502	-0,202	0,3