

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



**Prueba de Selección Universitaria para
aspirantes a la Licenciatura en Antropología,
Ciudad Universitaria, UNAH, 2018**

Presentada por
Keyla Yanelly Rivera Reyes

TESIS
Para obtener el grado de
Máster en Psicometría y Evaluación Educativa

Asesor temático
Mario Alas Solís, Ph.D.

Asesor metodológico
German Eduardo Moncada Godoy, Ph.D.

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

Agosto, 2019

Cesión de Derechos

Tegucigalpa, M.D.C., 5 de agosto de 2019

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: “Prueba de Selección Universitaria para aspirantes a la Licenciatura en Antropología Ciudad Universitaria, UNAH, 2018“, el cual he presentado como parte de los requisitos para optar por la Maestría en Psicometría y Evaluación Educativa.

Atentamente,

Keyla Yanelly Rivera Reyes
Cuenta No. PSE100123

AUTORIDADES UNIVERSITARIAS

Doctor Francisco José Herrera Alvarado

Rector

Máster Belinda Flores de Mendoza

Vice-Rectora Académica

Abogada Emma Virginia Rivera Mejía

Secretaria General

Máster Magda Elsy Hernández

Directora de Docencia

Doctor Armando Euceda

Director Sistema de Estudios de Postgrado

Doctora Martha Lorena Suazo Matute

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

Tegucigalpa, M.D.C., Honduras

Dedicatoria

A Dios, quien en su infinita misericordia me ha permitido culminar esta faceta de mi vida académica, siendo ese bálsamo de fe y perseverancia en los momentos que creía desistir. Gracias Padre Celestial, por estar presente en mi vida con sus infinitas bendiciones.

A mi abuela Libia Estela Aguilar, quien me enseñó mis primeras letras, quién siempre me ha impulsado y apoyado en mis sueños. Ha sido mi ejemplo a seguir, es la persona que más admiro.

A mis padres, Livia Estela Reyes Aguilar y Reynaldo Rivera Fúnez, por su esfuerzo, amor y dedicación. Gracias a su orientación y que me inculcaron a que en el camino de la formación era la mejor opción. Gracias a su sacrificio y su apoyo incondicional he llegado hasta acá.

A mi hermana, Anayenci Rivera Reyes, quien es mi inspiración para ser mejor persona cada día y a quien trato de darle la mejor versión de mí.

A mi hermana y mejor amiga, Daniela Alejandra Álvarez Rivera, con quien he compartido muchos momentos de la vida y siempre me ha animado para nunca desistir.

Agradecimiento

A la Coordinadora académica de la Maestría, MSc. Lina María Mendoza Recarte, quien gestionó el programa de manera eficiente y satisfactoria, proporcionándonos docentes con reconocimiento nacional e internacional y por la disposición de ayuda brindada durante el desarrollo de nuestra maestría.

Al Coordinador de Investigación-vinculación de la Maestría, German Eduardo Moncada Godoy, Ph.D., por compartir con nosotros su expertis, brindándonos las orientaciones y motivación, para terminar la tesis.

A mi asesor temático, Mario Alas Solís, Ph.D., por sus consejos, comprensión y valioso tiempo otorgado en el asesoramiento de este trabajo.

A mis maestros, por ayudarme en mi crecimiento profesional. En especial al Dr. Dennis Fernando Cáceres Valladares por su apoyo académico y por motivarme a terminar mi trabajo de investigación.

A mis compañeras de maestría, Nusly Moreno y Tessy Solórzano, con quien conformamos un equipo interdisciplinario y de ayuda mutua, para lograr culminar esta etapa de nuestra formación académica.

A la Jefa y Coordinadora del Departamento de Antropología MSc. Carmen Fajardo, por su apoyo proporcionado para el desarrollo esta prueba, tanto con la disponibilidad de los docentes para la validez de contenido como con los estudiantes de primer ingreso para llevar a cabo el grupo focal.

A los directores, docentes y estudiantes de los centros educativos participantes, que sin ellos, la culminación de esta prueba no hubiese sido posible.

Resumen

El objetivo general del presente estudio fue el diseño de una prueba de selección universitaria basada en las expectativas de logro y habilidades cognitivas establecidos en los espacios curriculares de Antropología y Fundamentos de Investigación Social del Plan de Estudios y Programa Curricular del Bachillerato de Ciencias y Humanidades, para ser aplicada con aspirantes a ingresar a la Licenciatura en Antropología de la Facultad de Ciencias Sociales de la Universidad Nacional Autónoma de Honduras.

En correspondencia con lo señalado por la literatura especializada en evaluación educativa, utilizando la versión actualizada que Anderson y Krathwohl desarrollaron de la taxonomía de objetivos educativos de B. Bloom, se elaboró una Tabla de Especificaciones conteniendo las expectativas de logro y los contenidos en relación con las habilidades cognitivas que más aportan al perfil de salida de los espacios curriculares de Antropología y Fundamentos de Investigación Social de nivel medio. A partir de dicha Tabla, se desarrolló el proceso de construcción y de ensamblaje para que la Prueba Oficial de Selección Universitaria para aspirantes a ingresar a la Licenciatura de Antropología, de manera que cumpla con los indicadores psicométricos de calidad bajo el análisis de la Teoría Clásica de los Test (TCT).

De acuerdo con lo señalado por la bibliografía especializada, los índices más usados al hacer el análisis psicométrico de una prueba de selección universitaria de acuerdo a la Teoría Clásica de los Test, son el índice de dificultad, el índice de discriminación, análisis de distractores, validez, confiabilidad y el índice de funcionamiento diferencial del ítem. Estos índices se convierten en indicadores de la calidad de una prueba en la medida que se encuentren dentro de los valores aceptables.

Al analizar los resultados de la prueba operativa se encontró que el rendimiento de los estudiantes en referencia a las habilidades cognitivas asociadas a conocimientos específicos de Antropología y Fundamentos de Investigación Social se encontró que el porcentaje promedio de estudiantes que contestaron correctamente ambas sub pruebas, es claramente decreciente en la medida que se consideran niveles cognitivos más complejos en la Taxonomía de Bloom. Los estudiantes aportaron valiosa información cualitativa que ayuda a explicar estos resultados. El Análisis de los criterios vertidos por los estudiantes de primer ingreso de la Licenciatura en Antropología, con respecto al desarrollo de los contenidos de Antropología y Fundamentos de Investigación Social en el Bachillerato, identifica que recibieron mejor formación en la segunda, mientras que se quedaron con muchos vacíos respecto a los contenidos propios de Antropología.

Palabras clave: Habilidades cognitivas, Expectativas de logro, Teoría Clásica de los Test (TCT) y La Clasificación Internacional Normalizada de la Educación (CINE).

Índice de contenido

Índice de tablas.....	vii
Índice de figuras.....	vii
Capítulo 1. Planteamiento del Problema	1
1.1 Situación Problemática: Prueba de Selección Universitaria para aspirantes a la Licenciatura en Antropología.....	1
1.2 Objetivos de la Investigación	7
1.2.1 Objetivo General	7
1.2.2 Objetivos Específicos	8
Capítulo 2. Marco Teórico.....	10
2.1 Las Pruebas Estandarizadas y su uso en educación.	10
2.1.1 Antecedentes de la evaluación como mecanismo de selección	11
2.1.2 La Evaluación Estandarizada	13
2.1.3 Las pruebas de admisión universitarias	17
2.1.3.2 Pruebas estandarizadas de aptitudes generales.	20
2.1.3.3 Pruebas estandarizadas de logro o conocimientos	22
2.2 Usos de las pruebas estandarizadas de selección universitaria.....	24
2.2.1 Pruebas de Admisión	24
2.2.2 Pruebas de Diagnóstico	30
2.2.3 Pruebas de Ubicación	32
2.3 Pruebas estandarizadas de selección/ubicación para Ciencias Sociales.....	33
2.3.1 Los Campos de la Educación y Capacitación CINE	36
2.3.2 Ámbito de acción de las Ciencias Sociales: Antropología y la Investigación Social	39
2.3.3 Situación de las Ciencias Sociales en Educación Superior	42
2.3.4 Pruebas basadas en expectativas de logro asociadas a contenidos específicos de las Ciencias Sociales.....	43

2.3.5 Taxonomía de objetivos de educación de Bloom	50
2.3.6 Las Habilidades cognitivas y las expectativas de logro	54
2.4 Bases para el desarrollo de las pruebas o Teorías de la Medición.....	55
2.4.1 Teoría de Medición que sustenta el estudio: Teoría Clásica de los Test	58
2.4.2 Parámetros de la TCT aplicada en la presente investigación	61
2.4.2.1 Confiabilidad.....	61
2.4.2.2 La Validez desde la TCT	63
2.4.2.3 Transformación e interpretación de los puntajes en la TCT	67
2.4.3 Análisis de los ítems en la TCT.....	70
2.4.2 Teoría de Respuesta al Ítem	74
2.4.2.1 Características generales de los Modelos TRI	75
2.4.2.2 Los Parámetros de la Curva Característica del Ítem	75
2.4.2.3 Los supuestos de la TRI	76
2.4.2.3 Modelos de la TRI.....	77
Capítulo 3. Marco Contextual	79
3.1 Honduras y su sistema educativo.	79
3.1.1 El Sistema Educativo.....	79
3.2 El Sistema de Educación Superior Universitario	81
3.2.1 Instituciones universitarias	83
3.3 Políticas de Admisión en la Educación Superior Universitaria	85
3.3.1 Normativa de ingreso a educación universitaria	85
3.4 El proceso de admisión y de selección de la UNAH.....	87
3.4.1 Prueba de Aptitud académica (PAA), Prueba de Ciencias Naturales y de la Salud (PCCNS) y otras específicas.	88
3.4.2 Políticas de inclusión en el proceso de admisión UNAH.	90
3.5 Facultad de Ciencias Sociales	92
3.5.1 Proceso de ingreso a la Facultad de Ciencias Sociales.....	92
3.5.2 Carrera de Antropología	93

Capítulo 4. Metodología	95
4.1 Enfoque, alcance y Diseño	95
4.2 Variables de investigación	96
4.3 Población y Muestra.....	98
Tipo de Muestreo	98
4.4 Instrumentos	100
4.5 Plan de análisis.....	101
4.5.1 Procedimiento.....	101
 Capítulo 5. Análisis y Discusión de los Resultados de Investigación.....	 108
Objetivo específico 1.....	108
Objetivos específicos 2 y 3	112
Objetivo específico 4.....	130
Objetivo específico 5.....	141
 Capítulo 6. Síntesis de los Resultados: Conclusiones	 144
Recomendaciones	152
 Bibliografía	 154
Declaración de Ética de la Investigación	164
Anexos	165

Índice de tablas

Tabla 1	<i>Niveles cognitivos evaluados en PSU Historia, Geografía y Ciencias Sociales ...</i>	34
Tabla 2	<i>Clasificación del CINE</i>	38
Tabla 3	<i>Diferencias entre el Aprendizaje Basado en contenidos y en Resultados</i>	48
Tabla 4	<i>Niveles cognitivos de la Taxonomía de Bloom</i>	51
Tabla 5	<i>Revisión de la Taxonomía de Bloom.....</i>	53
Tabla 6	<i>Universidades de Honduras.....</i>	83
Tabla 7	<i>Operacionalización de las variables</i>	96
Tabla 8	<i>Ficha técnica de la prueba operativa (final)</i>	100
Tabla 9	<i>Tabla de especificaciones de la Prueba de Selección Universitaria para la Licenciatura en Antropología.....</i>	109
Tabla 10	<i>Clasificación de los Ítems por el Índice de Dificultad.....</i>	114
Tabla 11	<i>Poder de Discriminación de los ítems según el valor D.....</i>	114
Tabla 12	<i>Categorías el Funcionamiento Diferencial del ítem.....</i>	115
Tabla 13	<i>Relación entre los Indicadores de calidad de los ítems de la Aplicación Piloto de la Prueba de Selección Universitaria para la Licenciatura en Antropología.....</i>	120
Tabla 14	<i>Aspectos psicómetros de la Aplicación Piloto de la Prueba de Selección Universitaria para la Licenciatura en Antropología.....</i>	121
Tabla 15	<i>Ejemplo de análisis de los ítems en la Aplicación Operativa de la Prueba de Selección Universitaria para la Licenciatura en Antropología.....</i>	127
Tabla 16	<i>Aspectos psicómetros de la Aplicación Operativa de la Prueba de Selección Universitaria para la Licenciatura en Antropología.....</i>	129
Tabla 17	<i>Métrica para representar los resultados de la Prueba de Selección Universitaria para la Licenciatura en Antropología</i>	135
Tabla 18	<i>Tabla de especificaciones para la Prueba de Selección Universitaria de la Licenciatura en Antropología.....</i>	145

Índice de figuras

<i>Figura 1.</i> Ejemplo de ítem del SAT verbal (pre-2005)	25
<i>Figura 2.</i> Ejemplo de ítem de cálculo aritmético de la ACT	27
<i>Figura 3.</i> Organigrama del Consejo Nacional de Educación	80
<i>Figura 4.</i> Estructura del Sistema Educativo de Honduras	81
<i>Figura 5.</i> Organigrama del Sistema de Educación Universitario	82
<i>Figura 6.</i> Puntajes mínimo requerido para las carreras de la Facultad de Ciencias Sociales, UNAH.....	92
<i>Figura 7.</i> Etapas de la Construcción de la Prueba de Selección Universitaria para la Licenciatura en Antropología	107
<i>Figura 8.</i> Ítems validados por el procedimiento de Juicio de Expertos utilizando Índice de validez de contenido	113
<i>Figura 9.</i> Índice de Dificultad de los ítems de la Aplicación Piloto de la Prueba de Selección Universitaria para la Licenciatura en Antropología.....	116
<i>Figura 10.</i> Índice de Discriminación de los ítems de la Aplicación Piloto de la Prueba de Selección Universitaria para la Licenciatura en Antropología	117
<i>Figura 11.</i> Análisis de Distractores del ítem 39 de la Aplicación Piloto de la Prueba de Selección Universitaria para la Licenciatura en Antropología	118
<i>Figura 12.</i> Funcionamiento Diferencial del Ítem de la Aplicación Piloto de la Prueba de Selección Universitaria para la Licenciatura en Antropología	119
<i>Figura 13.</i> Índice de Dificultad de los ítems de la Aplicación Operativa de la Prueba de Selección Universitaria para la Licenciatura en Antropología	123
<i>Figura 14.</i> Índice de Discriminación de los ítems de la Aplicación Operativa de la Prueba de Selección Universitaria para la Licenciatura en Antropología	124
<i>Figura 15.</i> Análisis de Distractores del ítem 16 de la Aplicación Operativa de la Prueba de Selección Universitaria para la Licenciatura en Antropología.	125

<i>Figura 16.</i> Funcionamiento Diferencial del Ítem de la Aplicación Operativa de la Prueba de Selección Universitaria para la Licenciatura en Antropología	126
<i>Figura 17.</i> Comparación del Alfa de Conbrach las SubPruebas de Antropología y Fundamentos de Investigación Social de la Prueba de Selección Universitaria para la Licenciatura en Antropología	128
<i>Figura 18.</i> Porcentaje promedio de estudiantes que contestaron correctamente en la SubPrueba de Antropología según los Niveles Cognitivos de la Taxonomía de Bloom...	131
<i>Figura 19.</i> Porcentaje promedio de estudiantes que contestaron correctamente en la SubPrueba de Fundamentos de Investigación Social según los Niveles Cognitivos de la Taxonomía de Bloom.....	132
<i>Figura 20.</i> Porcentaje de estudiantes que contestaron correctamente en la SubPrueba de Antropología según las Expectativa de Logro de la Malla Curricular del Bachillerato en Ciencias y Humanidades	134
<i>Figura 21.</i> Porcentaje promedio de estudiantes que contestaron correctamente en la SubPrueba de Fundamentos de Investigación Social según las Expectativa de Logro de la Malla Curricular del Bachillerato en Ciencias y Humanidades.....	135
<i>Figura 22.</i> Grupos según categorías por género en la Prueba de Selección Universitaria para la Licenciatura en Antropología. Elaboración propia.....	138
<i>Figura 23.</i> Frecuencia de puntajes en Escala Estandarizada en la Prueba de Selección Universitaria para la Licenciatura en Antropología.....	139
<i>Figura 24.</i> Comparabilidad de ingreso a la carrera con la Prueba de Selección Universitaria para la Licenciatura en Antropología teniendo como referencia el puntaje mínimo requerido en la PAA.....	140
<i>Figura 25.</i> Grupo focal estudiantes de primer ingreso de la Licenciatura en Antropología	143

Capítulo 1. Planteamiento del Problema

En este capítulo se mostrará el problema de esta investigación científica, abordando tres elementos fundamentales: situación problemática, objetivos de investigación y la justificación de la investigación.

1.1 Situación Problemática: Prueba de Selección Universitaria para aspirantes a la Licenciatura en Antropología

El tema de las pruebas estandarizadas de selección tiene gran relevancia en la agenda pública nacional, dada la importancia que reviste el hecho de poder ingresar o no a una institución universitaria, en el futuro de los jóvenes hondureños. En noviembre del 2016 se presentó ante el Congreso Nacional una propuesta para eliminar la Prueba de Aptitud Académica PAA como requisito de ingreso a la Universidad Nacional Autónoma de Honduras (UNAH), principal casa de estudios del país. La propuesta se apoyaba en el supuesto de que, mediante esa prueba, se está discriminando a los grupos de población de menores recursos y áreas rurales de nuestra sociedad, dado que han recibido una educación de menor calidad durante su formación en Básica y Media (El Herald, 2017).

Ante este tipo de afirmaciones, en fecha reciente, en ese entonces Director de Educación Superior en Honduras, defendió en los medios de comunicación el valor y la importancia de las pruebas de admisión que se vienen aplicando en el país.

“Estas pruebas de admisión llegaron para quedarse y no se aplican para dañar a los estudiantes, como se ha querido dejar entrever por sectores reacios a los cambios para la mejora de la calidad en educación superior y que se oponen y atacan a la UNAH por la realización de este tipo de pruebas. Con la aplicación de la Prueba de Aptitud Académica (PAA), la Universidad Nacional Autónoma de Honduras (UNAH) asumió el reto de aplicar un instrumento confiable y de alta calidad técnica, que le permitiera seleccionar cuáles estudiantes admitir en función de las competencias necesarias para cursar estudios de nivel universitaria.. Es una propuesta apresurada la de eliminar las PAA” (El Herald, 2015).

El tema de posibles resultados discriminatorios con ciertos grupos de estudiantes es un problema que ha sido estudiado en Honduras, dado que las desigualdades de nivel socioeconómico, el género y la pertenencia a grupos étnicos minoritarios aparecen en un

estudio nacional de la UNAH respecto a quiénes son los excluidos en este proceso de admisión (Moncada, Martínez y Castro 2012). Pero en el mismo estudio se señala que la recomendación internacionalmente aceptada es utilizar, además de la PAA, otros criterios complementarios que ayuden a disminuir estos efectos negativos sobre ciertos grupos de población, pero en ningún caso se recomienda eliminar los procesos de selección ni las pruebas como instrumento de medición (Moncada et al. 2012).

La aplicación de una prueba de admisión universitaria en países que, como Honduras padecen crónicamente de escasos recursos del erario público, puede considerarse como necesaria y conveniente para apoyar la buena inversión de esos recursos en jóvenes que realmente van a poder cursar exitosamente las carreras universitarias seleccionadas. Ello no debe interpretarse como un apoyo indirecto para ensanchar la brecha de desigualdad social, sino que por el contrario, al identificar aspirantes con aptitudes académicas universitarias puede permitir graduar profesionales calificados sin recargar las universidades de “eternos” estudiantes universitarios que, además de constituir una pesada carga al presupuesto público, van desarrollando sentimientos de frustración por no poder terminar sus estudios universitarios. Ningún país, aún los de mayores recursos, pueden admitir a todos los aspirantes a las universidades (Trost, 1992).

En este sentido, debe aclararse que estas pruebas de admisión no están entrando en contradicción con el derecho que tiene todo ser humano a la educación. Toda la normativa internacional establece que se debe dar la oportunidad a cada persona de poder educarse en diferentes niveles y modalidades de educación formal.

Pero tampoco se puede desconocer que Latinoamérica y especialmente Honduras, están atravesando por un conjunto de transformaciones sociales y educativas debido a las demandas sociales y económicas que afectan la educación. Actualmente, es una sociedad muy fragmentada a raíz de la pobreza y específicamente de la distribución desigual de las riquezas; es en las zonas rurales y las poblaciones indígenas, donde se concentran más los colectivos de exclusión educacional, por ende estos grupos tienen menor probabilidad de desarrollar su capital humano (UNESCO, 2007).

Expuesto lo anterior, también se aclara que este tipo de instrumento tiene como objetivo principal, discriminar sujetos por habilidades no por condiciones socioeconómicas.

Por lo tanto, no se debe etiquetar como un mecanismo de discriminación, sino que se debe tratar de enriquecer el proceso de admisión con políticas de inclusión en un contexto como el nuestro. Para ello, es necesario que en la agenda política educativa en materia de calidad, equidad y eficiencia educativa, principales factores en la competencia pública y estatal, encuentren un equilibrio en las nuevas propuestas a formular (Gajardo y Puryear, 2003).

Porqué aceptar a “todos” tampoco parece ser la solución, tomando en consideración las aptitudes, habilidades cognitivas y disciplina que este nivel de estudios demanda, como por la estructura misma del mercado laboral. Dado que se produciría una saturación de profesionales universitarios en ciertos campos, y un déficit de recurso humano en una serie de actividades técnicas y de servicios que son fundamentales para la buena marcha de la sociedad (Tirado, Backhoof, Larrazolo y Rosas, 1997).

Par exponer un ejemplo en la región Latinoamericana, el de Argentina es uno de los pocos países que reivindican el ingreso irrestricto como un política democratizadora, que permite el acceso a todos en Educación Superior. Sin embargo, el 20% son desertores antes de realizar el primer examen, y muchos aprueban alguna materia pero no continúan al año siguiente (El Día, 2010).

“La pregunta es: a la vista de los resultados, ¿se puede considerar que estas políticas son equitativas y correctas? Juan Carlos del Bello, actual rector de la nueva Universidad Nacional de Río Negro (UNRN), dijo lo que nadie se atreve a decir.

El principal problema de la universidad argentina es el ingreso, pero hablar de esto parece ser tabú. Las demandas democratizadoras de la sociedad llevaron al ingreso irrestricto. Y plantear algún tipo de selección resulta políticamente incorrecto. Entonces no se habla, se esconde el problema debajo de la alfombra” (El Día, 2010).

Si los estudiantes no tienen las competencias mínimas para los aprendizajes que las universidades deben desarrollar, entonces se producen problemas de reprobación, repetición, deserción o los docentes se ven obligados a disminuir el nivel de exigencias u omitir temáticas complejas, debilitando la formación profesional de estos jóvenes (Agencia Ejecutiva en el Ámbito Educativo, Audiovisual y Cultura, EACEA, 2009).

Todo ello va en desmedro de la calidad de la educación universitaria, y con ello, de la calidad de profesionales que el país está formando para sostener la

producción nacional. Por ello, dado que la formación universitaria es tan importante para la economía de un país, en las dos vías, por sus altos costos y por el desempeño que se espera de sus graduados, es importante que se seleccione rigurosamente quiénes ingresan a las universidades y también, a cuáles carreras ingresan, de acuerdo a sus aptitudes y actitudes. Adicionalmente, una selección rigurosa por capacidades y aptitudes, es más democrática y equitativa que el libre acceso masivo o la selección por capacidad de pago (Agència per a la Qualitat del Sistema Universitari de Catalunya, AQU, 2009, p.17).

Autoridades de la UNAH han señalado que esta institución ha tenido casos de estudiantes que han cursado por 20 años una carrera, y ello genera entre otros problemas, que los alumnos de nuevo ingreso no puedan matricularse en las clases que necesitan para avanzar en sus planes de estudios. “Esta situación tiene que cambiar porque nos encontramos en una institución pública y es necesario administrar de mejor manera los recursos. Casi 70% de los egresados de 2014 tardaron más de 8 años en graduarse” (Presencia Universitaria, 2014). Mirando retrospectivamente, puede señalarse que la UNAH ha logrado, mediante la aplicación de la PAA y otros instrumentos de medición, avances significativos en la selección de cuáles estudiantes deben admitirse en relación con las aptitudes académicas que han alcanzado, en comparación con el período en el que no se contaba con dichos instrumentos.

Tomando en consideración que las “decisiones extremas” (“muy selectiva o admitir a todos”) en el acceso a la universidad no son las más pertinentes para nuestra situación socioeconómica. Por lo tanto, se considera pertinente “apostar” a mantener el proceso de admisión, enriquecerlo y acompañarlo de políticas de discriminación positiva, de cuotas que respondan a una política destinada a propiciar el acceso a los colectivos “históricamente discriminados”.

Porque tal como lo señala la experiencia internacional, estos procesos de selección deben estar en continua mejora, ya que se ha venido planteando la necesidad ya no solo de identificar habilidades cognitivas genéricas que sirven para cualquier carrera universitaria, sino también la necesidad de identificar y evaluar habilidades cognitivas relacionadas con la currícula de campos educativos específicos. Ello con el propósito de ubicar a los aspirantes en las carreras más pertinentes según sus propias habilidades cognitivas y apoyar así la culminación exitosa de los estudios seleccionados por ellos, mejorando aspectos como la

permanencia, aprobación y graduación de los estudiantes (EACEA, 2009). Para ello se han venido desarrollando en diferentes países del mundo toda una serie de pruebas complementarias a las que miden las habilidades cognitivas generales, como es el caso de la PAA que se aplica en Honduras.

Este debate ha planteado el tema de decidir entre pruebas de aptitudes o habilidades cognitivas genéricas, o el uso de pruebas de conocimiento o logro. Los argumentos en favor de la primera opción se centran en que es más importante la capacidad de aprender, las aptitudes del sujeto, en particular en el manejo de las lenguas básicas del aprendizaje formal: Lenguaje verbal y lenguaje lógico matemático. Planteando que un buen desempeño en este tipo de pruebas ha demostrado ser un buen predictor del desempeño académico futuro de los estudiantes universitarios al cursar sus carreras (AQU, 2009).

Sin embargo, en algunos países se ha optado por aplicar pruebas de conocimientos o de habilidades cognitivas asociadas a campos específicos de la educación, como es el caso de la Prueba de Selección Universitaria (PSU) de Chile o la Prueba de Conocimiento de Ciencias Naturales y de la Salud que se aplica en la UNAH. En Chile se cambió de una prueba tipo aptitudes generales PAA, a una del tipo de conocimientos referidos al currículo de educación media. Al respecto se ha argumentado que esta es una forma de incentivar a los estudiantes de niveles previos al superior, a esforzarse por aprender y cubrir su currículo. Indicando que este es un criterio más “justo” respecto a los egresados de nivel medio, que saben que serán evaluados con base a lo que debieron aprender en sus estudios previos (DEMRE, 2005).

A partir de estas pruebas generales de ingreso y su discusión entre pruebas de aptitudes generales o pruebas de logro, se ha planteado también la necesidad de aplicar pruebas complementarias que ayuden a ubicar más acertadamente a sus aspirantes universitarios. En general estas pruebas complementarias están relacionadas a los Campos de Educación establecidos según la Clasificación Internacional Normalizada de la Educación (CINE), que han servido para establecer ordenamientos de validez internacional. La CINE forma parte de la familia internacional de Clasificaciones Económicas y Sociales de las Naciones Unidas, las cuales son empleadas a nivel mundial en la elaboración de estadísticas con el objetivo de acopiar y analizar datos comparables a nivel internacional de manera

consistente. La CINE clasifica los programas educativos de acuerdo a sus contenidos en función de dos variables de clasificación cruzadas: Niveles de educación y Campos de educación. (UNESCO, 2014)

La actual clasificación de los Campos de Educación y capacitación sigue un enfoque de contenido temático. Este es el mismo método que se utiliza en las versiones anteriores de la CINE (1976, 1997 y 2011). Cuando el contenido está estrechamente relacionado, las asignaturas se agrupan para formar campos amplios, específicos y detallados de clasificación, sobre la base de la similitud de los contenidos temáticos. La relación entre los Campos de Educación de CINE y habilidades cognitivas específicas es muy estrecha porque cada campo implica ciertas aptitudes y habilidades cognitivas específicas para su aprendizaje y su desempeño profesional. De manera que un estudiante que aspira a ingresar a una carrera de Artes, de Salud o de Ciencias Sociales, va a necesitar diferentes aptitudes, actitudes y habilidades cognitivas, tanto para cursar exitosamente la carrera en su condición estudiante, como para desempeñarse exitosamente como profesional en su área específica.

A partir de estas premisas, las universidades y empresas especializadas en este tipo de pruebas de admisión y selección para educación superior, han hecho la transición de exámenes de aptitudes generales a exámenes de aptitudes y habilidades orientadas a campos de conocimiento como complementos a las primeras (AQU, 2009). Ejemplos destacados de esos procesos son la Scholastic Aptitude Test (SAT) en Estados Unidos de Norteamérica, que cuenta con la SAT Subject Tests II que se centran en áreas académicas específicas (SAT en Historia de Estados Unidos y en Historia Universal), Prueba de Acceso a la Universidad y su fase específica divididas en ramas de conocimiento; así como la Prueba de Selección Universitaria (PSU) de Chile, que cuenta con dos pruebas complementarias electivas, Historia y Ciencias sociales por un lado, y Ciencias por el otro Campo de Educación. Cada una de estas pruebas utiliza los campos de CINE en sus diferentes niveles, ya sea en el nivel amplio, el detallado o el específico.

Dentro del debate de utilizar pruebas de aptitudes generales o de logro, y el hecho de adicionar pruebas complementarias para seleccionar con mayor rigor a quienes ingresan a determinadas carreras universitarias, a lo largo del siglo XX ha venido surgiendo la discusión respecto al uso de pruebas contextualizadas, elaboradas por especialistas en cada país. Chile

fue uno de los países que inició esta tendencia en América Latina, pero que ha sido continuada en toda la región, incluyendo Costa Rica y El Salvador en Centroamérica (DEMRE, 2012). Honduras en su mayor centro de estudios universitarios, la UNAH, viene aplicando durante la última década una prueba internacional de rigurosidad psicométrica, pero se plantea ya la posibilidad de construir una prueba contextualizada, tanto para admisión general como para selección/ubicación por campos de la educación.

En la UNAH hace poco más de una década (2006) se inició el proceso de selección de estudiantes admitidos mediante la aplicación de una prueba de aptitudes académicas generales, la PAA, elaborada por un reconocido centro internacional especializado en pruebas estandarizadas de este tipo. Sin embargo, en la actualidad se plantea la conveniencia de contar con pruebas complementarias que, tal como lo indica la experiencia internacional, permita seleccionar cuáles estudiantes tienen las habilidades cognitivas específicas que van a necesitar para cursar determinadas carreras universitarias. Y este es precisamente el vacío de conocimiento que el presente estudio pretende llenar: La falta de pruebas complementarias a la PAA que permitan identificar las habilidades cognitivas específicas relacionadas con los Campos de Educación en las que se inscriben determinadas carreras universitarias. Por ello, se propone una investigación orientada por los siguientes objetivos de investigación:

1.2 Objetivos de la Investigación

1.2.1 Objetivo General

Diseñar una prueba de selección universitaria basada en las expectativas de logro y habilidades cognitivas establecidos en los espacios curriculares de Antropología y Fundamentos de Investigación Social del Plan de Estudios y Programa Curricular del Bachillerato de Ciencias y Humanidades, para aspirantes a ingresar a la Licenciatura en Antropología de la Facultad de Ciencias Sociales de la Universidad Nacional Autónoma de Honduras, Tegucigalpa M.D.C., 2018.

1.2.2 Objetivos Específicos

1.2.2.1 Identificar las expectativas de logro y contenidos en relación con las habilidades cognitivas que más aportan al perfil de salida de los espacios curriculares de Antropología y Fundamentos de Investigación Social del Plan de Estudios y Programas Curriculares del Bachillerato en Ciencias y Humanidades, para la redacción de ítems para el pilotaje.

1.2.2.2 Desarrollar el proceso de construcción de instrumentos psicométricos para que la Prueba de Selección Universitaria para aspirantes a ingresar a la Licenciatura en Antropología, cumpla con los indicadores psicométricos de calidad bajo el análisis de la Teoría Clásica de los Test (TCT).

1.2.2.3 Determinar el Funcionamiento Diferencial de los ítems en las pruebas piloto y operativa, en relación con el género, para evitar discriminación en las oportunidades de selección y con el fin de garantizar la calidad de los resultados.

1.2.2.4 Analizar el rendimiento de los estudiantes en la Aplicación Operativa de la Prueba de Selección Universitaria para aspirantes a ingresar a la Licenciatura en Antropología.

1.2.2.5 Analizar los criterios vertidos por los estudiantes de primer ingreso de la Licenciatura en Antropología, respecto al desarrollo de los contenidos de Antropología y Fundamentos de Investigación Social en el Bachillerato.

Justificación

Un estudio de esta naturaleza puede justificarse desde varias perspectivas. En primer lugar su valor por la **relevancia social** que tiene el producto final de esta investigación: Una Prueba de Selección Universitaria para aspirantes a la Licenciatura en Antropología. Aporte de gran valor si se considera que la UNAH es la mayor institución universitaria del país con cobertura en las seis regiones geográficas del mapa nacional, y con la mayor matrícula a nivel nacional en este campo del saber (Hernández, Fernández & Baptista, 2014).

También puede justificarse este estudio por la **pertinencia** dado que viene a resolver una necesidad de la institución universitaria más importante del país. Se ha planteado que la UNAH requiere ir desarrollando pruebas que apoyen el proceso de admisión y selección de

estudiantes para su ingreso en determinadas facultades, escuelas y carreras que oferta la institución, como una medida que apoye la eficiencia interna de los estudiantes en su tránsito para completar sus planes de estudio.

También se justifica por su **aporte metodológico** para el país en general y la UNAH en particular, dado que sería la primera vez que se elaboran pruebas estandarizadas de selección universitaria para una carrera en específico con este tipo con el rigor metodológico que señalan los procedimientos psicométricos de la Teoría Clásica de los Test, aplicada internacionalmente para este tipo de tareas. El desarrollo de una prueba específica para medir habilidades cognitivas asociadas a contenidos específicos del programa curricular de Educación Media aporta una experiencia y práctica metodológica aún no realizada en Honduras (Hernández, Fernández & Baptista, 2014).

Su **aporte teórico** también está dado por la implicación de la Teoría Clásica de los Test para la elaboración y validación de una prueba específica al contexto hondureño, evaluando el desarrollo de habilidades cognitivas en relación con el Currículo desarrollado en el nivel medio.

Finalmente, la investigación puede justificarse por sus **implicaciones prácticas** dado que en la actualidad se está analizando la necesidad de que la UNAH pase del uso de pruebas internacionales de gran prestigio, a construir sus propias pruebas siguiendo los parámetros y requisitos que la experiencia internacional señala. Es muy conveniente para la UNAH como institución, empezar este tránsito desde pruebas internacionales hacia pruebas desarrolladas a nivel nacional.

Capítulo 2. Marco Teórico

En este capítulo se expone la fundamentación teórica que apoyará la elaboración y cálculo de los parámetros pertinentes en la Prueba de Selección Universitaria para aspirantes a la Licenciatura en Antropología, aunque las referencias sobre este tipo serán de las Ciencias Sociales en general, a falta de pruebas específicas a dicha carrera. La estructura de este capítulo contiene dos grandes temáticas, la primera referida a las pruebas estandarizadas y su uso como mecanismo de admisión y selección en las instituciones de educación superior, y una segunda que aborda las teorías que dan sustento a la elaboración e interpretación de los resultados basados en la Teoría Clásica de los Test (TCT).

Respecto a la primera temática, en un primer apartado se presentan las pruebas estandarizadas y su aplicación en educación, hasta analizar su uso en las instituciones universitarias, los argumentos a favor y en contra de las pruebas de admisión para los estudios terciarios. En un segundo apartado se aborda el tema de los tipos de pruebas estandarizadas aplicadas en las instituciones universitarias, las de aptitudes y las de conocimiento o logro, señalando sus usos diferenciados por las instancias evaluadoras en diversos países a nivel internacional.

En relación con la segunda temática abordada, las bases teóricas para el desarrollo de las pruebas, se expone primero una noción general de las teorías de medición vigente. Después se centra en la Teoría Clásica identificando los supuestos, deducciones del modelo así como su indicador de Fiabilidad, Validez, transformación de puntuaciones y análisis de ítems.

2.1 Las Pruebas Estandarizadas y su uso en educación.

Suele considerarse que la evaluación educativa en general, y la evaluación estandarizada de los aprendizajes en particular, son parte constitutiva de los procesos educativos desde la antigüedad. Pero esto no es cierto ni para la evaluación individual, ni menos la estandarizada de grandes grupos de examinandos. A continuación se presenta un breve resumen de los “exámenes” en general, para luego presentar las evaluaciones estandarizadas y las pruebas de admisión universitarias.

2.1.1 Antecedentes de la evaluación como mecanismo de selección

La historia de la evaluación en educación muestra que ha cambiado mucho a través del tiempo, y que la evaluación como actividad de selección propiamente dicha, se inició fuera del ámbito educativo. Suele pensarse que el examen tal y como hoy lo conocemos es propio de toda acción educativa desde la historia remota, pero no es así. El examen como instrumento de selección de talentos era comúnmente aplicado en la antigua China a los jóvenes que de mostraban públicamente poseer las cualidades necesarias para ocupar puestos en la administración pública, y que eran capaces de sostener constantes pruebas de su erudición para mantener el puesto (Garduño, 2000).

Este milenario sistema de selección, abandonado a principios de este siglo con la Revolución Maoísta, cumplía el propósito de estabilizar el poder jerárquico porque servía de motor para un ascenso social exento de favoritismos. La formación de los mandarines era cuidadosamente vigilada por respetados funcionarios que no tenían el menor inconveniente en ver a sus aprendices sometidos a la acreditación pública, ya que su éxito, casi seguro, significaba un galardón más para su propio prestigio.

Esta competitividad asociada a la necesidad de medir las destrezas se remonta a la antigüedad. Se ha señalado que en el mundo antiguo occidental la competitividad influía determinadamente en las actividades espirituales a partir de la imitación de la pericia física que se hacía alarde en las contiendas atléticas griegas. Señalando que del mismo modo en que públicamente se daba reconocimiento por las muestras de habilidades alcanzadas en las artes gimnásticas, así también el ejercicio de la retórica como expresión del desarrollo intelectual se convertía en evidencia socialmente aceptada (Judges, en Díaz Barriga, 1993).

La capacidad para sostener debates y controversias heredadas de la tradición griega, se mantiene como actividad importante durante la Edad Media. Por ejemplo, en el cultivo de las resurgidas siete ciencias liberales (trivium y cuadrivium) se realizaban ejercicios constantes de persuasión y de elegancia de estilo. En esa época los exámenes no eran una sola prueba sino constantes ejercicios de erudición con propósitos de discriminación de estudiantes (Garduño, 2000).

Rashdall (1936) “ha señalado que en la universidad medieval la palabra examen incluyó cierto proceso de investigación sobre la idoneidad de los candidatos, así como una

prueba directa de sus “realizaciones escolásticas”. Este proceso de investigación al que se refiere este autor era llevado a cabo constante y minuciosamente por una fraternidad profesional a cargo de la formación de los jóvenes, hasta la eventual consecución de un doctorado. En la medida en que más se elevaban los estudios más crecía la red de sociedades académicas interesadas en el estudiante (Garduño, 2000).

Es destacable que las prácticas de los profesores no abarcaban a masas o grupos sino a individuos. La prueba escrita bajo una reglamentación, fue una innovación incorporada por los jesuitas en 1599, carecía de un sentido de estandarización. El concepto de evaluación del grupo surgió en la Universidad de Cambridge, a mediados del siglo dieciocho, cambiando la mecánica de la educación por otros nuevos objetivos para los exámenes: obtener información y estandarizar el rendimiento escolar.

La Universidad de Oxford instituyó por primera ocasión el Estatuto de Exámenes (Examination Statute) en 1800, lo cual planteó nuevas exigencias y normas, bajo acuerdos generales, para permitir el ascenso académico sólo a una minoría privilegiada. Ante estos hechos, Garduño (2000) se plantea si “¿Podemos sospechar que el examen se empieza a convertir en ese momento en un instrumento para el ejercicio del poder?, ¿Se configura el examen como instrumento estructural de la política?” (p.12).

A partir de ese momento, los exámenes pasaron a tener una importancia mayor en el ámbito educativo, cuando fue separado de su articulación metodológica de la enseñanza, como lo concibe por ejemplo Comenio, para quien los exámenes y la didáctica constituían una unidad indisoluble. Al respecto se ha planteado que este puede ser considerado el rasgo que divide los sistemas pedagógicos medievales con la edad moderna, pues en sus aplicaciones medievales el examen no asumía una calificación o número, ni significaba por sí mismo la patente de promoción.

Esta idea se indica en el texto de Emilio Durkheim Historia de la educación y de las doctrinas pedagógicas, en el cual afirma que el sistema de exámenes, establecido en la época medieval, difería sustancialmente del actual, debido a que por su naturaleza, los exámenes no eran determinantes para la obtención de un título porque sólo los presentaban quienes tenían asegurado el éxito. Los profesores desarrollaban un esquema de motivación selectiva en relación directa con el grado de avance de los estudiantes (Garduño, 2000).

A partir de la modernidad el examen se fue convirtiendo en un instrumento de promoción social al mismo tiempo que opera como un mecanismo para la selección de los “más aptos”, para que puedan articularse “naturalmente” a los procesos sociales en desarrollo. Hay un cambio, ya que el rol del profesor, que antes se concentraba en la promoción “ad intra”, es decir, dentro del contexto aúlico, en la evaluación parcial de los aprendizajes. Sin embargo, en la evaluación “ad extra”, en referencia a la evaluación sumatoria como criterio de certificación de estudios para optar a posiciones sociales diferenciadas, las consecuencias de los exámenes comienzan a trascender el ámbito estrictamente educativo.

2.1.2 La Evaluación Estandarizada

Aunque desde que empezó a utilizarse en educación la medición ha estado marcada por el propósito de lograr la posibilidad de comparar en forma válida sus resultados entre los diferentes sujetos, escuelas y países, propósito que se evidencia en el desarrollo de la condición estandarizada de los resultados, la difusión mundial de esta forma de medición se ha acentuado mucho en los últimos veinte años. Las formas previas, aun las más antiguas con casi un siglo de existencia, se limitaron casi exclusivamente a los Estados Unidos, surgiendo inicialmente en el campo de la psicología y trasladándose luego al de la educación.

Las primeras acciones para internacionalizar la evaluación estandarizada surgieron en la década del setenta con el primer estudio realizado por la Asociación Internacional para la Evaluación del Rendimiento Educacional (IEA) siendo ésta entidad la que llevó a cabo la primera serie de estudios internacionales formales en las áreas de ciencia, literatura, comprensión lectora, Inglés y Francés como lenguas extranjeras y educación cívica, involucrando en esos entonces, a diecinueve (19) países del todo mundo (Comber y Keeves 1973).

Se puede plantear entonces que la evaluación estandarizada se plantea como una necesidad en la medida que los sistemas educativos nacionales se van volviendo de mayor tamaño y creciente complejidad. Ya a mediados del siglo XX varios países de Europa occidental comienzan a utilizar “Pruebas nacionales” como un medio de conocer el desempeño de los estudiantes de su sistema educativo. Para evaluar a poblaciones tan amplias se requirió el

desarrollo de pruebas estandarizadas de aprendizaje aplicadas en forma externa a los centros educativos (EACEA, 2009).

Las denominadas “pruebas estandarizadas” tienen la característica particular de que contienen ítems iguales o equivalentes para todos los alumnos que responden a una aplicación realizada en un momento específico en un espacio determinado (sea este un aldea, municipio, departamento, país.). Adicionalmente se cumple que sus aplicaciones posteriores (usualmente referida a otros años), mantienen un nivel de dificultad similar (American Institutes For Research, AIR, 2015).

Esta última característica se logra mediante un riguroso análisis de los ítems individuales en pruebas diferentes para asegurar que evalúen las mismas destrezas y tengan una complejidad similar (generalmente se aplican ensayos con muestras de estudiantes para comprobar que la proporción de respuestas correctas a dos ítems equivalentes sea la misma para muestras de alumnos seleccionados al azar).

Actualmente, el procedimiento usual para estandarizar las pruebas de esta manera, se fundamenta en el uso de ítems con respuestas a un conjunto de alternativas, con un número fijo de “distractores” y una sola respuesta “correcta”. Son las denominadas pruebas de selección única en las que el alumno escoge la respuesta correcta entre un conjunto de opciones (a menudo cinco), es el típico diseño de las pruebas estandarizadas (McMeekin, 2006). Sin embargo, en la actualidad se construyen pruebas de composición escrita que también se aplican en forma estandarizada.

Las preguntas que constituyen la mayoría de las pruebas actuales son de selección múltiple, con cuatro o cinco opciones de respuesta, de las cuales solo una es la correcta (clave). En las distintas pruebas, los ítems se encuentran constituidos de la siguiente forma: 1. Una pregunta o afirmación 2. Cuatro o cinco opciones de respuesta, que incluyen: tres o cuatro distractores, los cuales deben ser completamente incorrectos, claros y reflejar posibles errores (es decir, ser plausibles). 3. Una respuesta correcta, la cual debe ser precisa, completa e irrefutable. 4. Un estímulo: puede ser una textual o/y visual (ilustración, tabla, gráfico, texto, u otro elemento). El estímulo visual es una condición variable, pues algunos ítems incluyen directamente la pregunta o afirmación, sin necesidad de considerar un estímulo visual.

Para lograr que se cumplan estas condiciones, es necesario que la redacción del enunciado y las opciones sea clara, precisa y no dificulte innecesariamente el ítem. De ese modo, el ítem en su conjunto, debe tener un alto nivel de rigurosidad conceptual y no dar lugar a dobles interpretaciones o cuestionamientos.

En relación con los criterios generales que los constructores consideran para elaborar cada ítem, se destacan los siguientes. En primer lugar cada ítem debe ser fácilmente identificable en cuanto al contenido mínimo y la habilidad que evalúa. Segundo, los ítems deben ser totalmente originales, lo cual significa que no pueden ser copiados o reproducidos total o parcialmente de ningún tipo de literatura, internet u otro medio de difusión público. Tercero, los ítems deben ser imparciales, evitar en su formulación elementos con significados u orientaciones ideológicas, políticas, de género, religiosas, emotivas; así como prejuicios o estereotipos que pudiesen favorecer a un grupo específico. Finalmente, se requiere que sean disciplinariamente correctos, presentando información fidedigna y exacta y no pueden contener errores conceptuales.

Con estos requisitos técnicos, a mediados del siglo XX se comenzó a usar “pruebas nacionales” en Europa, Islandia 1946, Portugal 1947, Irlanda del Norte 1947, Escocia 1962, Luxemburgo 1968, Países Bajos 1970, Malta y Dinamarca 1975, son ejemplos de estos inicios. Pero es hasta mediados de los años 80, en el contexto de la difusión de la ideología neoliberal desde el área económica a las demás áreas sociales, que se comienza a plantear el problema de la calidad educativa y la necesidad de evaluarla (Cano, 1999).

Esta nueva perspectiva se ve apoyada por las investigaciones de los economistas publicados en la década de los años 90, señalando que para que la educación incida en el desarrollo social no es suficiente plantearlo como un problema de cantidad basado en años de escolaridad o número de graduados, sino más bien un tema de calidad y aprendizajes desarrolladas por los estudiantes (Hanushek & Wöbmann, 2007).

La importancia de evaluar los aprendizajes que los alumnos están alcanzando, en el contexto sociopolítico neoliberal de la “Nueva administración pública” de los años ochenta, brindó un nuevo impulso a la tendencia de aplicar prácticas de Accountability o Rendición de Cuentas en los sistemas educativos nacionales. Este tipo de aplicaciones cuenta con una amplia

experiencia en los países europeos occidentales pero aún es incipiente en la América Latina y otras regiones subdesarrolladas.

Es a partir de este nuevo contexto sociopolítico internacional, que afectó las estructuras gubernamentales especialmente en Europa occidental, América y parte de Asia, que se impulsó el desarrollo de sistemas nacionales de evaluación estandarizada para conocer los resultados de aprendizaje en los sistemas educativos, y esta tendencia se ha venido acrecentando en los últimos treinta años, en prácticamente todas las regiones del planeta. Al mismo tiempo ha aumentado la participación de los países en las evaluaciones internacionales estandarizadas tales como el Programme for International Student Assessment (PISA), Progress in International Reading Literacy Study (PIRLS), Trends in International Mathematics and Science Study (TIMSS), Tercer Estudio Regional Comparativo y Explicativo (TERCE) y el Southern Africa Consortium for Monitoring Educational Quality (SACMEQ), etc. (Secretaría de Educación de Honduras, SE, 2015).

Norma o Criterio en la interpretación de resultados de las pruebas estandarizadas

Los primeros instrumentos de medición estandarizada se elaboraron a partir de la necesidad de clasificar u ordenar sujetos entre sí y se denominan pruebas referidas a normas. Una forma simple de visualizar el mecanismo a través del cual éstas operan es considerar que lo que hacen es estructurar la distribución de los resultados obtenidos por los sujetos en un instrumento, de forma relativa y luego establecer el ordenamiento de estos últimos de manera acorde. Es importante precisar que en este enfoque la distribución de los resultados necesariamente mostrará proporciones de sujetos en las distintas categorías de resultados que coinciden con una distribución típica, como es la normal, ello es que porcentajes menores de los resultados corresponderán a sujetos con muy altos o muy bajos puntajes y que la mayor proporción de los resultados de los examinados se ubicarán en las áreas de puntajes intermedios. Un ejemplo de este tipo de instrumentos son las pruebas de selección para el ingreso a la educación superior, conocidas, en los Estados Unidos, como SAT.

El segundo enfoque, que se desarrolló posteriormente al anterior, incluye instrumentos cuyo propósito es determinar la posición de los sujetos respecto de un referente externo, absoluto y pre-establecido, no considerando especialmente relevante la posición relativa de los mismos. En esencia lo que interesa establecer, en este caso, es cuáles sujetos

pasan una determinada valla, criterio o puntaje, para establecer si son o no competentes de manera absoluta en un aspecto determinado. Este tipo de pruebas se denominan referidos a criterios y en su desarrollo e interpretación ha contribuido significativamente la elaboración teórica especializada denominada Teoría de Respuesta al Ítem (TRI). Las pruebas de rendimiento corresponden a este tipo de instrumentos y un buen ejemplo lo constituyen las usadas por los estudios internacionales mencionados anteriormente. Es importante señalar que la distribución de los resultados de este tipo de pruebas no tiene por qué corresponder a la curva normal y más aún, es esperable que no lo haga, dado que si se considera que la educación es un proceso de crecimiento equitativo para todos los estudiantes, sería esperable que su aprendizaje no se distribuyera normalmente.

De esta manera, una vez establecidas las diferencias iniciales entre ambos tipos de evaluación, los primeros trabajos en el ámbito de la evaluación criterial se vinculan con el estudio de los métodos de definición del universo de medida, intentando establecer técnicas que permitan estimar la puntuación de un sujeto definida a partir de su ejecución en una muestra de elementos del universo.

2.1.3 Las pruebas de admisión universitarias

En los últimos años el ingreso a la universidad, esencialmente en el ámbito público, aparece como tema problemático. Las políticas de admisión que buscan regular el acceso a este nivel mediante diversos procedimientos, constituye un componente cada vez más controvertido en las políticas de educación superior (Martínez, 2005). Estos procedimientos se supeditan, de algún modo, a las condiciones contextuales que en función de la relación demandante - poder público establecen acuerdos mínimos como los requisitos exigibles a los aspirantes. En este ámbito se sitúa actualmente la discusión acerca de la educación superior como derecho o como privilegio, es decir, si estos requisitos exigibles se materializan en la sola acreditación de la aprobación del nivel medio o si estos requisitos exigibles se materializan en la superación de determinados dispositivos de selección (Juarros, 2006).

Hace más de dos décadas Trust señaló que:

“Difícilmente existe algún país en el mundo donde el número de candidatos para la educación superior -al menos con respecto a ciertos cursos- no excede al número de lugares disponibles. Ninguna sociedad, ni aun las más ricas, puede tener la capacidad para ofrecer un número ilimitado de lugares para satisfacer las demandas académicamente orientada a los jóvenes” (1992, p.2).

La argumentación está relacionada en particular con los cursos para ciencias naturales y médicas que requieren de equipo caro de laboratorio; pero el razonamiento puede extenderse para otras áreas de las humanidades que requieren de salas de lectura y bibliotecas entre otros servicios. Pero en ambos casos, la argumentación presentada plantea la necesidad de alguna forma de proceso de selección.

Este planteamiento se basa en un argumento económico referido a la escases de recursos, por lo que no se puede ni debe aceptar a todas las personas que deseen ingresar a las instituciones de nivel superior universitario, es en principio incompatible con el derecho fundamental de elección libre de profesión vigente en algunos países (Juarros, 2000). Por ejemplo, en la República Federal de Alemania, la Ley Básica Constitucional plantea que este derecho es interpretado unánimemente como el derecho de escoger libremente los cursos en todas las etapas del sistema educacional, pero únicamente sí se cumple con los requerimientos necesarios para la admisión en la universidad se necesita haber obtenido el certificado Abitur o su equivalente.

Generalmente, en la mayoría de los países, la selección se realiza al ingresar a educación superior. Hay una amplia variedad de criterios en los cuales se basan las decisiones de selección. Pueden ser clasificadas de acuerdo con las siguientes categorías: 1. Logros escolares, 2. Aptitudes generales (referidas al manejo de los dos idiomas básicos del aprendizaje, el verbal y el lógico matemático), y 3. Criterio formal, como el “tiempo de espera” (oportunidad).

Actualmente, los logros escolares es el criterio más utilizado para la selección cuando se comienza la educación superior. En la mayoría de los países denominados como subdesarrollados, por ejemplo, es el único criterio, pero aun en sistemas donde se aplican otros criterios, como el de aptitudes escolares, son aplicados también los logros escolares. Generalmente se considera principalmente en alguna de estas tres formas: 1. En términos de

promedios o puntos promedios de calificaciones en la educación media superior, 2. En términos de posición (primero, segundo lugar, entre otros) dentro de su grupo al final de año, 3. Por medio de exámenes orientados a diferentes materias (Trust, 1992).

Ejemplo de países en los cuales se ha venido implementando estos procesos de selección por exámenes de conocimientos en los últimos veinte y más años están Brasil, República Popular de China, Grecia, Japón, Corea del Sur, Turquía y los Estados Unidos. De manera similar, el procedimiento de selección para la admisión a las universidades más prestigiosas incluye una evaluación oral y escrita de conocimientos adquiridos.

Ocurre también que en otras naciones el procedimiento aplicado para seleccionar a los estudiantes que ingresan por medio de puntajes mínimos en ciertas materias escolares o una combinación equilibrada de puntajes individuales. Sin embargo, varios estudios han señalado que el uso de puntajes aislados o la combinación equilibrada de puntajes seleccionados, no mejora la predicción de éxito como estudiantes, en comparación con el uso de un promedio del puntaje de pruebas específicas para ingreso (Lissmann, 1997). En general, los procedimientos de admisión incluyen exámenes de conocimientos generales; un claro ejemplo de esta situación lo constituyen las universidades italianas.

Cuando se aplican estos exámenes de admisión, generalmente la evaluación de los logros se refiere a la etapa previa de los ciclos educativos. Esto contrasta con la evaluación de aptitudes, la cual está orientada hacia los requerimientos de la etapa educacional hacia la cual están aspirando ingresar los estudiantes. Es decir, las aptitudes escolares son usualmente evaluadas utilizando exámenes estandarizados de opción múltiple, aunque también ocurre en algunos casos que este proceso se desarrolla por medio de exámenes tipo ensayo. Adicionalmente, la entrevista se utiliza algunas veces para valorar mejor la aptitud escolar de los candidatos. A continuación se presentan algunos de los argumentos principales que se esgrimen a favor de las pruebas de aptitudes generales y las pruebas de logro o conocimiento.

2.1.3.2 Pruebas estandarizadas de aptitudes generales.

De acuerdo con la literatura internacional, las pruebas que se utilizan en educación pueden dividirse en dos grandes grupos según las implicaciones que se le den a sus resultados para el estudiante: de altas y bajas consecuencias. Los exámenes de admisión a la universidad se consideran de altas consecuencias, ya que a partir de sus resultados se toman decisiones trascendentales para las personas, los estudiantes que aplican para ganar su ingreso.

Al irse extendiendo el uso de este tipo de pruebas ha sido necesario regular el proceso de construcción de las mismas y garantizar un uso adecuado de los resultados, la comunidad de expertos en el tema ha elaborado estándares que prescriben lo que se considera una buena práctica en el ámbito del desarrollo y uso de instrumentos de evaluación. El documento que los recoge es el *Standards for Educational and Psychological Testing*, elaborado por la American Educational Research (Eyzaguirre, 2003).

De acuerdo con estas directrices, en la etapa inicial de la planificación de una prueba se debe definir la suma de sus características. Las especificaciones contienen el propósito de la prueba, la fundamentación que avala la selección de contenidos, destrezas y habilidades a ser medidas, la tabla de especificaciones de los contenidos a evaluar, la definición del formato de las pruebas y una anticipación de las consecuencias que se pueden esperar. “Idealmente las especificaciones deben ser tan completas que dos constructores de prueba operando independientemente sobre la base de ellas debieran producir instrumentos comparables e intercambiables, que difieran tan sólo en la muestra de preguntas utilizadas” (Eyzaguirre, 2003, p.48).

Revisando un poco la historia de este tipo de exámenes se tiene que fue en Estados Unidos donde se inició el proceso de elaboración de pruebas de aptitudes para seleccionar los jóvenes que serían admitidos en las universidades. La bibliografía señala que fue el nacimiento del College Board, en noviembre del año 1900, se relaciona con un momento en que el incremento de jóvenes que terminaban la educación media superior, con la proliferación de colleges, que ocurrió en esa época en los Estados Unidos, lo cual volvía muy complejos los procesos de selección para el ingreso en una institución universitaria (Martínez, 2015).

Fuess señala que la fundación del College Board fue el primer intento organizado de “introducir la ley y el orden en una anarquía educativa que, a fines del siglo XIX, había llegado a ser exasperante, sin duda casi intolerable, para los directores de escuelas” (Citado en Donlon, 1984). En esa época había “un consenso preocupantemente reducido entre los colleges en cuanto al tipo de preparación en ciertas áreas de contenido y en cuanto a los estándares de desempeño que debían pedirse a los aspirantes” (Donlon, 1984, p.1).

Fue hasta 1925 cuando, gracias al desarrollo de las técnicas psicométricas, el College Board decidió desarrollar pruebas de aptitud, en lugar de pruebas de conocimientos, pretendiendo no quedarse con la medición de simple memorización de datos aislados, tratando de acercarse a la evaluación de habilidades intelectuales de tipo general. La prueba que todavía hoy se llama Scholastic Aptitude Test (SAT) se gestó a partir de abril de 1925, y vio la primera luz el 23 de junio de 1926, cuando se administró a 8,040 sustentantes. Se inició con nueve subpruebas (definiciones, aritmética, clasificaciones, lenguaje artificial, antónimos, series, analogías, inferencias lógicas y lectura), que posteriormente se redujeron a siete en 1928 y a seis en 1929, agrupadas en dos secciones, de aptitud verbal y de aptitud numérica (Donlon, 1984, citado por Martínez, 2015).

La bibliografía señala que fue en el año 1937 que se inició a hacer una segunda aplicación anual que, a diferencia de la tradicional que seguía siendo de tipo ensayo, consistía en pruebas íntegramente compuestas por preguntas de opción múltiple. Ambas se calificaban en una escala con media de 500 puntos y desviación estándar de 100. La igualación de versiones y el cuidado de la estabilidad de la prueba comenzaron a hacerse en 1941, cuando se estandarizaron los puntajes del SAT con la población de 10,654 sustentantes de la aplicación de abril de dicho año. A partir de 1942 todas las aplicaciones usaron exclusivamente preguntas de opción múltiple. El número de sustentantes anuales llegó desde fines de los años sesenta al millón y medio de personas, cifra que se ha mantenido desde entonces (Donlon, 1984, p.3-8, citado por Martínez, 2015).

Actualmente se reconoce que a partir de los trabajos pioneros del College Board, estas pruebas de aptitudes generales tienen las siguientes características: 1. Evalúa la capacidad para los estudios superiores y las habilidades desarrolladas durante el proceso educativo, 2. Aunque no mide conocimientos específicos, supone los conocimientos generales y destrezas

propias del nivel de escolaridad, 3. El componente de razonamiento verbal incluye razonamiento analógico en contexto. El componente de razonamiento matemático incluye temas fundamentales de álgebra intermedia, y 4. El enfoque de la redacción es morfosintáctico y semántico.

En relación con el contenido, puede plantearse que la Prueba de Aptitud Académica, consta de dos secciones, una “verbal” y otra de “matemáticas”, con tiempo separado para cada una de ellas. Las preguntas verbales miden la habilidad del candidato para entender lo que lee y la riqueza y dominio de su vocabulario. Por su parte las preguntas de matemáticas miden la habilidad para resolver problemas relacionados con el razonamiento aritmético, algebraico y geométrico.

Diferentes estudios han mostrado que ambas habilidades se relacionan con el éxito en las materias que se estudian a nivel universitario. Por esta razón se asume que el conocimiento del nivel de desarrollo de las habilidades matemáticas y de comprensión lectora es muy importante, ya sea para el afianzamiento de los currículos o para proponer sus reformulaciones, tanto en secuencia, pertinencia y profundidad de contenidos como en metodologías de enseñanza, pues está fuera de discusión el hecho de que estas dos habilidades, pensamiento lógico y comprensión lectora, constituyen la piedra angular del aprendizaje desde el punto de vista cognitivo, sin soslayar el hecho de que la motivación, intrínseca o extrínseca, permitirá, o no, que dicho aprendizaje se concrete.

2.1.3.3 Pruebas estandarizadas de logro o conocimientos

El diseño de pruebas estandarizadas de logros requiere un trabajo altamente sofisticado y complejo por las técnicas y conocimientos que implica. Lo que se pretende hacer es crear herramientas de evaluación que permitan hacer una inferencia válida sobre los conocimientos y/o las destrezas que posee un estudiante determinado en un área particular de contenidos. Más precisamente, esa inferencia tiene que referirse a normas, de manera que los conocimientos y/o las destrezas relativas de un estudiante puedan ser comparados con las poseídas por una muestra nacional de estudiantes de la misma edad o grado escolar (Popham, 1999).

De manera que se hacen inferencias respecto al dominio que el estudiante tiene sobre conocimientos y/o destrezas en un área temática o asignatura particular pueden dar mucha información a los padres y educadores. Por ejemplo, un alumno de 4to. Grado se está desempeñando realmente bien en lenguaje (percentil 94) y matemáticas (percentil 89), pero más bien insatisfactoriamente en ciencias (percentil 39) y estudios sociales (percentil 26). Esa información, al echar luz sobre las fortalezas y las debilidades de un niño, puede ser útil no solo para abordar el tema con el profesor, sino también para brindarle ayuda en casa. De igual manera, si los profesores saben cómo están sus alumnos en comparación con otros del país, pueden usar esta información para planificar una enseñanza apropiada en el salón de clase.

Sin embargo, la tarea es compleja por la enorme cantidad de conocimientos y/o destrezas que se espera que los estudiantes dominen antes de ingresar a un bachillerato o a una universidad en cualquier país. El elevado tamaño del ámbito de contenidos que se supone que una prueba estandarizada representa, plantea grandes dificultades a la labor de los diseñadores de esas pruebas. Los especialistas han planteado al respecto que, si una prueba cubriera en realidad todo los conocimientos y las destrezas de un dominio, sería excesivamente extensa (DEMRE, 2015).

Es por este importante motivo que las pruebas estandarizadas de logros a menudo tienen que cumplir su labor de medición con un conjunto mucho menor de ítems que el que se usaría si el tiempo que requiere una prueba no fuera problema. Para salir de esta complicación las pruebas estandarizadas de logros muestrean los conocimientos y/o las destrezas de una esfera o dominio de contenidos. Frecuentemente esas pruebas intentan cumplir su labor evaluadora con solo 40 o 50 ítems (a veces menos) en el campo de una asignatura (DEMRE, 2015).

Por lo anterior se tiene que los diseñadores de las pruebas estandarizadas de logros se ven obligados a crear un instrumento de evaluación que, con una relativamente pequeña cantidad de ítems, proporcione interpretaciones válidas, referidas a normas, sobre la situación de un estudiante respecto a una porción sustancial de contenidos. En términos generales se plantea que los ítems que cumplen mejor la labor de diferenciar o discriminar entre los estudiantes son aquellas que aproximadamente la mitad de ellos responden

correctamente. En este sentido los evaluadores deben evitar los ítems que son respondidos correctamente por demasiados estudiantes o por muy pocos de ellos.

De manera que en el diseño de las pruebas debe hacerse un muestreo cuidadoso de los contenidos y concentrarse en los ítems que mejor discriminan entre los estudiantes. Para ello los diseñadores de pruebas han producido herramientas de evaluación que son muy buenas para comparar el dominio relativo de contenidos que ha logrado un escolar con los demás escolares del país. Si se asume que el grupo normativo (promedio) nacional es representativo del total del país, los educadores y los padres pueden hacer inferencias útiles sobre los estudiantes. Y ellos son particularmente importante a la hora de seleccionar quiénes van a ingresar a la universidad.

2.2 Usos de las pruebas estandarizadas de selección universitaria

Otro tema fundamental que debe discutirse respecto a las pruebas estandarizadas es si deben aplicarse o no pruebas de selección para las universidades, y vinculado con ello, está la problemática sobre el fin u objetivo con el cual se aplican las mismas. A este respecto suele diferenciarse entre las pruebas orientadas a establecer la admisión, las consideradas como pruebas diagnósticas y las pruebas de ubicación. A continuación se exponen brevemente las características principales de cada una de ellas, así como las principales bondades y limitaciones que señala la literatura especializada.

2.2.1 Pruebas de Admisión

En relación con esta última problemática, la bibliografía especializada identifica que fue en Estados Unidos donde se inició el proceso de aplicación de pruebas estandarizadas para establecer procesos de admisión universitarios. La evaluación estandarizada para la admisión universitaria comenzó en 1926 con el SAT, en ese entonces conocido como la Prueba de Aptitud Académica (Scholastic Aptitude Test). Hasta ese momento, las pruebas de admisión en Estados Unidos habían involucrado exámenes escritos intensivos sobre diferentes materias. El SAT ofrecía algo completamente nuevo: una prueba de opción múltiple fácil de calificar para medir la capacidad general o aptitud de los estudiantes para aprender.

Esta prueba innovadora (la SAT), se originó a partir de la experiencia de las pruebas de Coeficiente Intelectual (CI) aplicadas a los soldados durante la Primera Guerra Mundial. Los creadores de ambas pruebas compartían varios supuestos que la mayoría de los especialistas hoy consideran problemáticos: que la inteligencia es un rasgo unitario y heredado; que no está sujeta a cambios a lo largo de la vida; y que puede medirse con un número único.

Por diversas razones esta prueba resultó ser atractiva para los institutos y universidades estadounidenses. Era estandarizada, a diferencia de las calificaciones de la escuela secundaria y, por lo tanto, podía usarse para comparar a los postulantes de diferentes escuelas. De acuerdo a sus promotores, podía identificar a estudiantes prometedores de escuelas más pobres que de otro modo podrían no ser aceptados. Pero sobre todo, el SAT ofrecía una herramienta de predicción que les entregaba a los encargados de admisión un medio para distinguir entre postulantes que se desempeñarían bien o mal en la universidad. Es fácil entender por qué la prueba se hizo popular en los años que siguieron a la Segunda Guerra Mundial.

En buena medida influida por su vínculo inicial con las pruebas de evaluación del CI, el SAT durante largo tiempo ha sido conocido por sus ítems complicados, similares a los de un puzzle. La Figura 1 muestra un tipo de ítem muy discutido, denominado analogía verbal, que se usó hasta hace unos pocos años.

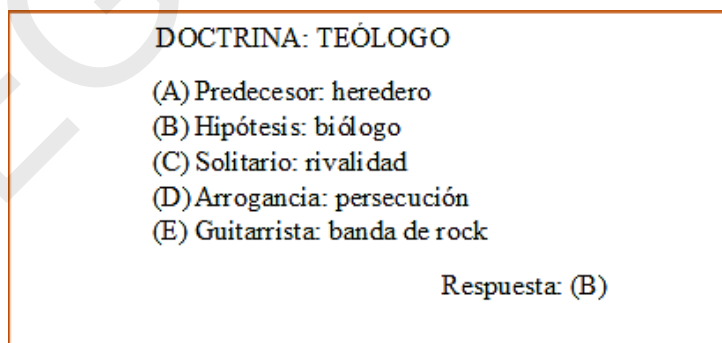


Figura 1. Ejemplo de ítem del SAT verbal (pre-2005). Geiser (2016). Tomado del College Board.

Aun cuando estos ítems presentan bastante atractivo por su capacidad discriminatoria, fueron muy criticados señalando que los estudiantes normalmente no aprenden analogías verbales como parte de sus estudios regulares en la educación secundaria. Aunque tenía la intención de medir el razonamiento verbal, sus críticos afirmaron que solo

medía la capacidad de los estudiantes de pagar servicios de preparación para la prueba, en los que se les enseñaban analogías verbales. Este tipo de ítem fue retirado del SAT en 2005.

El American College Testing (ACT) entra en escena

Unos treinta años más tarde apareció una competencia para la prueba del SAT, el ACT (un acrónimo de American College Testing), Exámenes de Ingreso a las Universidades Estadounidenses fue presentado en 1959 como un competidor del SAT. El ACT encarnaba un enfoque alternativo para las pruebas de admisión. En lugar de evaluar la capacidad o aptitud general de los estudiantes, el ACT fue diseñado para medir el logro de los estudiantes, es decir, cuánto habían aprendido en la escuela y el grado en el cual habían logrado dominar las diferentes asignaturas. La percepción común es que, mientras el SAT mide la capacidad innata de los estudiantes, el ACT mide cuánto aprendieron en sus estudios escolares.

Aplicando esta perspectiva diferenciadora respecto al SAT, el ACT se alinea más estrechamente con los currículos de las escuelas secundarias que el SAT. Los ítems de las pruebas se desarrollan a partir de los cursos ofrecidos a nivel secundario en todos los Estados Unidos. El ACT parece ser menos “entrenable”, y el consenso es que el ACT enfatiza menos las habilidades para rendir pruebas y más el conocimiento acabado del currículo. Siguiendo la filosofía del ACT, los ítems incluidos en las pruebas tienden a ser más directos que los del SAT. La Figura 2 muestra un ítem representativo de la sección matemática del ACT.

Para defender este tipo de ítems se han argumentado varias razones. Por ejemplo, que responder el ítem presentado en la figura 2, requiere solamente un cálculo directo. El mayor desafío para los estudiantes es lo que se conoce como la restricción en el tiempo de la prueba, es decir, la necesidad de responder un gran número de ítems, como este, en un tiempo acotado. Es decir, se toma en cuenta la velocidad con la cual el sujeto es capaz de hacer algunos cálculos aritméticos sencillos. La velocidad de respuesta pasa a ser un factor considerado a evaluar.

En promedio, un automóvil rinde 27 millas por galón (11,5 kilómetros por litro). Si la gasolina cuesta \$4, 04 por galón (\$1,07 por litro), ¿Cuál de las siguientes alternativas se acerca más al costo de la gasolina necesaria para que el automóvil viaje 2.727 millas (4.373 kilómetros) en una autopsia típica?

- (A) \$44, 44
- (B) \$109,08
- (C) \$118,80
- (D) \$408,04
- (E) \$444,40

Respuesta: (D)

Figura 2. Ejemplo de ítem de cálculo aritmético de la ACT. Geiser (2016) tomado de ACT Inc.

Con el paso del tiempo ambas pruebas se han venido utilizando para seleccionar a los nuevos estudiantes universitarios. El SAT ha sido la prueba de admisión dominante en EE.UU. durante un largo tiempo, pero el ACT ha logrado avances en los últimos años; en 2012, por primera vez, hubo más estudiantes que rindieron el ACT que el SAT. El SAT predomina en las Costas Este y Oeste, así como en Texas, mientras que el ACT predomina en los estados del medio oeste. California es en su mayoría un estado SAT, aunque la Universidad de California también acepta el ACT. De todos los estados, California es el mayor usuario del SAT.

Convergencia del SAT y el ACT. Los especialistas han señalado que pese a que ambas pruebas fueron muy diferentes al principio, con el tiempo se han ido acercando en sus propuestas. El SAT ha eliminado varios de sus tipos de ítem más complicados, como las analogías verbales y las comparaciones cuantitativas. Por su parte, el ACT se ha vuelto más restringido en el tiempo, haciendo mayor hincapié en las habilidades de manejo del tiempo de los estudiantes y en el recuerdo rápido, con lo que se ha ido asemejando al SAT en ese aspecto.

Otro aspecto en el que han convergido es que ambas han añadido también una prueba de escritura, respondiendo en parte a las investigaciones realizadas en la Universidad de California que muestran que la prueba de escritura era uno de los mejores predictores de rendimiento en la universidad. Las dos pruebas tienen aproximadamente la misma capacidad predictiva, y hoy casi todas las universidades estadounidenses aceptan las dos y ven las

puntuaciones del SAT y del ACT como intercambiables. Además, el punto central es que ambas son pruebas con referencia a normas.

Pruebas de admisión universitaria con referencia a normas versus pruebas con referencia a criterios

Otro aspecto fundamental en el cual existen diferentes tipos de pruebas estandarizadas es respecto a si están referidas a norma o a criterio. Las pruebas con referencia a normas buscan ordenar a los estudiantes, es decir, compararlos con otros estudiantes que rinden la misma prueba. Están diseñadas para producir una distribución normal. La distribución normal es producto de la manera en que se construye la prueba y de la forma en que se crea la escala. Al desarrollar la prueba, el diseñador descarta los ítems que demasiados estudiantes pueden responder correctamente; el ítem ideal es el que divide en partes iguales a la población que rinde la prueba. Luego, en el escalamiento de los puntajes, los puntajes originales se convierten en puntuaciones escaladas de una manera que producen «colas» con pendientes a ambos lados de la distribución.

Se ha identificado que hay una importante diferencia entre las pruebas con referencia a normas versus las pruebas con referencia a criterios en este aspecto. De manera que las pruebas con referencia a criterios miden cuánto saben los estudiantes sobre un tema dado, más que cómo se comparan con otros. En una evaluación con referencia a criterios, es posible que la mayoría o incluso todos los estudiantes logren un buen desempeño, suponiendo que saben el tema sobre el cual se les pregunta. Esto no es posible en una prueba con referencia a normas debido al modo en que se califica el rendimiento en una curva.

Actualmente se usan tanto la prueba SAT como el ACT. En algunos estados en EE.UU. usan el SAT o el ACT para medir el logro académico en sus escuelas K-12 (desde kindergarten hasta el duodécimo grado). Sin embargo, debido a que las pruebas con referencia a normas están diseñadas para producir siempre una distribución normal, no son apropiadas para medir cambios en el logro académico dentro de cualquier población de estudiantes o escuelas. Los diseñadores de pruebas regularmente eliminan los ítems que demasiados estudiantes responden correctamente, por lo que es imposible que las pruebas detecten mejoras en logro académico a través del tiempo.

Debe recordarse que las evaluaciones con referencia a norma como forma de medir lo que han aprendido los estudiantes (a diferencia de cómo se comparan con otros estudiantes) las instituciones de educación superior selectivas las consideran muy útiles para fines de admisión. Esta situación ocurre con mayor frecuencia en aquellos institutos y universidades de alta selectividad que reciben muchas postulaciones. En estos casos, las puntuaciones en estas pruebas entregan una medición breve y aparentemente objetiva de la capacidad de los estudiantes que hace fácil compararlos y ordenarlos.

También es importante considerar que las pruebas con referencia a normas destacan pequeñas diferencias en el extremo superior de la distribución de puntuaciones, dentro del cual las universidades seleccionan a sus estudiantes. Esto permite que los funcionarios de admisión establezcan diferencias sutiles dentro de grupos de postulantes donde casi todos los estudiantes han alcanzado un alto rendimiento en la educación secundaria.

Características esenciales de las pruebas de admisión universitaria. En síntesis, a partir de las investigaciones de la Universidad de California, se puede concluir que las siguientes características son las más importantes a considerar cuando se utilizan pruebas de admisión universitaria:

- El foco principal no debiera ser cómo se compara un postulante con otros, sino más bien si ha logrado dominar el conocimiento y las habilidades básicas necesarias en la universidad.
- Las pruebas de admisión debieran exhibir tanto validez aparente como validez predictiva, de modo que el vínculo entre el conocimiento y las habilidades puestas a prueba y las necesarias para la universidad, sea transparente.
- Las pruebas de admisión debieran alinearse con el trabajo realizado en la educación secundaria y reforzar la enseñanza y el aprendizaje de un riguroso currículo en nuestras escuelas.

Las pruebas de admisión debieran minimizar la necesidad de prepararse, premiando el conocimiento de los contenidos por sobre las habilidades para rendir pruebas, de modo tal que la mejor preparación para la prueba sea la instrucción normal en el aula.

- Finalmente, las pruebas de admisión debieran enviar una señal clara a los estudiantes: trabajar duro y dominar las asignaturas en la educación secundaria son el camino más directo a la universidad.

A manera de resumen de los estudios realizados por la Universidad de California, los hallazgos muestran que los criterios de admisión que miden el dominio de los contenidos curriculares (como las calificaciones en la educación secundaria y las pruebas basadas en estándares) predicen el éxito de los estudiantes en la universidad al menos igual de bien que las pruebas con referencia a normas como el SAT o el ACT. Estos indicadores (especialmente el promedio de calificaciones en la enseñanza secundaria) también son más “justos y más inclusivos” para los postulantes de bajos ingresos y pertenecientes a minorías.

Adicionalmente, se señala que otro aspecto igualmente importante es que estos criterios tienen un efecto de señalización más amplio y más beneficioso durante todo el proceso educativo. Ellos envían un potente mensaje a los estudiantes: trabajar duro y dominar las asignaturas académicas en la educación secundaria (no la capacidad innata) es la clave para ingresar a la universidad. De este modo, ayudan a alinear la enseñanza, el aprendizaje y la evaluación durante el camino entre la escuela secundaria y la universidad.

2.2.2 Pruebas de Diagnóstico

Además del uso de las pruebas estandarizadas para orientar los procesos de selección universitaria o de otras instituciones educativas es la búsqueda de diagnósticos en diferentes áreas como salud o educación. Las pruebas de diagnóstico son un componente fundamental de la práctica clínica de salud, pero también en educación; no solamente para el diagnóstico, como lo indica su nombre, sino también para el manejo diario de los pacientes o estudiantes, para la toma de decisiones relacionadas con el pronóstico, y aun para definir políticas en el ámbito de las indicaciones y la utilidad de las pruebas de tamizaje. El ejercicio del diagnóstico no es un fin en sí mismo. Éste obtiene sentido solamente cuando conduce a un curso de acción específico o a una decisión específica basada en este conocimiento alcanzado.

La información disponible señala que es en el área de la salud donde las pruebas estandarizadas de tipo diagnóstico tienen una aplicación más frecuente. El médico en ejercicio clínico, es decir, en contacto permanente con pacientes, enfrenta diariamente un

sinnúmero de situaciones en las que pone en juego su conocimiento y su razonamiento. Desde que se escucha un motivo de consulta –un síntoma, una queja, una preocupación– se desencadena un proceso por medio del cual se debe discernir continuamente entre diferentes cursos de acción.

Pruebas diagnósticas en educación. Respecto al uso de las pruebas diagnósticas en el campo específico de la educación, estas pruebas se enfocan en poder indagar respecto a la posible existencia de determinados problemas de aprendizaje y cuáles son sus características concretas. El objetivo es, pues, delimitar si se da un determinado problema en el proceso normal de adquisiciones que desarrolla un sujeto y poder extraer información sobre la cualidad de tal problema, de modo que se pueda orientar mejor la subsiguiente intervención. En este grupo se consideran dos grandes tipos de pruebas: referidas al currículum y las de diagnóstico propiamente dichas.

Respecto a las primeras, tienen por objeto indagar acerca de la posición de un sujeto respecto a un diseño curricular dado (que actúa como Dominio Educativo). Se han clasificado así porque su propósito generalmente es de tipo diagnóstico, dado que se trata de recabar información independiente de las calificaciones académicas y sin finalidad de uso en el contexto de aula, acerca de si los sujetos tienen adquisiciones básicas correspondientes con su desarrollo curricular, o bien presenta disfunciones.

Usualmente estas pruebas referidas al currículum pretenden abarcar Dominios amplios, correspondientes a dimensiones que son identificables a través de un Diseño Curricular de largo alcance (como por ejemplo, Numeración, Cálculo Mental, Interpretación de Datos o Resolución de Problemas). Sin embargo, en el análisis del Dominio prevalece la identificación de los elementos clave que asocian a diferentes etapas de adquisición. Esta identificación es la que permite situar al sujeto en un nivel de aprendizaje en el Dominio.

Adicionalmente, las pruebas diagnósticas propiamente dichas tienen como finalidad no solo determinar la posición del sujeto en el Dominio de referencia, sino describir adecuadamente los elementos deficitarios con el fin de planificar la intervención. En este sentido, las unidades del Dominio deben estar claramente definidas y previamente analizadas respecto a su asociación con unidades de intervención.

Por su parte, las pruebas Diagnósticas propiamente dichas tienen como finalidad no solo determinar la posición del sujeto en el Dominio de referencia, sino describir adecuadamente los elementos deficitarios con el fin de planificar la intervención. En este sentido, las unidades del Dominio deben estar claramente definidas y previamente analizadas respecto a su asociación con unidades de intervención. Es habitual que en estas situaciones se parta de un marco teórico que define dimensiones respecto de las cuales se puede concentrar la información significativa para la toma de decisiones, aunque no siempre tenga la consistencia teórica-metodológica debida.

Como ha señalado Oosterhof (1994), las pruebas de propósito diagnóstico deben ser utilizadas con cautela, pues la investigación básica acerca de los constructos en ellas implicados, todavía es escasa. En relación directa con esto, el objetivo de la evaluación con estas pruebas es mixto, en el sentido que pueden estar orientadas al proceso educativo, al producto o cualquier combinación de ambos objetivos. Esto conlleva que las pruebas estén basadas en unas reglas de conexión con el universo tan específicas como lo permitan las características y la amplitud del mismo. Finalmente, el estándar en que se basa la decisión suele ser una combinación de indicadores absolutos y normativos.

Esto último se justifica porque para la determinación de la existencia de un problema suele ser tan útil emplear definiciones absolutas que reflejen las claves de su identificación (nivel en que se produce un problema, patrón procesual del mismo entre otros) como la información relativa al grupo de pertenencia para situar la dimensión característica del mismo, por ejemplo, un problema de inversiones en la lectura dependerá tanto de una determinada frecuencia concreta como de la situación relativa dependiendo del grupo de edad al que pertenece el sujeto.

2.2.3 Pruebas de Ubicación

Cuando las pruebas estandarizadas se utilizan para ubicación se presentan algunas situaciones adicionales que deben analizarse. Es posible que el sistema de selección sea único para todos los centros que constituyen una universidad. En este caso los alumnos una vez admitidos pueden elegir la Facultad o Centro de enseñanza que prefieran entre los que constituyen la universidad con tal de que no sobrepasen el cupo establecido. Si esto ocurre son preferidos los candidatos que han recibido mejor puntuación en la prueba de selección.

Sin embargo, ocurre que muchas universidades utilizan procedimientos selectivos que en todo caso o en parte son distintos para las distintas Facultades o Centros de enseñanza. La parte común, si existe, consiste como en el caso anterior en una valoración de las calificaciones de la enseñanza media (sólo pueden aspirar a ingresar en la universidad los que han alcanzado ciertas calificaciones) o en un examen de ingreso común a toda la universidad. La parte específica propia de cada Facultad puede consistir en determinada ponderación de las calificaciones recibidas en Educación Media (por ejemplo la Facultad de Ciencias suele dar más importancia a las calificaciones recibidas en las asignaturas de ciencias). O puede consistir, como es muy frecuente, en un examen distinto en cada Facultad o para cada grupo de Facultades.

Para estos casos en los cuales el proceso de selección es distinto según las Facultades, hay que reglamentar el procedimiento por el que los que han sido admitidos para ingresar en una Facultad y a los que, sin embargo, se consideran con derecho o con aptitud para ingresar en la universidad, pueden pretender el ingreso en otra Facultad.

Pero este tipo de normativa admite diversas variantes, ninguna de ellas plenamente satisfactoria. Es posible que una vez realizada la primera vuelta del proceso de selección se anuncien las Facultades y Centros los que quedan vacantes para que puedan aspirar a ella los que no han sido aceptados en las Facultades más solicitadas. También puede ocurrir que esta redistribución lo haga la propia universidad que previamente ha pedido a los candidatos que indiquen si, caso de no poder ser admitidos en la Facultad que pretenden, aceptarían el estudiar en otra Facultad o Centro y cuál sería su segunda elección.

2.3 Pruebas estandarizadas de selección/ubicación para Ciencias Sociales

Es a partir de esta clasificación internacional del CINE que en el presente estudio se plantea la necesidad de construir una prueba que evalúe habilidades cognitivas asociadas a las competencias que un egresado tiene que haber desarrollado para estudiar una carrera en específica. En este capítulo, se habla de pruebas referidas a las Ciencias Sociales, porque en las diferentes referencias consultadas, las experiencias en selección universitaria aun no hacen pruebas de disciplinas individualizadas, sino en conjunto, como Ciencias Sociales.

Un ejemplo de las habilidades cognitivas que se evalúan en las pruebas universitarias de selección para Ciencias Sociales es la aplicada en Chile. La siguiente tabla desglosa las

habilidades cognitivas en relación con contenidos particulares de este Campo de la Educación.

Tabla 1

Niveles cognitivos evaluados en PSU Historia, Geografía y Ciencias Sociales

Comprensión: entender hechos, procesos e ideas. Junto con comprender la información y aprehender su significado, implica también trasladar el conocimiento a contextos nuevos.

Aplicación: resolver o solucionar problemas empleando el conocimiento adquirido, hechos, técnicas y reglas de manera diferente, es decir, implica utilizar la información aprendida, conceptos, métodos y principios, a situaciones concretas.

Análisis, Síntesis y Evaluación: son habilidades cognitivas de nivel superior e implican examinar y fragmentar la información en diferentes partes, realizar inferencias y encontrar evidencias que apoyen generalizaciones, reunir información y relacionarla de manera diferente combinando elementos y exponer y sustentar opiniones realizando juicios sobre distinto tipo de informaciones.

Nota: Tomada de del Departamento de Evaluación, Medición y Registro Educacional (DEMRE), de la Universidad de Chile, 2016.

Un ejemplo de pruebas específicas de Ciencias Sociales es la aplicada en Chile, denominada Prueba de Historia, Geografía y Ciencias Sociales del Departamento de Evaluación, Medición y Registro Educacional (DEMRE), de la Universidad de Chile. Esta prueba se elabora sobre la base del currículum vigente, en función de la noción de referencia curricular. Por lo tanto, las preguntas que componen el test, combinan los Contenidos Mínimos Obligatorios (CMO) con habilidades de razonamiento asociadas a dicho currículum (Objetivos Fundamentales). La prueba está compuesta por 80 preguntas, de las cuales 5 corresponden a preguntas que se están piloteando en cada aplicación, por lo que no están incluidas en el cálculo del puntaje (DEMRE, 2016^a).

La prueba se orienta a evaluar habilidades cognitivas y ejes temáticos. Se orienta a una evaluación de la comprensión de diversos procesos geográficos, económicos, sociales, políticos y culturales, la integración disciplinaria, el trabajo con fuentes, el análisis de documentos, la contrastación historiográfica, como también a la realización de inferencias,

interpolaciones y síntesis de información disciplinaria. Asimismo, se evalúan las habilidades cognitivas de comprensión, aplicación y análisis, síntesis y evaluación.

El mayor peso, dentro de los ejes temáticos, lo tienen las dimensiones relacionadas con el estudio de la historia de Chile por sobre ámbitos geográficos, económicos, cívicos y de la historia mundial. Ello es coherente con los CMO del Ajuste Curricular de la Enseñanza Media utilizado referencialmente para la elaboración de la PSU, toda vez que si bien existe en el currículum de la Enseñanza Media una integración disciplinaria entre diversas temáticas, es clara la mayor presencia de contenidos que apuntan explícitamente al estudio histórico nacional de Chile.

Otro ejemplo lo constituye la prueba de la Universidad Complutense de Madrid. En relación con la estructura de la prueba y programa de las materias se aplica una fase general que comprende tres grandes ámbitos, Comentario de Texto, Lengua Castellana y Lengua Extranjera (Alemán, Francés, Inglés, Italiano y Portugués), este último se elige en el momento de realizar la inscripción. Para la fase específica orientada el estudiante deberá examinarse en dos materias vinculadas a la opción u opciones de su elección. La opción u opciones elegidas tendrán carácter preferente para su admisión y entre ellas se presenta la de Ciencias Sociales y Jurídicas. Para esta opción la prueba incluye contenidos denominados Economía de Empresa, Geografía de España, Historia de España, Historia de la Filosofía y Matemáticas aplicadas a las Ciencias Sociales II.

Otro ejemplo de institución universitaria que aplica pruebas específicas para Ciencias Sociales es la Benemérita Universidad Autónoma de Puebla (BUAP) que implementa una prueba del College Board, la cual es una organización internacional que desarrolla programas para llevar a cabo procesos de evaluación y admisión universitaria. Actualmente, el College Board sirve a destacadas instituciones educativas localizadas en once países de América Latina. Los exámenes utilizados en esta universidad para este proceso de selección son: el PAA y PAC. La Prueba de Aptitud Académica (PAA) es una prueba que evalúa las habilidades y los conocimientos requeridos para comenzar los estudios universitarios. Este instrumento se encarga de predecir, junto con otros criterios, el éxito en el primer año de estudios superiores. Esta prueba tiene una duración de tres horas y cinco minutos, y consiste en 125 ejercicios.

Por su parte la Prueba por Área de Conocimiento (PAC), como su nombre lo indica, va enfocada dependiendo del área de la carrera a la cual deseas ingresar y por lo tanto son cuatro exámenes con estructuras distintas. Para la Prueba de conocimiento denominada Ciencias Sociales y Humanidades se incluyen 90 ejercicios de selección múltiple con las siguientes materias: Ciencias Sociales, Historia de México, Historia Contemporánea de México y Humanidades.

2.3.1 Los Campos de la Educación y Capacitación CINE

Debe tenerse presente que la Clasificación Internacional Normalizada de la Educación (CINE) es un marco de referencia para recopilar, compilar y analizar estadísticas comparables a nivel internacional en el ámbito de la educación. La CINE forma parte de la familia internacional de Clasificaciones Económicas y Sociales de las Naciones Unidas y es una clasificación de referencia que permite ordenar los programas educativos y sus respectivas certificaciones por niveles de educación y campos de educación. Desarrollada por primera vez a mediados de la década de 1970 por la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO), la clasificación CINE fue revisada en dos ocasiones; la más reciente, en el año 2011. La clasificación CINE es producto de un acuerdo internacional y la Conferencia General de los Estados Miembros de la UNESCO la adoptó formalmente.

En este largo proceso, la última revisión realizada en el año 2011, CINE 2011, se concentró principalmente en cambios a los niveles educativos de los programas (CINE-P) e introdujo, por primera vez, una clasificación de niveles de logro educativo sobre la base de certificaciones (CINE-A). Durante el proceso de análisis que llevó a la revisión del año 2011, se decidió que los campos de la educación debían ser examinados en un proceso separado con el fin de establecer una clasificación independiente pero relacionada que podría ser actualizada siguiendo una frecuencia distinta, si fuera pertinente, de cualquier revisión futura de los niveles de educación y logro educativo.

En este sentido se puede plantear que la clasificación de los niveles y campos forma parte de la misma familia de clasificaciones. En consecuencia, se denomina a esta nueva clasificación como la clasificación CINE de campos de educación y capacitación (CINE-F). Las unidades básicas de clasificación de la CINE 2011 son los programas educativos y las

respectivas certificaciones. Estas son las mismas unidades de clasificación en los Campos de Educación y Capacitación de la CINE (CINE-F).

a. Un programa educativo es una ‘secuencia coherente de actividades educativas diseñadas y organizadas para lograr objetivos predeterminados de aprendizaje o para llevar a cabo un conjunto específico de tareas educativas a lo largo de un periodo sostenido de tiempo’. Las actividades educativas son ‘actividades intencionadas que implican alguna modalidad de comunicación destinada a producir aprendizaje’.

b. Una certificación es la ‘confirmación oficial de la conclusión de un programa educativo, que por lo general se oficializa mediante un documento. Los créditos concedidos por completar satisfactoriamente cursos individuales (por ejemplo: módulos o asignaturas) no se consideran certificaciones dentro de la CINE. En dichos casos, un número suficiente de créditos o signaturas que sea equivalente en duración y/o cubra el currículo de un programa completo representaría una certificación.

Los Campos de Educación y Capacitación de la CINE clasifican a los programas de educación y las certificaciones que ellos otorgan por campos de estudio: Un campo es la ‘esfera amplia, la rama o el área de contenido cubierto por un programa de educación o por una certificación’.

Los campos de educación y capacitación, y los niveles de educación o logro educativo son variables de clasificación cruzada dentro de la CINE y, por lo tanto, son independientes la una de la otra. A diferencia de la clasificación original CINE (CINE 1976), los campos no se definen dentro de los niveles de educación. Esto se debe a que, aunque en algunos niveles de educación los programas y las certificaciones están únicamente disponibles en algunos campos de educación y capacitación, ello puede variar entre países o con el transcurso del tiempo. Por ejemplo, se ofrece formación docente o enfermería en diferentes niveles de la educación en distintos países. En los últimos años ha habido una tendencia a eliminar estos programas en los niveles más bajos de la educación y la formación para ser ascendidos a un nivel de CINE más alto.

Alcance y estructura de la clasificación de la CINE. La CINE tiene una alta valoración a nivel internacional. Esta clasificación ha sido diseñada principalmente para describir y categorizar los campos de educación y capacitación en los niveles secundario,

postsecundario y terciario de la educación formal, tal como se definen en la CINE 2011, aunque se puede usar para la clasificación de los programas y certificaciones que se ofrecen en otros niveles. La clasificación también puede ser utilizada en otros contextos, por ejemplo, para clasificar el contenido de la educación no formal, la formación vocacional inicial y continua o el aprendizaje informal.

Tabla 2

Clasificación del CINE

Nivel	Categorías de campos	Números de campos
1er nivel	Campo amplio	11
2do nivel	Campo específico	29
3er nivel	Campo detallado	80

Nota: Tomada de CINE, 2011.

El proceso de revisiones ha sido gradual y acumulativo de manera que se basa en las versiones anteriores de la clasificación con el fin de garantizar la comparabilidad en el tiempo. Se ha diseñado como una jerarquía de tres niveles entre campos amplios (el nivel más alto), campos específicos (el segundo nivel) y los campos detallados (el tercer nivel), y utiliza un esquema de codificación de cuatro dígitos. Hay 11 campos amplios, 29 campos específicos y alrededor de 80 campos detallados de educación y capacitación.

Al respecto debe considerarse que los campos detallados (el tercer nivel jerárquico de la clasificación) están destinados principalmente para su uso en el nivel terciario de la educación y para los programas de educación vocacional y capacitación, y las certificaciones, en los niveles secundarios y postsecundarios no terciarios. La clasificación puede utilizarse también para los programas y las certificaciones de educación general en los que haya una especialización de contenidos. Adicionalmente, los programas de educación general y certificaciones que cubren una amplia gama de asignaturas con poca o ninguna especialización en un determinado campo o campos, normalmente se clasifican en el campo amplio 00 "Programas y certificaciones genéricos".

Justificación de la clasificación. La clasificación de los campos de educación y capacitación sigue un enfoque de contenido temático. Este es el mismo método que se utiliza en las versiones anteriores de la CINE (1976, 1997 y 2011). Cuando el contenido está estrechamente relacionado, las asignaturas se agrupan para formar campos amplios, específicos y detallados de clasificación, sobre la base de la similitud de los contenidos temáticos. El objetivo es clasificar los programas de educación y sus respectivas certificaciones por campos de educación y capacitación sobre la base del contenido del programa y no según las características de los participantes previstos.

El **contenido temático principal** determina el campo de educación y capacitación en el que se debe clasificar un programa o una certificación. El contenido temático es el conocimiento fáctico, práctico y teórico impartido durante el programa y que se reconoce mediante la respectiva certificación. Este conocimiento se aplica a determinados tipos de problemas o para fines específicos que pueden ser de naturaleza abstracta (por ejemplo, filosofía), naturaleza práctica (por ejemplo, ingeniería) o de ambas naturalezas (por ejemplo, arquitectura).

A efectos prácticos, el contenido temático principal de un programa o certificación se determina por el campo detallado en el que se cursa la mayor parte (más del 50%) o la parte claramente predominante de los créditos de aprendizaje o del tiempo de aprendizaje previsto para los alumnos. Para el cálculo, los créditos de aprendizaje deberán utilizarse donde estén disponibles. Si esta información no está disponible, deberá realizarse una evaluación aproximada del tiempo de aprendizaje previsto. El tiempo de aprendizaje incluye el tiempo invertido en exposiciones orales y seminarios, como también en laboratorios o proyectos especiales. El tiempo de estudio privado está excluido (ya que es difícil de medir y varía según los estudiantes). Los programas y las certificaciones se clasifican en el campo detallado que contiene su temática principal.

2.3.2 Ámbito de acción de las Ciencias Sociales: Antropología y la Investigación Social

En este estudio particular interesan las disciplinas de las ciencias sociales que la UNAH ofrece como campos especializados del saber, es decir la antropología, la historia, la sociología y las ciencias psicológicas. A continuación se presenta una breve caracterización

de la Antropología y la Investigación Social ya que son los ámbitos de acción de interés del estudio.

La Antropología. Actualmente se le considera una ciencia reciente, pero sus orígenes se remontan desde la antigüedad con “el padre de la historia”, Heródoto, analizó y observó las formas de vida e informó las diferencias existentes. Él se interesó por el matriarcado y las prácticas religiosas. Tácito también contribuyó, con estudios, al describir el carácter, las costumbres y la distribución de la geografía de los pueblos germánicos, que para esa época se consideraban muy diferentes al pueblo romano. En la Edad Media, aproximadamente por el siglo XIII Marco Polo aportó algunas descripciones de las costumbres del lejano oriente, así como la forma ideológica de los europeos. Y para los siglos XV Y XVI, uno de los precursores antropológicos fue fray Bernardino de Sahagún (siglo XVI), nacido en el Reino de León, España; que hacia 1540 recogiera material para su obra magna “historia general de las cosas de Nueva España”; su método consistía en conocer el lenguaje de esta gente mexicana (Corbetta, 2010).

Se considera que la Antropología se convirtió en una ciencia social a partir de las ideas pregonadas en el siglo XVIII, el siglo de las luces o la ilustración. Ideas que se interesaron en explicar las variaciones raciales, las formas culturales, la multiplicidad del lenguaje y las manifestaciones religiosas. De manera que sus inicios se remontan a finales del siglo XVIII y principios del XIX. El contexto en la que nace la Antropología, es en Europa desarrollada económicamente, gracias a la Revolución Industrial y al comercio (Corbetta, 2010).

La bibliografía identifica como los fundadores de la ciencia antropológica al norteamericano Lewis Morgan y el inglés Edward B. Tylor. Cabe citar que ambos pertenecen a la corriente antropológica evolucionista. Morgan influyó y se dejó sentir en Federico Engels en su obra “el origen de la familia”, “la propiedad privada” y “el Estado”.

El término tiene origen en el idioma griego y proviene de *anthropos* (“hombre” o “humano”) y *logos* (“conocimiento”). Reunidos estos términos se concluye que la Antropología significa ciencia que estudia lo relativo al hombre etimológicamente, o el ser humano propiamente dicho. Aunque no se debe olvidar que existen otras ciencias que también tienen este objeto de estudio, tales como: a biología humana, la ética, el derecho, la

medicina y otras. La Antropología es relativamente reciente, se puede decir que la sociología es el antecedente de ésta disciplina. Por lo tanto la antropología estudia al ser humano en su evolución biológica y cultural.

Investigación social. En el campo de las Ciencias Sociales los paradigmas que han imperado desde sus comienzos han sido el positivismo y el interpretativo. Entre los elementos que devienen del paradigma del Positivismo se distinguen el monismo metodológico, al cual hace referencia que el mismo método para la Ciencias Naturales y se puede utilizar para las Ciencias Sociales; la formulación de leyes o generalizaciones de las ciencias naturales se pueden adaptar al campo de las ciencias sociales. Estos elementos también tienen su raíz en la noción durkheimiana de la Sociología que consideraba como objeto de estudio a los hechos sociales analizados objetivamente y de modo similar a las ciencias naturales, es decir en tanto cosas. En el curso del siglo XX la inspiración positivista sufrió ciertas transformaciones para superar sus propias limitaciones (Corbetta, 2010).

El paradigma interpretativo muchos lo han considerado como una alternativa a las limitaciones del paradigma positivista en el campo de las Ciencias Sociales, porque considera que las Ciencias Sociales tienen características diferentes a las Ciencias Naturales. Este paradigma tiene sus antecedentes históricos en la fenomenología, el interaccionismo simbólico interpretativo, la etnografía, la antropología, entre otros. Este paradigma nace producto de la escuela alemana y se considera a Husserl su fundador. Entre sus referentes principales destacan: Dilthey, Baden, Berger, Shutz, Mead, Blumer, Lukman, etc. Ese enfoque cualitativo, define al paradigma interpretativo como el modelo con el cual se busca profundizar en la investigación, planteando diseños abiertos y emergentes desde la globalidad y contextualización (Corbetta, 2010).

Estos dos planteamientos referenciales en el ámbito de la investigación de las Ciencias Sociales: positivista e interpretativo. Es innegable que a lo largo de la historia estos dos primeros enfoques han definido una dicotomía metodológica: cuantitativa / cualitativa proveniente de dos modos diferentes de concebir, afrontar e interpretar los fenómenos y hechos sociales.

2.3.3 Situación de las Ciencias Sociales en Educación Superior

En la segunda década del siglo XXI, las ciencias sociales se enseñan en la gran mayoría de las universidades. El número de estudiantes, conferencistas, profesores e investigadores en las ciencias sociales se ha incrementado rápidamente, así como el número de libros y artículos escritos en diferentes idiomas. Como resultado de esta producción, un gran número de científicos sociales no sólo trabaja como académicos e investigadores, sino también como expertos en administración pública nacional, son asesores de los gobiernos y a veces conducen el desarrollo de sus economías.

El desarrollo de las tecnologías de la información ha ayudado a los científicos sociales a comunicarse más a menudo y rápidamente, tanto entre ellos, como con la sociedad civil. En las primeras décadas del siglo XXI, el conocimiento experto de las ciencias sociales es ampliamente demandado por parte de hacedores de políticas, medios de comunicación y público en general (UNESCO, 2010).

Actualmente, se espera que los científicos sociales tengan habilidades y conocimientos que son necesarios para identificar, analizar y descifrar tanto las estructuras y cambios en la sociedad, como las tendencias del cambio futuro. Se expresan grandes expectativas respecto del conocimiento y la experiencia de las ciencias sociales cuando se busca resolver retos como pobreza, cambio climático y crisis alimentaria, por nombrar sólo algunos. No obstante, el éxito y el crecimiento de las ciencias sociales también han evidenciado sus limitaciones.

Limitaciones que no son pequeñas en su significado. Los economistas son a menudo señalados por haberse engolosinado con modelos sofisticados y abstractos y perder de vista la realidad social. Demasiado confiados en el valor del mercado, no levantaron advertencias en contra de las malas prácticas financieras y no previeron, y mucho menos previnieron, las mayores crisis económicas y financieras de la presente era globalizada (UNESCO, 2010).

A los profesionales de las Ciencias Políticas por su parte, se les ha acusado de no anticipar los cambios profundos en la opinión pública, de no predecir adecuadamente los resultados electorales, o de tener opiniones comprometidas con la industria de las encuestas. Los sociólogos son culpados por su dificultad para identificar las tendencias sociales importantes, o por la lentitud con la que lo hacen (UNESCO, 2010).

Desde una visión panorámica se puede señalar que las ciencias sociales han atravesado, por décadas, una crisis de reconocimiento y por amplios debates epistemológicos. Mientras los tomadores de decisiones y la sociedad en general requerirían de una mayor aportación de las ciencias sociales para resolver problemas locales y globales, algunos científicos sociales prefieren el análisis distanciado y la observación crítica, y evitan involucrarse en la acción.

Adicionalmente, en la actualidad a los expertos de las ciencias sociales se les ha acusado de sobre especializarse, de desarrollar modelos teóricos y de dirigirse sólo al medio académico. A otros se les acusa de ser locales y no producir suficiente teoría con lo cual pierden relevancia global. Por muchos años estas tensiones han alimentado los debates entre científicos sociales, pero recientemente se han agudizado como consecuencia de los cambios en el contexto general en el que las ciencias sociales se desarrollan (UNESCO, 2010).

2.3.4 Pruebas basadas en expectativas de logro asociadas a contenidos específicos de las Ciencias Sociales

Es a partir del CINE que en el presente estudio se plantea la necesidad de construir una prueba que evalúe habilidades cognitivas asociadas a las Expectativas de Logro que un egresado de Educación Media debe haber cumplido en Antropología y Fundamentos de Investigación Social.

En las primeras décadas del siglo XXI se ha cambiado el foco del proceso educativo, destacándose los Resultados de Aprendizaje como punto de referencia para la vanguardia de la innovación educativa. Estos representan el cambio en el foco de la formación desde uno centrado en la "enseñanza" a uno centrado en el "aprendizaje". Esta tipificación hoy se conoce como la adopción de un enfoque centrado en el estudiante, en contraste con la tradición "centrado en el profesor" (Calderón Patier & Barruso Castillo, 2009; García & Gómez, 2010; Gómez & Fernando, 2010; Iraeta, Isabel, Sánchez, & others, 2010; Kember, 2009; Means, Learning, Policy, & Service, 2009; Reichert, 2010; Wojtczak, 2010; Zubizarreta, 2009).

En este sentido, el cambio de foco implica que el proceso se focalice en la integración entre "la enseñanza-aprendizaje-evaluación y los vínculos fundamentales entre el diseño, ejecución y medición del aprendizaje" (Adam, 2007). En otras palabras, son los Resultados

de Aprendizaje los que permiten llevar adelante este proceso integrativo al interior del proceso formativo.

Sin embargo, debe recordarse que el enfoque del aprendizaje basado en logro o resultados, es de larga data. Ya con los objetivos educativos de Tyler en 1950, se identificaron relevantes desafíos en el desarrollo y la planificación de la instrucción, tanto en relación con los propósitos, contenidos, organización y evaluación (Wiburg, 1995). A su juicio, los objetivos son esenciales para la planificación sistemática y la identificación del comportamiento requerido en el estudiante luego de la instrucción, así como los contenidos y el contexto de aplicación. Este enfoque de diseño curricular ha influído en la enseñanza durante varias décadas y la filosofía básica del enfoque basado en resultados está arraigada desde mediados del siglo pasado (Malan, 2000). Tyler enumeró cuatro cuestiones básicas de su medio-fin o razón de ser orientadas a los productos para el diseño curricular (Flinders & Thornton, 2004):

¿Qué Expectativas de Logro deseo aspirar en el centro educativo?

¿Cómo seleccionar esas estrategias de enseñanza-aprendizaje que son pertinentes para ser utilizadas en la consecución de esas expectativas?

¿Cómo deben ser las experiencias de aprendizaje organizadas para una instrucción efectiva?

¿Cómo afectaran las estrategias de enseñanza-aprendizaje a los resultados?

Adicionalmente, se plantea que los objetivos deben contemplar dos aspectos, uno de comportamiento y otro de contenido (Tyler, 1949). En la misma década de los cincuenta, Bloom con su taxonomía, ayudó a determinar si los objetivos instruccionales eran comparables con los resultados del aprendizaje, a nivel de la educación universitaria, los primeros destinatarios de su trabajo (Booker, 2007). Una de sus premisas se basó en asumir que suficientes oportunidades y el apoyo en un entorno de aprendizaje apropiado, provocaría que la mayoría de los estudiantes deberían tener éxito en sus tareas de aprendizaje.

En esta misma línea de análisis, en 1963, Glaser desarrolla el tema de la medición basada en el criterio, localizando al estudiante en un continuo que va desde "no aptitud" a

"rendimiento perfecto". Para Él, el objetivo de la instrucción y la evaluación se basa en la consecución de resultados específicos, a partir de pruebas para la competencia en términos del criterio establecido (Glaser, 1967, 1968; Glaser & Cox, 1968), comparando un resultado de aprendizaje o dominio de las competencias con un patrón externo predeterminado (Glaser & Cox, 1968). En esta perspectiva, el éxito se mide por la demostración de los logros, y cuando sea necesario durante el proceso, la aplicación de acciones correctivas.

Unos años antes, en el año de 1962, Mayer, planteó cambiar el término objetivos “educacionales” a “instruccionales”, vinculando el logro de los objetivos pre- especificados con resultado directo de la enseñanza que recibe un estudiante (Mager, 1962). Se plantea que el objetivo debe establecer lo que el estudiante "será capaz de hacer" después de la experiencia de aprendizaje que no podía hacer antes. En consecuencia, de acuerdo con Mager (1962), los objetivos instruccionales deben considerar: ser inequívocos, ya que deben estar compuestos por verbos que describen el desempeño, las acciones observables que el alumno se espera que haga; tener las condiciones importantes en las que el rendimiento se produzca con claridad, e indicar el “criterio de desempeño aceptable, describiendo qué tan bien el estudiante debe realizar para ser considerado aceptable”.

En esta discusión, Eisner(1979) planteó que existe consenso claro entre Mager y Tyler, y transversalmente con Gagne, Bobbitt, y Bloom, en el sentido que los objetivos de instrucción deben considerar: la descripción del comportamiento del estudiante, no el comportamiento de los profesores; el describir tanto el comportamiento que se muestra y el contenido en el que el comportamiento es que se produzca; es decir, sus condiciones; y deben señalar, en un nivel de especificidad que lo hace posible de reconocer el comportamiento en caso de que se muestre.

Esta perspectiva de estudio que se centra en logros o Resultados de Aprendizaje fue crecientemente aceptado en el ámbito de las autoridades educativas en diversos países como por ejemplo, en el Reino Unido por la Agencia para asegurar la calidad en la educación superior (Quality Assurance Agency for Higher Education - QAA), y por las Autoridades para la Titulación (Qualification Authorities) en Australia, Nueva Zelanda y Sudáfrica, entre otros (Gosling y Moon, 2002)

Pero no hay consenso en todos los aspectos pues existen diversidad de enfoques y énfasis en torno a la definición de los Resultados de Aprendizaje. Algunos de ellos se examinan a continuación. Los Resultados de Aprendizaje son enunciados acerca de lo que se espera que el estudiante sea capaz de hacer como resultado de una actividad de aprendizaje (Jenkins y Unwin, 1996). Un resultado de aprendizaje es una declaración escrita de lo que se espera que el estudiante sea capaz de hacer en el final del módulo, curso de unidad, o la calificación (Adam, 2007).

En este tipo de análisis se considera que los Resultados de Aprendizaje son enunciados que especifican lo que el aprendiz va a saber o lo que él será capaz de hacer como resultado de una actividad de aprendizaje (American Association of Law Libraries, s.d.). Los estudiantes aprenden “Resultados de Aprendizaje” que están bien definidos en términos de conocimientos, destrezas y habilidades logrados por el aprendiz al final del proceso por su participación en un conjunto particular de experiencias educativas de nivel superior (The Council for Higher Education Accreditation, 2006). Los Resultados de Aprendizaje son una descripción explícita acerca de lo que un estudiante debe saber, comprender y ser capaz de hacer como resultado del aprendizaje (Kennedy, Hyland, & Ryan, 2007).

En esta perspectiva de análisis se considera que los resultados de aprendizaje son enunciados acerca de lo que se espera que un aprendiz deba saber, comprender y / o ser capaz de demostrar una vez finalizado el proceso de aprendizaje (Users’Guide, 2005), en términos similares se pronuncia Gosling & Moon (2002). Estas definiciones tienen en común dos elementos identificables. Por una parte, el estar centradas en lo que el estudiante puede transferir de lo que ha aprendido integradamente; y por otra, que es “condición necesaria” el demostrar su logro al término de una actividad o proceso de aprendizaje.

Resumiendo acerca de las fortalezas de esta perspectiva de análisis se pueden plantear algunas interrogantes tales como ¿Cuáles son los principales atributos de la formación basada en Resultados de Aprendizaje? Según Malan (2000), las respuestas podrían sintetizarse en las siguientes características:

Está impulsado por las necesidades. Los planes de estudios están diseñados en términos de los conocimientos, habilidades y actitudes esperados en los graduados y tienen por objeto dotar a los estudiantes para un aprendizaje permanente o a lo largo de la vida.

Es impulsado por los resultados. El modelo tiene una línea que considera las necesidades de formación para establecer un propósito para el programa, las metas, los resultados de Aprendizaje y, finalmente, el cómo evaluar esos resultados.

Tiene un enfoque de diseño descendente. Relaciona las necesidades y la finalidad del programa, estableciendo los contenidos luego que se han especificado los resultados de Aprendizaje. Los contenidos se convierten en un vehículo para alcanzar los resultados de Aprendizaje.

En él se especifican los resultados y sus niveles. Los objetivos de aprendizaje se describen a partir de una taxonomía de aprendizaje y de acuerdo a las directrices para la formulación de objetivos de Mayer.

La atención se desplaza desde la enseñanza al aprendizaje. El modelo tiene un enfoque de aprendizaje centrado en el estudiante donde los profesores actúan como facilitadores. Guías de estudio para ayudar a los alumnos a organizar sus actividades de aprendizaje y trabajo en grupo. La evaluación continua y autoevaluación son las principales herramientas.

El marco es holístico en sus resultados. Los Resultados de Aprendizaje están vinculados a metas y objetivos en niveles superiores. El logro de aprendizaje, no es un fin en sí mismo, sino que provee de bloques de construcción complejos para lograr un mayor nivel dominio e integración con otros saberes.

Para Spady (1994), un aprendizaje basado en contenidos y otro basado en resultados, cuentan con diferencias significativas. Estas se pueden apreciar en la siguiente (Tabla 3).

Tabla 3*Diferencias entre el Aprendizaje Basado en contenidos y en Resultados*

Basado en Contenidos	Categoría	Basado en Resultados
Basado en la adquisición de contenidos.	Sistema de Aprendizaje	Desarrollo de Resultados de Aprendizaje.
Estudiantes Pasivos, receptores frente al proceso.	Rol del Estudiante	Los aprendices activos, sobre el proceso y los propios resultados.
Por medio de exámenes parciales o finales desintegrados.	La Evaluación	La evaluación continua e integrada.
Aprendizaje memorístico.	Tipo de Aprendizaje que privilegia	El pensamiento crítico, el razonamiento, la reflexión, la movilización y la acción
Temas y contenidos aislados.	Qué privilegia el proceso de Aprendizaje	Integración del conocimiento, del aprendizaje pertinentes /
Basado en Contenidos	Categoría	Basado en Resultados
		situaciones de la vida conectada

El Libro de texto, hoja de cálculo y el docente.	Recursos de Aprendizaje	Centrado en la interacción entre educando y educador / facilitador de grupo, y el uso del trabajo en equipo.
Es rígido y no permite que los docentes se involucren no pudiendo ajustar ni negociar.	Programa de Curso	Visto como guías que permiten a los educadores a ser innovadores y creativos en el diseño de los programas y actividades.
Los profesores y formadores son exclusivos responsables del aprendizaje, dependen de las propias motivaciones.	Responsabilidad del proceso de enseñanza y aprendizaje	Los estudiantes asumen la responsabilidad de su aprendizaje. Los estudiantes se motivan por la retroalimentación constante a partir de juicios de valor preestablecidos.
Subraya lo que el docente espera lograr en sus estudiantes en términos de intención formativa.	Énfasis del proceso Formativo	El énfasis en los que el estudiante debe evidenciar como Resultados de Aprendizaje, el cual comprende y se vincula.
Los tiempos son más rígidos en los cuales se deben cubrir con	Plazos Formativos	Considera plazos flexibles, en donde cada estudiante trabaja

Basado en Contenidos	Categoría	Basado en Resultados
Contenidos.		Considerando sus propios ritmos.
Estancia durante el proceso de formativo en una sola institución, hasta terminar.	Movilidad	Los estudiantes obtienen créditos de diferentes instituciones, hasta alcanzar la certificación.
El conocimiento previo y experiencia en el aprendizaje de campo, es ignorado. Cada vez que asiste a curso, es de manera completa.	Los Aprendizajes Previos	Reconocimiento de aprendizajes previos: después de una preevaluación. Los estudiantes abonarán los resultados demostrados o créditos, transfiriéndolos, a otro lugar

Nota: Tomada de Spady, 1994

2.3.5 Taxonomía de objetivos de educación de Bloom

Debe recordarse que el concepto taxonomía implica una clasificación, categorización u ordenamiento, que para el caso específico que se aborda en el presente estudio, la educación y el aprendizaje, incluye caracterizar cómo se construye a partir de lo conocido; reflejando los diferentes niveles de complejidad de los aprendizajes a desarrollar por parte de los docentes.

La taxonomía de los objetivos de la educación, más conocida como la taxonomía de Bloom, consiste en un esfuerzo por lograr una clasificación de los diferentes objetivos y habilidades que los docentes tratan de desarrollar en sus alumnos. La bibliografía especializada señala que esta idea surgió en una reunión de la Asociación Americana de Psicología en el año de 1948, como un intento de aclarar conceptos y poder facilitar la discusión académica e intercambio de ideas en el colectivo de profesionales. Se formó una comisión para establecer esta clasificación y la misma estuvo dirigida por Benjamin Bloom, psicólogo de la educación de la Universidad de Chicago.

Como producto del trabajo de esta comisión se tuvo una clasificación que fue propuesta por Bloom en el año 1956, la cual incluía tres "dominios": cognitivo, afectivo y psicomotor, aun se desarrollaron en esa época solamente los dos primeros. Esta taxonomía propuesta por Bloom tiene la cualidad de ser jerárquica, es decir, asume que el aprendizaje a niveles superiores depende de la adquisición del conocimiento y habilidades de ciertos niveles inferiores. Otra característica importante, esta clasificación muestra una visión global del proceso educativo, promoviendo una forma de educación a partir de esta visión global. Hay tres dimensiones en la taxonomía de objetivos de la educación propuesta por Benjamin Bloom: Dimensión afectiva, Dimensión psicomotora y Dimensión cognitiva

En la dimensión cognitiva que interesa para el presente estudio, esta se define como la habilidad para pensar las cosas. Los objetivos cognitivos giran en torno del conocimiento y la comprensión de cualquier tema dado. Hay seis niveles en la taxonomía.

Tabla 4

Niveles cognitivos de la Taxonomía de Bloom

Conocimiento. Está referido al recuerdo de materiales previamente aprendidos por medio de hechos evocables, términos, conceptos básicos y respuestas. Incluye el Conocimiento de terminología o hechos específicos; Conocimiento de los modos y medios para tratar con convenciones, tendencias y secuencias específicas, clasificaciones y categorías, criterios, metodología, Conocimiento de los universales y abstracciones en un campo: principios y generalizaciones, teorías y estructuras.

Comprensión. Se conceptualiza como el entendimiento demostrativo de hechos e ideas por medio de la organización, la comparación, la traducción, la interpretación, las descripciones y la formulación de ideas principales: Traducción, Interpretación y Extrapolación.

Aplicación. Está definido como el uso de conocimiento nuevo. Resolver problemas en nuevas situaciones aplicando el conocimiento adquirido, hechos, técnicas y reglas en un modo diferente.

Análisis. Este nivel se considera de orden superior y se considera como el examen y discriminación de la información identificando motivos o causas. Hacer inferencias y

encontrar evidencia para fundamentar generalizaciones, incluye el Análisis de los elementos, el Análisis de las relaciones y el Análisis de los principios de organización.

Síntesis. Este nivel de orden superior se define como la compilación de información de diferentes modos combinando elementos en un patrón nuevo o proponiendo soluciones alternativas. Incluye la Elaboración de comunicación unívoca, la Elaboración de un plan o conjunto de operaciones propuestas, así como la Derivación de un conjunto de relaciones abstractas.

Evaluación. En su formulación original es el último nivel de aprendizaje y está referido a la presentación y defensa de opiniones juzgando la información, la validez de ideas o la calidad de una obra en relación con un conjunto de criterios, incluyendo Juicios en términos de evidencia interna y Juicios en términos de criterios externos.

Nota: Tomada de Bloom, 1956.

Por años muchos expertos han revisado la taxonomía de Bloom haciéndole algunos cambios de diferente nivel de profundidad. Una de ellas es La taxonomía de Anderson y Krathwohl de inicios del presente siglo.

“La cual es una secuencia ordenada y lógica que permite al docente una visión compleja de las habilidades que desea alcanzar en sus estudiantes, asociado a acciones concretas que ellos pueden realizar en el aula o fuera de ella, como consecuencia de directrices ad hoc propuestas por el profesor” (C. Bancayán, 2013, p.1).

Anderson y Krathwohl proponen una taxonomía para procesos de enseñanza-aprendizaje y evaluación, diseñada en dos dimensiones: la del conocimiento y la de categorías del conocimiento. Esta última, en una clara analogía con la taxonomía de Bloom, plantea la existencia de seis niveles de conocimiento.

Tabla 5*Revisión de la Taxonomía de Bloom*

Reconocer	Comprender	Aplicar	Analizar	Evaluar	Crear
Reconocer y traer a la memoria información relevante de la memoria de largo plazo.	Habilidad de Construir significado a partir de material educativo, como la lectura o las explicacion es del docente.	Aplicación de un proceso aprendido, ya sea en una situación familiar o en una nueva.	Descomponer el conocimiento en sus partes y pensar en cómo estas se relacionan con su estructura global	Ubicada en la cúspide de la taxonomía original de 1956, evaluar es el quinto proceso en la edición revisada. Consta de comprobación y crítica.	Nuevo en esta Taxonomía. Involucra reunir cosas y hacer algo nuevo. Para llevar a cabo tareas creadoras, los aprendices generan, planifican y producen.

Nota: Tomada de Anderson & Krathwohl, 2001.

La dimensión de los procesos cognitivos de la taxonomía revisada de Bloom (1956), comprende seis habilidades cognitivas, como la versión original. Ellas son, de las más simples a las más complejas: recordar, comprender, aplicar, analizar, evaluar y crear. La revisión citada incorpora cambios nada despreciables que le han añadido actualidad y valor conceptual, entre los que destaca el cambio de sustantivos por verbos. Se incorporó para cada nivel conceptual los verbos y, por ende, las acciones asociadas a esa habilidad cognitiva. Esta modificación en la clasificación debe asumirse como un paso hacia el modelo educativo que fija cada vez más su mirada en el estudiante antes que en el docente.

2.3.6 Las Habilidades cognitivas y las expectativas de logro

Las taxonomías anteriormente descritas hacen referencia a habilidades cognitivas. Las habilidades se relacionan directamente con las aptitudes y el pensamiento, por lo que se requiere trabajar en el estudiante una estimulación adecuada que desarrolle la capacidad de pensar y aprender situando en primer lugar la exigencia de un aprendizaje continuo, o como es también llamado el aprender a aprender (King & Schneider, 1991).

Las habilidades cognitivas son procesos y métodos que puede emplear el estudiante para obtener, retener y recuperar diferentes tipos de conocimientos. El estudiante supone capacidades de representación como: La lectura, imágenes, habla, escritura y dibujo, las capacidades de selección: atención e intención, y las capacidades de autodirección: auto programación y autocontrol (Rigney, 1978, p. 165).

En este sentido, se ha elaborado una clasificación de las habilidades cognitivas dividiéndolas en habilidades cognitivas generales y habilidades cognitivas específicas, según el dominio de conocimiento (Van Lehn, 1996). En relación con las primeras, como su nombre lo indica, tienen un carácter más general, ya que actúan de forma similar a través de los diferentes dominios del conocimiento; mientras que en el caso de las segundas, se van desarrollando en relación con el aprendizaje de contenidos específicos o que han sido estudiadas dentro de un dominio particular (Gilar Corbi, 2003).

Relacionadas con estas habilidades cognitivas se encuentran las expectativas de logro que se plasman en un currículo determinado. “Las expectativas de logro explicitan las intencionalidades educativas y expresan el grado de desarrollo de las competencias del área de tipo cognitivo, procedimental y valorativo/actitudinal que la Educación Media debe garantizar equitativamente a los alumnos y alumnas” (Secretaría de Educación, DCNB Tercer Ciclo, 2003, p.32).

En el contexto de la educación hondureña, para la educación media se han elaborado expectativas de logro por asignatura y nivel, de manera de constituirse en insumo para construir los perfiles de egreso de los estudiantes. En este marco, se considera a la competencia como la facultad de movilizar los saberes (conocimientos, habilidades, actitudes, valores) y a las expectativas de logro como repertorios de saberes (conocimientos, habilidades, destrezas, actitudes, valores, etc.) que constituyen los requerimientos básicos

para cada competencia en cada nivel. De este modo, se elaboraron las expectativas de logro por nivel y por espacio curricular así como las relacionadas con las competencias transversales de Lectura, Escritura y Pensamiento crítico.

2.4 Bases para el desarrollo de las pruebas o Teorías de la Medición

La historia de la psicología está fuertemente asociada a las pruebas tipo test. La bibliografía plantea que los test constituyen seguramente la tecnología más sofisticada de la que disponen los psicólogos para ejercer su profesión, por eso no es infrecuente que la sociedad identifique a los psicólogos con los test. Naturalmente, unos psicólogos utilizan los test más que otros, dependiendo de su campo profesional y de su forma de trabajar. Los test son muestras de conducta que permiten llevar a cabo inferencias relevantes sobre la conducta de las personas. Bien utilizados son herramientas claves en la profesión del psicólogo.

En este análisis debe tenerse presente que los test nacen con un afán de objetividad y justicia, para evaluar a las personas por lo que realmente valen, evitando evaluaciones sesgadas por aspectos tales como la condición económica, la clase social, origen étnico, el sexo, las creencias, las cartas de recomendación, y otros sistemas de evaluación subjetivos. Unas veces estos nobles fines se han alcanzado mejor que otras, pero ésa era y sigue siendo la idea central, evaluar a todos por el mismo estándar.

El desarrollo de los test en la historia y el surgimiento de las Teorías de Medición

Como se señaló en la introducción del presente documento, suele citarse como el origen remoto de los test unas pruebas que los emperadores chinos ya hacían allá por el año 3000 antes de Cristo para evaluar la competencia profesional de los oficiales que iban a entrar a su servicio. Otras muchas huellas antiguas pueden rastrearse, pero los test actuales tienen sus orígenes más cercanos en las pruebas senso-motoras utilizadas por Galton (1822-1911) en su laboratorio antropométrico. Pero será James McKeen Cattell (1860-1944) el primero en utilizar el término test mental, en 1890. Pronto quedó claro (Wissler, 1901) que estos primeros test senso-motores no eran buenos predictores de las capacidades cognoscitivas de las personas.

Pero fueron otros autores, Binet y Simon (1905) los que dieron un cambio fundamental en el tema de los test al introducir en su nueva escala tareas cognoscitivas para

evaluar aspectos como el juicio, la comprensión y el razonamiento. Terman llevó a cabo la revisión de la escala en la Universidad de Stanford, la cual se conoce como la revisión Stanford-Binet (Terman, 1916), utilizando por primera vez el concepto de Cociente Intelectual (CI) para expresar la puntuación de las personas. La idea del CI había sido propuesta originariamente por Stern, dividiendo la Edad mental por la Edad Cronológica y multiplicando el resultado por 100 para evitar decimales.

Estos trabajos marcaron una nueva época en el tema. La escala de Binet abre una tradición de escalas individuales que llega hasta nuestros días. En 1917 los test reciben otro gran impulso al aparecer los test colectivos Alfa y Beta a raíz de la necesidad del ejército estadounidense de reclutar rápidamente soldados para la primera guerra mundial. El test Alfa iba dirigido a la población general y el Beta a las poblaciones analfabetas o que no dominaban el inglés. Las pruebas tuvieron mucho éxito y terminada la guerra las empresas y otras instituciones adoptaron de forma entusiasta el uso de los test para distintos menesteres.

Con estos trabajos dio inicio una expansión creciente en el uso y creación de test de todo tipo. La aparición de la técnica del análisis factorial va a suponer un gran avance en la construcción y análisis de los test, permitiendo la aparición de las baterías de tests, cuyo representante más genuino serían las Aptitudes Mentales Primarias (PMA) de Thurstone (Thurstone, 1938; Thurstone & Thurstone, 1941).

Este tipo de análisis marca una nueva concepción de la inteligencia. La división de la inteligencia en sus distintos factores o dimensiones dio lugar a la aparición de dos grandes líneas de estructuración de las dimensiones cognoscitivas, lo que ha dado en llamarse la escuela inglesa y la escuela americana. En la primera se da más importancia a un factor central de inteligencia general, que coronaría una estructura en la que luego vendrían dos amplias dimensiones, la verbal-educativa y la mecánico-espacial, en las que se articularían otros muchos factores más específicos.

Esta perspectiva teórica para abordar la inteligencia asume una serie de dimensiones no jerarquizadas que compondrían el perfil cognoscitivo, que por ejemplo en el caso del PMA serían: la comprensión verbal, la fluidez verbal, aptitud numérica, aptitud espacial, memoria, rapidez perceptiva y razonamiento general. Ambos enfoques son compatibles, y tienen mucho que ver con la tecnología estadística utilizada, sobre todo el análisis factorial.

Toda esta línea de investigaciones psicométricas sobre la inteligencia culmina en la obra magna de Carroll (1993), donde se sintetizan los grandes avances alcanzados.

En parte como una consecuencia del avance en el campo de los test cognoscitivos, los test de personalidad aprovechan de los avances que se producen en la psicometría. Suele citarse la hoja de datos personales utilizada por Woodworth en 1917 para detectar neuróticos graves como el pionero de los test de personalidad. Por su parte el psiquiatra suizo Rorschach propone en 1921 su test proyectivo de manchas de tinta, al que seguirán otros muchos test basados en el principio de la proyección, que asume que ante un estímulo ambiguo, la persona evaluada tenderá a producir respuestas que de algún modo reflejan aspectos importantes de su personalidad.

Estos avances en psicometría necesitaron fundamentos teóricos que guiaran su construcción y que condicionan los test según los avances teóricos y estadísticos de cada momento. Las teorías estadísticas de los test van a permitir la estimación de las propiedades psicométricas de los test para de ese modo garantizar que las decisiones tomadas a partir de ellos son las adecuadas. Esas teorías permiten estimar la fiabilidad y la validez de los test, lo cual es imprescindible para poder usar los test de forma rigurosa y científica. Por supuesto, aparte de estas teorías estadísticas sobre los test, la construcción de una prueba debe guiarse por un modelo o teoría psicológica sustantiva que dirige su construcción.

En general, los test pueden describirse como los procedimientos de recolección de información sobre un individuo o grupo utilizados habitualmente en Evaluación Psicológica. La construcción de los mismos se basa en modelos psicométricos que permiten evaluar la calidad de la medida y dar garantías de la misma. En el campo de la Psicometría, la Teoría de los Test constituye el marco de referencia teórico y metodológico que reúne los modelos que subyacen a la elaboración y uso de test (Muñiz, 1997).

La bibliografía especializada señala que los modelos que componen la Teoría de los Test formalizan las interrelaciones de tres componentes que intervienen en la medición mediante test: a) la puntuación observada tras la administración del test (el puntaje total en un test o la respuesta de un individuo a un ítem), b) un valor inobservable del dominio o rasgo psicológico que se pretende medir y; c) el error de medida que conlleva todo proceso de medición. Hay dos grandes enfoques o teorías a la hora de construir y analizar los test,

son la Teoría Clásica de los Test (TCT) y el enfoque de la Teoría de Respuesta al Ítem (TRI). A continuación se presentan los elementos principales de cada una de estas teorías, sus semejanzas y diferencias.

Como una idea inicial se debe resaltar la idea de que tanto la TCT como la TRI persiguen el mismo objetivo: estimar el error que se comete al intentar medir un fenómeno psicológico específico. Son construcciones teóricas (con menor o mayor grado de complejidad y profundidad) respecto de un mismo hecho. Ambas teorías plantean un modelo y un conjunto de supuestos, que si se cumplen, garantizan la precisión de la medida. También cabe destacar que la coexistencia de ambas teorías no implica su incompatibilidad. Lejos de competir, estos modelos se complementan en la práctica psicométrica para realizar un análisis más profundo y exhaustivo de la calidad y/o del funcionamiento del test.

2.4.1 Teoría de Medición que sustenta el estudio: Teoría Clásica de los Test

Es la teoría que sirvió de fundamento teórico para la primera generación de test. Desde su primera formulación (Spearman, 1904), la Teoría Clásica de los Test (TCT) ha servido como modelo para dar una interpretación a los puntajes de las personas en los test.

En relación a sus fundamentos, la TCT surgió del modelo lineal de puntuaciones formulado por Spearman (1904) y alcanzó su formalización más precisa en la obra de Novick (1966). Su formulación matemática es bastante simple y supone que el puntaje observado de un sujeto en un test es el resultado de la suma del valor real (puntaje verdadero) y el error de medición.

Características del Modelo de la TCT. Las raíces de la Teoría Clásica están fuertemente vinculadas a la tradición positivista en las Ciencias Sociales. De hecho, la propuesta de Spearman se asemeja a la forma de medir en las “ciencias duras”. Si se realizan múltiples mediciones de una longitud, la mejor estimación de la “longitud verdadera” resultará del promedio de todas las observaciones efectuadas. Si cada una de las mediciones que se ejecutan son independientes y el error de medida en cada una de ellas es aleatorio, el promedio de éste tenderá a cero (dado que las observaciones que subestimaron a la puntuación verdadera se cancelarían con aquéllas que la sobreestimaron). De esta forma, tras infinitas mediciones, el promedio de las puntuaciones empíricas obtenidas podría considerarse igual a la verdadera.

En relación con estos planteamientos básicos de Spearman se puede decir propone un modelo muy simple, de sentido común, para las puntuaciones de las personas en los test, y que ha dado en llamarse modelo lineal clásico. Consiste en asumir que la puntuación que una persona obtiene en un test, que denominamos su puntuación empírica, y que suele designarse con la letra X , está formada por dos componentes, por una parte la puntuación verdadera de esa persona en ese test (V), sea la que sea, y por otra, un error (e), que puede ser debido a muchas causas que se nos escapan y que no controlamos.

La fundamentación teórica plantea respecto a los errores que al medir alguna variable con un test (e) puede deberse a muchas razones, que pueden estar en la propia persona, en el contexto, o en el test, una clasificación bastante exhaustiva de las fuentes posibles de error puede consultarse en Stanley (1971). Para poder avanzar Spearman añade tres supuestos al modelo y una definición, que se exponen a continuación.

Supuestos de la TCT. La Teoría Clásica de los Test se fundamenta sobre 3 supuestos básicos. El primero de ellos es definir la puntuación verdadera (V) como la esperanza matemática de la puntuación empírica, que formalmente puede escribirse así: $V = E(X)$. Lo que esto significa conceptualmente es que se define la puntuación verdadera de una persona en un test como aquella puntuación que obtendría como media si se le pasase infinitas veces el test. Se trata de una definición teórica, nadie va a pasar infinitas veces un test a nadie, por razones obvias, pero parece plausible pensar que si esto se hiciese la puntuación media que esa persona sacase en el test sería su verdadera puntuación.

Ya en relación con el segundo supuesto. En este, Spearman asume que no existe relación entre la cuantía de las puntuaciones verdaderas de las personas y el tamaño de los errores que afectan a esas puntuaciones. En otras palabras, que el valor de la puntuación verdadera de una persona no tiene nada que ver con el error que afecta esa puntuación, es decir, puede haber puntuaciones verdaderas altas con errores bajos, o altos, no hay conexión entre el tamaño de la puntuación verdadera y el tamaño de los errores. De nuevo se trata de un supuesto en principio razonable, que formalmente puede expresarse así: $r(v,e) = 0$.

Y finalmente, el tercer supuesto establece que los errores de medida de las personas en un test no están relacionados con los errores de medida en otro test distinto. Es decir, no hay ninguna razón para pensar que los errores cometidos en una ocasión vayan a covariar

sistemáticamente con los cometidos en otra ocasión. Formalmente este supuesto puede expresarse así: $r(e_j, e_k)=0$.

La bibliografía especializada señala que estos supuestos parecen razonables y sensatos, pero no se pueden comprobar empíricamente de forma directa, serán las deducciones que luego se hagan a partir de ellas las que permitan confirmarlas o falsearlas. Tras cien años formulados y con muchos resultados empíricos detrás, se puede plantear que las ideas de Spearman han sido de gran utilidad para la psicología. Además del modelo y de estos tres supuestos, se formula una definición de lo que son test Paralelos, entendiendo por ello aquellos test que miden lo mismo exactamente pero con distintos ítems. Las puntuaciones verdaderas de las personas en los test paralelos serían las mismas, y también serían iguales las varianzas de los errores de medida.

Modelos de la TCT. La Teoría Clásica de los Test marca un punto fundamental en la historia de la psicometría, al grado que las teorías posteriores se construyen en relación con ella y sus supuestas debilidades. Este enfoque clásico ha generado diversas variantes sobre todo en función del tratamiento dado al error de medida. Ha habido numerosos intentos de estimar los distintos componentes del error, tratando de descomponerlo en sus partes. De todos estos intentos el más conocido y sistemático es la Teoría de la Generalizabilidad (TG) propuesta por Cronbach y sus colaboradores (Cronbach, Gleser, Nanda y Rajaratnam, 1972). Se trata de un modelo de uso complejo, que utiliza el análisis de varianza para la mayoría de sus cálculos y estimaciones.

Además de las anteriores, otro desarrollo psicométrico surgido en el marco clásico ha sido el de los Test Referidos al Criterio (TRC). En esta elaboración se aborda principalmente el caso de test utilizados fundamentalmente en el ámbito educativo y en la evaluación en contextos laborales. Su objetivo es determinar si las personas dominan un criterio concreto o campo de conocimiento, por tanto no pretenden tanto discriminar entre las personas, como la mayoría de los test psicológicos, sino evaluar en qué grado conocen un campo de conocimiento denominado criterio, de ahí su nombre. Estos test se desarrollan a partir de la propuesta de Glaser (1963) y han tenido una gran influencia sobre todo en el ámbito educativo.

Además de las elaboraciones teóricas derivadas directamente de la Teoría Clásica como mejoras posibles a la misma, también fue surgiendo una teorización alternativa. En la década del 60, surgieron los primeros desarrollos de la TRI, un enfoque emergente del campo educativo que se propuso profundizar el estudio de las propiedades psicométricas de los ítems y de los test muy general, como los más relevantes que diferencian a la TRI de la TCT.

2.4.2 Parámetros de la TCT aplicada en la presente investigación

Dado que en el presente estudio se aplicó la Teoría Clásica de los Test TCT para la elaboración de la prueba de selección universitaria, a continuación se exponen en forma resumida 4 parámetros que se utilizaron en su desarrollo. Comenzando por la Confiabilidad, para luego examinar la Validez, posteriormente se trata el tema de la transformación e interpretación de los puntajes, para finalmente abordar el análisis de los ítems.

2.4.2.1 Confiabilidad

En este apartado se aborda la “Confiabilidad” o “Fiabilidad”, el cual constituye un aspecto fundamental para las mediciones realizadas mediante test: La obtención de puntuaciones consistentes y precisas. Para ello debe tenerse presente que, como sucede con cualquier procedimiento de medición, el desarrollo y la administración de pruebas requieren el conocimiento del error que se puede cometer. De manera que si este error de medida es grande, de modo que las puntuaciones obtenidas no representan fielmente los niveles de aprendizaje o fenómenos psicológicos objeto de interés, las pruebas no proporcionan la confianza necesaria para cumplir con su objetivo.

Dada esta importancia, en este apartado se aborda la confiabilidad a partir de la perspectiva de la teoría clásica de los test, que es la que se implementó en el presente estudio, empezando con una descripción del modelo lineal clásico, derivando el coeficiente de fiabilidad y prestando una especial atención tanto a su interpretación como a las diferentes estrategias que se han ido desarrollando para calcularlo.

Se iniciará recordando que, de acuerdo con la teoría clásica de los test, la fiabilidad de los mismos está relacionada con los errores de medida aleatorios presentes en las puntuaciones obtenidas a partir de su aplicación. Por ello se considera que una prueba será más fiable cuantos menos errores de medida contengan las puntuaciones obtenidas por los

sujetos a quienes se les aplica. Dicho de otro modo, la fiabilidad de un test será su capacidad para realizar medidas libres de errores.

En este sentido, se puede decir que cualquier proceso de medida de una característica de los objetos o de los sujetos lleva inherente un cierto error en su medición. Por esa razón se pueden encontrar instrumentos de medición con mayor o menor capacidad para minimizar estos errores, pero es casi imposible obtener alguno que los elimine del todo. En el ámbito de la educación y la psicología, donde las variables que se miden habitualmente son características propias de los sujetos, relacionadas con sus rasgos de personalidad, sus capacidades cognitivas, sus estados de ánimo, entre otros, de manera que los instrumentos que se usan generalmente son pruebas estandarizadas tipo test, aún resulta más evidente que las medidas que se hacen de estos atributos están también afectadas por ciertos errores.

Esta problemática significa que las puntuaciones obtenidas con las administraciones de estos test no se correspondan exactamente con los verdaderos niveles de los sujetos en la característica medida. De cualquier forma, algunos de estos errores propios de toda medición pueden responder a factores sistemáticos que tendrán una posible causa en el propio proceso de medida, en el instrumento utilizado o en ciertas características de los objetos o sujetos medidos. Algunos errores no tienen un componente sistemático originado en un error del instrumento de medición directamente, sino que son aleatorios, indeterminados y no responden a ningún factor que pueda ser conocido, y por lo tanto eliminado. Estos errores aleatorios son los que están implicados en el concepto de fiabilidad.

Relacionado con el concepto de fiabilidad se encuentra el de consistencia interna de los test. Un instrumento de medida fiable se caracteriza por una elevada estabilidad temporal y por una adecuada consistencia interna. La consistencia interna está referida al grado en que cada una de las partes de las que se compone el instrumento es equivalente al resto. Esta característica aplicada al caso de los test está determinada por el grado en el que cada ítem, como elemento básico de este, evidencia un adecuado nivel de equivalencia respecto a los demás ítems, o sea, que mide con el mismo grado el constructo medido.

Así, si hay una elevada equivalencia entre los ítems del test, es de suponer que las respuestas de los sujetos a estos diferentes ítems estarán altamente correlacionadas, y las diferentes partes en las que se pueda dividir el test siempre mostrarán esta elevada

covariación. Existen en la actualidad varias técnicas para medir la consistencia interna de un test, y en la presente investigación se utilizó el coeficiente de Alfa de Cronbach.

Este método en particular, el coeficiente alfa (D) de Cronbach (1951) representa la consistencia interna de un test a partir de la covariación entre sus ítems. Cuanto más elevada sea la proporción de la covariación entre estos ítems respecto a la varianza total del test, más elevado será el valor del coeficiente alfa (D) de Cronbach, y más elevada su fiabilidad. Existen diferentes fórmulas para obtener el valor del coeficiente D, la más ampliamente utilizada de las cuales es la que se deriva del cálculo de las varianzas de cada ítem y de la varianza de las puntuaciones totales en el test.

Para el cálculo de este coeficiente debe considerarse que la fórmula del coeficiente D también se puede expresar en función del cociente entre la media de las covarianzas y la media de las varianzas de los diferentes ítems del test. Este cociente, que designamos como r_1 , constituye una estimación de la fiabilidad de cada ítem. En este sentido, la fórmula del coeficiente D a partir de r_1 es una aplicación de la corrección de Spearman-Brown, a partir de la estimación de la fiabilidad de cada ítem, teniendo en cuenta que si se tienen n ítems es como si se hubiese alargado n veces el ítem inicial.

2.4.2.2 La Validez desde la TCT

En este apartado denominado “Validez”, se aborda otro aspecto básico en la medición de los fenómenos educativos y psicológicos mediante pruebas estandarizadas: su adecuación a los objetivos para los cuales han sido construidos y son utilizados en la práctica. Como se explica a continuación, en la actualidad no se considera la validez como una propiedad intrínseca de los test, sino que es más bien producto del análisis de su adecuación al propósito específico al que sirven.

Pero este concepto ha venido evolucionando a través del tiempo. Empezando en la primera mitad del siglo XX. La utilización de cuestionarios se vio impulsada por la primera y segunda guerras mundiales. Durante esos momentos se tuvo la necesidad de incorporar al ejército a la población civil, destinándola al puesto más adecuado. Tras rellenar los cuestionarios se comprobaba en el campo de entrenamiento si los sujetos rendían satisfactoriamente o no en el puesto al que se les había destinado. Dado que en primer lugar se hacía la medición y posteriormente se evaluaba el éxito, se hablaba de validez predictiva.

Es decir, se entiende que un test muestra validez predictiva si es capaz de predecir las variaciones de un constructo que será evaluado posteriormente a la aplicación del cuestionario.

Años más tarde se modificó esta concepción de validez. Cuando se trató de evaluar la relación existente entre las características de las personas que realizaban un trabajo y su éxito en él. De este modo, se trataba de conocer qué características podrían predecir el éxito laboral y buscarlas cuando se realizaba una selección de personal. Dado que el estudio se realizaba sobre personas que ya tenían el puesto y se valoraba su ejecución, se hablaba de validez concurrente, ya que ambas mediciones se hacían a la vez. Es decir, un test posee validez concurrente si sirve para predecir el comportamiento en un constructo que es evaluado simultáneamente a la aplicación del cuestionario (Meneses et al, 2013).

Pero el desarrollo de la teoría psicológica de la inteligencia con un auge de las teorías que trataban de conocer la estructura factorial de la misma, incidieron para modificar esta concepción de validez. Planteando que un test será considerado como válido cuando represente de manera fidedigna el constructo psicológico que pretende medir, así como las relaciones esperadas entre los diferentes constructos. En este marco nace la validez de constructo (Cronbach & Meehl, 1955). Y fueron surgiendo paralelamente las técnicas estadísticas empleadas para poder comprobar dicha validez, el análisis factorial exploratorio y las matrices multirrasgo-multimétodo (Campbell & Fiske, 1959), y más recientemente el análisis factorial confirmatorio.

En este proceso de evolución del concepto, hasta los años ochenta se podía hablar de validez predictiva, validez concurrente, validez de contenido y validez de constructo de un cuestionario, si bien las dos primeras en los estándares de los test y manuales educativos y psicológicos publicados por la APA, AERA y NCME en 1966 y 1974 se englobaban como validez de criterio.

En otro ámbito de trabajo, Cronbach (1971) puntualizó que en un test que pretende medir un rasgo de personalidad no existe solo un criterio relevante que predecir, ni un contenido que muestrear (validez predictiva y de contenido respectivamente). Se dispone, por el contrario, de una teoría acerca del rasgo y de sus relaciones con otros constructos y variables (validez de constructo). Si se asume que la puntuación del test es una manifestación

válida del atributo, se puede contrastar la asunción analizando sus relaciones con otras variables.

Como consecuencia de esta concepción de validez, comenzó a desarrollarse una tendencia en la que se considera la validez como algo unitario, siendo la validez de constructo la científicamente más admisible y estando la validez de criterio y de contenido incluidas en esta (Messick, 1989). De manera que se impone la concepción de que la validación de constructo constituye un marco integral para obtener pruebas de la validez incluyendo las procedentes de la validación de criterio y de contenido. En este nuevo enfoque, ya no se consideran las diferentes categorías de validez para comenzar a hablar de diferentes evidencias implicadas en los tres tipos tradicionales de validez (criterio, contenido y constructo).

Actualmente, se plantea que la validez hace referencia al grado en el que la evidencia empírica y la teoría apoyan la interpretación de las puntuaciones de los test relacionada con un uso específico. En este sentido, se concibe la validez como un concepto unitario. Para comprobar la validez se debe atender a cinco evidencias de esta:

- El contenido de test: Los ítems que constituyen el test son relevantes y representativos del constructo psicológico que se desea medir.*
- El proceso de respuesta: El proceso que siguen las personas al contestar al test permite extraer respuestas indicadoras de lo que se quiere evaluar.*
- La estructura interna: Las relaciones de los ítems entre sí son congruentes con el modelo teórico empleado a la hora de definir el constructo que evaluar.*
- La relación con otras variables: Las relaciones que se establecen entre el constructo que se evalúa y otros constructos son las esperadas según el marco teórico en el que se haya definido el constructo que evaluar.*
- Las consecuencias de la aplicación del cuestionario: Las consecuencias tanto positivas como negativas que se extraen al emplear un test son las previstas. (Meneses, et al, 2013, p.145)*

Para el caso de la presente investigación, la elaboración de un test de admisión universitaria, la validez de contenido es fundamental. La evidencia de la validez de contenido hace referencia a la relación que existe entre los ítems que componen el test y lo que se pretende evaluar con él, prestando atención tanto a la relevancia como a la representatividad de los ítems. Este tipo de evidencia se recoge principalmente en el momento de la elaboración del test.

En el caso específico de las pruebas educativas, las evidencias de validez basada en el contenido son fundamentales. Si no se comprueba que el test es consistente con los objetivos curriculares perseguidos (relevancia), es decir, que está libre de material irrelevante y que el que está representa adecuadamente el dominio que se pretende evaluar (representatividad), la utilidad del test se verá seriamente afectada y, por tanto, las conclusiones que se obtengan serán erróneas. En estas situaciones se suele recomendar, dado que el dominio que se quiere evaluar está claramente definido, emplear los diferentes métodos estadísticos de muestreo para obtener una muestra representativa de los contenidos que deben constituir el test (Muñiz, 2003).

El procedimiento recomendado es que después de elaborar los ítems se debe acudir a un grupo de expertos en la materia, que harán las veces de jueces. Para evitar cualquier sesgo, dichos jueces no deben estar implicados en la elaboración del cuestionario. Estos deben analizar cada uno de los ítems valorando en qué medida son representativos y relevantes para evaluar el dominio de interés, tomando como definición de este la aportada por los autores del test.

En este sentido se plantea que existen tres aspectos bien diferenciados que se deben tener en cuenta a la hora de comprobar las evidencias de la validez de contenido: la definición del dominio, la representación de los ítems que evalúan el dominio y su relevancia (Sireci, 1998). Es recomendable que la valoración de los ítems la realice cada juez por separado para, de este modo, evitar posibles sesgos a la hora de responder. Una vez que se poseen las valoraciones de todos los expertos, se deben buscar aquellos ítems en los que haya concordancia, seleccionándolos para formar parte del cuestionario.

La descripción anterior hace referencia al procedimiento más habitual a la hora de valorar los indicios de validez de contenido, si bien no está libre de críticas. El principal problema que se plantea en la utilización de expertos es que estos son altamente competentes en el contenido que se evalúa, por lo que pueden pasar por alto un texto cuyo nivel no sea adecuado para la comprensión de los sujetos que hay que evaluar o que puede ser fácilmente malinterpretado. Es decir, aunque el experto nos puede proporcionar información muy relevante, lo que realmente importa es cómo percibe y reacciona ante el test o el ítem la persona que está respondiendo (Leighton, 2004).

2.4.2.3 Transformación e interpretación de los puntajes en la TCT

Los puntajes alcanzados en las pruebas estandarizadas, con frecuencia requieren ser transformados para su interpretación. Interpretar el valor obtenido por una persona en una evaluación en educación o psicológica, plantea que se debe conocer con cierto detalle las características técnicas del instrumento que es empleado, pero también conocer el constructo al que se hace referencia y entender cómo este se estructura y se explica en un marco teórico que tiene sentido, a la vez, en un marco disciplinario concreto, ya sea educacional, clínico o de cualquier otro tipo (Meneses et al, 2013).

En este sentido el significado de interpretar una puntuación, se debe considerar desde una perspectiva eminentemente metodológica y matemática y, lo más importante, inferida de acuerdo con la respuesta de un conjunto de individuos que, evaluados con una misma medida, proporcionarán el modelo de comportamiento que será dispuesto a modo de norma, de marco de referencia respecto al cual la puntuación de una persona concreta será interpretada. Desde este enfoque es que tienen sentido las puntuaciones transformadas. Los percentiles, las puntuaciones estandarizadas y las normas cronológicas se abordan a continuación en forma resumida, exponiendo sus características más importantes, la manera de calcularlas y, donde se considere oportuno, sus limitaciones más relevantes.

Una propiedad necesaria de las puntuaciones transformadas es que no alteran el escalamiento de las puntuaciones directas, es decir, respetan la diferente disposición de los distintos sujetos según la magnitud de sus puntuaciones. Y este mantenimiento de la ordenación original se lleva a cabo, como se ha mencionado, otorgando a las nuevas puntuaciones un sentido eminentemente práctico. Los percentiles, las puntuaciones estandarizadas y las puntuaciones estandarizadas derivadas serán las distintas estrategias de transformación que serán tratadas a continuación.

Los percentiles. Esta primera medida estadística está asociada a la mediana, puesto que son una generalización de la mediana. La transformación basada en los percentiles consiste en asignar a cada puntuación directa una puntuación porcentual, según la posición de las observaciones dentro del conjunto de observaciones. Si el valor porcentual es del 25%, se estará hablando del percentil 25. Y bajo esta denominación se estará identificando aquella puntuación directa que deja por debajo una cuarta parte de todas las observaciones. O

expresado de otra manera, aquel valor que con él y por encima de él se encuentra el 75% de todas las observaciones. Si el porcentaje asciende a un 50%, se hablará del percentil 50, que se corresponderá con el valor que ocupará la posición central de la distribución.

El percentil 50 es, efectivamente, la mediana. La primera formulación práctica de los percentiles se remonta a finales del siglo XIX, y proviene del explorador y científico Francis Galton, primo de Charles Darwin, que entre otras aficiones estaba interesado en la medida de multitud de rasgos antropométricos (Meneses et al, 2013). Hoy en día, y sin abandonar el ámbito de la medida de las propiedades del cuerpo, los percentiles continúan siendo una medición de interés.

Los percentiles han resultado una medida muy empleada no solo en el contexto de la medida del peso o de la altura, sino también en el contexto de las medidas psicológicas. Su simplicidad y universalidad, dado que no hay que tener amplios conocimientos en matemáticas ni de estadística en particular, ha reforzado su utilización como estrategia interpretativa.

La primera de las ventajas del uso de los percentiles, como ya se ha indicado, es su simplicidad conceptual. Una segunda ventaja es que en el cálculo y la correcta interpretación de este, dado que centra su atención solo en la orden en el que las observaciones son dispuestas, resulta irrelevante el modo como se distribuyen los valores objeto de análisis. Esto implica que, a pesar de que hay estrategias de transformación que requieren una distribución de valores esencialmente ajustada a una curva normal, en el caso de los percentiles es posible abordar cualquier tipo de distribución.

Esta característica es especialmente interesante en educación y en psicología, dado que con frecuencia los fenómenos medidos no se ajustan a una distribución normal de valores. Como se mencionó anteriormente en este apartado, los percentiles consisten en dividir una distribución de valores en cien posiciones ordinales. No obstante, existen otros abordajes que simplifican la cantidad de órdenes posibles y que pueden ser suficientes para resumir el comportamiento de un conjunto de valores.

Puntuaciones estandarizadas. Aun cuando la media es la medida de tendencia central más utilizada y que permite expresar el comportamiento medio de un conjunto de observaciones, se considera a la desviación típica (o estándar) como un complemento que permite cuantificar el grado de acercamiento de los datos respecto a su media. Este índice de dispersión expresa la discrepancia promedio entre una observación cualquiera y la media del conjunto de observaciones. La media y la desviación típica constituyen los dos datos a partir de los cuales se puede calcular una puntuación estandarizada. Si bien los percentiles expresan la posición que ocupa un individuo respecto a un grupo de referencia, en el caso de las puntuaciones estandarizadas el interés se focaliza en la discrepancia que presenta este individuo respecto al modo de comportamiento promedio del grupo, su media.

La puntuación estandarizada o puntuación típica es aquella que permite expresar cuántas desviaciones típicas por encima o por debajo de una media se sitúa una observación. Es importante considerar que este proceso de estandarización tiene su sentido cuando la distribución de valores objeto de interés se ajusta esencialmente a una curva normal. El valor resultante de esta transformación se constituye en una nueva puntuación, que recibe el nombre de puntuación estandarizada z .

Y aquí descansa una de las consideraciones más relevantes que hay que tener presente cuando se interpretan puntuaciones estandarizadas: las inferencias a las que podemos llegar mediante el uso de estas puntuaciones dependen de la muestra que es empleada, dado que es de ella de donde se obtienen la media y desviación típica que permiten el cálculo de las puntuaciones transformadas. Elegir cuidadosamente la muestra de referencia resulta imprescindible (Meneses et al, 2013).

Puntuaciones estandarizadas derivadas. Como su nombre lo señala, estas puntuaciones se derivan de las anteriormente expuestas, y su uso está relacionado a una característica matemática. Aunque intrínsecamente no es un problema, las puntuaciones z se expresan con valores negativos y decimales. El sentido básico de las puntuaciones estandarizadas derivadas es transformar las puntuaciones típicas según un sistema de codificación que elimina estos valores negativos y decimales. Una de las puntuaciones típicas derivadas más comúnmente empleadas es la puntuación T de McCall (Meneses et al, 2013).

Consiste en partir de una constante, en este caso 50, y sumarle diez veces el valor de la puntuación típica z . Para interpretar debidamente esta puntuación solo hay que tener presente que la media de esta puntuación es 50 y su desviación típica 10. De este modo, por ejemplo, una puntuación z de 2 tendrá un valor de 70 una vez transformada en puntuación T . En ambos casos (z y T), y necesariamente, sería indicador de una puntuación directa que se ubica por encima de la correspondiente media en dos desviaciones típicas (Meneses et al, 2013). La puntuación escala o scale score utilizada en este estudio pertenece a esta categoría.

Sin embargo, es importante tomar en cuenta que es posible encontrar puntuaciones típicas derivadas de distinta naturaleza, y es que estrictamente cualquiera puede crearse la suya, dado que una puntuación derivada se formula según criterios arbitrarios. Ejemplos diferentes los encontraremos en el test de inteligencia Stanford-Binet, que emplea una media de 100 y una desviación típica de 16, la Wechler Adult Intelligence Scale (WAIS), con una media también de 100, pero con una desviación típica de 15, o el Minnesota Multiphasic Personality Inventory (MMPI), que, coincidiendo con la escala de McCall, se interpreta en torno a una media de 50 y una desviación típica de 10.

2.4.3 Análisis de los ítems en la TCT

Un último apartado respecto a los parámetros de la TCT aplicados en la presente investigación es el referido al análisis mismo de los ítems. Se abordan las propiedades de los ítems de acuerdo con la formulación de la teoría clásica de los test y se discuten los diferentes procedimientos disponibles para evaluar la dificultad y la discriminación de los ítems. En concreto se presenta una breve explicación del índice de dificultad, índice de discriminación, análisis de los distractores y del funcionamiento diferencial del ítem.

Índice de dificultad. Este parámetro representa la proporción de personas que contestan bien un ítem dado, dado que es una proporción que frecuentemente se expresa como porcentaje, sus valores oscilan entre 0 y 1, o entre 0 y 100%. Contrario a lo que señala su nombre, índice de dificultad, los valores cercanos a 1 indican una baja dificultad y valores cercanos a 0 indican dificultad máxima. Pero debe tenerse presente un problema: que una parte de los aciertos se dan por puro azar. Al tratarse de preguntas con alternativas cerradas, es lógico pensar que una parte de los acertantes no conocían la respuesta y que si han acertado

es “solo” porque han elegido una de las alternativas, no porque “sepan” la respuesta. El problema que se plantea es desconocer cuántos lo han hecho.

En este contexto de análisis de la dificultad, un ítem que todos aciertan o que todos fallan no sirve para nada más que para perder el tiempo contestándolo. Si todo aquel que responde acierta, es como si regaláramos a todos los alumnos una parte de la puntuación. Y si todos lo fallan, es como si los penalizáramos porque nadie puede contestar acertadamente.

Una vez que se conoce la dificultad de un ítem, cabe plantearse la interrogante, ¿cómo deberían ser las dificultades de todos los ítems de una prueba? La bibliografía especializada señala que (Moreno, Martínez y Muñiz, 2004), la dificultad de un ítem debe relacionarse con la del concepto que recoge. Esto es, si un contenido es fácil, el ítem debe ser fácil. Por tanto, una prueba que mide contenidos diversos debería tener ítems de todas las dificultades, y éstas deberían corresponderse a la dificultad de los conceptos medidos.

En el presente trabajo, para utilizar los valores que se han venido utilizando en el país por las instituciones que han venido desarrollando pruebas estandarizadas para aplicación con estudiantes nacionales, se tomará: “El rango de valores del índice de dificultad es de 0 a 1.0, los valores aceptables del índice de dificultad oscilan entre 0.3 y 0.8” (AIR, 2015, p.9).

Índice de Discriminación. Los especialistas plantean que no es suficiente saber si un ítem es fácil o difícil para decidir si es adecuado o no. Aun cuando podría parecer suficiente, en la bibliografía especializada se una propiedad psicométrica de los ítems sobre el resto, y es precisamente la de discriminación. Si un ítem no discrimina, no es útil para la medición, y ese es el objetivo para el que fue redactado.

Como su nombre lo indica, se debe entender por discriminación la capacidad de un ítem de distinguir entre las personas que tienen un buen rendimiento en el test, respecto a las que lo tienen malo. Cabe plantearse entonces ¿Quiénes deben contestar correctamente una pregunta de examen? No es tan importante si son muchos o pocos alumnos, sino más bien que los acertantes sean, en general, “de los buenos alumnos”. Pero cabe preguntarse entonces ¿A qué se refiere al señalar “los buenos”? Y la respuesta sería a aquellos que tienen una alta puntuación en la prueba. Es decir, un ítem debe ser más acertado entre aquellos que han obtenido una alta puntuación en la prueba que entre los que no la tienen. Obviamente, una

pregunta no puede ser buena si solo la aciertan los peores alumnos: debe ocurrir exactamente lo contrario.

Pero cabe preguntarse entonces cuál es el punto para diferenciar entre alumnos de alto y bajo rendimiento. Existen varias maneras de definir el punto de corte de la puntuación total en la prueba para hacer esta clasificación. Por un lado, es frecuente utilizar la media de la puntuación total en la prueba, lo que crea dos grupos de igual tamaño. Esta estrategia tiene como ventaja que todos los participantes participan en el cálculo, pero tiene como claro inconveniente que los grupos son poco extremos. Intuitivamente comprobamos que dos personas con rendimiento muy semejante pueden estar en grupos diferentes solo por una pequeña diferencia. Es preferible utilizar grupos más extremos para poder estudiar correctamente este índice. Kelley (1939) recomienda utilizar los percentiles superior e inferior del 27%.

En el presente trabajo, para utilizar los mismos valores estadísticos que se han venido aplicando por las instituciones que han desarrollado pruebas estandarizadas para evaluar estudiantes hondureños, se tomará como referencia: El rango de valores del índice de discriminación es de -1.0 a 1.0 con valores aceptables mayores a 0.30. Cuando se está interpretando el índice de discriminación es importante tener en mente que existe una relación con el índice de dificultad. Cuando un ítem es demasiado fácil o demasiado difícil no se puede esperar que tenga un alto poder de discriminación (AIR, 2015, p.9).

Análisis de los distractores. Un parámetro fundamental para analizar la calidad de un ítem y su buen funcionamiento, es que sus distractores realmente operen como tales. Una opción de un ítem que no es seleccionada por nadie o casi nadie, no está funcionando como distractor para las personas que responden, y por tanto no está cumpliendo la función para la cual se ha elaborado. Y en sentido contrario: una alternativa incorrecta que es muy elegida por los mejores quizá no es tan incorrecta como pensó el que la redactó. Entonces cabe plantearse cómo se estudia el comportamiento de los distractores. El índice habitual es el denominado índice D, pero no se calcula a partir de quienes aciertan y fallan, sino que se calcula a partir de quienes eligen cada una de las alternativas de respuestas (AIR, 2015).

Pasando al análisis estadístico de la calidad de los distractores, la bibliografía especializada plantea varios elementos importantes. En primer lugar se señala que debe tener

discriminación negativa, es decir, ser más elegido entre los peores estudiantes que entre los mejores. Adicionalmente, es deseable que todos los distractores tuvieran una discriminación parecida, ya que indicaría que sus capacidades de atracción son semejantes. Conseguir esto es especialmente difícil, y esta dificultad crece exponencialmente con las alternativas de respuesta. En resumen: es mucho más difícil redactar tres distractores que sean efectivos que dos. Por ello, la mayoría de los estudios realizados recomiendan usar como mucho tres alternativas de respuesta (AIR, 2015).

Respecto a los valores estadísticos aceptados para incluir un distractor como “aceptable”, en el presente trabajo se utilizaron los valores recomendados por el proyecto (MIDEH) que trabajó por más de una década en pruebas estandarizadas en el país: En el caso de que la frecuencia de un distractor sea menor al 2% o nula proporción, se puede decir que el distractor no está cumpliendo la función que corresponde, que es la de ser una posible respuesta para aquel estudiante que no ha logrado esa expectativa de logro, por ende el ítem puede ser contestado fácilmente hasta por sentido común. También en el caso inverso, en el que la selección de un distractor es mayor que la respuesta correcta, se entiende que esta opción está confundiendo a los estudiantes y es recomendable reestructurar el ítem (AIR, 2015).

Funcionamiento diferencial del ítem (FDI). En la elaboración de pruebas estandarizadas es importante analizar, tanto un ítem como un instrumento de medida, que estos no sean sesgados. Pero primero hay que aclarar a qué se refiere el sesgo en un instrumento de medida. En el contexto de las pruebas estandarizadas, se señala que un ítem o un test están sesgado cuando algún grupo, por ejemplo, al comparar entre hombres y mujeres o entre estudiantes ricos y pobres, que tienen el mismo conocimiento sobre la materia evaluada, no obtienen resultados similares, sino que uno de los grupos es sistemáticamente “perjudicado”.

En este sentido, los instrumentos que tienen sesgo a favor de uno u otro grupo pueden tener implicaciones sociales importantes y no deseadas. Por ejemplo, si un examen de admisión a una institución universitaria, favoreciera sistemáticamente a los alumnos con nivel socioeconómico alto frente a otros que no lo tienen, y partiendo del supuesto de que ambos saben exactamente lo mismo sobre el tema, la selectividad no sería socialmente justa.

¿Cómo se calcula el sesgo de una prueba estandarizada? La documentación actualizada utiliza el concepto de funcionamiento diferencial de los ítems y el principal índice que se deriva es el FDI. Se dice que un ítem “tiene o presenta” FDI cuando se dan diferencias estadísticamente significativas en la puntuación de un ítem en dos grupos diferentes que de acuerdo con lo que se conoce previamente, el mismo nivel. En esta investigación se calculará el FDI utilizando la variable género (femenino y masculino).

Para evaluar el FDI pueden utilizarse diferentes procedimientos matemáticos. En el marco de TCT, el más utilizado por su sencillez de cálculo y sus buenos resultados del método de Mantel-Haenszel (1959) o la diferencias de medias. Como valor de referencia para aceptar un FDI aceptable en el presente estudio se utiliza el valor de .075 como el límite máximo. Por lo tanto, los valores aceptables oscilan entre 0 a .075 de acuerdo con lo señalado por Zenisky, Hambleton y Robin (2004).

Con esta amplia perspectiva teórica sobre las pruebas estandarizadas, sus usos, los argumentos respecto a sus usos como pruebas de admisión y selección, así como de sus fundamentos psicométricos y diversas técnicas de análisis que se vienen aplicando para los resultados generados en estas pruebas, se procede a hacer la presentación de los datos generados tanto por la prueba piloto como por la operativa, de acuerdo con los objetivos de investigación planteados, haciendo la discusión de los resultados a la luz de esta teoría, en el capítulo 5 del presente informe.

2.4.2 Teoría de Respuesta al Ítem

La denominación TRI agrupa líneas de investigación psicométricas independientes iniciadas por Rasch en 1960) y Birnbaum en 1968. El factor común, es que establecen una relación entre el comportamiento de un sujeto frente a un ítem y el rasgo responsable de esta conducta (rasgo latente). Para ello, recurren a funciones matemáticas que describen la probabilidad de dar una determinada respuesta al ítem para cada nivel del rasgo medido por este.

El objetivo sustancial de la TRI es la construcción de instrumentos de medición con propiedades invariantes entre poblaciones. Si dos individuos presentan idéntico nivel de rasgo medido ambos tendrán igual probabilidad de dar la misma respuesta,

independientemente de la población de pertenencia. Esto conlleva un gran beneficio respecto de la TCT en tanto que es posible evitar el uso de un grupo normativo.

2.4.2.1 Características generales de los Modelos TRI

Mientras que en la TCT se modeliza sobre el puntaje verdadero en una prueba particular, en la TRI se toma al ítem como unidad de análisis y se modeliza directamente sobre el rasgo latente. El nivel de rasgo latente que presenta un individuo es fruto de una estimación a partir del patrón de respuestas manifestado en un conjunto de ítems. Si se varía el conjunto de ítems utilizado se mantiene la puntuación estimada aunque eventualmente hayan cambiado las propiedades psicométricas de los reactivos. Por lo tanto, la TRI permite mediciones invariantes más allá de los ítems que componen el instrumento.

Hambleton y Swaminathan (1985) también rescatan la importancia de las medidas locales de precisión que proporciona la TRI. Desde la TCT se indica la fiabilidad como un valor global y constante para todos los niveles del rasgo. Sin embargo, se sabe que los tests suelen ser más precisos para discriminar en un determinado rango de la variable y menos en otros. La TRI provee información respecto del 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 desarrolladas por Birnbaum en 1968.

Expuestas las características básicas de la TRI, es pertinente decir que obtener garantías de precisión en la medición de un constructo no es una tarea simple para un investigador. La fuerza de esta teoría se sostiene en un conjunto de supuestos exigentes a los que la mayoría de los datos empíricos difícilmente se acomodan, y que por ende, condicionan su aplicabilidad.

2.4.2.2 Los Parámetros de la Curva Característica del Ítem

En el marco de la TRI se postula la existencia de una relación directa entre el comportamiento de un individuo frente a un ítem y el rasgo que genera esta conducta. La formalización de esta relación adopta la forma de una función matemática que vincula la probabilidad de dar una determinada respuesta (opción-clave o respuesta clave) a un ítem con cada nivel del rasgo latente. En los ítems de habilidades la clave es la respuesta correcta y en los de personalidad es aquella opción que indica la presencia de un nivel mayor de rasgo

en el individuo. El gráfico de dicha función matemática se denomina Curva Característica del Ítem (CCI).

Si se supone que el rasgo medido es una habilidad, de un modo intuitivo se podría conjeturar que cuanto más hábil es la persona mayor probabilidad tendrá de contestar correctamente el ítem que mide esa habilidad. De igual manera podría entenderse la respuesta a un ítem dicotómico de un test de personalidad. Por ejemplo, cuanto más alto sea el nivel de Extraversión de un individuo, mayor probabilidad tendrá de dar una respuesta afirmativa al ítem “Soy una persona conversadora”. Esta descripción racional de la respuesta a un ítem ayuda a entender que la curva de la CCI deberá ser siempre creciente.

Dentro de la TRI existen distintas familias de modelos matemáticos probabilísticos que cumplen con este requisito de monotonía creciente, pero sólo dos fueron empleados con mayor frecuencia como base para la CCI: las funciones de distribución de probabilidad normal acumulada y la logística. Ambas curvas adoptan una forma de S suavizada con un punto de inflexión en el que tienen su máxima pendiente. Presentan dos asíntotas horizontales (superior a derecha e inferior a izquierda). La superior corta al eje vertical en el valor de máxima probabilidad (igual a 1) mientras que la inferior lo corta en un valor comprendido entre 0 y 1 (Hambleton y Swaminathan, 1985).

Además del tipo de función adoptada (normal o logística), la CCI de cada ítem queda determinada cuando se especifican los parámetros que describen al reactivo. Estos parámetros se refieren a la localización del ítem en la escala del rasgo medido (b), a su potencia discriminatoria (a) y la probabilidad de dar la respuesta clave para un nivel muy bajo de rasgo (c). Por lo tanto, cada ítem tendrá una CCI propia según los valores que adopten estas propiedades psicométricas (Muñiz, 2010).

2.4.2.3 Los supuestos de la TRI

La aplicación de los modelos se asienta sobre conjunto de hipótesis que deben suponerse de partida. Asumir que la conducta de una persona ante un ítem sigue un modelo probabilístico con una forma determinada (CCI) es en sí mismo un supuesto de la TRI. La cantidad de rasgos latentes que intervienen al momento de contestar a un ítem establece una importante distinción clasificatoria de los modelos.

Si se supone que la respuesta al ítem está en función de un único rasgo los modelos son denominados unidimensionales. En la práctica es difícil que este supuesto de unidimensionalidad del rasgo latente se satisfaga totalmente ya que múltiples factores pueden afectar a la respuesta a un ítem. Por este motivo, para los modelos unidimensionales sólo se exige un rasgo fundamental (factor dominante) que explique las respuestas de los sujetos. Así también, para situaciones en las que se contempla la incidencia de dos o más rasgos afectando en la respuesta al ítem se han generado modelos multidimensionales. No obstante, estos desarrollos son relativamente incipientes y su puesta en práctica está resultando verdaderamente compleja para los investigadores.

Otro supuesto de los modelos de la TRI es la independencia local de los ítems. Este requisito supone que conocido el nivel θ de un sujeto, las respuestas a cualquier subconjunto de ítems no agregan ninguna información para el cálculo de probabilidad de respuesta a un ítem en particular. Es decir, que las respuestas a distintos ítems son estadísticamente independientes. La unidimensionalidad del rasgo latente y la independencia local de los ítems son las dos caras de un mismo requerimiento. Esto es, si se confirma que dos ítems no son independientes significa que otro factor ajeno al que se pretende medir incide en la estimación del nivel del rasgo violando el supuesto de unidimensionalidad (Lord & Novick, 1968).

2.4.2.3 Modelos de la TRI

La respuesta de una persona a los ítems obedece, por un lado, a la cantidad del rasgo que tiene dicha persona y, por otra parte, a las características propias del ítem que está contestando. Por ende, todos los modelos de la TRI vinculan el nivel del rasgo del evaluado con las propiedades psicométricas que describen al ítem y la probabilidad de optar entre las opciones del reactivo. Los tres parámetros de la CCI mencionados arriba se corresponden con los definidos en la primera generación de modelos de la TRI, los cuales son unidimensionales y suponen que la respuesta al ítem sólo admite dos opciones. Estos modelos dicotómicos se utilizan tanto para ítems que evalúan rendimiento o habilidad (correcto – incorrecto) o rasgos de personalidad (Acuerdo – Desacuerdo/ Si – No).

De la combinación del tipo de función matemática adoptada para la CCI (logística o normal) y el número de parámetros considerados (uno, dos o tres) es posible definir seis

modelos diferentes. Si un ítem es modelizado con el Modelo de Rasch, la CCI sólo se describe a partir del parámetro b y considera que el parámetro a es constante y el c es nulo para todos los ítems. El Modelo de Dos Parámetros (ML2p) contempla, además del b , el parámetro a , mientras que el c sigue considerándose nulo (Muñiz, 2010).

Por último, el Modelo de Tres Parámetros (ML3p) utiliza los parámetros b , a e incorpora el parámetro c . A modo de ejemplo, a continuación se desarrollarán las particularidades del ML3p. Su formulación es la siguiente: El símbolo θ corresponde al nivel del rasgo latente que se desea medir con el ítem i y $P_i(\theta)$ es la probabilidad de dar la respuesta clave al ítem i para un nivel dado de θ . Para el lector acostumbrado a trabajar con la TCT es importante aclarar que no se debe confundir el nivel θ con el puntaje total (observado) de una persona en un test. Θ es el equivalente de la puntuación verdadera de la TCT (Muñiz, 2010).

Esto significa que la puntuación total de una persona en un test (puntaje bruto) es una estimación de θ , de la misma manera que lo es de la puntuación verdadera en el marco de la TCT. La escala adoptada para medir θ tiene un rango teórico de $-\infty$ a $+\infty$ y su origen está determinado por consenso según la escala estandarizada, con media 0 y desvío típico 1 (Muñiz, 2010).

Después de haber expuesto las dos teorías, cabe señalar que el presente estudio está sustentado en la TCT. Como señala Muñiz (2010), lo que le ha proporcionado una vigencia prolongada a la TCT, a pesar de teorías más “recientes”, son su sencillez matemática e importancia psicológica. También su éxito se debe a que proporciona solución a una amplia gama de problemas de medida, sus supuestos son más sencillos y el análisis requiere menor complejidad estadística. En resumen, la sencillez, claridad y flexibilidad de sus conceptos, junto a la simplicidad de sus supuestos y procedimientos han posibilitado que pueda ser más viable para este tipo de estudio (tesis), en el caso de la muestra mínima requerida es 3 veces menor a lo que exige en TRI, agregando que es necesario en cuestión de disponibilidad de los examinados, ya que no es tan sencillo lograr aplicar en muchos centros educativos y también por lo del presupuesto.

Capítulo 3. Marco Contextual

En este capítulo se presenta una panorámica del contexto educativo nacional de Honduras, con énfasis en el nivel superior. Con ello se pretende dar un marco general para comprender la razón de ser de las pruebas de selección universitarias en la mayor casa de estudios superiores del país, en la que se forman y han formado a través de la historia nacional, la gran mayoría de los profesionales hondureños.

3.1 Honduras y su sistema educativo.

Honduras tiene una extensión territorial de 112,492 km² ocupado con una población de 8.6 millones de personas de las cuales el 54 por ciento vive en la ciudad y un 51 por ciento son mujeres, según el XVII Censo de Población y VI de Vivienda 2015 (Instituto Nacional de Estadísticas (INE)). Según datos del Banco Mundial en 2016, Honduras está categorizado como un país de ingreso medio-bajo que se enfrenta a desafíos significativos, con cerca del 66 por ciento de la población viviendo en la pobreza. En zonas rurales aproximadamente uno de cada 5 hondureños vive en pobreza extrema o con menos de US\$1.90 al día (INE, 2017).

En lo que concierne al financiamiento de la educación en Honduras se plantea una situación crítica. La economía del país presenta problemas de crecimiento y endeudamiento, por lo que resulta poco probable volver a disponer del 8% del PIB para invertir en educación como se contó a fines de la década pasada. El país tiene una deuda social pendiente en Educación (baja cobertura y de mejora de la calidad educativa), y el Estado tiene el compromiso de buscar fuentes alternativas de financiamiento y de hacer uso eficiente de ellos (FEREMA, 2017).

3.1.1 El Sistema Educativo

El Sistema Educativo de Honduras está conformado administrativamente de la siguiente manera: Según el Artículo 28 de la Ley Fundamental de Educación y el artículo 37 de la Ley de Educación Superior, el Consejo Nacional de Educación se creó con la finalidad elaborar y dar seguimiento a la política educativa de Honduras. Estableciendo los mecanismos de articulación horizontal y vertical del Sistema Educativo Nacional.

En el caso de la Secretaría de Estado en el Despacho de Educación en el Artículo 157 de nuestra carta magna deja bajo su administración (los centros de dicho sistema que sean

totalmente financiados con fondos públicos), la educación en todos los niveles del sistema educativo formal (Prebásico, básico y media), excepto el nivel superior.

En lo que respecta a la Educación Superior, la Universidad Nacional Autónoma de Honduras es una Institución autónoma del Estado, con personalidad jurídica, goza de la exclusividad de organizar, dirigir y desarrollar la educación superior y profesional, está contemplado en el artículo 160 de nuestra Constitución.

El Sistema Educativo Nacional de Honduras comprende cuatro niveles educativos: Educación Pre- básica, Básica, Media y Superior.

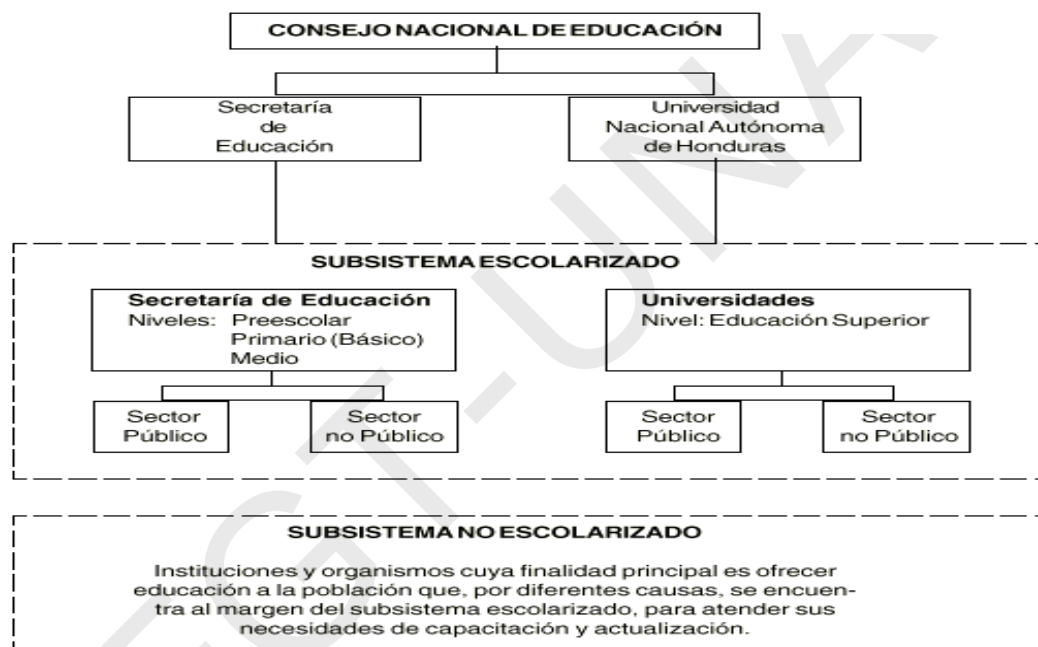


Figura 3. Organigrama del Consejo Nacional de Educación. Tomado de la Organización de Estados Iberoamericanos, OEI.

- La Educación Pre básica, con 6 años de duración, para la atención de la población de una edad comprendida entre 0 y 6 años; compuesta por dos Ciclos (0 – 3 años Primer Ciclo y 3 – 6 el Segundo Ciclo).
- La Educación Básica, con una duración de 9 años, para una edad promedio, de 6 a 15 años, compuesta por tres Ciclos.
- La Educación Media, de 2 a 3 años de duración, para una edad promedio de 15 a 17 o 18 años. Se ofrece en dos modalidades: Bachillerato Científico Humanista y Bachillerato Técnico Profesional.

En el Artículo 171 de la Constitución de la Republica de Honduras expresa lo siguiente La educación impartida será gratuita y obligatoria. Por una un año en

el nivel pre- básico y en su totalidad en los niveles básico y medio, totalmente costeados por el Estado, quien establecerá los mecanismos de compulsión para hacer efectiva esta disposición.

- Educación Superior es aquel nivel en el cual se consolida la formación integral de las personas. En la cual puede optar a los siguientes niveles: técnico, licenciatura, maestrías y doctorados. En el cual se debe producir conocimiento, desarrollar investigación e innovación, formar profesionales en el más alto nivel de especialización y perfeccionamiento, en todos los campos del saber, el arte, la cultura, la ciencia y la tecnología, a fin de atender la demanda de la sociedad y contribuir al desarrollo sostenible del país (Reglamento de la Ley Fundamental de Educación 2014 p.9)

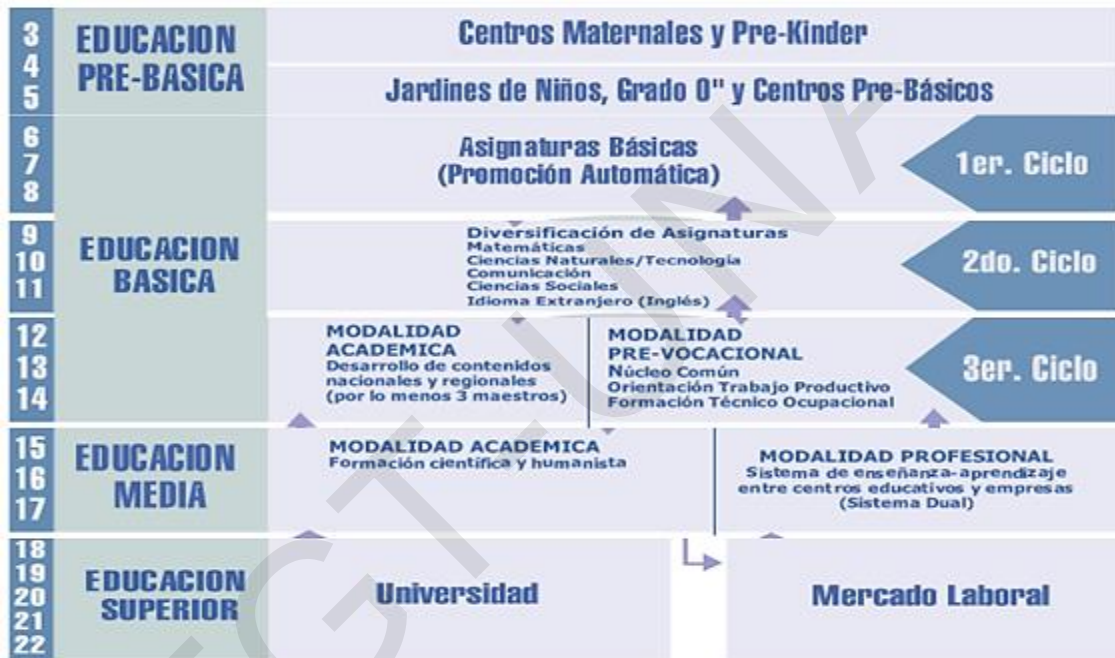


Figura 4. Estructura del Sistema Educativo de Honduras. Tomado del Currículo Nacional Básico, 2002.

3.2 El Sistema de Educación Superior Universitario

La Constitución de la República mediante el Artículo 160 otorga a la Universidad Nacional Autónoma de Honduras la dirección del Sistema de Educación Superior. La organización, dirección y desarrollo del este nivel corresponde a la UNAH, mediante los siguientes órganos:

- a) Claustro Pleno organismo superior del Nivel, conoce de segunda instancia las resoluciones de El Consejo y conoce del recurso de apelación contra las resoluciones del Consejo Educación Superior y es también un órgano de consulta para establecer criterios de doctrina académica en los asuntos en los cuales se le solicite.

b) Consejo de Educación Superior es el órgano encargado de la dirección y decisión del sistema mediante sus resoluciones y reglamentos. Esta instancia está integrado por el Rector de la UNAH quien la preside, seis miembros representantes de la UNAH, seis rectores o Directores o autoridades de otras universidades, tres corresponderán a las instituciones privadas y los otros tres a las universidades públicas.

c) El Consejo Técnico Consultivo entre sus atribuciones más destacadas está, la de elegir a los rectores o autoridades superiores de la respectiva institución, para que sean los representantes en el Consejo de Educación Superior. También a participar en la resolución de cualquier asunto de carácter general o en el caso de que el Consejo de Educación Superior solicite su opinión. Sus dictámenes no son vinculantes.

d) La Dirección de Educación Superior es el ente encargado de aplicar las resoluciones del Consejo de Educación Superior. Además actúa como secretaría del nivel y su director es el medio de comunicación y el enlace con los centros de educación superior (Ley de Educación Superior, 1994)

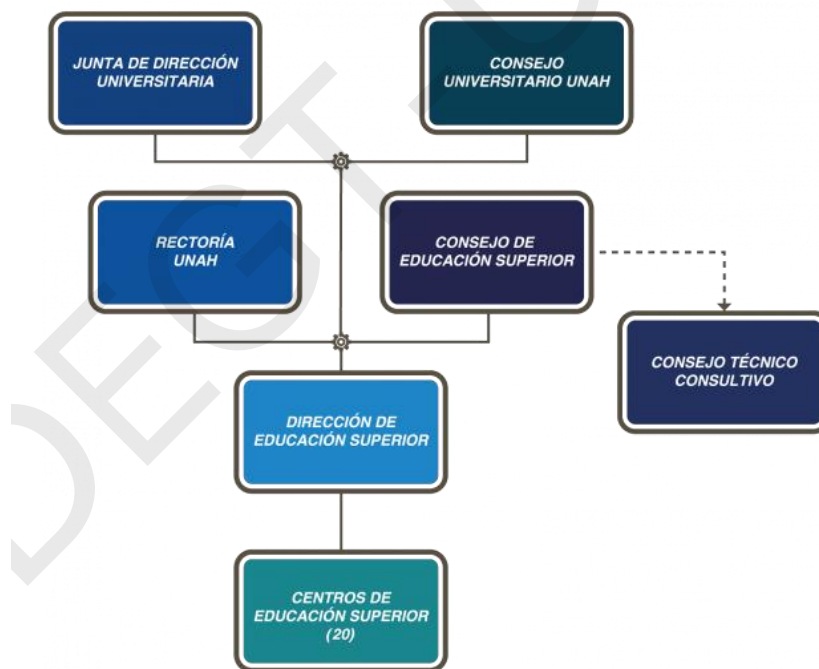


Figura 5. Organigrama del Sistema de Educación Universitario. Tomado de Dirección de Educación superior 2017).

3.2.1 Instituciones universitarias

El sistema de educación superior universitario en Honduras consta en la actualidad de 20 instituciones, 6 de administración pública y 14 de tipo privado, alcanzando una cobertura bruta de 15.65% (INE, 2015), que se clasifica como muy baja, menores del 20%, en el contexto latinoamericano (UPNFM, 2012). Presentando su matrícula la peculiaridad de que es predominantemente femenina desde finales del siglo XX. Hasta el año 1978 la única universidad del país fue la Universidad Nacional Autónoma de Honduras UNAH que señala el año 1847 como la de su fundación.

En 1978 se aprobó por el Congreso Nacional la Ley de Universidades Privadas, y ese mismo año se creó la Universidad José Cecilio del Valle, y años más tarde la Universidad de San Pedro Sula. A mediados de los años ochenta surgió en Tegucigalpa M.D.C. la Universidad Tecnológica de Honduras UNITEC, y operaban ya a nivel superior el Seminario Mayor Nuestra Señora de Suyapa SMNSS y la Escuela Agrícola Panamericana EAP. A partir de la emisión de la Ley de Educación Superior en 1989, se expande el nivel terciario de educación en Honduras (UPNFM, 2012).

En este contexto se transformaron en universidades la antigua Escuela Superior del Profesorado Francisco Morazán en la Universidad Pedagógica Nacional Francisco Morazán (1989), y la Escuela Nacional de Agricultura en Universidad Nacional de Agricultura (2003). Y surgieron más tarde universidades privadas tales como Universidad Católica de Honduras Nuestra Señora Reina de la Paz UNICAH, la Universidad Tecnológica de Honduras UTH y el Centro de Diseño, Arquitectura y Construcción CEDAC. Más tarde la Universidad Cristiana Evangélica Nuevo Milenio y la Universidad Metropolitana de Honduras (2003).

Tabla 6

Universidades de Honduras

Públicas	Privadas
▪ Universidad Nacional Autónoma de Honduras ▪ Universidad Pedagógica Nacional Francisco Morazán ▪ Universidad Nacional de Agricultura	▪ Universidad Cristiana Evangélica "Nuevo Milenio" ▪ Universidad Cristiana de Honduras ▪ Universidad José Cecilio del Valle ▪ Universidad Metropolitana de Honduras

-
- | | |
|---|---|
| <ul style="list-style-type: none"> ▪ Universidad Nacional de Ciencias Forestales ▪ Universidad de Defensa de Honduras ▪ Universidad Nacional de la Policía de Honduras | <ul style="list-style-type: none"> ▪ Universidad Católica de Honduras "Nuestra Señora Reina de la Paz" ▪ Universidad Tecnológica Centroamericana ▪ Universidad Politécnica de Honduras ▪ Universidad Politécnica de Ingeniería de Honduras ▪ Universidad Católica de Honduras "Nuestra Señora Reina de la Paz" ▪ UNITEC: Universidad Tecnológica Centroamericana ▪ Universidad Politécnica de Honduras ▪ Universidad Politécnica de Ingeniería de Honduras ▪ Universidad de San Pedro Sula ▪ Universidad Tecnológica de Honduras ▪ Escuela Agrícola Panamericana ▪ Universidad Jesús de Nazareth ▪ Centro de diseño, Arquitectura y Construcción |
|---|---|
-

Nota: Elaboración propia.

3.2.2 La normativa de la educación universitaria

La educación superior es aquel nivel educativo que contempla la última fase del proceso de formación académico es decir, aquella que viene luego de la etapa de Educación Media. Es el caso de Honduras es impartida en las universidades. La enseñanza que ofrece la educación superior es a nivel profesional. A continuación se detallará su fundamento legal.

Constitución de la República de Honduras

Mediante la máxima ley del país, en su artículo 160, le otorga a la Universidad Nacional Autónoma de Honduras personalidad jurídica para organizar, dirigir y desarrollar la educación superior y profesional.

Ley de Educación Superior

La Ley de Educación Superior, aprobada por El Congreso Nacional el 14 de septiembre de 1989, mediante el Decreto Número 142-89 y está en vigencia desde su publicación en el Diario Oficial La Gaceta No. 25961 del 17 de octubre del mismo año.

Esta Ley tiene la finalidad de fijar la organización, funcionamiento y atribuciones de la Universidad Nacional Autónoma de Honduras, para el ejercicio de sus competencias (Dirección de Educación Superior, 1989). Esta ley cuenta con su respectivo Reglamento, el cual fue aprobado en diciembre del mismo año.

Normas Académicas de Educación Superior

Con lo dispuesto en la Ley de Educación Superior y su Reglamento General en 1992 emiten las Normas Académicas del Nivel de Educación Superior. Las Normas Académicas son de obligatorio cumplimiento para las universidades. Cada universidad deberá emitir sus propias normas tomando lo aquí establecido como marco legal general que deberá regular el proceso académico. En el caso de la UNAH, las normas académicas vigentes son las del 2015.

3.3 Políticas de Admisión en la Educación Superior Universitaria

Es el proceso de admisión, mediante el cual cada institución de Educación Superior, comprueba si el aspirante al ingreso, reúne los requisitos establecidos. Al final de este proceso, la institución dará su resolución sobre su aceptación o no al nivel. Los aspirantes al primer ingreso, presentarán solicitud de admisión de conformidad a las normas de cada universidad (Normas Académicas, 2004).

3.3.1 Normativa de ingreso a educación universitaria

Según las Normas Académicas de la Educación Superior (2004), detalla, los requisitos de ingreso al Nivel Superior en el título E de Ingreso, permanencia y promoción:

Artículo 91, Son requisitos para admisión de estudiantes de pregrado: a) Solicitud formal de admisión. b) Título original y fotocopia cotejada o legalizada del mismo, que acredite los estudios cursados en el ciclo diversificado del Nivel Medio. En el caso de reconocimiento de estudios del nivel medio realizados en el extranjero, se deberá presentar acuerdo del Poder Ejecutivo. c) Certificado de aprobación de los requisitos de admisión para el nivel superior. d) Los demás que establezcan la reglamentación de cada centro.

En el caso de la UNAH, en su artículo 198 de las normas académicas, indican los requisitos generales de admisión para los aspirantes que por primera desean ingresar a dicho centro de educación de Educación Superior, aplica en aquellos estudiantes tiene su título de educación media y que no han estado matriculado en otra universidad, deberán de aplicar a la Prueba de Aptitud Académica (PAA) o la prueba institucional de admisión vigente (UNAH, 2015).

Pruebas de admisión en instituciones universitarias: Escuela Agrícola Panamericana de El Zamorano, pionera en Honduras.

En las experiencias de admisión universitaria, la universidad pionera en aplicar examen de admisión para ingresar al Nivel Superior, fue la Escuela Agrícola Panamericana Zamorano (EAP), en el año 1943, después aplicó examen en 1952 la Escuela Granja Demostrativa mejor conocida en 1968 como Escuela Nacional de Agricultura (ENA) y en 2002 como Universidad Nacional de Agricultura, tiempo después, en 1989, aplicó examen la Escuela Superior del Profesorado Francisco Morazán actualmente conocida como la Universidad Pedagógica Nacional Francisco Morazán (UPNFM).

Escuela Nacional de Agricultura (ENA) y la Escuela Superior del Profesorado Francisco Morazán

La Universidad Nacional de Agricultura (UNAG), inició en 1955 como Escuela Granja Demostrativa (EGD), en el periodo presidencial del Dr. Juan Manuel Gálvez, y el único requisito para su ingreso era tener completa la educación primaria y aprobar un examen de admisión, para así lograr su cupo. Dichos estudios estaba orientados a brindarles formación a los hijos de los campesinos y hacendados de la región.

Además de aprobar el examen de admisión, los estudiantes debían aprobar los exámenes de cada semestre para asegurar su permanencia en la Escuela, el nivel que se les otorgaba al finalizar los tres años de estudio era de Perito Agrícola. En 1968, cuando se convierte en Escuela Nacional de Agricultura, entra en vigencia la reforma académica y la denominación de Escuela Nacional de Agricultura, la institución aplica un proceso de selección riguroso, tratando de reclutar sobre todo aquellos jóvenes provenientes de la zona rural con el propósito de formarlos y mejorar su realidad socioeconómica (Portillo, 2011).

Desde el 2002, este centro de formación pasó a denominarse Universidad Nacional de Agricultura (UNAG) y en la actualidad los aspirantes a ingresar deberán cumplir los siguientes requisitos: título de educación media, aprobar el examen de admisión, certificado de buena salud y los pagos que establece cada programa (Portillo Sáenz, 2011). En la actualidad, la Universidad Nacional de Agricultura y la Escuela del Zamorano, tienen establecido como examen de admisión la Prueba de Aptitud Académica (PAA) aplicada por la Dirección del Sistema de Admisión UNAH.

En la antes denominada Escuela Superior del Profesorado Francisco Morazán, en 1958, el proceso de selección de los aspirantes, era riguroso y tenía un límite de cupo. Los estudiantes seleccionados tenían un estipendio de L.110.00 lempiras mensuales, pero mantener vigente su beca tenía una serie de requisitos: debían aprobar todas las clases y se exigía dedicación a tiempo completa, de no cumplir con los requerimientos de la beca, los estudiantes no podrían continuar sus estudios (Portillo, 2011).

El 14 de diciembre de 1989, se convierte en la Universidad Pedagógica Nacional Francisco Morazán (UPNFM). Actualmente la UPNFM, diseña y administra su propio examen de admisión el cual comprende una serie de apartados: 1) Español, 2) Matemáticas, 3), Ciencias Naturales 4) Ciencias Sociales y 5) Vocación docente. Una vez aprobado el examen, el estudiante deberá cumplir con los requisitos de ingreso que la institución establece (Universidad Pedagógica Nacional Francisco Morazán, s.f.)

3.4 El proceso de admisión y de selección de la UNAH

Los aspirantes a ingresar a la Universidad Nacional Autónoma de Honduras (UNAH) deben cumplir con los pasos del proceso de admisión correspondiente al período académico al cual desean ingresar. Cada año, la UNAH aplica tres pruebas. La primera, para ingresar en el tercer período académico del mismo año; la segunda y la tercera, para ingresar en el primer período académico del año siguiente. La Prueba de Aptitud Académica tiene dos componentes: Razonamiento Verbal y Razonamiento Matemático.

Los aspirantes a las carreras de Medicina, Odontología, Química y Farmacia y Microbiología deben realizar, tiene con requisito adicional someterse a la Prueba de Conocimientos de Ciencias Naturales y de la Salud (PCCNS) diseñada también por el College Board. La PCCNS evalúa los conocimientos en las áreas de Biología, Química y

Física / Matemática. Los aspirantes deben cumplir con el índice de admisión mínimo establecido para su carrera en la PAA y el mayor puntaje en la PCCNS para competir por un cupo.

Existen otras carreras con requisitos específicos para su admisión. Entre ellas se encuentra: Arquitectura, Psicología y Música entre otras. (Dirección de Sistema de Admisión UNAH, 2017)

3.4.1 Prueba de Aptitud académica (PAA), Prueba de Ciencias Naturales y de la Salud (PCCNS) y otras específicas.

La UNAH, principal centro de educación superior de Honduras que acumula más del 40% de la matrícula total del sistema, operó durante más de un siglo, desde el año 1847 hasta el 2002, sin establecer mecanismos de ingreso y/o selección para sus aspirantes de primer ingreso. Solo recientemente, a inicios del presente siglo se comienzan a dar pasos para regular este proceso, y en el año 2002 se estableció el acuerdo No. 948-143-2002, mediante el cual el Consejo de Educación Superior reforma los artículos 90, 91, y 94 del capítulo VI de las Normas Académicas, donde elimina las excepciones de requisitos de admisión para la UNAH.

Cuatro años más tarde, en el año 2006 la Comisión de Transición creó la Dirección del Sistema de Admisión para organizar y dirigir el proceso de admisión de la UNAH llevada a cabo a través de la Prueba de Aptitud Académica del College Board, la cual se ha aplicado en 32 procesos hasta el 2016. Esta Prueba de Aptitud Académica está diseñada para evaluar a los solicitantes al sistema universitario, al igual que se ha venido aplicando en Estados Unidos desde hace más de 100 años. Desde sus inicios ha sido utilizada por las universidades tanto de Estados Unidos como de varios países de América Latina, como criterio externo para sistematizar y uniformar en lo posible el proceso de admisiones a los centros de educación superior (College Board, 2013)

En relación con su adaptabilidad al contexto hondureño de esta prueba internacional, se realizó una evaluación al respecto y según el informe interpretativo de los análisis estadísticos y psicométricos de las pruebas de Aptitud Académica aplicada en la UNAH (2013), se concluye que en todas las pruebas estadísticas y psicométricas realizadas se muestran evidencias suficientes para concluir que la batería de pruebas de la PAA tiene

características psicométricas apropiadas para la población examinada, entendiendo por lo anterior que es útil para la evaluación de los aspirantes a la UNAH (College Board, 2013)

Sin embargo, gradualmente ha venido surgiendo el cuestionamiento de si esta prueba por sí misma es suficiente como criterio de admisión a todas las carreras. Una revisión de la literatura sobre el poder predictivo de las pruebas de admisión a las universidades señala que estas pruebas, por lo general, ayudan a predecir el éxito de los estudiantes en su carrera universitaria, sin embargo el informe presenta la recomendación de no utilizar la PAA como único criterio de admisión (College Board, 2013)

En la misma UNAH han surgido peticiones al respecto. Ya en el año 2011 la carrera de Medicina de la UNAH decide incorporar la Prueba de Conocimiento de las Ciencias Naturales y de la Salud PCCNS, con el propósito de mejorar sus criterios de admisión y tratar de escoger a aquellos aspirantes con mayores posibilidades de cursar con éxito esta carrera de medicina (Sosa, 2016). Siguiendo el ejemplo de la Carrera de Medicina las carreras de Microbiología, Odontología, Química y Farmacia han analizado el desempeño de sus estudiantes y han incorporado como criterio de admisión la PCCNS. (Acuerdos del Consejo Universitario, 2014, 2015). De forma análoga, en el año 2014 la Carrera de Arquitectura realiza una investigación con personal de la carrera de Psicología para utilizar pruebas complementarias que midan habilidades específicas que puedan predecir de forma más eficaz el futuro desempeño universitario de sus aspirantes.

Todo lo anterior plantea la necesidad de que la UNAH desarrolle una estrategia de ampliar y enriquecer el proceso de selección de sus aspirantes. Tanto en relación con la admisión en general, como respecto a su ingreso en determinadas carreras que la institución oferta. Siempre bajo el supuesto de que una mejor selección ayudará a mejorar la eficiencia terminal de sus estudiantes. Es en este contexto que se ubica la conveniencia de desarrollar, entre otras, una prueba de habilidades cognitivas de orden superior, asociadas al campo de la Educación de las Ciencias Sociales.

Otras pruebas específicas

Arquitectura

Basados en el análisis de la capacidad Instalada de la carrera y considerando criterios de inclusión, diversidad y equidad, será un total de 164 estudiantes los admitidos en el año.

El ingreso estará definido a través de un ranking; en donde como primer requisito será el obtener 900 puntos en la PAA, sin embargo, los 164 mayores puntajes, del ponderado entre la Prueba de Aptitud Académica (PAA) y las Pruebas Psicológicas Específicas para la Carrera de Arquitectura (PPEA), de las diferentes categorías de selección son quienes estarían admitidos a Arquitectura (UNAH, 2017)

Música

En el caso de la carrera de Música solicita al aspirante, Aprobar la Prueba de Conocimientos Teóricos e Instrumentales (PCTI) comprendida por pruebas de solfeo, armonía, historia e instrumenta (Cada aspecto debe aprobarlo con los porcentajes mínimos requeridos) y aprobar la PAA con un puntaje mínimo de 700 puntos. (UNAH, 2017).

3.4.2 Políticas de inclusión en el proceso de admisión UNAH.

En este apartado se presentan los datos estadísticos y el análisis de los resultados de las 3 Pruebas de Aptitud Académica aplicadas en el 2014 en comparación con años anteriores, con el objetivo de representar las políticas de inclusión en la UNAH. Un informe estadístico proporcionado por la Dirección del Sistema de Admisión proporciona la siguiente información.

“Un dato importante que se debe destacar es que el proceso de admisión no discrimina a los aspirantes que tienen bajos recursos económicos ya que se puede ver en la siguiente gráfica que la mayor concentración de aspirantes reportaron ingresos familiares entre 5,000 a 10,000 lempiras y menos de 5,000 lempiras. Desde el 2006 hasta el 2014 alrededor de 79,044 aspirantes reportaron tener ingresos entre los 5 y 10 mil lempiras, 29232 entre 10 mil y 15 mil lempiras, 16040 aspirantes reportaron ingreso entre los 15 y 20 mil lempiras, 14414 entre los 20 y 25 mil lempiras, solo 1704 reportaron más de 25 mil lempiras, unos 997 reportaron tener ingresos arriba de los 30 mil a 35 mil lempiras y una pequeña parte más de 35 mil lempiras en adelante (Dirección de Sistema de Admisión, UNAH, 2014, pp.11).

Expuesto lo anterior, en lo que respecta a la admisión tomando en consideración los aspirantes admitidos según ingreso familiar mensual, no representa un factor de no admisión. En el caso de la admisión de aspirantes con Necesidad Especiales en los procesos que comprenden desde el 2006 hasta el 2014 han aplicado a la PAA un total de 306 aspirantes, de los cuales 183 (60%) han sido admitidos y 123 (40%) no han sido admitidos. Siendo los de baja visión severa los que han logrado una mayor admisión 34%, seguido por los que

tienen ceguera total (25.1%). Pero tomando en consideración este intervalo de años de aplicación de la PAA, presenta un desafío en lo que respecta a los aspirantes sordos, ya que de 39 aspirantes solamente 5 han logrado la admisión a la UNAH, por ello la Dirección de Sistema de Admisión junto Programa de Servicios a Estudiantes Con Necesidades Especiales (PROSENE) propondrían enriquecer los criterios de admisión (Dirección de Sistema de Admisión, UNAH, 2014).

En lo que respecta a la admisión de los pueblos indígenas y afrodescendientes, la Dirección de Sistema de Admisión comenzó a solicitar desde el año 2011 en su página web de admisión como una petición del Comisionado Universitario para conocer sobre la inclusión de los Pueblos Indígenas y su acceso a la educación superior en el proceso de admisión de la UNAH. A partir del 2011, unos 9,590 aspirantes que realizaron la PAA se identificaron perteneciente a un pueblo indígena o afrohondureño, lo que representa aproximadamente un 10% de la población total de aspirantes que realizan la Prueba para ingresar a la UNAH. De los 8 pueblos aspirantes a la PAA, el Pueblo Lenca es el que más solicitudes de admisión a la UNAH han reportado con un 6%, seguido por el pueblo Garífuna con 1.3%, los Misquitos 0.8%, los Chortí 0.7%, Tolupanes 0.5%, Pech 0.4%, Isleños 0.2%, y Tawakhas 0.1%.

La UNAH ha tratado de enriquecer su proceso de admisión y de permanencia acompañándolo de políticas de inclusión y de discriminación positiva, como las siguientes:

1. En el año 2016 se creó el Área de Inclusión con el propósito de que los estudiantes con necesidades especiales y los estudiantes procedentes de pueblos indígenas y afrodescendientes, puedan gozar de un entorno inclusivo. Procurando proporcionarles un asesoramiento que les permita un desenvolvimiento pleno en el ámbito de la vida estudiantil y posteriormente integrarse a la sociedad con igualdad de oportunidades.
2. El Área de Inclusión presta servicios de Accesibilidad a los espacios, tutorías y adaptaciones en general para el proceso de la PAA
3. Mediante el Programa de Atención Socioeconómica y Estímulos Educativos (PASEE), la UNAH ofrece becas de equidad social (becas de inclusión) que son para pueblos originarios o personas con discapacidad.

3.5 Facultad de Ciencias Sociales

La Facultad de Ciencias Sociales (FCCSS) fue creada por la Comisión de Transición de la UNAH, mediante Acuerdo Número Trescientos Sesenta y Ocho (No.368-2008), de fecha quince de mayo de dos mil ocho (15-05-2018). En la actualidad la Facultad, ha diversificado su oferta académica a nivel de grado y postgrado, distribuidos de la siguiente manera: 8 carreras de pregrado (Periodismo, Psicología, Historia, Antropología, Sociología, Desarrollo Local y el Técnico en Desarrollo Municipal). Respecto a los programas de maestría, en el 2016 la Facultad amplió su oferta académica, iniciando las promociones de los másteres de Estado y Políticas Públicas, Sociología, Historia Cultural y Social y Psicometría y Evaluación Educativa, teniendo un total de 7 programas de maestría. Actualmente solo ofrece un doctorado.

Además de esos programas académicos de pregrado y posgrado, el Instituto Universitario en Democracia Paz y Seguridad (IUDPAS) está adscrito a la FCCSS desde su creación el 14 de febrero del año 2008, en años más recientes se ha creado el Instituto de Investigaciones Sociales, el Observatorio Demográfico Universitario y el Observatorio Universitario en Seguridad Alimentaria y Nutricional. También están bajo la responsabilidad de la facultad las asignaturas generales de Sociología, Historia de Honduras, Ciencia Política, Derechos Humanos en Honduras, Estudios de la Mujer, Psicología General y Antropología General.

3.5.1 Proceso de ingreso a la Facultad de Ciencias Sociales

FACULTAD DE CIENCIAS SOCIALES OFERTA ACADÉMICA Índice PAA requerido					
Centros	Modalidad	CARRERA	Duración	TÉCNICO	Índice PAA
CU	Presencial	Antropología	5.0 años		900
CU UNAH-VS	Presencial	Periodismo	4.0 años		700
CU UNAH-VS	Presencial	Psicología	4.5 años		900
CU	Presencial	Trabajo Social	5.0 años		700
CU	Presencial	Historia	5.0 años		700
CU UNAH-VS	Presencial	Sociología	5.0 años		700
CU CUROC UNAH TEC-TELA	Presencial	Desarrollo Local	5.0 años		700
CU	Presencial		2.5 años	Desarrollo Municipal	700

Figura 6. Puntajes mínimo requerido para las carreras de la Facultad de Ciencias Sociales, UNAH. Tomado de Facultad de Ciencias Sociales, UNAH.

En la figura 6, se puede apreciar el puntaje mínimo requerido en la PAA para poder ingresar a cada carrera de la FCCSS. De ocho, dos programas de pregrado como Antropología y Psicología son las licenciaturas que mayor índice requieren, siendo 900 puntos los necesarios para ser admitido. Las 6 carreras restantes de esta facultad, requieren un puntaje mínimo de 700. De estos pregrado, la licenciatura de Psicología, tiene como requisito adicional de admisión, un estudio vocacional en el Centro de Asistencia Psicológica (CAPs) de la Escuela de Ciencias Psicológicas.

3.5.2 Carrera de Antropología

La carrera para la cual el presente estudio elabora la prueba de selección es precisamente la licenciatura de Antropología adscrita a la Facultad de Ciencias Sociales. Esta disciplina es el estudio científico y humanístico de la especie humana como un todo, el pasado, el presente y el futuro; es comparativa y transcultural, analizando datos de poblaciones períodos diferentes. La carrera de Antropología de la UNAH tiene como finalidad formar profesionales en las áreas particulares de antropología sociocultural y arqueología, formándolos para participar en todas las etapas del proceso de investigación antropológica y con capacidades para elaborar y difundir conocimiento antropológico en sus diferentes niveles.

Esta carrera dio comienzo a funcionar en el 2012, se matricularon aproximadamente 50 estudiantes. En la graduación de marzo de 2018, esta carrera tuvo su primera y hasta el momento la única graduada, se trata de Ridel Giselle Morales Ochoa, quien estudió Licenciatura en Antropología, con orientación en Arqueología. La duración de la Carrera de Antropología es de 5 Años (15 períodos académicos), para alcanzar el grado de Licenciatura y el título a obtener es el de Licenciado (a) en Antropología con Orientación en Antropología Sociocultural o Arqueología. Los requisitos de Graduación establecidos son aprobar el mínimo de 176 unidades valorativas Aprobar todas las asignaturas, talleres y laboratorios, prácticas y actividades co-programáticas del plan de estudios, 800 horas de práctica profesional supervisada. Además, cumplir con las 40 horas de trabajo comunitario interno de la UNAH Presentar constancia de examen TOEFL con mínimo de 75 puntos o una prueba equivalente, y elaborar y defender la correspondiente tesis de grado.

En una entrevista, a la Jefa y Coordinadora del Departamento de Antropología MSc. Carmen Fajardo opinó lo siguiente respecto a la prueba: Si pienso que son importantes porque para estudiar ciertas carreras se requieren ciertas habilidades y aptitudes y es de beneficio tanto para las carreras como para los estudiantes conocer el grado de compatibilidad con ciertas áreas de estudio. Con las pruebas de admisión y las pruebas específicas es posible identificar los potenciales de los estudiantes, al igual que determinar el nivel de conocimiento sobre la temática que tienen los estudiantes. De esa forma los estudiantes pueden tomar decisiones informadas sobre la carrera que desean estudiar.

Me parece muy importante las pruebas de selección universitaria específicas a la carrera de Antropología porque ayuda a seleccionar estudiantes que ingresan a la carrera sabiendo que tienen un interés genuino en la Antropología y que han mostrado su potencial para cursar la carrera exitosamente al poseer aptitudes y habilidades compatibles con el quehacer de la Antropología. Respecto al porcentaje a la prueba le asignaríamos 60% PAA y 40% prueba específica.

Capítulo 4. Metodología

El estudio se orienta a diseñar una prueba basada en expectativas de logro en relación con habilidades cognitivas para aspirantes a ingresar a la Licenciatura en Antropología de la Facultad de Ciencias Sociales de la UNAH, complementaria a la PAA.

El presente capítulo ofrece el proceso que se siguió para el diseño de esta prueba estandarizada de selección universitaria. El propósito de este apartado consiste en ofrecer al lector algunas estrategias para desarrollar este tipo de prueba de acuerdo con los estándares actuales en materia de medición y psicometría.

4.1 Enfoque, alcance y Diseño

Según el nivel de profundidad del estudio se tiene que el producto será un instrumento de medición, un test, y en este sentido, toda medición es en sí misma **descriptiva**. Sin embargo, para la elaboración de este instrumento de medición, un test de niveles cognitivos, será necesario hacer análisis de tipo relacional, y en este sentido el estudio es **relacional**.

Según la temporalidad el presente estudio orientado a la elaboración de un test de habilidades cognitivas es un estudio **transversal**. Dado que la investigación se centra en analizar cuál es el nivel o estado de una o diversas variables en un momento dado o bien en cuál es la relación entre un conjunto de variables en un punto en el tiempo. En este tipo de diseño se recolectan datos en un solo momento, en un tiempo único y no se realizan comparaciones de la o las variables principales a través del tiempo (Hernández, Fernández y Baptista, 2014).

Según el tipo de fuente el presente estudio de elaboración de un test de niveles cognitivos, es **de campo**. La investigación de campo se presenta mediante la manipulación de una variable externa no comprobada, en condiciones rigurosamente controladas, con el fin de describir de qué modo o porque causas se produce una situación o acontecimiento particular (Hernández, et al. 2014).

Según su finalidad la investigación es **aplicada**. Para Murillo (2008), la investigación aplicada recibe el nombre de “investigación práctica o empírica”, que se caracteriza por que busca la aplicación o utilización de los conocimientos adquiridos, a la vez que se adquieren otros, después de implementar y sistematizar la práctica basada en investigación. El uso del

conocimiento y los resultados de investigación que da como resultado una forma rigurosa, organizada y sistemática de conocer la realidad.

4.2 Variables de investigación

El estudio se enfoca en la elaboración de un test basado en las expectativas de logro de los espacios curriculares de Antropología y Fundamentos de Investigación Social desarrollados en el Bachillerato de Ciencias y Humanidades clasificadas según habilidades cognitivas. Tal como fueron definidas en el Marco Teórico, las expectativas de logro son entendidas como las intencionalidades educativas y expresan el grado de desarrollo de las competencias del área de tipo cognitivo, procedimental y valorativo/actitudinal que la Educación Media debe garantizar equitativamente a los alumnos y alumnas (Secretaría de Educación, DCNB Tercer Ciclo, 2003).

Tabla 7

Operacionalización de las variables

Variables	Definición conceptual	Definición Procedimental	Dimensiones	Indicadores
1. Expectativas de logro	Las expectativas de logro explicitan las intencionalidades educativas y expresan el grado de desarrollo de las competencias del área de tipo cognitivo, procedimental y valorativo/actitudinal que la Educación Media debe	Por medio de una prueba de selección construida bajo los parámetros de la TCT	Relación entre Expectativas de Logro y habilidades cognitivas: Reconocer Comprender Aplicar Analizar Evaluar	Dominio en cada nivel: Bajo Promedio Alto

	garantizar equitativamente a los alumnos y alumnas (Secretaría de Educación, DCNB Tercer Ciclo, 2003, p.32).	Crear
2. Habilidades Cognitivas	Las habilidades cognitivas son procesos y métodos que puede emplear el estudiante para obtener, retener y recuperar diferentes tipos de conocimientos. El estudiante supone capacidades de representación como: • La lectura, imágenes, habla, escritura y dibujo. • Capacidades de selección: atención e intención. • Capacidades de autodirección: auto programación y	

autocontrol (Rigney,
1978, pág. 165).

Nota: Elaboración propia.

Se utilizó las expectativas de logro según habilidades cognitivas de los espacios curriculares de Antropología y Fundamentos de Investigación Social contemplados en la Malla Curricular del Bachillerato de Ciencias y Humanidades que los aspirantes a la carrera de Antropología de la Facultad de Ciencias Sociales UNAH requieren como una formación base. Las variables principales del estudio son, expectativas de logro y habilidades cognitivas, que se clasificaron según contenido y niveles cognitivos. Este procedimiento se analizó bajo la TCT y para la interpretación de los resultados de la prueba se utilizó la métrica de scale score.

4.3 Población y Muestra

Población. La población principal del estudio son los estudiantes de undécimo grado del Bachillerato en Ciencias y Humanidades que a nivel de Tegucigalpa y Comayagüela están cursando este grado durante el año 2018. Se consideran principalmente los estudiantes de este Bachillerato, porque esta modalidad de Educación Media es la que tiene un currículo diseñado para estudiar las carreras de la Facultad de Ciencias Sociales y específicamente la Licenciatura de Antropología, ya que contemplan ambos curriculares.

Tipo de Muestreo

Se aplicó un muestreo no aleatorio intencional o de conveniencia, tomando en consideración los siguientes criterios:

- Los centros educativos que ofertaran el Bachillerato de Ciencias y Humanidades.
- Se seleccionó los centros educativos de Educación Media de un listado de donde proceden la mayoría de aspirantes (1ra y 2da opción) a la carrera de Antropología en el Distrito Central, en nueve procesos de admisión (2016, 2017 y 2018). Los centros educativos seleccionados fueron los siguientes: Instituto Gubernamental Central Vicente Cáceres, Instituto Gubernamental Mixto Hibueras, Centro Educativo No Gubernamental Católico Emiliani, Instituto Gubernamental San José del Pedregal, Instituto

Gubernamental Héctor Pineda Ugarte e Instituto Gubernamental Blanca Adriana Ponce Ponce (Dirección de Sistema de Admisión, 2018).

- Los centros educativos que participaron en el Taller de Redacción de Ítems, coinciden con varios centros educativos del listado proporcionado por la Dirección de Sistema de Admisión, UNAH.
- Se eligió centros educativos con población estudiantil de ambos géneros.

Una vez seleccionados los centros educativos, se visitó a los directores y se les entregó una solicitud para aplicar el instrumento en la cual se informaba los detalles de la aplicación de la forma Piloto y después la Operativa. El proceso se aplicará en la modalidad estratificada por sexo,

Tamaño de la muestra. A partir de los centros educativos de Educación Media seleccionados, se enfocó en los estudiantes de undécimo grado del Bachillerato de Ciencias y Humanidades. Para el desarrollo de una prueba utilizando como base la TCT, según la literatura especializada en Psicometría, se considera aceptable un mínimo de 75 sujetos (Martínez, M. R., Hernández M.J. y Hernández, M.V., 2006).

Aplicación Piloto:

Se aplicó la Prueba de Selección Universitaria para la Licenciatura en Antropología a 204 estudiantes, desglosados de la siguiente manera:

Instituto Gubernamental Central Vicente Cáceres (48 estudiantes), Instituto Gubernamental Mixto Hibueras (62 estudiantes), Centro Educativo No Gubernamental Católico Emiliani (27 estudiantes), Instituto Gubernamental San José del Pedregal (13 estudiantes), Instituto Gubernamental Héctor Pineda Ugarte (54 estudiantes).

Prueba de Aplicación Operativa:

Se aplicó la Prueba de Selección Universitaria para la Licenciatura en Antropología a 158 estudiantes, desglosados de la siguiente manera:

Centro Educativo No Gubernamental Católico Emiliani (29 estudiantes), Instituto Gubernamental, Instituto Gubernamental Héctor Pineda Ugarte (72 estudiantes) y Blanca Adriana Ponce Ponce (57 Estudiantes).

4.4 Instrumentos

En esta investigación cuantitativa aplicamos un *instrumento* para *medir* las variables definidas en este estudio. En la cual se trató de acercarse a lo más posible a la representación fiel de las variables a observar, mediante el instrumento de medición que se desarrolló.

Tabla 8

Ficha técnica de la prueba operativa (final)

Característica	Detalle
Grado o edades de la población meta	Estudiantes de undécimo grado del Bachillerato de Ciencias y Humanidades
Tiempo de duración de la evaluación	80 minutos
Número de formas operativas	1
Número total de ítems en la aplicación operativa	60
Tipos de ítems y valor (puntaje) de cada uno	Selección Única
Tipo de administración	Autoadministrado (papel y lápiz)
Sub-grupos de interés en el análisis	Estudiantes por género.
Validez	Validez de Contenido por Juicio de Expertos.
Confiabilidad	Alfa de Cronbach.

Nota: Adaptado de American Institutes For Research (AIR, 2015).

En la Tabla 8, se describió el instrumento que se diseñó. En esta investigación se utilizó la técnica de recolección de información denominada encuesta. Según y Meneses y Rodríguez (2016). “Un cuestionario es, por definición, el instrumento estandarizado que utilizamos para la recogida de datos durante el trabajo de campo de algunas investigaciones cuantitativas, fundamentalmente, las que se llevan a cabo con metodologías de encuestas” (p.9). Se utilizaron cuestionarios auto administrados en las dos aplicaciones (piloto y operativas) de la Prueba de Selección Universitaria para la Carrera de Antropología, la primera constaba de 100 ítems y de dos horas de aplicación y la segunda de 60 ítems y 80 minutos de tiempo para contestarla. La tabulación de datos y calificación se realizó con el programa Startical Product and Service Solutions (SPSS), también para el análisis y además se utilizó Microsoft Excel para esta etapa de la construcción de la prueba.

Para todo profesional de la psicometría, es importante ofrecer un instrumento válido y confiable. Por esta razón, se realizó el análisis de propiedades psicométricas de validez de Contenido por Juicio de Expertos y la Confiabilidad Alfa de Cronbach. Puerta y Marin (2015) definen que “La validez de contenido se refiere al grado en que un instrumento refleja un dominio específico de contenido de lo que se mide” (párr.5). Esta prueba está basada en los contenidos de los espacios curriculares de Antropología y Fundamentos de Investigación Social, por lo tanto se requirió de expertos en el área, para que emitieran juicios sobre la Tabla de Especificaciones. Debido a la finalidad de esta prueba, se eligió utilizar el método de consistencia interna Alfa de Cronbach, porque se tienen grupos de ítems que miden dos campos de contenido, Antropología y Fundamentos de Investigación Social y la literatura especializada recomienda utilizar en este caso.

4.5 Plan de análisis

Se aplicó los procedimientos y estándares establecidos en la Teoría Clásica de los Test para medir, los índices de dificultad, discriminación, análisis de distractores, validez de contenido por juicio de expertos, confiabilidad de Alfa de Cronbach y el Funcionamiento Diferencial del Ítem de cada uno de los ítems de la prueba.

4.5.1 Procedimiento

En este apartado se presentó de forma sintética las etapas que se sugieren para construir un instrumento de medida con garantías de calidad. Este proceso de construcción de una prueba, es un proceso complejo que para la finalidad de este estudio se adaptó a trece pasos, si bien estos no son universales, pueden variar en función del propósito del instrumento de medida, del modelo psicométrico, del tipo de respuesta exigida por los ítems, del formato de aplicación, por exponer algunos casos. Todo el proceso de construcción debe aplicarse de forma rigurosa y objetiva, estableciendo estándares de calidad, para así maximizar la validez de las inferencias hechas a partir de los puntajes de los estudiantes, así como la equidad en la prueba de las personas evaluadas (Dorans y Cook 2016; Downing, 2006; Lane, Raymond y Haladyna, 2016). En la Figura 1, se detallan las Etapas de la Construcción de la Prueba de Selección Universitaria para la Licenciatura en Antropología, que se explicará a continuación:

Tabla de Especificaciones

La elaboración de esta tabla de especificaciones requirió de una fase previa de selección de los propósitos y contenidos representativos contemplados en los programas curriculares de Antropología y Fundamentos de Investigación Social y en el taller los docentes especialistas de Ciencias Sociales darán validez a esos contenidos o propondrán otros. Estos elementos fueron la base para determinar el nivel cognitivo a valorar según la Taxonomía de Bloom revisada por Anderson & Krathwohl, el número y tipo de reactivo (selección única). La cantidad de ítems que se le asignará a cada expectativa de logro y contenido, será tomando el tiempo que se ha destinado para su desarrollo y el verbo de cada expectativa determinará a qué nivel cognitivo pertenece. El formato de la tabla puede variar y agregar, incluso, información adicional a los aspectos mencionados en la Tabla 9 (Sans Martín, 2008)

Redacción de Ítems

En esta etapa se necesitan dos documentos importantes para la construcción de ítems:

- Marco conceptual para la construcción de una prueba.
- Tabla de Especificaciones para la construcción de una prueba.

En esta etapa, se desarrolló un Taller de Redacción de Ítems de la línea de investigación Pruebas de Selección Universitaria dirigido y organizado por las maestrandas Keyla Rivera, Tessy Solórzano y Nusly Moreno. En lo que respecta a los ítems de esta prueba, se invitó a 13 docentes del área de Ciencias Sociales que imparten clases en las modalidades de Educación Media, a los cuales se les capacitó en criterios generales y específicos para la redacción de Ítems, esta actividad se desarrolló en el Instituto de Profesionalización y Superación Docente de la UNAH. En este taller se enfatizó la importancia de que los ítems redactados deben estar alineados a las expectativas de logro y habilidades cognitivas de la taxonomía de Bloom revisada por Anderson y Krathwohl, incluidas en la tabla de especificaciones de la prueba.

Validación de Contenido por juicio de Expertos

En esta etapa se realizó una presentación sobre el trabajo académico a realizar a los Docentes de la Carrera de Antropología UNAH-CU. Ellos, como especialistas de Contenido conjuntamente con la maestranda, revisaron detenidamente cada uno de los ítems redactados. Tenían que expresar sus juicios en una ficha, en la cual tenían que evaluar los ítems de acuerdo a los siguientes criterios: veracidad, claridad, coherencia y relevancia. Y posteriormente la investigadora calculó el índice de validez de contenido (CVI), para así seleccionar los ítems para el ensamblaje de la forma piloto.

Lawshe (1975) propuso un índice de validez basado en la valoración de un grupo de expertos de cada uno de los ítems del test como innecesario, útil y esencial. El índice se calcula a través de la siguiente fórmula:

$$IVC = \frac{n_e - N/2}{N/2}$$

Donde n_e es el número de expertos que han valorado el ítem como esencial y N es el número total de expertos que han evaluado el ítem.

Ensamble de Forma Piloto

Una vez que se incorporó las recomendaciones proporcionadas por el juicio de expertos, se procedió a ensamblar las formas piloto. El propósito de ensamblar la forma piloto y aplicarla fue obtener un banco de ítems, tener suficientes ítems validados (de acuerdo a los parámetros y sus valores aceptados expuestos en las tablas 10, 11 y 12) para la forma final (forma operativa), para así poder construir la prueba aplicando la Teoría Clásica de los Test (TCT). En este caso se partió de cero, no se contaba con un banco de ítems, para realizar esta investigación.

Revisión de la Prueba Ensamblada

Con la prueba piloto diagramada, se presentó el instrumento a la Jefa del Departamento MSc. Carmen Fajardo y a los docentes que participaron en la validez de

contenido, para que corroboraran que las recomendaciones incorporadas en la prueba ensamblada.

Aplicación de Forma Piloto

En todo sistema de evaluación es necesario crear y validar ítems para poder obtener un banco de ítems. El pilotaje que se hizo nos permitió identificar:

- a) Problemas en la logística de la aplicación.
- b) Errores de redacción de los ítems.
- c) Si el tiempo estimado para resolver la prueba operativa sería suficiente, entre otros.

En este sentido la prueba se aplicó a una cantidad requerida de estudiantes (según el tamaño de la muestra en la TCT), los datos obtenidos se utilizaron en el análisis psicométrico y posteriormente se ensambló la prueba operativa con aquellos ítems que cumplieron con los criterios tanto de calidad estadística como de contenido.

Una vez finalizado el proceso de ensamble e impresión de la prueba piloto, esta fue aplicada a una muestra de 204 estudiantes, distribuidos de la siguiente manera: Instituto Gubernamental Central Vicente Cáceres (48 estudiantes), Instituto Gubernamental Mixto Hibueras (62 estudiantes), Centro Educativo No Gubernamental Católico Emiliani (27 estudiantes), Instituto Gubernamental San José del Pedregal (13 estudiantes), Instituto Gubernamental Héctor Pineda Ugarte (54 estudiantes).

Una vez finalizada la aplicación piloto se analizaron los ítems para determinar si cumplían con los criterios para poder ser incluidos en la prueba operativa. Durante este proceso se aplicó los principios de la Teoría Clásica de los Test (TCT). También se analizó el Funcionamiento Diferencial del Ítem según género (femenino y masculino).

Análisis de ítems

El análisis de los ítems se realizó con la Teoría Clásica de los Test por medio de programas estadísticos, se calcularon los índices de dificultad, discriminación, el análisis de distractores así como el Funcionamiento Diferencial del Ítem (FDI) según la variable de género.

Revisión de Estadísticas

Durante la fase de revisión de estadísticas de la Aplicación Piloto se analizó la calidad de cada uno de los ítems en función de los indicadores psicométricos de calidad según la TCT. En este proceso los ítems fueron clasificados en una de estas tres categorías:

- a) Aprobado: el ítem puede ser usado sin reservas en la aplicación operativa.
- b) Revisión, necesita enmiendas: estos ítems pueden incluirse en una prueba operativa después de hacerles pequeños cambios de contenido y de repilotarlos para obtener nuevas estadísticas.
- c) Rechazado: el ítem no cumple con los criterios de calidad psicométricos.

Ensamble de Forma Operativa

En esta etapa se seleccionaron aquellos ítems que fueron clasificados como “Aprobados” durante el proceso de Revisión de Estadísticas. La cantidad de ítem que debe incluirse en cada prueba depende de la Tabla de Especificaciones de la prueba, en este caso fueron 60 ítems. El Ensamble de Forma Operativa requirió un minucioso análisis de los ítems, porque era necesario que tuvieran los valores aceptables en dificultad, discriminación, análisis de distractores y Funcionamiento Diferencial del Ítem y otros valores indispensables de la prueba en general. En el caso de la fiabilidad se calculó el Alfa de Cronbach para ambas subpruebas.

Aplicación de la Prueba Operativa

Se aplicó la Prueba de Selección Universitaria para la Licenciatura en Antropología a 158 estudiantes, desglosados de la siguiente manera:

Centro Educativo No Gubernamental Católico Emiliani (29 estudiantes), Instituto Gubernamental, Instituto Gubernamental Héctor Pineda Ugarte (72 estudiantes) y Blanca Adriana Ponce Ponce (57 Estudiantes).

Análisis de los resultados

El análisis de la prueba operativa comprendió los mismos pasos que para el análisis de la prueba piloto. A diferencia del pilotaje, los ítems operativos si contribuyen al puntaje del examinando. Una vez que se obtuvieron los puntajes se procedió a analizar el rendimiento

de los estudiantes de acuerdo a clases estadísticas, género, métrica scale score y puntaje mínimo de selección.

Grupo Focal

Esta técnica de recolección de información cualitativa fue un aporte enriquecedor al proceso de construcción de la Prueba de Selección Universitaria para la Licenciatura en Antropología, para tener una mejor comprensión de los resultados de la misma. La finalidad de esta etapa fue la de conocer la experiencia educativa en el bachillerato de estos actores, respecto al desarrollo de los contenidos de Antropología y Fundamentos de Investigación Social.

Difusión de resultados

Esta última etapa representa la difusión de los resultados por medio de la revista que publicará la Maestría en Psicometría y Evaluación Educativa y/o de la presentación de la prueba a la comunidad académica de la Universidad Nacional Autónoma de Honduras, especialmente a la de Antropología. En esta fase se dará a conocer todo el proceso que implica el proceso de construcción de pruebas y el análisis de los resultados de la aplicación operativa.

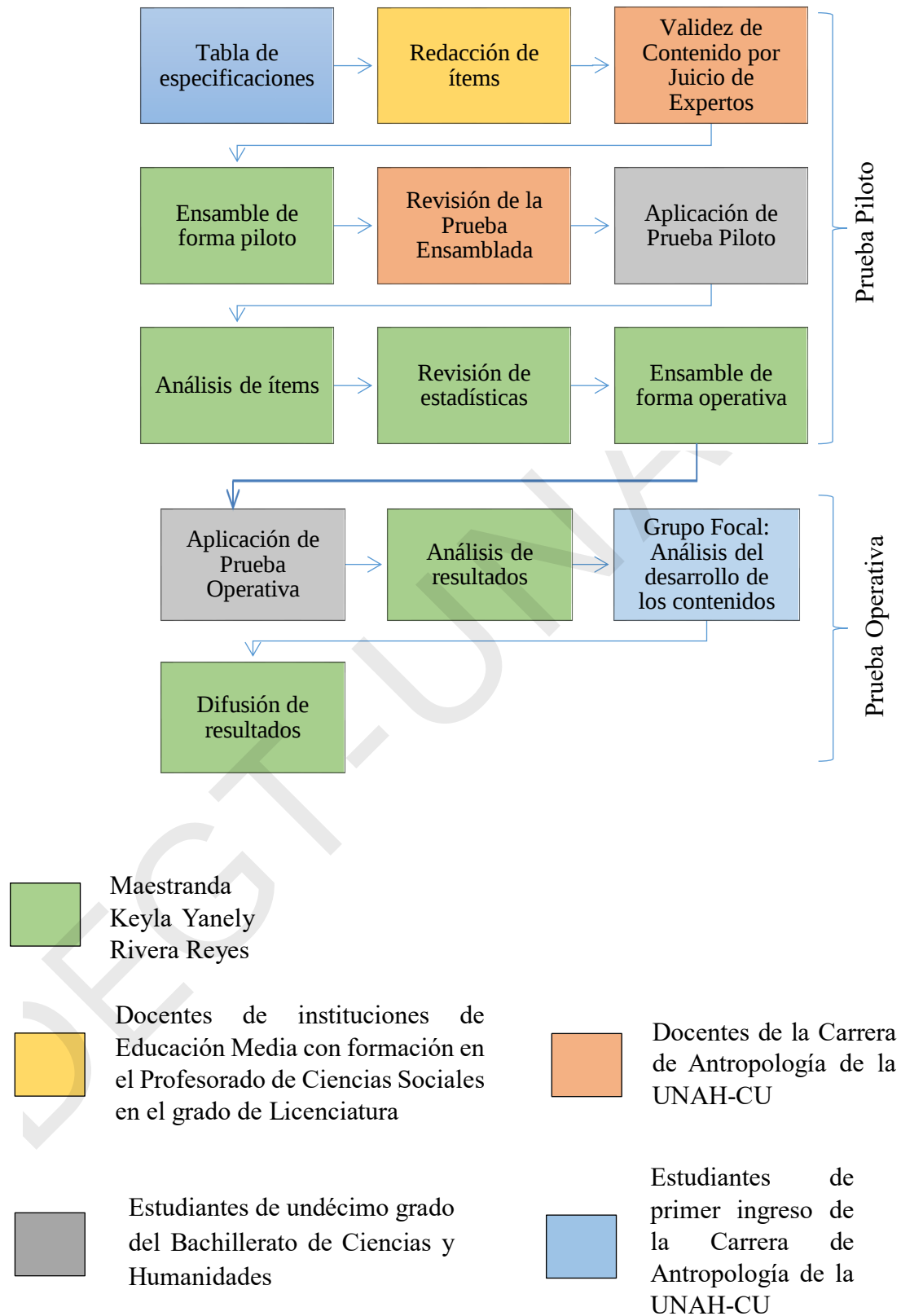


Figura 7. Etapas de la Construcción de la Prueba de Selección Universitaria para la Licenciatura en Antropología. Adaptada de AIR (2015).

Capítulo 5. Análisis y Discusión de los Resultados de Investigación

El propósito de esta investigación fue diseñar una prueba de selección universitaria basada en las expectativas de logro y habilidades cognitivas que se requieren en los espacios curriculares de Antropología y Fundamentos de Investigación Social establecidos en el Plan de Estudios y Programa Curricular del Bachillerato de Ciencias y Humanidades, para aspirantes a ingresar a la Licenciatura en Antropología de la Facultad de Ciencias Sociales de la Universidad Nacional Autónoma de Honduras. Este estudio es de enfoque cuantitativo con análisis de tipo relacional, en el cual se aplicó un muestreo no aleatorio intencional a muestras de 204 y 158 estudiantes de undécimo grado del Bachillerato de Ciencias y Humanidades.

En esta investigación se realizó las etapas de la Construcción de pruebas que habría que seguir, para construir un instrumento de medida con garantías de calidad. Las pruebas piloto y operativa se construyeron bajo Teoría Clásica de los Test (TCT), también se analizó el Funcionamiento Diferencial del Ítem. En concordancia con los objetivos específicos de investigación, en este capítulo se presentan los resultados de los análisis efectuados, discriminados según su tipo.

Objetivo específico 1

Identificar las expectativas de logro y contenidos en relación con las habilidades cognitivas que más aportan al perfil de salida de los espacios curriculares de Antropología y Fundamentos de Investigación Social del Plan de Estudios y Programas Curriculares del Bachillerato en Ciencias y Humanidades para la Prueba de Selección Universitaria de la Licenciatura en Antropología.

La redacción de ítems es la primera etapa del proceso de construcción de pruebas, a partir de la cual se genera el insumo para poder generar los ensamblajes, cuya finalidad es someter a prueba el funcionamiento de los ítems, para finalmente seleccionar aquellos que conformaron la prueba oficial, por ello es importante que los ítems redactados estén alineados a las Expectativas de Logro incluidas en la Tabla de Especificaciones.

Tabla 9

Tabla de especificaciones de la Prueba de Selección Universitaria para la Licenciatura en Antropología

Contenido a evaluar			Nivel cognitivo						Total
Espacio	Contenido	Código/Expectativa de logro	A	B	C	D	E	F	
1. Fundamentos de Investigación Social	Método científico	111				4			4
	Metodología de la investigación científica	121				5			5
	Tipos de investigación según el enfoque	131		4					4
	Etapas de una investigación.	141	7						7
	Divisiones de la Antropología	211				7			7
	Teorías del origen y evolución del ser humano.	221	7						7
	Hominización	231					6		6

2. Antropología	Contenidos básicos de la Antropología Social	241	Reconocer la existencia de diversos tipos de organización económica y social.	7					7
	Poblamiento de América	251	Conocer en forma resumida los principales datos sobre el poblamiento del continente americano y Honduras.	7					7
	Pueblos indígenas y afrodescendientes de Honduras.	261	Diferenciar los Pueblos indígenas y afrodescendientes de Honduras.	6					6
Número total de ítems por nivel cognitivo:				21	11	6	16	6	60
Porcentaje de ítems por nivel cognitivo:				35	18	10	27	10	100

Nota: El significado de las letras según el nivel cognitivo A= Reconocer B= Comprender C= Aplicar D= Analizar E= Evaluar F=Crear. Elaboración propia.

Para cumplir con este objetivo de investigación, las expectativas de logro que contiene esta Tabla de Especificaciones se seleccionaron del Plan de Estudios y Programas Curriculares del Bachillerato en Ciencias y Humanidades, particularmente de los espacios curriculares de Fundamentos de Investigación Social y Antropología, se apoyó en la experiencia latinoamericana en este tipo de instrumentos y se tomaron en consideración los siguientes aspectos: Correlacionar las expectativas de logro con los contenidos y su dosificación, tipos y cantidad de ítems, nivel cognitivo según la Taxonomía de Bloom y pesos de cada espacio curricular en la carrera para la cual está destinada dicha Prueba (DEMRE, Universidad de Chile, 2016). El código en esta tabla de especificaciones, está compuesto de la siguiente manera: este código cuenta de tres dígitos, el primero hace

referencia al espacio curricular, el segundo dígito al número de posición del contenido y el último a la posición de la expectativa de logro.

Los docentes que imparten clases en el nivel de Educación Media y que participaron en el Taller de Redacción de Ítems, confirmaron que estas expectativas de logro, son las que más aportan al perfil a cumplir en cada espacio curricular. Según Sans Martín (2008) “La clave para organizar correctamente una evaluación consiste en definir una buena tabla de especificaciones que plasme la relación entre los contenidos del programa de la asignatura y los objetivos de aprendizaje que pretendemos que asuman nuestros alumnos” (p.8).

Dando como resultados en esta primera etapa, cuatro expectativas para Fundamentos de Investigación Social y seis para Antropología, que son la base de la Prueba de Selección Universitaria para la Licenciatura en Antropología y vinculadas con la habilidades cognitivas, la prueba oficial quedó estructurada con el 63% de los ítems de habilidades de orden inferior y 37 de orden superior.

Objetivos específicos 2 y 3

Desarrollar el proceso de construcción de instrumentos psicométricos para que la Prueba de Selección Universitaria para aspirantes a ingresar a la Licenciatura en Antropología, cumpla con los indicadores psicométricos de calidad bajo el análisis de la Teoría Clásica de los Test (TCT).

Determinar el Funcionamiento Diferencial de los ítems en las pruebas piloto y operativa, en relación con el género, para evitar discriminación en las oportunidades de selección y con el fin de garantizar la calidad de los resultados.

Análisis de la Aplicación Piloto. En función del análisis psicométrico se enlazaron los objetivos específicos 2 y 3, para una mejor comprensión de los resultados en la Aplicación Piloto y el porqué de la elección de cada ítem para conformar la prueba operativa. Los índices más usados al hacer el análisis psicométrico de una prueba de selección universitaria de acuerdo a la Teoría Clásica de los Test, son el índice de dificultad, el índice de discriminación, análisis de distractores, validez y confiabilidad. Estos índices se convierten en indicadores de la calidad de una prueba en la medida que se encuentren dentro de los valores aceptables. Se agregó el análisis del Funcionamiento Diferencial del Ítem, producto del importante papel que juega este tipo de prueba en la toma de decisiones para ser admitido o no en la universidad o en una carrera en particular. Este análisis ayuda a crear evaluaciones más objetivas. Por eso, en este apartado se agregó los datos del FDI, para asegurarse que la diferencia por género en el desempeño en la Prueba es real y no está siendo causada por algún(os) ítem(s).

La Figura 8, trata sobre la validez de contenido. Puerta y Marín (2015) definen que “La validez de contenido se refiere al grado en que un instrumento refleja un dominio específico de contenido de lo que se mide” (párr.5). Se realizó mediante el juicio de expertos (docentes de la carrera de Antropología UNAH) utilizando como referencia un instrumento de medición que indagó 4 criterios: veracidad, claridad, coherencia y relevancia en cada uno de 192 ítems pero de ellos, pero solo 100 ítems fueron elegibles para conformar la prueba piloto. Fueron tenidas en cuenta las observaciones cualitativas de los expertos, antes de aplicar esta versión de la prueba. Las observaciones apuntaban principalmente a mejorar la

redacción de algunos ítems. Se determinó el grado de acuerdo entre expertos con el Índice de validez de contenido (Pedrosa, Suárez y García, 2014).

Se obtuvo como resultado una razón de validez de contenido (CVR) donde el valor mayor fue de 0.96 y el valor menor de 0.59, con un índice de validez de contenido (CVI) de 0.83, el cual excede el punto de corte mínimo de 0.58, por lo que se considera que la prueba validándose era aceptable (Vargas, Máyne, Cavazos y Cervantes, 2016). Por lo tanto, se puede decir que según los criterios de los expertos, este realmente instrumento mide las variables que se busca medir.

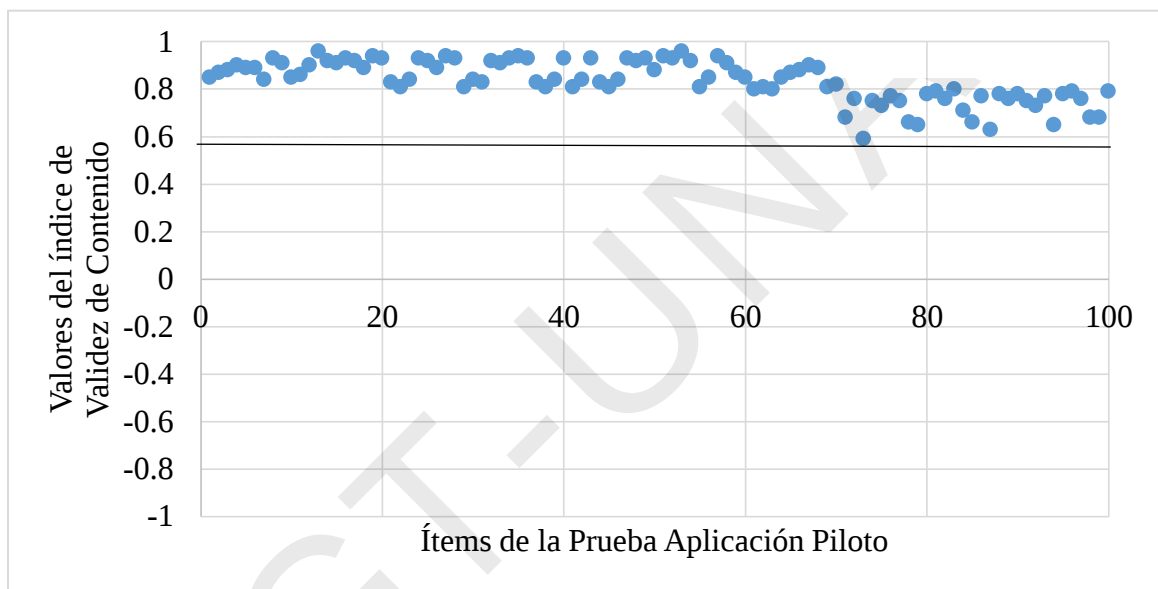


Figura 8. Ítems validados por el procedimiento de Juicio de Expertos utilizando Índice de validez de contenido. Elaboración propia.

Para el análisis de ítems bajo la Teoría Clásica de los Test, se utilizó las tablas de valores de los indicadores psicométricos de calidad, para determinar que ítems cumplieran con los valores aceptables y también los criterios a seguir en el análisis de distractores, ésta información es aplicable para la revisión de estadísticas tanto de la aplicación piloto como de la operativa.

Tabla 10

Clasificación de los Ítems por el Índice de Dificultad

Clasificación del Índice de Dificultad	Rango de valores p-value
Muy fácil	De 0.81 a 1.00
Relativamente fácil	De 0.66 a 0.80
Dificultad intermedia	De 0.51 a 0.65
Relativamente difícil	De 0.31 a 0.50
Difícil	De 0.11 a 0.30
Muy difícil	De 0.00 a 0.10

Nota: En la tabla 10 se puede apreciar que los valores para el Índice de Dificultad oscilan entre 0 y 1. De tal manera que cuanto más cerca de 1, más fácil será ese ítem, mientras que cuánto más se aproxime a 0, más difícil será. Tomado de Cortada (1999).

Tabla 11

Poder de Discriminación de los ítems según el valor D

D	Calidad	Recomendaciones
0.40>	Excelente	Conservar
0.30-0.39	Buena	Posibilidades de mejorar
0.20-0.29	Regular	Necesidad de mejorar
0.00-0.19	Pobre	Descartar o revisar a profundidad
<-0.01	Pésima	Descartar definitivamente

Nota: La Tabla 11 proporciona los valores para determinar la calidad de los ítems, en términos del Índice de Discriminación., muestra los valores D y su correspondiente interpretación. También proporciona las recomendaciones a seguir según sea el caso en cada ítem. Tomado de Ebel y Frisbie (1986).

Tabla 12*Categorías el Funcionamiento Diferencial del ítem*

Categoría	Valores FDI
FDI pequeño o inexistente	0 y .05
FDI bajo:	>.05 a .075
FDI moderado	.>075 y .10
FDI Alto	>.10

Nota: En la tabla 12 utiliza el valor de .075 como el límite para un FDI moderado. Por lo tanto, los valores aceptables oscilan entre 0 a .075. Tomado Zenisky, Hambleton y Robin (2004).

La Figura 9, representa los valores p-value de cada ítem en la aplicación piloto de la Prueba de Selección Universitaria para la Licenciatura en Antropología. “El índice de dificultad representa la cantidad de estudiantes que contestaron el ítem correctamente” (AIR, 2015, p.9). Según las categorías especificadas en la Tabla 10, se clasificó a los 100 ítems de la siguiente manera: 15 ítems difíciles (0.11 a 0.30), 34 ítems relativamente difíciles (0.31 a 0.50), 34 ítems con dificultad adecuada (0.51 a 0.65) y 17 ítems relativamente fáciles (0.66 a 0.80).

En realidad este valor, más que representar el índice de dificultad del ítem, representa qué tan fácil es contestar al ítem, debido a que nos proporciona un cociente, producto de los estudiantes que lo contestaron correctamente entre el total de esa muestra. “El rango de valores del índice de dificultad es de 0 a 1.0, los valores aceptables del índice de dificultad oscilan entre 0.3 y 0.8” (AIR, 2015, p.9). El propósito de este análisis gráfico, fue determinar que ítems sí cumplen con los criterios del índice de dificultad para poder ser incluidos en la prueba operativa. Por lo tanto, de los 100 ítems, 71 fueron elegibles para poder conformar la Prueba Operativa y 29 tienen que ser revisados para ser incluidos en un próximo pilotaje.

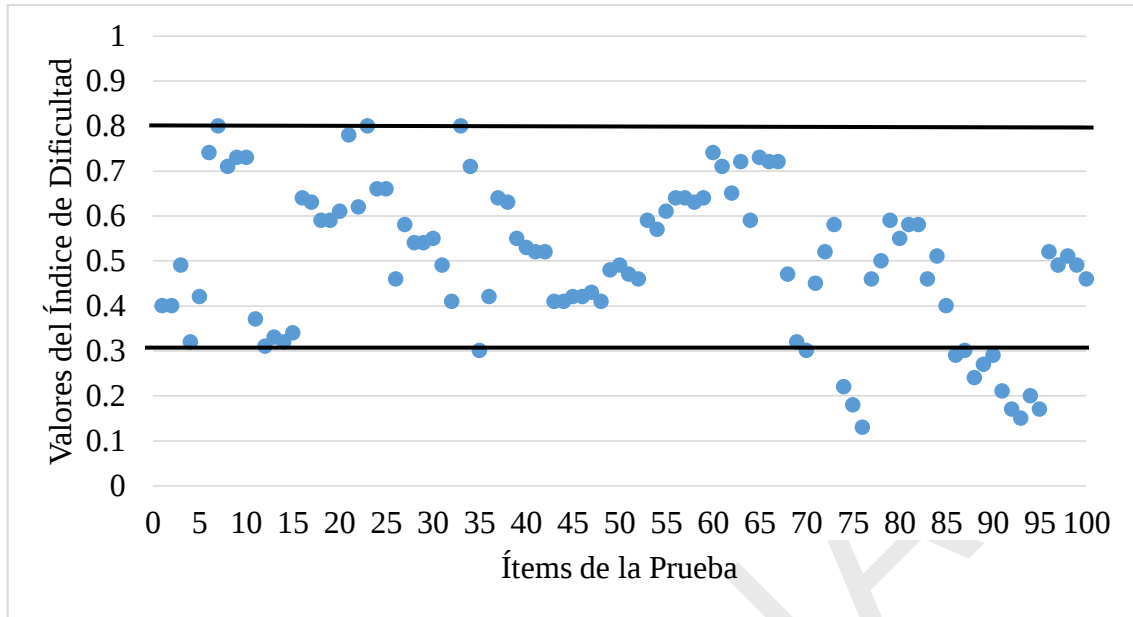


Figura 9. Índice de Dificultad de los ítems de la Aplicación Piloto de la Prueba de Selección Universitaria para la Licenciatura en Antropología. Elaboración propia.

La Figura 10, representa los valores D de cada ítem en la aplicación piloto de la Prueba de Selección Universitaria para la Licenciatura en Antropología. “El índice de discriminación es la medida en que tan bien un ítem permite distinguir entre aquellos estudiantes que han alcanzado las expectativa de logro y aquellos que no” (AIR, 2015, p.9). Según las categorías especificadas en la Tabla 11, se clasificó a los 100 ítems de la siguiente manera: 56 ítems con poder de discriminación excelente ($0.40 >$), 14 ítems con poder de discriminación buena ($0.30-0.39$), 4 ítems con poder de discriminación regular ($0.20-0.29$), 19 ítems con poder de discriminación pobre ($0.00-0.19$) y 7 ítems con poder de discriminación pésima (<-0.01).

Los ítems que tiene un coeficiente de discriminación alto indicaron que los estudiantes que lo respondieron correctamente también obtuvieron puntajes altos en Prueba de Selección Universitaria para la Licenciatura en Antropología, mientras que aquellos examinandos que contestaron incorrectamente también obtuvieron puntajes bajos en la prueba. Sin embargo, los ítems con un índice de correlación menor a 0 (negativa) indicaron que los estudiantes de alto desempeño lo respondienddo incorrectamente y los examinandos de bajo desempeño lo respondienddo correctamente, por lo tanto estos ítems fueron definitivamente descartados.

El rango de valores del índice de discriminación es de -1.0 a 1.0 con valores aceptables mayores a 0.30. Cuando se está interpretando el índice de discriminación es importante tener en mente que existe una relación con el índice de dificultad. Cuando un ítem es demasiado fácil o demasiado difícil no se puede esperar que tenga un alto poder de discriminación (AIR, 2015, p.9)

El propósito de este análisis gráfico, fue determinar que ítems sí cumplen con los criterios del índice de discriminación para poder ser incluidos en la prueba operativa. Por lo tanto, de los 100 ítems, 70 ítems fueron elegibles para poder conformar la Prueba Operativa, 23 ítems tienen que ser revisados para ser incluidos en un próximo pilotaje y 7 ítems definitivamente fueron descartados.

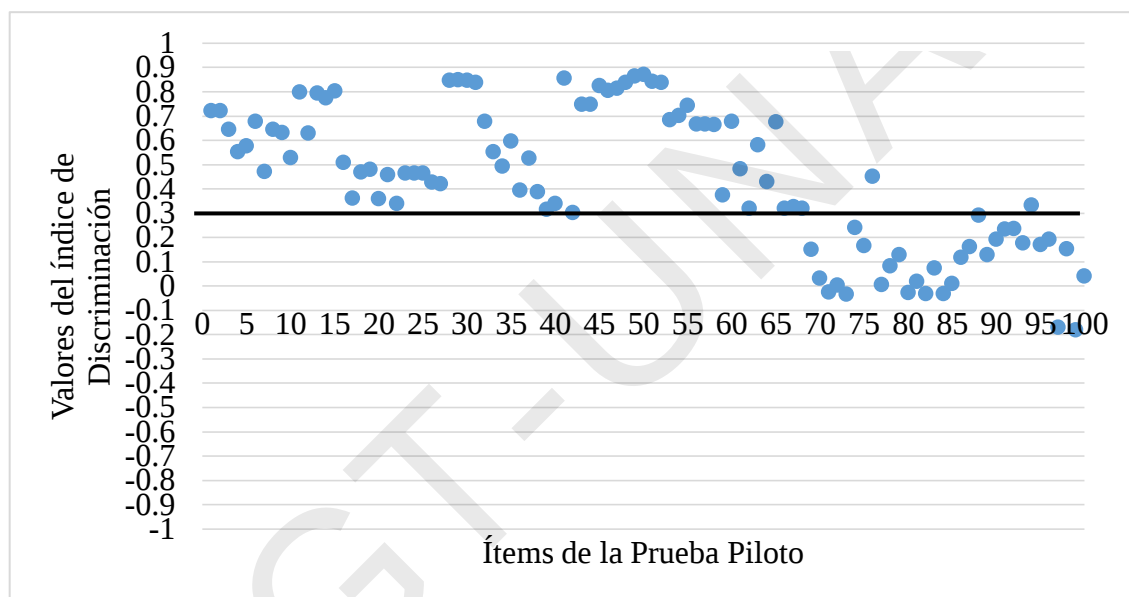


Figura 10. Índice de Discriminación de los ítems de la Aplicación Piloto de la Prueba de Selección Universitaria para la Licenciatura en Antropología. Elaboración propia.

Tal como indica la literatura internacional, para ítems de opción múltiple es importante realizar el análisis de distractores, ya que permite entender la calidad de los mismos, de los cuales también depende la calidad del ítem. El criterio que el Proyecto Mejorando el Impacto al Desempeño Estudiantil en Honduras (MIDEH), utilizaba en este tipo de análisis, consistía en revisar en cada ítem, que cualquiera de los distractores fuese seleccionado al menos por el 2% de los estudiantes. (AIR, 2015).

Un aspecto esencial para el buen funcionamiento de un ítem es que sus distractores cumplan su función. Una opción que la selecciona menos del 2% de los estudiantes, significa

que ni siquiera aquellos estudiantes que no han alcanzado las Expectativas de Logro, consideran a esos distractores como una respuesta razonable y correcta al ítem. En este sentido, los distractores deben ser claramente incorrectos, pero plausibles.

El propósito de este análisis gráfico para el ítem 39, es mostrar que se realizó este procedimiento para los 100 ítems del pilotaje y los resultados muestran que sí cumplen con el criterio utilizado por MIDEH. Es decir, en los 100 ítems, todos sus distractores fueron elegidos al menos por el 2% de los estudiantes. Cabe aclarar que esta distribución de porcentaje en cada una de las opciones solo es aplicable para este ítem.

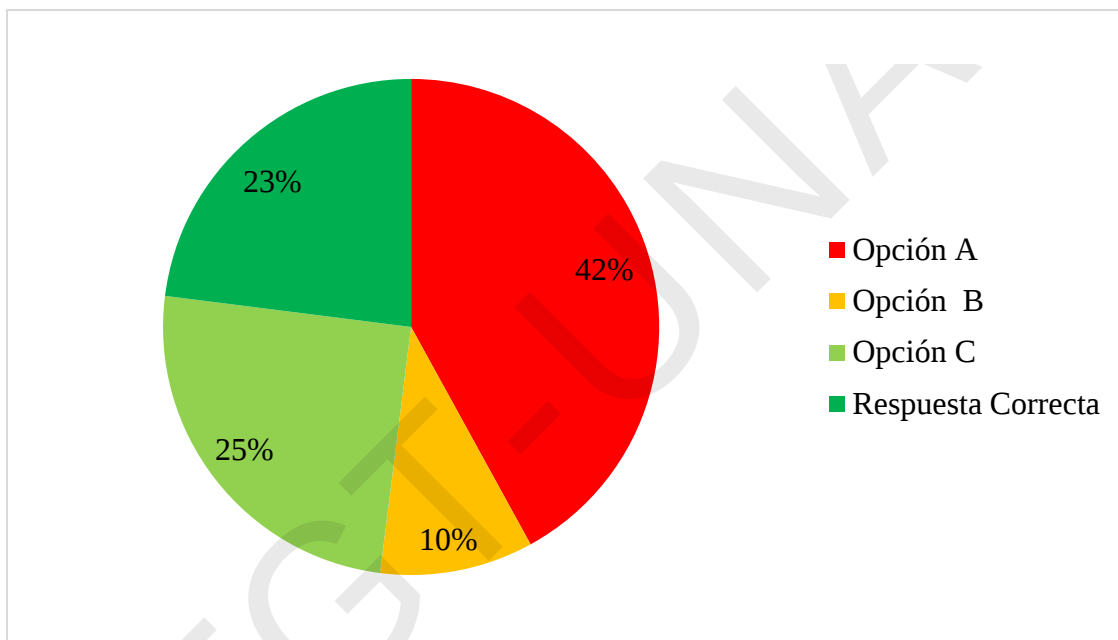


Figura 11. Análisis de Distractores del ítem 39 de la Aplicación Piloto de la Prueba de Selección Universitaria para la Licenciatura en Antropología. Elaboración propia.

La Figura 12, representa los valores FDI de cada ítem. El propósito de la aplicación del FDI en los 100 ítems que conformaron la aplicación piloto de la Prueba de Selección Universitaria para la Licenciatura en Antropología, fue analizar que el desempeño por género en cada ítem de la prueba, se debiera a diferencias reales entre ambos grupos en relación a los contenidos medidos y no porqué el contenido del algún ítem estuviera favoreciendo a uno de los grupos (Elousa, 2006). Según las categorías especificadas en la Tabla 12, se clasificó a los 100 ítems de la siguiente manera: 48 ítems con valores de FDI pequeño o inexistente (0 y .05), 32 ítems con valores de FDI bajo (>.05 a .075), 5 ítems con valores de FDI moderado (>.075 y .10) y 15 ítems con valores de FDI Alto (>.10).

El rango de valores del Funcionamiento Diferencial del Ítem es de 0 a >1.0 , con valores aceptables de 0 a 0.75 (Zenisky et al, 2004). El propósito de este análisis gráfico, fue determinar que ítems sí cumplen con los criterios del Funcionamiento Diferencial del Ítem según el género, para poder ser incluidos en la prueba operativa. Por lo tanto, de los 100 ítems, 80 ítems fueron elegibles para poder conformar la Prueba Operativa (tienen representatividad por cada contenido curricular), 5 ítems tienen que ser revisados para ser incluidos en un próximo pilotaje y 15 ítems definitivamente fueron descartados.

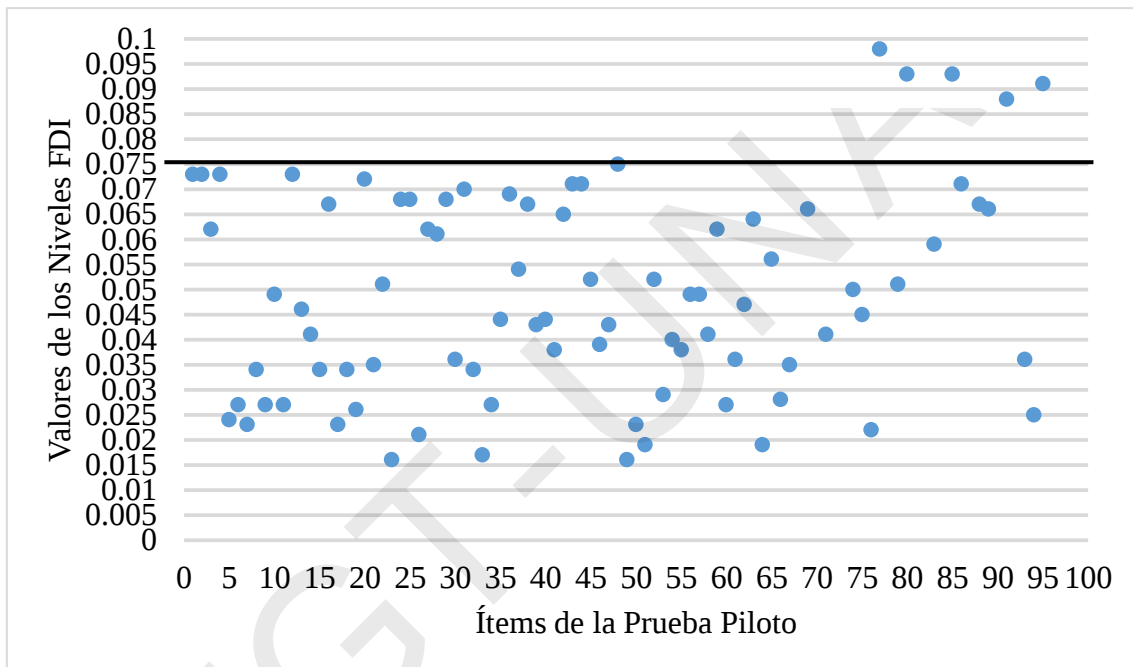


Figura 12. Funcionamiento Diferencial del Ítem de la Aplicación Piloto de la Prueba de Selección Universitaria para la Licenciatura en Antropología. Elaboración propia.

Tabla 13

Relación entre los Indicadores de calidad de los ítems de la Aplicación Piloto de la Prueba de Selección Universitaria para la Licenciatura en Antropología

Ítem	P-Value	Correlación	Porcentaje de frecuencia en las opciones de respuestas				FDI
			A	B	C	D	
22	0.62	0.34	11%	62%*	16%	11%	0.051
94	0.2	0.334	28%	27%	20%*	25%	0.025
80	0.55	-0.169	55%*	17%	14%	14%	0.143

Nota: Elaboración propia.

En la Tabla 13, se describieron 3 ejemplos del análisis de ítems (ítem aceptado, revisado y rechazado) y revisión de estadísticas del pilotaje, para poder seleccionar los mejores ítems estadísticamente hablando, y que conformaron la aplicación operativa de la Prueba de Selección Universitaria para la Licenciatura en Antropología. Para cada ítem se revisó que en cada indicador se cumpliera con los valores aceptados. Según las categorías especificadas en las Tablas 10, 11 y 12, se clasificó a los 100 ítems de la siguiente manera: 67 ítems podían ser elegibles para conformar la prueba operativa sin ningún problema, 13 ítems deben ser revisados para ser incluidos en un próximo pilotaje y 20 ítems definitivamente fueron descartados.

Tabla 14

Aspectos psicómetros de la Aplicación Piloto de la Prueba de Selección Universitaria para la Licenciatura en Antropología

Alfa de Conbrach de la		
Subprueba Antropología	Dificultad de la Prueba	Discriminación de la Prueba
.919		
Alfa de Conbrach de la	0.49	0.44
Subprueba Fundamentos de Investigación Social		
.920		

Nota: Elaboración propia.

La elección del método para la fiabilidad dependerá de la finalidad de la prueba, el más utilizado en medición es el de consistencia interna, el cual solo requiere que el instrumento se aplique una sola vez. El del Alfa de Conbrach, se obtuvo utilizando el paquete estadístico SPSS. Con este tipo de análisis de consistencia interna, se determinó el grado en el que están relacionados recíprocamente los ítems de la Aplicación Piloto de la Prueba de Selección Universitaria para la Licenciatura en Antropología. El coeficiente de Cronbach expresa el grado en que los ítems miden la misma variable: homogeneidad. De esta manera, como este instrumento consta de subpruebas individuales (Antropología y Fundamentos de Investigación Social), la recomendación es disponer de medidas de confiabilidad para cada subprueba (Brown, 1980).

Expuesto lo anterior, se siguió la recomendación de la literatura internacional, se calculó el coeficiente de Alfa de Conbrach para cada una de las Subpruebas, Antropología (.919) y de la Subprueba Fundamentos de Investigación Social (.920). Los investigadores Carmines y Zeller (1979) consideran, que como regla general, las confiabilidades no deben ser inferiores a 0.80. Tomando este criterio como referencia, entonces la valoración de fiabilidad en ambas subpruebas se consideraron como excelentes.

En resultados anteriores se abordó la dificultad y discriminación de cada ítem, en este apartado se hará referencia a la dificultad y discriminación pero de la prueba. En el caso de la dificultad promedio (0.49) de la prueba se categoriza como relativamente difícil y en lo

que respecta a la discriminación promedio (0.44), la prueba tiene un poder de discriminación excelente. En ambos casos los valores están dentro de lo que se consideran como aceptables.

Análisis de la Aplicación Operativa. En el análisis de la Aplicación Operativa de la Prueba de Selección Universitaria para la Licenciatura en Antropología, se incluyeron los mismos pasos y la revisión estadística de los indicadores psicométricos que para el análisis de las Aplicación Piloto. Sin embargo, debido a que el propósito de este análisis ya no es crear y validar ítems para poder obtener un banco de ítems, sino revisar el nivel de desempeño de los estudiantes en la prueba, se hizo un análisis adicional sobre los resultados. Pero este procedimiento se abordará en el desarrollo del objetivo específico 4.

La Figura 13, representa los valores p-value de cada ítem en la aplicación operativa de la Prueba de Selección Universitaria para la Licenciatura en Antropología. . “El índice de dificultad representa la cantidad de estudiantes que contestaron el ítem correctamente” (AIR, 2015, p.9). Según las categorías especificadas en la Tabla 10 de la sección del pilotaje, se clasificó a los 60 ítems de la siguiente manera: 29 ítems relativamente difíciles (0.31 a 0.50), 29 ítems con dificultad adecuada (0.51 a 0.65) y 2 ítems relativamente fáciles (0.66 a 0.80).

Se reitera lo anteriormente dicho, este valor más que representar la dificultad del ítem, representa qué tan fácil es contestar al ítem, debido a que representa la cantidad de estudiantes que lo contestaron correctamente. “El rango de valores del índice de dificultad es de 0 a 1.0, los valores aceptables del índice de dificultad oscilan entre 0.3 y 0.8” (AIR, 2015, p.9). El propósito de este análisis gráfico, fue determinar que ítems sí cumplen con los criterios del índice de dificultad para poder ser parte de la versión oficial de la Prueba de Selección Universitaria para la Licenciatura en Antropología. En este caso, los 60 ítems cumplen con los valores aceptados en la literatura especializada.

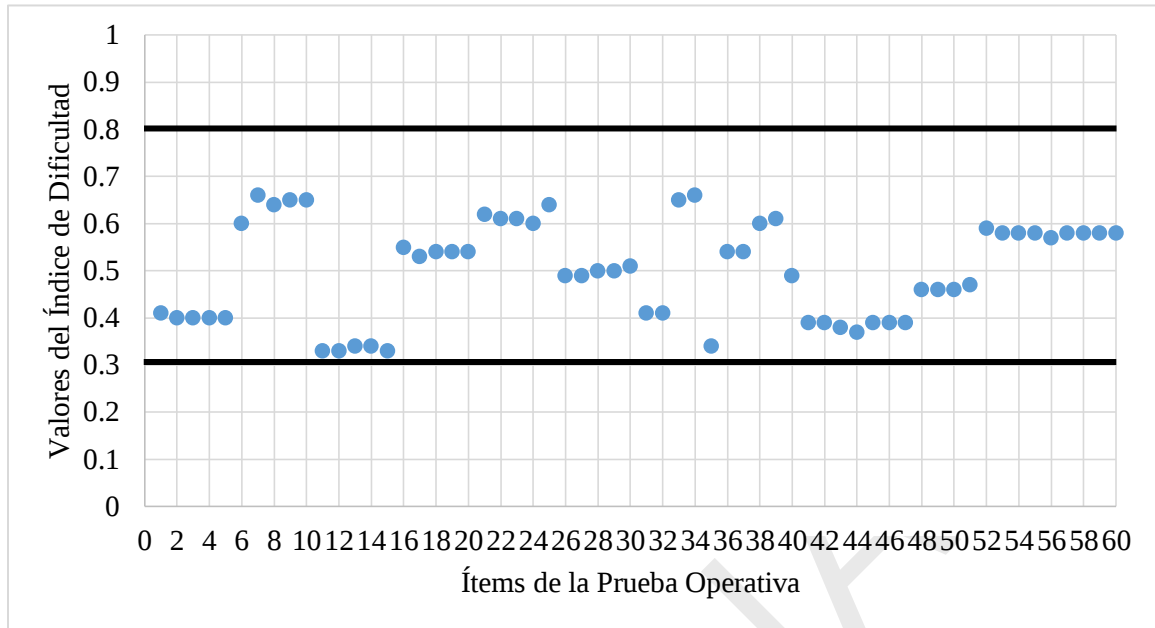


Figura 13. Índice de Dificultad de los ítems de la Aplicación Operativa de la Prueba de Selección Universitaria para la Licenciatura en Antropología. Elaboración propia.

La Figura 14, representa los valores D de cada ítem en la aplicación operativa de la Prueba de Selección Universitaria para la Licenciatura en Antropología. “El índice de discriminación es la medida en que tan bien un ítem permite distinguir entre aquellos estudiantes que han alcanzado las expectativa de logro y aquellos que no” (AIR, 2015, p.9). Según las categorías especificadas anteriormente en la Tabla 11, se clasificó a los 60 ítems de la siguiente manera: 53 ítems con poder de discriminación excelente ($0.40 >$) y 7 ítems con poder de discriminación buena ($0.30-0.39$).

“El rango de valores del índice de discriminación es de -1.0 a 1.0 con valores aceptables mayores a 0.30 ” (AIR, 2015, p.9). El propósito de este análisis gráfico, fue determinar que ítems sí cumplen con los criterios del índice de discriminación para poder ser incluidos en la versión oficial de la Prueba de Selección Universitaria para la Licenciatura en Antropología. Por lo tanto, de los 60 ítems, todos cumplen con los estándares esperados en esta etapa de la construcción de pruebas.

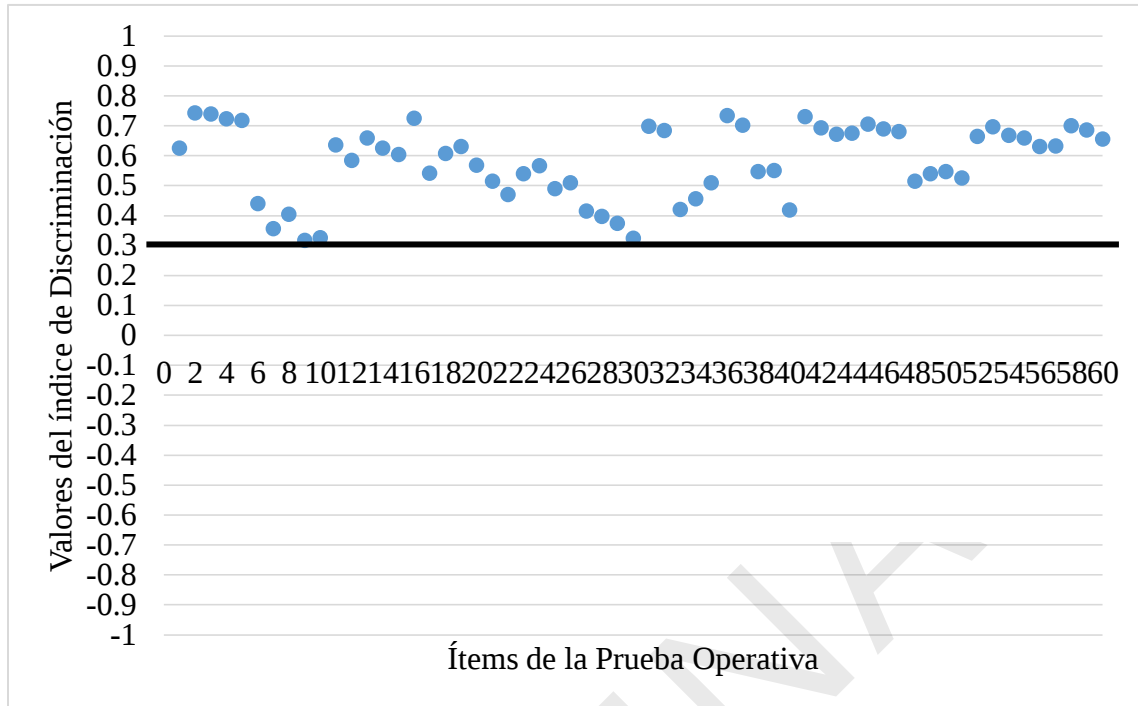


Figura 14. Índice de Discriminación de los ítems de la Aplicación Operativa de la Prueba de Selección Universitaria para la Licenciatura en Antropología. Elaboración propia.

La Figura 15, es sobre el ítem 16, se utilizó para mostrar un ejemplo del reporte del análisis de distractores, de la respuesta correcta y en general, del funcionamiento de este ítem. Para comprobar el propósito de esta recomendación psicométrica, uno de los procedimientos utilizados por MIDEH, fue analizar la frecuencia en que cada distractor es seleccionado. En el caso de que la frecuencia de un distractor sea menor al 2% o nula proporción, se puede decir que el distractor no está cumpliendo la función que corresponde, que es la de ser una posible respuesta para aquel estudiante que no ha logrado esa expectativa de logro, por ende el ítem puede ser contestado fácilmente hasta por sentido común. También en el caso inverso, en el que la selección de un distractor es mayor que la respuesta correcta, se entiende que esta opción está confundiendo a los estudiantes y es recomendable reestructurar el ítem (AIR, 2015).

Tomando en consideración lo anteriormente expuesto, el reporte que se realiza del ítem 16 y del resto de la Aplicación Operativa de la Prueba de Selección Universitaria para la Licenciatura en Antropología, son los siguientes aspectos técnicos:

- a) Cada distractor fue seleccionado al menos por el 2 % de los evaluados.

- b) La respuesta correcta en cada uno de los ítems fue la que tuvo mayor porcentaje de selección.
- c) La elección de los tres distractores fue aproximada en proporción.
- d) La “no respuesta” al ítem no debe superar el 20 % de los evaluados.

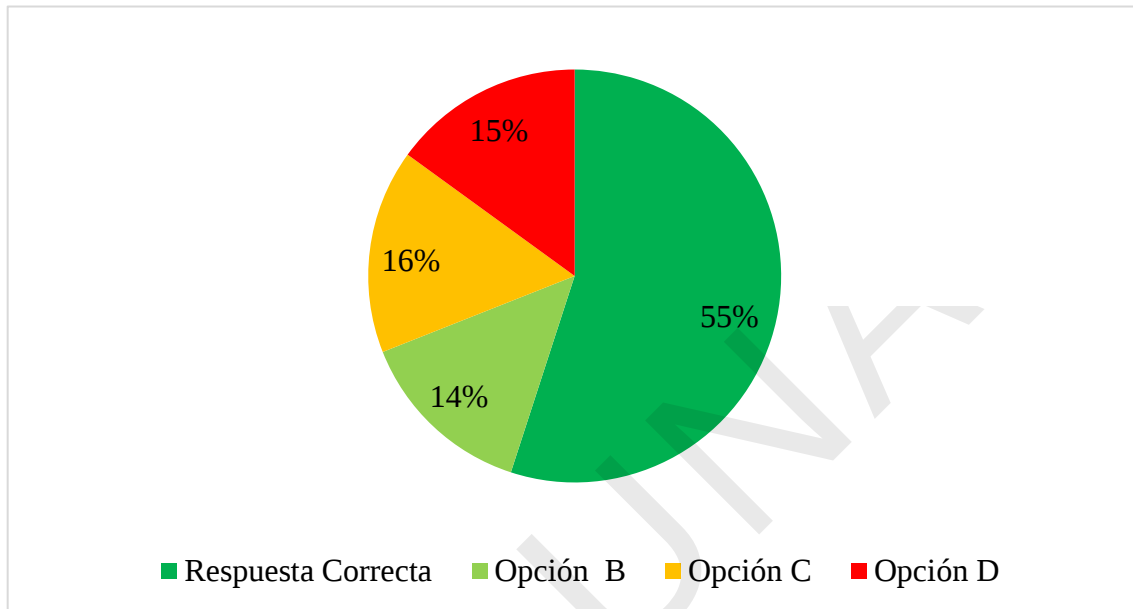


Figura 15. Análisis de Distractores del ítem 16 de la Aplicación Operativa de la Prueba de Selección Universitaria para la Licenciatura en Antropología. *Elaboración propia.*

Los estudios internacionales, recomiendan que para este tipo de instrumento, es necesario hacer el análisis del Funcionamiento Diferencial del Ítem (FDI), por lo tanto se realizó en esta investigación, ver la figura 16. Hay FDI, cuando estudiantes con el mismo nivel de habilidad cognitiva, difieren a la hora de contestar un ítem.

Las evaluaciones psicológicas deben garantizar la equidad y validez de las interpretaciones y decisiones adoptadas a partir de las mismas. Para ello es necesario la utilización de instrumentos libres de sesgo, y capaces de evaluar necesidades personales y sociales de individuos con diferentes características (Gómez, Hidalgo y Guilera, 2010, párr.1).

Tomando en cuenta esta peculiaridad, con el fin de garantizar la calidad de los resultados, se estudió la posible existencia de FDI en los ítems en la Prueba de Selección Universitaria para la Licenciatura en Antropología de acuerdo a la variable género (femenino y masculino), la cual se aplicó a 158 estudiantes (91 señoritas y 67 varones) de undécimo grado del Bachillerato en Ciencias y Humanidades. La prueba estaba compuesta de 60 ítems.

Para la detección del FDI en los ítems se realizó el método de Diferencias de Medias Estandarizadas. La prueba indicó la existencia de FDI en las siguientes categorías:

Según las categorías especificadas en la Tabla 12, se clasificó a los 60 ítems de la siguiente manera: 22 ítems con valores de FDI pequeño o inexistente (0 y .05) y 38 ítems con valores de FDI bajo (>.05 a .075). Con estos resultados se concluyó que el FDI es tan imperceptible que no afecta en ninguno de los ítems y por tanto los puntajes obtenidos no se encuentran afectados por FDI. La importancia del FDI en este ámbito, es porque no se debe olvidar, que un instrumento se administra con un objetivo concreto, generalmente para tomar decisiones (selección universitaria) que en este caso son relevantes para la vida de los aspirantes a ingresar a la carrera de Antropología.

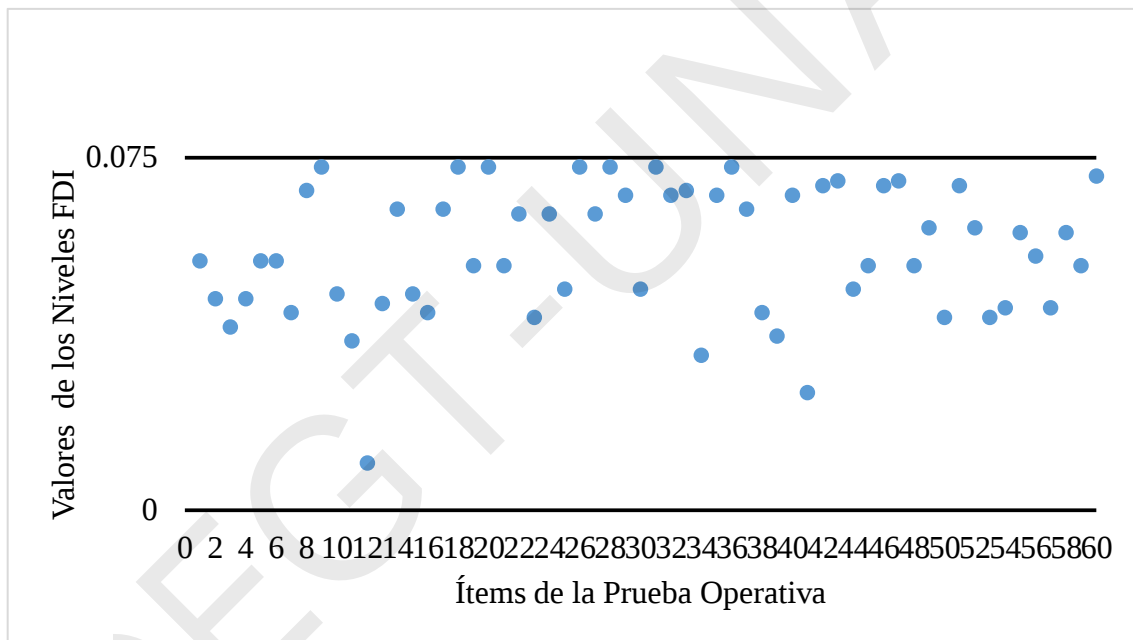


Figura 16. Funcionamiento Diferencial del Ítem de la Aplicación Operativa de la Prueba de Selección Universitaria para la Licenciatura en Antropología. Elaboración propia.

Tabla 15

Ejemplo de análisis de los ítems en la Aplicación Operativa de la Prueba de Selección Universitaria para la Licenciatura en Antropología

Ítem	P-Value	Correlación	Porcentaje de frecuencia en las opciones de respuestas				FDI
			A	B	C	D	
8	0.64	0.403	64%*	14%	11%	11%	0.068
35	0.34	0.509	23%	34%*	24%	19%	0.067
48	0.46	0.514	20%	16%	18%	46%*	0.052

Nota: Elaboración propia.

En la Tabla 15, se muestra un ejemplo de análisis estadístico de la Aplicación Operativa en la que se resume el análisis y revisión de estadísticas en los ítems 8, 35 y 48, lo cual se hizo para el resto de los ítems, en total 60 ítems. Se resume el reporte psicométrico basado en la TCT de la siguiente manera:

a) Respecto al Índice de Dificultad, los valores p-value de los 60 ítems, están dentro de los aceptables: 29 ítems relativamente difíciles (0.31 a 0.50), 29 ítems con dificultad adecuada (0.51 a 0.65) y 2 ítems relativamente fáciles (0.66 a 0.80).

b) Tomando como referencia el Índice de Discriminación, los valores D también oscilan dentro de los aceptables: 53 ítems con poder de discriminación excelente ($0.40 >$) y 7 ítems con poder de discriminación buena (0.30-0.39).

c) Y en el Análisis de Distractores, se revisó la frecuencia de selección en cada una de las opciones y se constató que al menos cada distractor fue elegido por un 2% de los estudiantes.

d) Por último, el análisis del Funcionamiento Diferencial del Ítem, se puede decir que los ítems de esta prueba, no se encuentran afectados por un funcionamiento diferencial; por ende, los puntajes obtenidos con dicha prueba no se encuentran a favor de ningún grupo.

En la aplicación de la prueba operativa también se confirmó la validez de la tabla de especificaciones que se propuso en esta investigación, la cual después de la aplicación

mantuvo la representatividad en cada uno de los contenidos (ver tabla 9) y la cantidad de ítems propuestos.

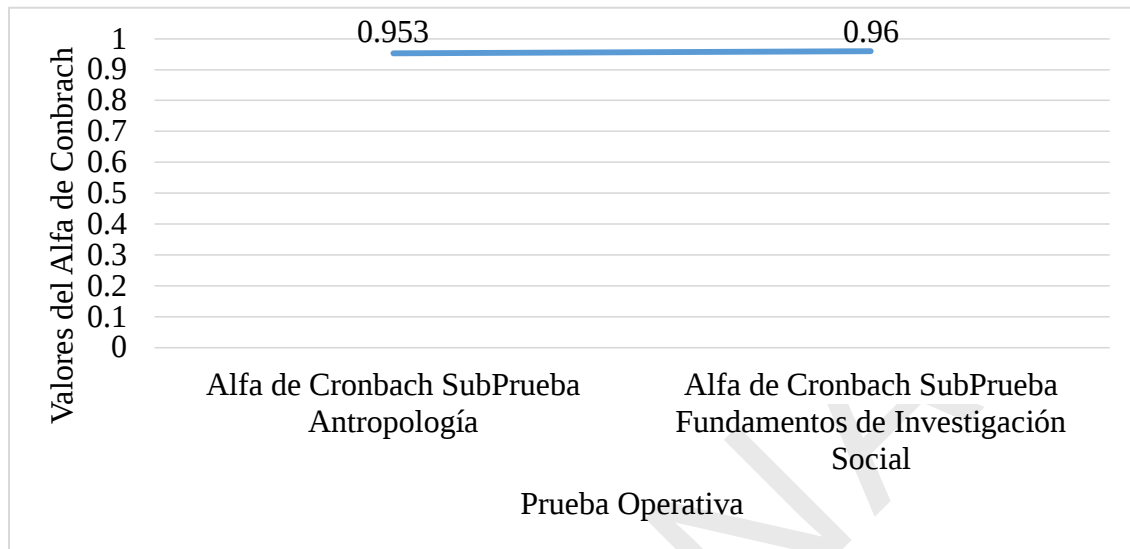


Figura 17. Comparación del Alfa de Conbrach las SubPruebas de Antropología y Fundamentos de Investigación Social de la Prueba de Selección Universitaria para la Licenciatura en Antropología. Elaboración propia.

Esta prueba se construyó bajo la Teoría Clásica de los Tests (TCT) y en esta teoría, el método de consistencia interna mediante Alfa de Cronbach es el más utilizado para estimar la fiabilidad de un instrumento psicométrico que está estructurada por dos conjuntos de ítems que se espera midan esos campos de contenidos. Una condición de este método, la cual es su principal ventaja, es que solo requiere una administración de la prueba; además su estimación es factible, está disponible como opción de análisis en SPSS (Muñiz, 2010). Por esas condiciones, este fue el método que se utilizó para analizar la fiabilidad de esta prueba.

En la Figura 17, del Alfa de Conbrach, se obtuvo utilizando el paquete estadístico SPSS. Con este tipo de análisis de consistencia interna, se determinó el grado en el que están relacionados recíprocamente los ítems de la Aplicación Operativa de la Prueba de Selección Universitaria para la Licenciatura en Antropología. El coeficiente de Cronbach expresa el grado en que los ítems miden la misma variable: homogeneidad. Igual que en la prueba del pilotaje, el instrumento consta de dos subpruebas (Antropología y Fundamentos de Investigación Social), la recomendación es disponer de medidas de confiabilidad para cada subprueba (Brown, 1980).

Expuesto lo anterior, se siguió la recomendación de la literatura internacional, se calculó el coeficiente de Alfa de Conbrach para cada una de las Subpruebas, Antropología (.953) y de la Subprueba Fundamentos de Investigación Social (.96). Tomando el criterio de un coeficiente no inferior a 0.80, entonces la valoración de fiabilidad en ambas subpruebas se consideraron como excelentes. Como información adicional a este análisis, se incluyó el Alfa de Conbrach de la prueba, aplicación piloto (0.957) y operativa (0.970), así como sus respectivos Errores de Medida utilizando el Nivel de confianza del 95%, 0.041 y 0.038.

Tabla 16

Aspectos psicómetros de la Aplicación Operativa de la Prueba de Selección Universitaria para la Licenciatura en Antropología

Dificultad promedio de la Prueba	Discriminación promedio de la prueba
0.50	0.58

Nota: Elaboración propia.

En el análisis de los aspectos psicométricos según la TCT se abordó estos indicadores a nivel de ítem, en la tabla 16 se hará referencia a la dificultad y discriminación pero de la prueba. En el caso de la dificultad promedio (0.50) de la prueba se categoriza como relativamente difícil y en lo que respecta a la discriminación promedio (0.44), la prueba tiene un poder de discriminación excelente. En ambos casos los valores están dentro de lo que se consideran como aceptables.

Objetivo específico 4

Analizar el rendimiento de los estudiantes en referencia a las habilidades cognitivas asociadas a conocimientos específicos de Antropología y Fundamentos de Investigación Social en la Aplicación Operativa de la Prueba de Selección Universitaria para aspirantes a ingresar a la Licenciatura de Antropología.

El ingreso al nivel de Educación Superior, sigue ocupando un lugar importante para las expectativas personales y profesionales de muchos estudiantes que están por egresar de Educación Media y que se encuentran en la disyuntiva ¿Qué quiero estudiar? Por ello, la importancia de este tipo de pruebas complementarias a la PAA, las cuales son útiles para la ubicación del estudiante en una determinada Facultad o en una carrera en específica, como es este caso.

En el presente apartado se expone un análisis detallado del desempeño mostrado por los estudiantes de undécimo grado del Bachillerato en Ciencias y Humanidades en la Prueba de Selección Universitaria para la Licenciatura en Antropología. Para interpretar los resultados de desempeño en esta prueba de selección, se aplicó la métrica de scale score, una escala de 200 a 800 (College Board, s.f).

Las habilidades cognitivas incluidas en la en la Prueba de Selección Universitaria para la Licenciatura en Antropología, se basaron en la Revisión de la Taxonomía de Bloom de Anderson y Krathwohl. En este caso, el verbo de cada Expectativa de Logro indicó a que habilidades de pensamiento se tenía que asociar cada ítem, de orden inferior hasta superior. Los 40 ítems de esta subprueba tienen las siguientes categorías: 14 ítems de recordar, 7 ítems de comprender, 6 ítems de aplicar, 7 ítems de analizar y 6 ítems de evaluar.

Esta Taxonomía tiene una estructura jerárquica que va de lo más simple a lo más complejo. (Anderson & Krathwohl, 2001) refieren que el proceso de aprendizaje del educando se construye desde las Habilidades del Pensamiento de Orden Inferior (LOTS) hacia las Habilidades de Pensamiento de Orden Superior (HOTS). Coincidentemente con este supuesto, se aprecia en la figura 18, que la mayor proporción de estudiantes se encuentran en las 3 primeras habilidades, o sea el 70% de la muestra, que son las de orden

inferior y solamente el 30% de estudiantes de esta muestra pueden contestar correctamente los ítems de habilidades de orden superior.

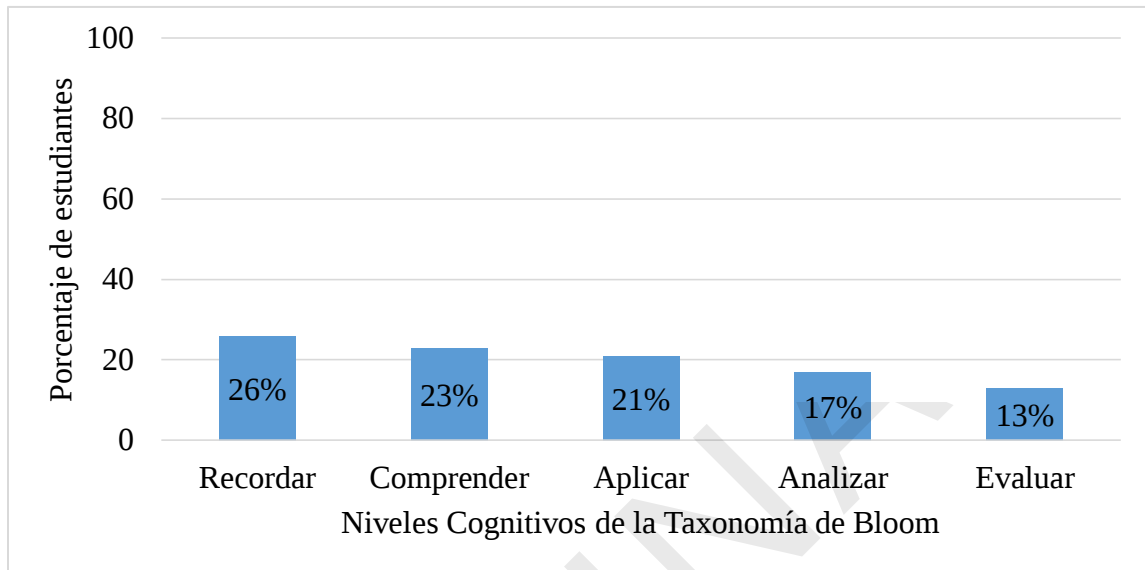


Figura 18. Porcentaje promedio de estudiantes que contestaron correctamente en la SubPrueba de Antropología según los Niveles Cognitivos de la Taxonomía de Bloom. Elaboración propia.

Para el caso de los 20 ítems de la SubPrueba de Fundamentos de Investigación Social los ítems están categorizados de la siguiente manera: 7 ítems de recordar, 4 ítems de comprender y 9 ítems de analizar, ver a figura 19. En esta subprueba, se utilizó el mismo procedimiento que en la de Antropología y la proporción de estudiantes en ambas subpruebas también se asemejan. En este caso, para los ítems asociados con las habilidades de pensamiento de orden inferior la proporción de estudiantes que contestó correctamente es del 73% y los contestaron los ítems de las habilidades de pensamiento de orden inferior representa el 27%, en concordancia con el postulado de la Taxonomía de Bloom.

Después de analizar ambas subpruebas, se infiere que el proceso de aprendizaje de los estudiantes de los centros educativos participantes en esta prueba, se construye sobre la base de recordar, comprender y aplicar para llevarlos a usar y aplicar habilidades como de analizar y evaluar procesos, pero solo la tercera parte de ellos logró contestar un ítem de habilidad de orden superior.

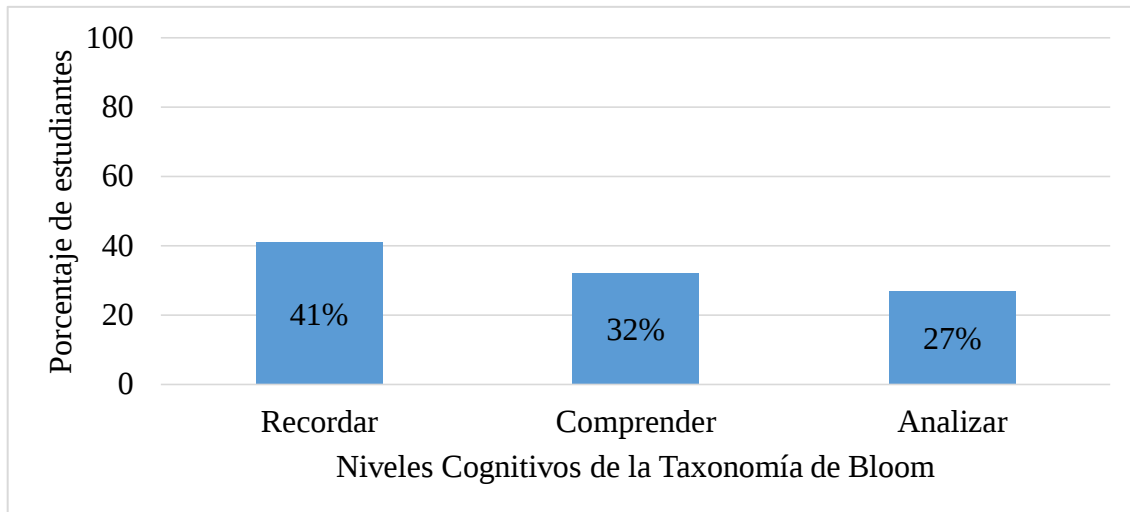


Figura 19. Porcentaje promedio de estudiantes que contestaron correctamente en la SubPrueba de Fundamentos de Investigación Social según los Niveles Cognitivos de la Taxonomía de Bloom. Elaboración propia.

Para la construcción de ítems se utilizó el modelo de referencia curricular proporcionado por la Secretaría de Educación, denominado Plan de Estudios y Programas Curriculares del Bachillerato en Ciencias y Humanidades. Es decir, cada ítem de la prueba siguió explícitamente a una serie de categorías que vinculaban contenidos, expectativas de logro y las habilidades cognitivas declaradas en el currículum vigente (DEMRE, Universidad de Chile, 2016).

Los contenidos seleccionados para esta prueba, corresponden a lo que los docentes invitados al Taller de Redacción de Ítems consideraron como prioritarios para alcanzar el perfil deseado en cada espacio curricular. En el caso de Antropología, su programa consta de 4 unidades, se procuró elegir como mínimo un contenido de cada unidad. En la figura 13, se puede apreciar que los tipos de ítems que los estudiantes más contestaron fueron los de los contenidos sobre las teorías del origen del ser humano, el 64% de los estudiantes y el poblamiento de América el 61% y el que menos contestaron fueron los relacionados con la Hominización el cual representa el 33%.

En el caso de los contenidos referentes a las teorías del origen del ser humano y el poblamiento de América, son ítems relacionados con habilidades de orden inferior, en las cuales los estudiantes contestan correctamente más ítems. En el caso de la hominización,

también se cita el análisis de la figura 13 y se puede inferir que al ser ítems de habilidad de orden superior, se entiende que haya menor proporción de estudiantes.

En resumen, los 40 ítems de la subprueba de Antropología respecto a la dificultad promedio, quedan de la siguiente manera: Divisiones de la Antropología 40% (relativamente difícil), Teorías del Origen del Ser humano 64% (relativamente fácil), Hominización 33% (relativamente difícil), Organización social y económica 54% (dificultad adecuada), Teorías sobre el poblamiento de América 61% (dificultad adecuada) y Pueblos indígenas y afrodescendientes 49% (relativamente difícil).

“El índice de dificultad representa la cantidad de estudiantes que contestaron el ítem correctamente” (AIR, 2015, p.9), en este caso sería los ítems organizados por contenido. Respecto a estos valores, los ítems de Organización social y económica que implica habilidades cognitivas de comprensión, en promedio está 10% de diferencia respecto a los contenidos de habilidad cognitiva recordar. Y teorías del origen del ser humano respecto al poblamiento de América, aunque son de la misma habilidad cognitiva, difieren en la categorización de dificultad.

También se revisó el Diseño Curricular Nacional Básico (DCNB) de Ciencias Sociales de primer a noveno grado (Educación Básica), para tener una noción de la base que tienen los estudiantes sobre estos contenidos cuando llegan a Educación Media y al inicio de este análisis se infería, que los ítems de los Pueblos indígenas y afrodescendientes hubiesen sido contestados con bastante familiaridad, porque la temática se ha abordado en la mayoría de los años anteriores de formación, pero se categorizan como relativamente difícil.

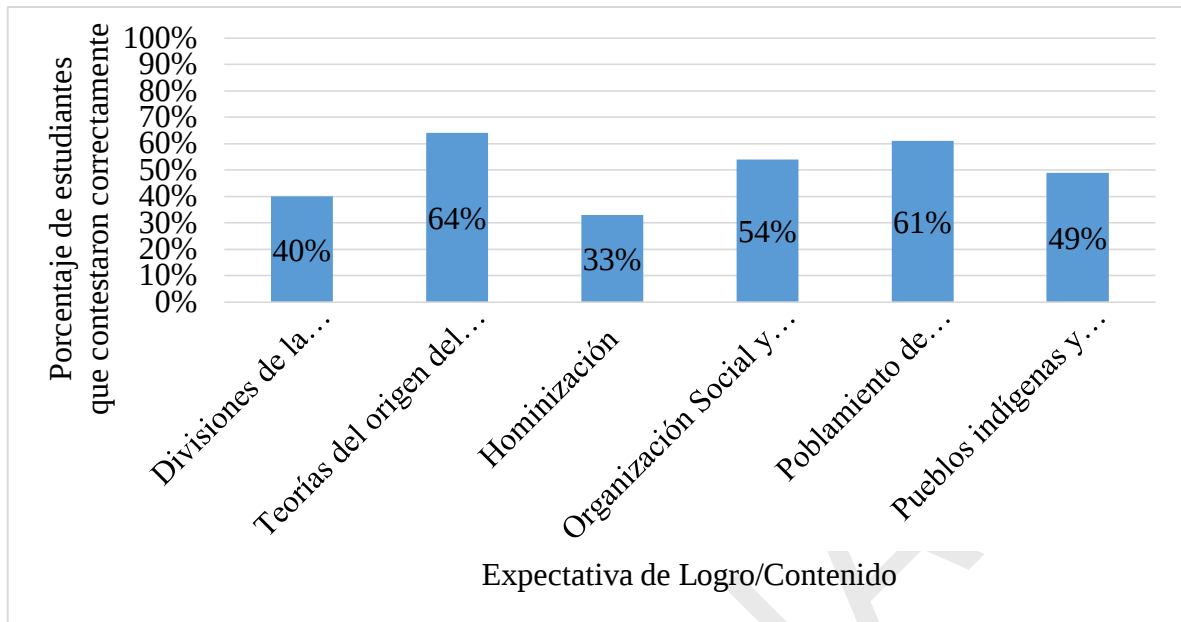


Figura 20. Porcentaje de estudiantes que contestaron correctamente en la SubPrueba de Antropología según las Expectativa de Logro de la Malla Curricular del Bachillerato en Ciencias y Humanidades. Elaboración propia.

Los contenidos seleccionados para esta subprueba, corresponden a lo que los docentes invitados al Taller de Redacción de Ítems consideraron como prioritarios para alcanzar el perfil deseado en cada espacio curricular. En el caso de Fundamentos de Investigación Social, su programa consta de 3 unidades, se procuró elegir como mínimo un contenido de cada unidad. En la figura 20, se puede apreciar que los tipos de ítems que los estudiantes más contestaron fueron los de los contenidos referentes a las Etapas de una Investigación y en los contestados en menor proporción fueron los ítems relacionados con el Método Científico. En este espacio curricular las diferencias de responder correctamente al ítem tomando como referencia el contenido, se puede deber a: profundidad en los contenidos o a las habilidades cognitivas asociadas a cada ítem, ya que como se dijo en la figura 18, el que contenido que tiene mayor porcentaje, está asociado con las habilidades cognitivas de orden inferior.

En resumen, los 20 ítems de la subprueba de Fundamentos de Investigación Social, respecto a la dificultad promedio, quedan de la siguiente manera: Método Científico 38% (relativamente difícil), Metodología de investigación científica 41% (relativamente difícil), Enfoque cualitativo y cuantitativo de la investigación 52% (dificultad adecuada) y Etapas de una investigación 57% (dificultad adecuada).

Analizando cada subprueba y su índice promedio de dificultad, para el caso de Antropología el valor p-value es de 0.51, lo que representa que tiene una dificultad adecuada y para Fundamentos de Investigación Social el valor p-value es de 0.48, lo que significa relativamente difícil. Por lo tanto, hay mayor proporción de estudiantes que contestan correctamente los ítems de la subprueba de Antropología.

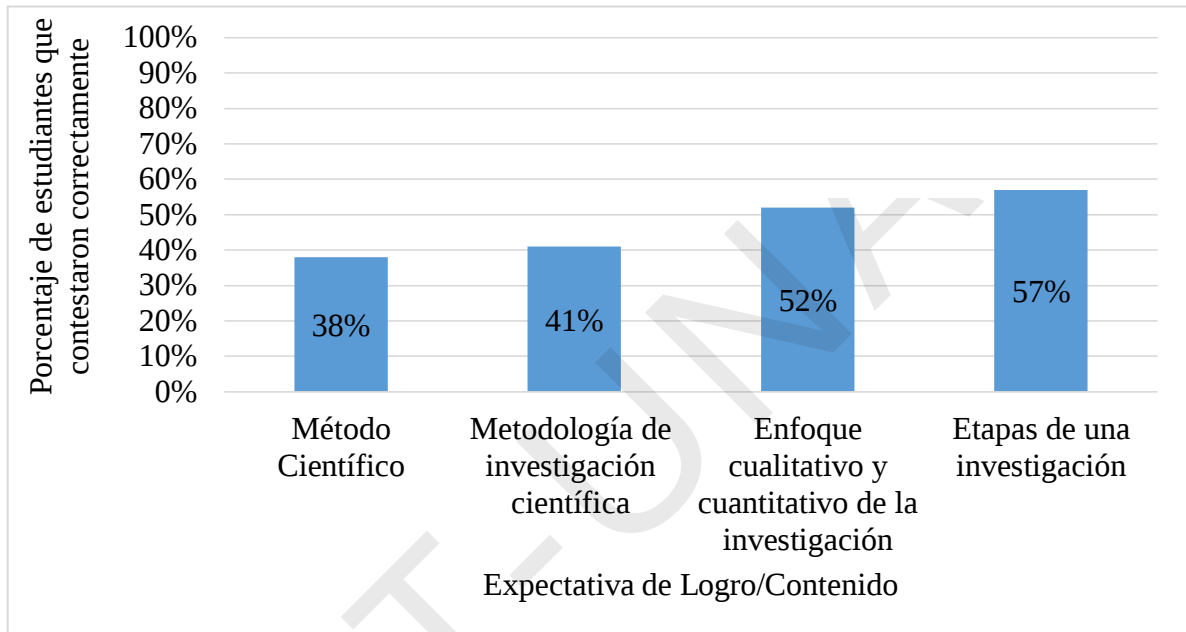


Figura 21. Porcentaje promedio de estudiantes que contestaron correctamente en la SubPrueba de Fundamentos de Investigación Social según las Expectativa de Logro de la Malla Curricular del Bachillerato en Ciencias y Humanidades. Elaboración propia.

Tabla 17

Métrica para representar los resultados de la Prueba de Selección Universitaria para la Licenciatura en Antropología

Puntuación a Escala
340
351
357
363
368
374
380
385

391
397
402
408
414
420
425
431
437
442
454
471
476
482
488
493
499
505
510
516
527
533
539
544
550
607
612
618
624
629
635
641
646
652
658
663
669

Nota: Elaboración propia.

Esta prueba como se indicó en el planteamiento del problema, es complementaria a la PAA, por lo tanto también se utilizó en esta investigación la mismas métrica que emplea el College Board para informar sus resultados de desempeño en la prueba: se realizó la transformación de puntuaciones para obtener las puntuaciones estandarizadas o Scale Score de la Prueba de Selección Universitaria para la Licenciatura en Antropología.

Para la transformación a esta métrica, se utilizó el procedimiento siguiente:

1. Se convirtió cada puntaje a puntaje Z. La fórmula fue la siguiente:

X= Puntaje en la Prueba

$\bar{x}$ = Media de la Prueba.

S= Desviación estándar

$$Z = \frac{X - \bar{X}}{S}$$

2. Después de convertir los puntajes a Puntajes Z. Después se convirtieron los Puntajes Z a Escala de puntuación estandarizada o Scale Score. La fórmula aplicada fue la siguiente:

$$Ti = a \cdot zi + b$$

a= desviación estándar.

zi= puntaje Z.

b=media.

Hay una serie de escalas derivadas de amplio uso con valores fijados para la media y desviación típica, como las puntuaciones T (50, 10), CI de desviación (100,15 ó 16), College Board (500,100), NAEP (250, 50), PISA (500, 100), etc. En este caso, como es una prueba complementaria a la PAA se utilizó la escala derivada con valores fijados para la media y desviación típica del College Board (500,100) AIDEI, 2011 (Tomado del College Board).

De la tabla 17, se infiere lo siguiente:

Puntaje más alto: 669
Puntaje más bajo: 340
Moda: 550
Media: 499
Mediana: 499

En la figura 16, se hizo mención al análisis de FDI y el resultado en los 60 ítems oscilaba entre inexistente y bajo, anticipando que las diferencias por género en esta investigación serían reales y no producto de que algún ítem esté funcionado a favor de un grupo. Referente al proceso de admisión en la UNAH “Curiosamente, aunque son más las mujeres que aspiraron entrar y las que lograron matricularse en la UNAH, fueron los hombres los que mostraron un mejor índice de admisión” (El Herald, 2014).

En la figura 22, se utilizó las Clases de Estadísticas para explicar los resultados por género, ya que si no se cuenta con Estándares de Desempeño, no se puede hacer uso de los niveles de desempeño. Para efecto de este análisis se agrupó los datos en 3 intervalos de clase y se categorizaron de la siguiente manera: bajo, promedio y alto. El Intervalo de clase se determinó conteniendo un límite inferior (2) y un límite superior (60), restando ambos datos y así obteniendo el tamaño para cada clase. En este gráfico se utiliza la media, es una medida de tendencia central que denota el promedio de un conjunto de datos. Se calculó dividiendo la suma de datos entre el total de ellos (Instituto Nacional de Estadística e Informática, 2006, pp. 42). Se puede apreciar que no hay diferencias significativas por género, mantienen un promedio similar en las 3 categorías.

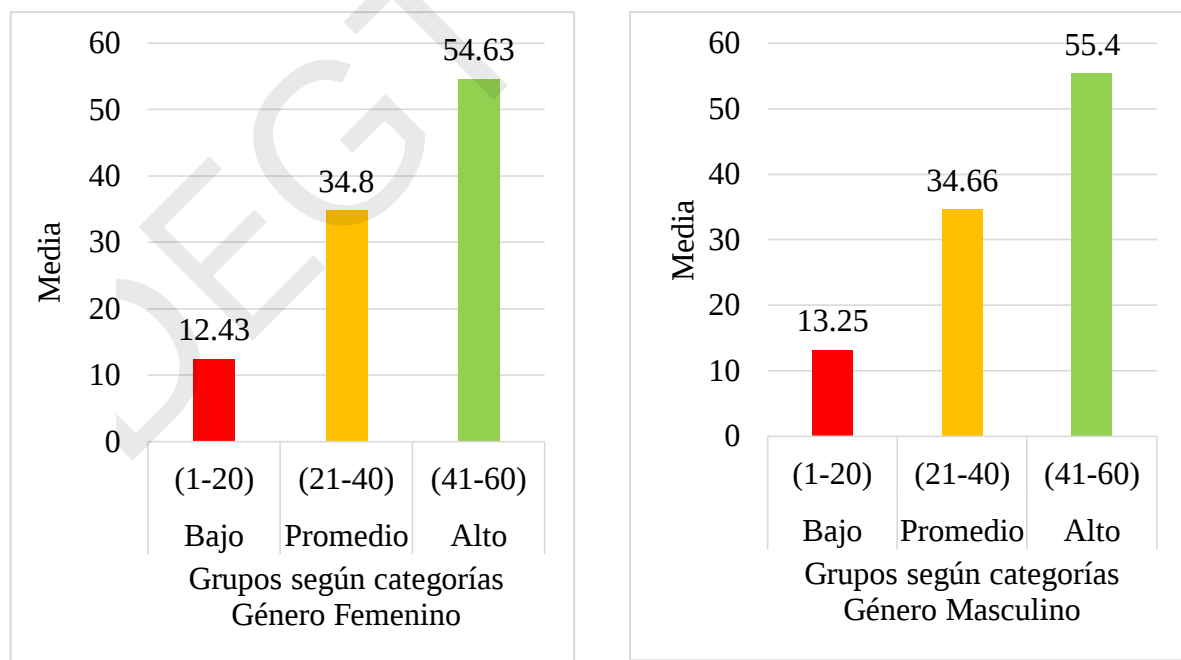


Figura 22. Grupos según categorías por género en la Prueba de Selección Universitaria para la Licenciatura en Antropología. Elaboración propia. Elaboración propia.

En la figura 23, podemos apreciar que el puntaje que más se repite de la Escala Estandarizada es el de 550 ósea que 17 de los 158 estudiantes que participó en la prueba obtienen ese puntaje. También este puntaje más frecuente es el que sirve de límite entre las categorías de clase promedio y alto.

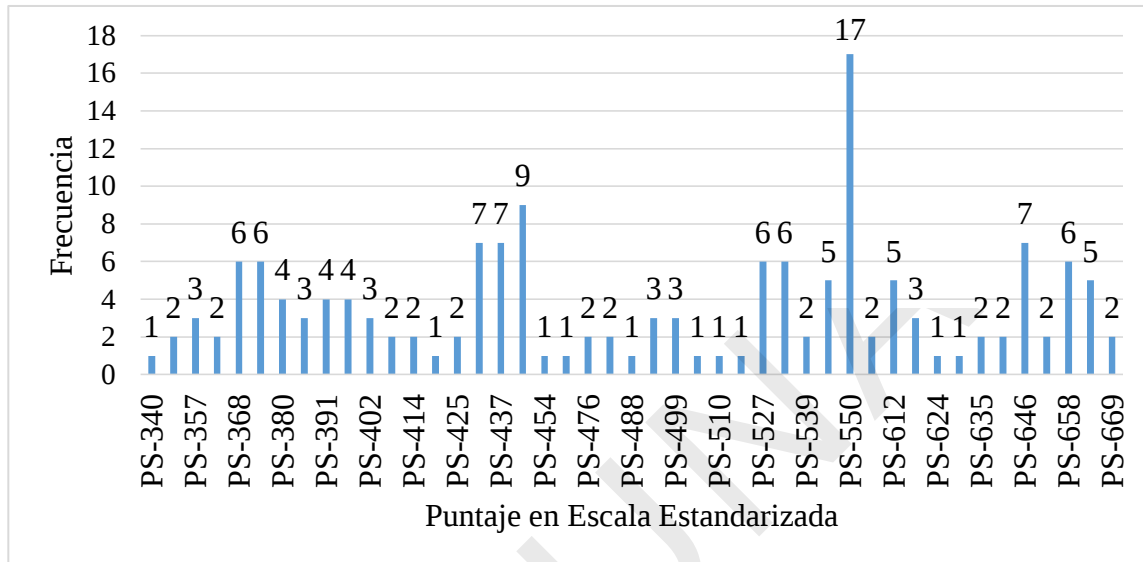


Figura 23. Frecuencia de puntajes en Escala Estandarizada en la Prueba de Selección Universitaria para la Licenciatura en Antropología. Elaboración propia.

“La puntuación mínima que debe obtener un candidato para entrar a la UNAH es de 700 puntos, de no lograrlo tiene dos oportunidades más para someterse a la PAA, misma que solo puede realizarse en tres oportunidades” (Blog UNAHVS, 2018). El propósito de hacer un ejemplo de cuantos ingresarían con el puntaje mínimo de 700 es solo para contrastar como cambia la proporción de ser seleccionado.

El índice PAA requerido para la Licenciatura en Antropología es de 900 puntos, lo que significa que un estudiante requiere contestar el 56% de la PAA para poder ser admitido a dicha carrera. Por lo tanto, se utilizó este criterio para determinar el puntaje mínimo para ser seleccionado con la Prueba de Selección Universitaria para la Licenciatura en Antropología, este en puntuación escala sería 516 y en puntaje bruto sería 33. Por lo tanto, como resultados obtendríamos que el 47% de los estudiantes de esta muestra aprobarían esta prueba.

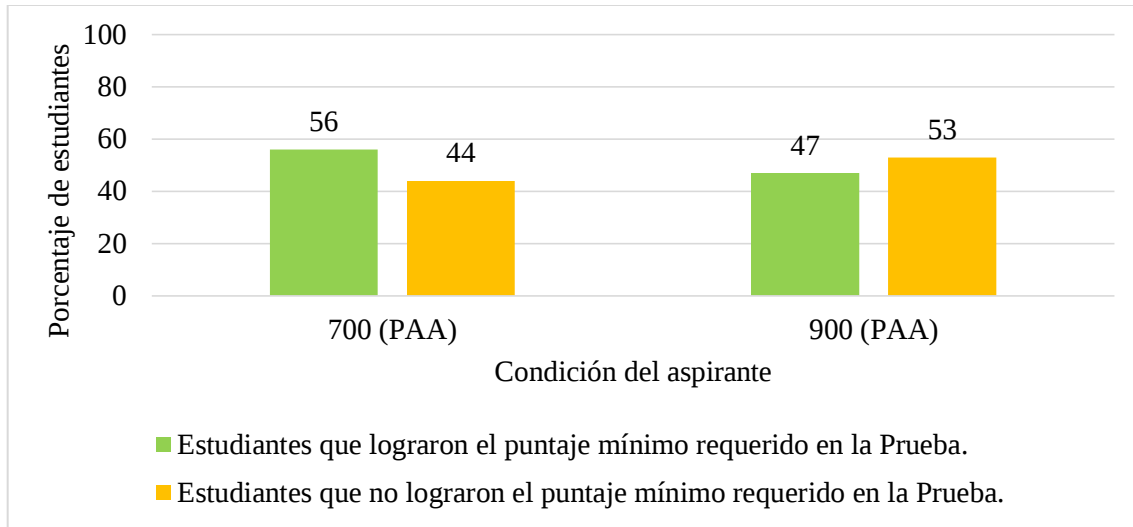


Figura 24. Comparabilidad de ingreso a la carrera con la Prueba de Selección Universitaria para la Licenciatura en Antropología teniendo como referencia el puntaje mínimo requerido en la PAA. Elaboración propia.

Objetivo específico 5

Analizar los criterios vertidos por los estudiantes de primer ingreso de la Licenciatura en Antropología, con respecto al desarrollo de los contenidos de Antropología y Fundamentos de Investigación Social en el Bachillerato.

Para el logro de este objetivo, se utilizó la técnica de recolección de información cualitativa, denominada grupo focal. Su uso permite hacer un aporte enriquecedor para comprender mejor el análisis de desempeño en la Prueba de Selección Universitaria para la Licenciatura en Antropología. El grupo focal, permitió tener una representación colectiva de estudiantes de primer ingreso de la carrera de Antropología a nivel microsocial, los cuales proporcionaron información desde su experiencia personal educativa, de cómo fue el desarrollo de los contenidos que conforman la Tabla de Especificaciones de Antropología y Fundamentos de Investigación Social (Mella, 2000).

Antes de la organización de este grupo focal además de los contenidos comprendidos en la Tabla de Especificaciones, también se revisó el Diseño Curricular Nacional Básico (DCNB) de Ciencias Sociales de primer a noveno grado (Educación Básica), para tener una noción de la base que tienen los estudiantes sobre estos contenidos cuando llegan a Educación Media. Los resultados de esta recolección de información son los siguientes:

1. Los participantes del grupo focal opinaron que el contenido de las divisiones de la Antropología: nos enseñaron de lo que se trataba cada una, a profundidad no nos lo enseñaron.
2. Los participantes del grupo focal opinaron que el contenido de las Teorías del origen del ser humano: nos enseñaron los términos más importantes.
3. Los participantes del grupo focal opinaron que el contenido del proceso de Hominización: Nos hablaron sobre las características principales de cada una de las especies del género Homo, no tan amplio, nos enseñaron lo que la mayoría de las personas conocen, no a profundidad.
4. Los participantes del grupo focal opinaron que el contenido de la organización social y económica: los temas mencionados en esta unidad no nos son familiares, no los recordamos que nos enseñaran.

5. Los participantes del grupo focal opinaron que el contenido del Poblamiento de América y Honduras: nos enseñaron en qué consistía cada teoría del poblamiento de América, los supuestos básicos.

6. Los participantes del grupo focal opinaron que el contenido de Pueblos indígenas y afrodescendientes: nos enseñaron lo básico como cualquier otra clase de Ciencias Sociales, no fue a profundidad. Abordábamos ese contenido haciendo una feria.

7. Los participantes del grupo focal opinaron que el contenido del Método Científico: lo recordamos en un proyecto ¿Por qué los estudiantes reprueban la clase de Química?

8. Los participantes del grupo focal opinaron que el contenido de la Metodología de investigación científica: si nos enseñaron de forma general, un poquito de cada una.

9. Los participantes del grupo focal opinaron que el contenido del Enfoque cualitativo y cuantitativo de la investigación: si lo vimos, nos hicieron elegir que enfoque íbamos a utilizar en esta investigación. Nos explicaron las diferencias entre los dos o si queríamos utilizar el mixto.

10. Los participantes del grupo focal opinaron que el contenido del Etapas de una investigación: abordábamos este tema realizando una investigación basado en un problema del centro educativo y nosotros teníamos que buscar una solución. Consideramos que nos dieron una breve preparación.

A manera de resumen, los participantes del grupo focal consideraron que su experiencia en el espacio curricular de Antropología como “regular”, no fue ni “buena” ni “mala”, como para llenar un espacio. Dijeron que muchas veces influye cómo el maestro desarrolla la clase, “porque si un maestro solo va a leer, a uno no le llama la atención”. Perciben que era una clase como requisito para graduarse, a la cual no le daban tanta importancia como a las Matemáticas o a la Química, por mencionar algunos ejemplos. Era “como aquella clase que les daba la maestra, tenían que sentarse y memorizar lo que les decía”. Si se hubiesen basado solamente con lo que les enseñaron en el colegio, no la hubiesen elegido como carrera en la universidad. Respecto a la clase de Fundamentos de Investigación Social, la consideraron como “buena”, “nos enseñaron lo elemental, pero nos gustó”.

La información recogida con el grupo focal no tiene la intención de generalizar, sino es un aporte enriquecedor para tener una noción sobre el contexto enseñanza-aprendizaje respecto a estos dos espacios curriculares, en esta investigación los contenidos con mayor proporción de estudiantes que contestaron correctamente esos ítems son los contenidos de Teorías del origen del ser humano, Poblamiento de América y Etapas de una investigación, coincide con las opiniones vertidas en el grupo focal, en el cual consideran que estas temáticas sí fueron abordadas.

Respecto a la Hominización y el Método Científico, que son los contenidos que muestran menor desempeño, los estudiantes comentaron que esos contenidos fueron abordados muy superficialmente en el aula de clase. En relación con que los ítems del contenido de los Pueblos indígenas y afrodescendientes hubiesen sido contestados con bastante familiaridad porqué se desarrollan en años anteriores de formación, en palabras de los estudiantes, dicen que este tema se aborda con una feria y no es explicado por el docente.



Figura 25. Grupo focal estudiantes de primer ingreso de la Licenciatura en Antropología. Fotografía propia.

Capítulo 6. Síntesis de los Resultados: Conclusiones

El objetivo general del presente estudio fue el diseño de una prueba de selección universitaria basada en las expectativas de logro y habilidades cognitivas que se requieren en los espacios curriculares de Antropología y Fundamentos de Investigación Social establecidos en el Plan de Estudios y Programa Curricular del Bachillerato de Ciencias y Humanidades, para aspirantes a ingresar a la Licenciatura en Antropología de la Facultad de Ciencias Sociales de la Universidad Nacional Autónoma de Honduras. Por ello, los hallazgos principales son los productos alcanzados para cada uno de los objetivos específicos planteados en el capítulo 1.

A continuación se exponen los hallazgos principales en el orden y correspondencia con los objetivos de investigación, siguiendo la secuencia de las etapas de la construcción de pruebas estandarizadas, tanto la piloto como la operativa, en el marco de la Teoría Clásica de los Test (TCT).

- 1. Se elaboró una Tabla de Especificaciones conteniendo las expectativas de logro y los contenidos en relación con las habilidades cognitivas que más aportan al perfil de salida de los espacios curriculares de Antropología y Fundamentos de Investigación Social del Plan de Estudios y Programas Curriculares del Bachillerato en Ciencias y Humanidades, para la Prueba de Selección Universitaria de la Licenciatura en Antropología.**

En correspondencia con lo señalado por la literatura especializada en evaluación educativa, la que señala que “La clave para organizar correctamente una evaluación consiste en definir una buena tabla de especificaciones que plasme la relación entre los contenidos del programa de la asignatura y los objetivos de aprendizaje que pretendemos que asuman nuestros alumnos” (Sans Martín, 2008, p.8), en el presente trabajo de investigación se construyó una detallada Tabla de Especificaciones detallando la información correspondiente.

Los datos contenidos señalan cuatro expectativas para Fundamentos de Investigación Social y seis para Antropología, que son la base de la Prueba de Selección Universitaria para la Licenciatura en Antropología y vinculadas con las habilidades cognitivas, la prueba oficial

quedaría estructurada con el 63% de los ítems de habilidades de orden inferior y 37 de orden superior, contando con 60 ítems en total. De los cuales el 35% corresponden al nivel Reconocer, 18% al de Comprender, 10% al de Aplicar, 27% al de Analizar y 10% al de Evaluar.

Tabla 18

Tabla de especificaciones para la Prueba de Selección Universitaria de la Licenciatura en Antropología

Contenido a evaluar				Nivel cognitivo						Total
Espacio	Contenido	Código/Expectativa de logro		A	B	C	D	E	F	
2. Fundamentos de Investigación Social	Método científico	111	Analizar el método científico y sus características.				4			4
	Metodología de la investigación científica	121	Caracterizar la metodología de investigación científica tomando en cuenta la profundidad y finalidad del estudio.				5			5
	Tipos de investigación según el enfoque	131	Reconocer las características del enfoque cualitativo y cuantitativo de la investigación		4					4
	Etapas de una investigación.	141	Comprender las etapas de una investigación.	7						7
	Divisiones de la Antropología	211	Caracterizar las principales divisiones de la Antropología				7			7
	Teorías del origen y evolución del ser humano.	221	Conocer las teorías acerca del origen de los seres humanos.	7						7

3. Antropología	Hominización	231	Explicar el proceso de hominización y socialización humana.	6					6
	Contenidos básicos de la Antropología Social	241	Reconocer la existencia de diversos tipos de organización económica y social.	7					7
	Poblamiento de América	251	Conocer en forma resumida los principales datos sobre el poblamiento del continente americano y Honduras.	7					7
	Pueblos indígenas y afrodescendientes de Honduras.	261	Diferenciar los Pueblos indígenas y afrodescendientes de Honduras.	6					6
	Número total de ítems por nivel cognitivo:			21	11	6	16	6	60
Porcentaje de ítems por nivel cognitivo:			35	18	10	27	10	100	

Nota: El significado de las letras según el nivel cognitivo A= Reconocer B= Comprender C= Aplicar D= Analizar E= Evaluar F=Crear. Elaboración propia.

2. Se desarrolló el proceso de construcción y de ensamblaje para que la Prueba Oficial de Selección Universitaria para aspirantes a ingresar a la Licenciatura de Antropología, de manera que cumple con los indicadores psicométricos de calidad bajo el análisis de la Teoría Clásica de los Test (TCT).

De acuerdo con lo señalado por la referencia especializada, los índices más usados al hacer el análisis psicométrico de una prueba de selección universitaria de acuerdo a la Teoría Clásica de los Test, son el índice de dificultad, el índice de discriminación, análisis de

distractores, validez y confiabilidad. Estos índices se convierten en indicadores de la calidad de una prueba en la medida que se encuentren dentro de los valores aceptables (J. Meneses et al, 2013).

Dado que en el análisis de la Aplicación Operativa de la Prueba de Selección Universitaria para la Licenciatura en Antropología, se incluyeron los mismos pasos y la revisión estadística de los indicadores psicométricos que para el análisis de las Aplicación Piloto, en esta síntesis de resultados, solo se comentarán los datos correspondientes a la primera.

El índice de dificultad de los ítems se mide por los valores p-value de cada ítem: “El índice de dificultad representa la cantidad de estudiantes que contestaron el ítem correctamente” (AIR, 2015, p.9). Según las categorías especificadas por la literatura especializada “El rango de valores del índice de dificultad es de 0 a 1.0, los valores aceptables del índice de dificultad oscilan entre 0.3 y 0.8” (AIR, 2015, p.9), y los ítems de la prueba operativa cumplieron ese requisito. En esta prueba se clasificó a los 60 ítems de la siguiente manera: 29 ítems relativamente difíciles (0.31 a 0.50), 29 ítems con dificultad adecuada (0.51 a 0.65) y 2 ítems relativamente fáciles (0.66 a 0.80).

Un segundo parámetro analizado fue el índice de discriminación, entendido como “...la medida en que tan bien un ítem permite distinguir entre aquellos estudiantes que han alcanzado las expectativas de logro y aquellos que no” (AIR, 2015, p.9). Según las categorías especificadas para la prueba operativa, se clasificó a los 60 ítems de la siguiente manera: 53 ítems con poder de discriminación excelente ($0.40 >$) y 7 ítems con poder de discriminación buena (0.30-0.39), considerando que la bibliografía especializada señala que “El rango de valores del índice de discriminación es de -1.0 a 1.0 con valores aceptables mayores a 0.30” (AIR, 2015, p.9).

Un tercer parámetro considerado en el análisis fue el coeficiente de discriminación de los distractores, definido como “...la medida en qué tan bien un ítem permite distinguir entre aquellos estudiantes que han alcanzado las expectativas de logro y aquellos que no” (AIR, 2015, p.9). Uno de los procedimientos recomendados por la literatura internacional, fue analizar la frecuencia en que cada distractor es seleccionado. En el caso de que la frecuencia de un distractor sea menor al 2% o nula proporción, se puede decir que el

distractor no está cumpliendo la función que corresponde, que es la de ser una posible respuesta para aquel estudiante que no ha logrado esa expectativa de logro, por ende, el ítem puede ser contestado fácilmente hasta por sentido común. También en el caso inverso, en el que la selección de un distractor es mayor que la respuesta correcta, se entiende que esta opción está confundiendo a los estudiantes y es recomendable reestructurar el ítem (AIR, 2015). Todos los ítems de la prueba operativa cumplieron estos valores en sus distractores. Respecto a la prueba en su conjunto, cumple con los valores aceptables del Alfa de Conbrach y con el valor requerido en la Validez de Contenido, estadísticamente hablando y de contenido, la prueba cumple con los requisitos psicométricos de calidad.

3. Determinar el Funcionamiento Diferencial de los ítems en las pruebas piloto y operativa, en relación con el género, para evitar discriminación en las oportunidades de selección y con el fin de garantizar la calidad de los resultados.

Los estudios internacionales, recomiendan que para los instrumentos de evaluación como el desarrollado en el presente trabajo de investigación, es necesario hacer el análisis del Funcionamiento Diferencial del Ítem (FDI): “Las evaluaciones psicológicas deben garantizar la equidad y validez de las interpretaciones y decisiones adoptadas a partir de las mismas. Para ello es necesario la utilización de instrumentos libres de sesgo, y capaces de evaluar necesidades personales y sociales de individuos con diferentes características” (Gómez, Hidalgo y Guilera, 2010, párr.1).

Los resultados del análisis FDI en relación con la variable género en los ítems de la prueba operativa indican que se aplicó a 158 estudiantes (91 señoritas y 67 varones), de undécimo grado del Bachillerato en Ciencias y Humanidades. La prueba estaba compuesta de 60 ítems. Para la detección del FDI en los ítems se realizó el método de Diferencias de Medias Estandarizadas. La prueba indicó la existencia de DIF en las siguientes categorías: 22 ítems con valores de FDI pequeño o inexistente (0 y .05) y 38 ítems con valores de FDI bajo (>.05 a .075).

Con estos resultados se concluyó que el FDI es tan imperceptible que no afecta en ninguno de los ítems y por tanto los puntajes obtenidos no se encuentran afectados por FDI. La importancia del FDI en este ámbito, es porque no se debe olvidar, que un instrumento se administra con un objetivo concreto, generalmente para tomar decisiones (selección

universitaria) que en este caso son relevantes para la vida de los aspirantes a ingresar a la carrera de Antropología.

4. Al analizar el rendimiento de los estudiantes en referencia a las habilidades cognitivas asociadas a conocimientos específicos de Antropología y Fundamentos de Investigación Social se encontró que el porcentaje promedio de estudiantes que contestaron correctamente ambas sub pruebas, es claramente decreciente en la medida que se consideran niveles cognitivos más complejos en la Taxonomía de Bloom.

Las habilidades cognitivas incluidas en la en la Prueba de Selección Universitaria para la Licenciatura en Antropología, se basaron en la Revisión de la Taxonomía de Bloom de Anderson y Krathwohl. En este caso, el verbo de cada Expectativa de Logro indicó a que habilidades de pensamiento se tenía que asociar cada ítem, de orden inferior hasta superior. Los 40 ítems de la subprueba de Antropología tienen las siguientes categorías: 14 ítems de recordar, 7 ítems de comprender, 6 ítems de aplicar, 7 ítems de analizar y 6 ítems de evaluar. Esta Taxonomía tiene una estructura jerárquica que va de lo más simple a lo más complejo. (Anderson & Krathwohl, 2001) refieren que el proceso de aprendizaje del educando se construye desde las Habilidades del Pensamiento de Orden Inferior (LOTS) hacia las Habilidades de Pensamiento de Orden Superior (HOTS). Coincidentemente con este supuesto, en la sub prueba de Antropología los resultados indican que la mayor proporción de estudiantes se encuentran en las 3 primeras habilidades, ósea el 70% de la muestra, que son las de orden inferior y solamente el 30% de estudiantes de esta muestra pueden contestar correctamente los ítems de habilidades de orden superior.

Para el caso de los 20 ítems de la subprueba de Fundamentos de Investigación Social los ítems están categorizados de la siguiente manera: 7 ítems de recordar, 4 ítems de comprender y 9 ítems de analizar. En esta subprueba, se utilizó el mismo procedimiento que en la de Antropología y la proporción de estudiantes en ambas subpruebas también se asemejan. En este caso, para los ítems asociados con las habilidades de pensamiento de orden inferior la proporción de estudiantes que contestó correctamente es del 73% y los contestaron los ítems de las habilidades de pensamiento de orden inferior representa el 27%, en concordancia con el postulado de la Taxonomía de Bloom.

Después de analizar ambas subpruebas, podemos inferir que el proceso de aprendizaje de los estudiantes de los centros educativos participantes en esta prueba, se construye sobre la base de recordar, comprender y aplicar para llevarlos a usar y aplicar habilidades como de analizar y evaluar procesos, pero solo la tercera parte de ellos logró contestar un ítem de habilidad de orden superior.

5. El Análisis de los criterios vertidos por los estudiantes de primer ingreso de la Licenciatura en Antropología, con respecto al desarrollo de los contenidos de Antropología y Fundamentos de Investigación Social en el Bachillerato, identifica que recibieron mejor formación en la segunda, mientras que se quedaron con muchos vacíos respecto a los contenidos propios de Antropología. Pero elemental, no siempre es sinónimo de cumplir con las Expectativas de Logro.

Para el logro de este objetivo, se utilizó la técnica de recolección de información cualitativa, denominada grupo focal. Su uso permite hacer un aporte enriquecedor para comprender mejor el análisis de desempeño en la Prueba de Selección Universitaria para la Licenciatura en Antropología. El grupo focal, permitió tener una representación colectiva de estudiantes de primer ingreso de la carrera de Antropología a nivel microsocial, los cuales proporcionaron información desde su experiencia personal educativa, de cómo fue el desarrollo de los contenidos que conforman la Tabla de Especificaciones de Antropología y Fundamentos de Investigación Social (Mella, 2000). Debe recordarse que la información recogida con el grupo focal no tiene la intención de generalizar, sino es un aporte enriquecedor para tener una noción sobre el contexto enseñanza-aprendizaje respecto a estos dos espacios curriculares.

De acuerdo con los resultados de la prueba, los contenidos con mayor proporción de estudiantes que contestaron correctamente esos ítems son los contenidos de Teorías del origen del ser humano, Poblamiento de América y Etapas de una investigación, coincide con las opiniones vertidas en el grupo focal.

A manera de resumen, los participantes del grupo focal consideraron que su experiencia en el espacio curricular de Antropología como regular, “no fue ni buena ni mala, como para llenar un espacio”. Dijeron que muchas veces influye “como el maestro da la clase, porque si un maestro solo va a leer, a uno no le llama la atención”. Sintieron que era

una clase “como requisito para graduarse”, no le daban tan importancia como a las Matemáticas o a la Química. Era como “aquella clase que les daba la maestra, tenían que sentarse y memorizar lo que les decía”. Si “se hubiesen basado con lo que les enseñaron en el colegio, no la hubiesen elegido” la Antropología como la carrera universitaria a cursar. Respecto a la clase de Fundamentos de Investigación Social la consideraron como “buena, nos enseñaron lo elemental, pero nos gustó”. Sin bien es cierto que la percepción respecto al espacio curricular de Fundamentos de Investigación Social es de satisfacción, ellos agregaron los términos “nos enseñaron lo elemental”, esta expresión con los resultados de la prueba que indican que la subprueba de este espacio curricular tiene mayor dificultad promedio que la de Antropología, categorizándose como relativamente difícil. La de Antropología en promedio tiene dificultad adecuada, pero coincide con los resultados respecto al contenido de hominización que fue el que contestaron menor proporción de los estudiantes, lo consideraron un tema no desarrollado a profundidad, cuando su expectativa demanda que sea asociado con la habilidad cognitiva de Evaluar.

Recomendaciones

1. Es necesario que la educación superior desempeñe un papel importante en el conjunto del sistema educativo hondureño y la UNAH como rectora de este nivel educativo, se sugiere que se considere conveniente ampliar el uso de las pruebas de selección universitaria complementarias a la PAA basadas en el Currículo de Educación Media, para que los estudiantes que están por egresar no vean su formación en las modalidades de Bachillerato, como una formación aislada, sino que la consideren como una formación base para su vida universitaria.

2. Revisar el Plan de Estudios y Programas Curriculares, específicamente el de Antropología y Fundamentos de Investigación Social. Se sugiere a los especialistas de la Secretaría de Educación revisar algunos aspectos. En el caso del espacio curricular de Antropología, una tendencia fuerte en el “mundo académico de Antropología” que se acentúa actualmente, consiste en rechazar la definición de etnia porque consideran que conlleva una herencia pesada, la de una práctica que hace de ella un simple sustituto de la noción de raza, por ello se sugiere utilizar los términos de pueblos indígenas y afrodescendientes, según sea el caso y sustituir los anteriormente mencionados. En lo que respecta al espacio de Fundamentos de Investigación Social, en la Unidad II, se sugiere revisar errores de contenido, como clasificar a los métodos Hermenéutico, Fenomenológico e Investigación Acción en el Enfoque Cuantitativo.

3. En la legislación del sistema educativo está contemplada la formación permanente. En el caso de la línea de investigación de selección universitaria (Ciencias Sociales, Física e Inglés) de la Maestría en Psicometría y Evaluación Educativa, se realizó un taller en el cual se desarrolló todo un programa de evaluación educativa (no solo de redacción de ítems), con los docentes de Educación Media. En este caso, la UNAH podría contribuir de manera significativa considerando conveniente replicar este tipo de iniciativas mediante un programa de Vinculación Académica de Capacitación, Talleres y Seminarios que sus docentes pueden impartir a los docentes de Educación Media, para ayudar a fortalecer sus competencias en dominio de contenido y en aquellos casos que llevan muchos años de servicios en la docencia, sería de actualización en las últimas tendencias de cada

disciplina. Incluso pueden sugerir referencias bibliográficas para utilizar en las planificaciones. O como en el caso de la Carrera de Antropología que elaboró un libro para bachillerato, esto lo podrían replicar el resto de las carreras y dar la requerida difusión para su uso.

4. La UNAH, debe seguir en el “camino” de la inclusión y enriquecer sus políticas de discriminación positiva. Porque en un contexto socioeconómico como el hondureño, requiere que la máxima casa de estudios del país, implemente políticas dirigida a mejorar la calidad de vida de grupos desfavorecidos y reducir las brechas de las desigualdades educativas con las que provienen muchos aspirantes debido a diversos factores de la realidad nacional.

DEGT-UNAH

Bibliografía

- Adam, A. M. (2007). *An introduction to learning outcomes*. Articl B, 2, 3 -1.
- Agencia Ejecutiva en el Ámbito Educativo, Audiovisual y Cultural, AQU. (2009). *Pruebas nacionales de evaluación del alumnado en Europa: objetivos, organización y utilización de los resultados*. Madrid: Secretaría de Educación de España. Recuperado de <http://redined.mecd.gob.es/xmlui/bitstream/handle/11162/66229/00820112013576.pdf?sequence=1>
- Agència per a la Qualitat del Sistema Universitari de Catalunya. (2009). *Guía para la evaluación de competencias en el área de Ciencias Sociales*. Barcelona: Àgata Segura Castellà. Recuperado de http://www.aqu.cat/doc/doc_14646947_1.pdf
- Alvarenga, C. (27 de Julio de 2014). *El 76% de los estudiantes de la unah tienen un índice de aprobación arriba del 60 %*. Presencia Univesitaria. Recuperado de <https://presencia.unah.edu.hn/investigacion-cientifica/articulo/el-76-de-los-estudiantes-de-la-unah-tienen-un-indice-de-aprobacion-arriba-del-60->
- American Institutes For Research AIR (2015). *Informe del Proceso de Evaluación del Rendimiento Academico 2014 en las Areas de Español y Matematicas de 1° a 9° grados*. Honduras.
- Anderson, L.W. y Krathwohl, D.R. (2001). *Una taxonomía para el aprendizaje, la enseñanza y la evaluación: una revisión de la taxonomía de los objetivos educativos de Bloom*. Edición completa. Nueva York: Longman.
- Bancayán C. (2013). *Operacionalización de la taxonomía de Anderson y Krathwohl para la docencia universitaria*. Paideia XXI, Vol. 3, N° 4, Lima, diciembre 2013, pp. 109-119.
- Bloom, B. S. (1956). *Taxonomy of Educational Objectives, the classification of educational goals– Handbook I: Cognitive Domain*. New York: McKay
- Blog UNAHVS. (9 de marzo de 2018). *El puntaje que necesitas para pasar la PAA en UNAH-VS*. Recuperado de <https://blogs.unah.edu.hn/unahvs/el-puntaje-que-necesitas-para/>
- Booker, M.J.(2007). *A Roof without Walls: Benjamin Bloom*. Academic Questions, 20(4), 9.
- Brown, S.R. (1980). *Subjetividad política: aplicación de la metodología Q en ciencias políticas*. Prensa de la Universidad de Yale, New Haven.

Campbell, DT y Fiske, DW (1959). *Validación convergente y discriminante por la matriz multitrait-multimethod*. Boletín psicológico, 56 (2), 81-105.

Cano, V. (1999). *Bibliometric overview of Library and Information Science Research in Spain*. Revista de la bandera de la Sociedad Estadounidense de Ciencias de la Información, Volumen 50, Issue8 1999, Pages 675-680.

Carmines, EG y Zeller, RA (1979). *Evaluación de confiabilidad y validez*. Vol. 17. Thousand Oaks, CA: Salvia. <https://doi.org/10.4135/9781412985642>

CollegeBoard. (2016). *CollegeBoard*. Recuperado de <https://latam.collegeboard.org/page/paa>

Comber, L.E. & Keeves, J.E., 1973. *Science education in nineteen countries*. (Almquist: Stockholm).

Consejo de Educación Superior. *Ley de Educación Superior, Reglamento General de la Ley, Normas Académicas del Nivel de Educación Superior*. Recuperado de https://portalunico.iaip.gob.hn/portal/ver_documento.php?uid=NzUwNTA4OTM0NzYzNDg3MTI0NjE5ODcyMzQy (1994).

Constitución de la República, Recuperado de <http://pdba.georgetown.edu/Parties/Honduras/Leyes/constitucion.pdf> (1982).

Corbetta, P. (2007). *Métodos y técnicas de investigación social*. Madrid, España: McGraw-Hill/Interamericana de España, S.A.U.

Cortada, N. (1999). *Teorías psicométricas y construcción de tests*. Buenos Aires: Editorial Lugar.

Cronbach, LJ y Meehl, PE (1955). *La validez de constructo en las pruebas psicológicas*. Boletín psicológico, 52 (4), 281-302.

Cronbach, LJ (1971). *Validación de prueba*. En R. Thorndike (Ed.), *Educational Measurement* (2nd ed., P. 443). Washington DC: Consejo Americano de Educación.

Cronbach, L.J, Gleser, G.C, Nanda, H. & Rajaratnam, N. (1972). *La fiabilidad de las mediciones de comportamiento. Teoría de la generalización de puntajes y perfiles*. Wiley, Nueva York, 1972. xx, 410 pp.

Departamento de Evaluación, Medición y Registro Educacional DEMRE. (2005). *Teoría Clásica de Medición TC y Teoría de Respuesta al Ítem TRI*. Chile.

Departamento de Evaluación, Medición y Registro Educacional DEMRE. (2013). *Informe Final Evaluación de la PSU Chile*. Chile.

Departamento de Evaluación, Medición y Registro Educacional DEMRE. (2016a). *Informe Técnico 2016, Volumen I, Características principales y composición*. Santiago: Universidad de Chile.

Departamento de Evaluación, Medición y Registro Educacional DEMRE. (2016b). *Informe Técnico 2016, Volumen II, Proceso de Construcción y Ensamblaje de Pruebas*. Santiago: Universidad de Chile.

Díaz Barriga, A. (1993). *Los procesos de frustración en la tarea docente, en Tarea docente. Una perspectiva didáctica, grupal y psicosocial*. México, Nueva Imagen. Cordié, A. (1998) *Malestar en el docente*. Buenos Aires, Nueva Visión.

Dirección de Sistema de Admisión, UNAH (2014). *Informe estadístico resultados de los procesos de admisión del año 2014. Tegucigalpa, Honduras*.

Donlon, T.T (1984). *The College Board Technical Handbook for the Scholastic Aptitude Test and Achievement Tests*. New York: College Entrance Examination Board.

Dorans N. J., y Cook, L. (2016). *Fairness in educational assessment and measurement*. New York: Taylor & Francis.

Downing, S. M. (2006). *Twelve steps for effective test development*. En S.M. Downing y T. M. Haladyna (Eds.), *Handbook of test development* (pp. 3-25). Mahwah, NJ: Lawrence Erlbaum Associates.

Ebel, R.L. y Frisbie, D.A. (1986). *Essentials of Education Measurement*. Englewood Cliffs, NJ: Prentice Hall.

Eisner, E. (1979). *The educational imagination: On the design and evaluation of educational programs*. New York: Macmillan.

El tabú del fracaso en la universidad argentina. (20 de mayo de 2010 - 03:00). El Día. Obtenido de <https://www.eldiaonline.com/el-tabu-del-fracaso-la-universidad-argentina-n266393>.

Elousa, P. (2006). *Funcionamiento diferencial del ítem en la evaluación internacional PISA. Detección y comprensión*. Relieve. Revista Electrónica de Investigación y Evaluación Educativa, vol. 12, núm. 2, 2006, pp.247-259Universitat de València, Valencia, España.

Escuela Agrícola Panamericana Zamorano. (s.f.). Zamorano. Obtenido de Zamorano: <https://www.zamorano.edu/admisiones/requisitos/>

Escuela Nacional de Ciencias Forestales. (s.f.). ESNACIFOR. Obtenido de <http://www.esnacifor.hn/>

Es una propuesta acelerada lo de eliminar las PAA. (05 de noviembre de 2015). El Heraldito.Obtenido de <http://www.elheraldo.hn/pais/898088-466/es-una-propuesta-apresurada-la-de-eliminar-las-paa>

Eyzaguirre, B. (2003). *Exigencias para la Construcción de una prueba de selección a la Universidad*. Centro de Estudios Públicos, 90, Santiago, Chile.

Fundación para la Educación Ricardo Ernesto Maduro Andreu (FEREMA). (2017). *Informe de Progreso Educativo Honduras, Educación: Una deuda pendiente*. Honduras.

Gajardo, M. y Puryear, J. (2003). *Formas y Reformas de la Educación en America Latina. Programa de Promoción Educativa en America Latina y el Caribe*, MECESUP. Santiago de Chile.

Garduño, F. (2000). *Los exámenes de admisión y la selección a la universidad pública ¿razón pedagógica o racionalidad técnica?* Red de Revistas Científicas de América Latina, el Caribe, España y Portugal, 4(14), 11-18.

Geiser, S.(2016). *Medición y evaluación para los procesos de admisión de la educación superior: Hallazgos desde California. Pensamiento Educativo*. Revista de Investigación Educativa Latinoamericana, 53(1), 1-18.

Glaser, R. (1967). *Evaluation of Instruction and Changing Educational Models*. Pittsburgh Univ Pa Learning Research and Development.

Glaser, R. (1968a). *Adapting the elementary school curriculum to individual performance* (pp.3-36).

Glaser, R, & Cox, R.C. (1968b). *Criterion-Referenced Testing for the Measurement of Educational Outcomes*.

Gilar-Corbi, R. (2003). *Adquisición de habilidades cognitivas: factores en el desarrollo inicial de la competencia experta*. (Tesis de maestría inedita). Universidad de Alicante, Alicante, España.

Gómez, J.,Hidalgo, M., y Guilera, G. (2016). *El sesgo de los instrumentos de medición. tests justos*. Universidad de Barcelona. & Universidad de Murcia. Papeles del Psicólogo, 2010. Vol. 31(1), pp. 75-84.

Gosling, D.,& Moon, J. (2002^a). *How to use learning outcomes and assessment criteria*.

Hambleton, Ronald K., & Hariharan Swaminathan (1985). *Item response theory: principles and applications*. Boston: Kluwer-Nijhoff Pub.

Hanushek, E.A. & Wößmann, L. (2007). *The Role of Education Quality in Economic Growth*. Policy Research Working Paper 4122, World Bank, Washington, D.C.

Hernández, Fernández & Baptista (2014). *Metodología de la Investigación* (6a ed.). México D.F: McGrawHill Education.

Instituto Nacional de Estadística. (12 de junio de 2015). INE. Obtenido de INE: http://www.ine.gob.hn/index.php?option=com_content&view=article&id=90

Jenkins, A., & Unwin, D. (1996). *How to write learning outcomes*. Retrieved February, 5, 2008.

Instituto Nacional de Estadística e Informática (2006). *Glosario básico de términos estadístico*, Lima Perú (pp. 42).

Jomet, J. M. (1986). *Una aproximación teórico-empírica a los métodos de medición de referencia criterial*. Tesis Doctoral. Universidad de Valencia, Facultad de Filosofía y Ciencias de la Educación.

Juarros, M. F. (2006). *¿Educación superior como derecho o como privilegio? Las políticas de admisión a la universidad en el contexto de los países de la región*. Andamios, Revista de Investigación Social, 3(5), 69-90.

Kennedy, D., Hyland, A, &Ryan, N. (2007). *Writing and using learning outcomes: a practical guide*. Cork: University College Cork.

King, A y Schneider, B. (1991). *La Primera Revolución Mundial. Informe del Consejo al Club de Roma*. Edit. Plaza y Janés.

Lane, S., Raymond, M.R., y Haladyna, T. M. (2016). *Handbook of test development* (2nd edition). New York, NY: Routledge.

Lawshe, C. H. (1975). *A quantitative approach to content validity*. Personnel Psychology, 28.

Lissmann, U. (1997): *Inhaltsanalyse von Texten*. Landau: Verlag Empirische Pädagogik.

Lord, FM y Novick, MR.(1968). *Teorías estadísticas de puntajes de pruebas mentales*. Addison-Wesley, Menlo Park.

Mager, R.F. (1962b). *Preparing Instructional Objectives*.

Malan, S.P.T. (2000). *The new paradigm of outcomes-based education in perspective*. Journal of Family Ecology and Consumer Sciences Tydskrif vir Gesinsekologie en Verbruikerswetenskappe, 28.

McMeekin R.W (2006). *Acreditación, accountability y mejoramiento de la calidad de la educación en escuelas vulnerables*. Rev. Pensamiento Educativo, Vol. 39, nº 2, 2006. pp. 237-253

Martínez, F. (2005). *Evaluación educativa y pruebas estandarizadas*. Elementos para enriquecer el debate. Aguascalientes : UAA-IEA.

Martínez, M. R., Hernández M.J. & Hernández, M.V. (2006). *Psicometría*. Madrid: Alianza Editorial.

Mcmeekin, R. (2006). *Hacia una comprensión de la accountability educativa y cómo puede aplicarse en los países de América Latina*. In: CORVALÁN, Javier (Ed.). Accountability Educacional: posibilidades y desafíos para América Latina a partir de la experiencia internacional. Santiago: Cide - Preal, 2006. P. 19-47.

Mejorando el Impacto al Desempeño Estudiantil en Honduras (MIDEH/AIR). (2010). *Estándares de Desempeño*. Honduras.

Mejorando el Impacto al Desempeño Estudiantil en Honduras (MIDEH)/ American Institutes for Research (AIR) (2012). *Proceso de Evaluación, las pruebas que se han elaborado son las diagnósticas, formativas y de fin-de-grado*, Tegucigalpa, Honduras.

Meneses, J., Barrios, M. Bonillo, A. (Eds.). (2013). *Psicometría*. Barcelona. (Universitat Oberta de Catalunya).

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *The American Council on Education/Macmillan series on higher education. Educational measurement*. New York, NY, England: Macmillan Publishing Co, Inc; : American Council on Education. (pp. 13-103)

Mella, O. (2000). *Grupos focales ("focus groups"). técnica de investigación cualitativa*. Publicado como Documento de Trabajo N° 3, CIDE, Santiago, Chile

Moncada, M. C. (2012). *El proceso de admisión de la UNAH: Caracterización y Trayectoria de Aspirantes Admitidos y No Admitidos*. Universidad Nacional Autónoma de Honduras, Escuela de Ciencias psicológicas UNAH & Escuela de Pedagogía y Ciencias de la Educación UNAH. Tegucigalpa: Dirección de Investigación Científica de la UNAH.

Moreno, R., Martínez, R.F., y Muñiz, J. (2004). *Directrices para la construcción de ítems de elección múltiple*. Universidad de Oviedo, Psicothema, vol. 16, núm. 3, 2004, pp. 490-497.

Muñiz, J. (1997). *Introducción a la teoría de respuesta a los ítems*. Madrid: Pirámide.

Muñiz, J. (2003). *La validación de los tests*. Metodología de las Ciencias del Comportamiento, 5, 119-139.

Muñiz, J. (2010). *Las teorías de los tests: Teoría Clásica y Teoría de Respuesta a los Ítems*. Papeles del Psicólogo, vol. 31, núm. 1, enero-abril, 2010, pp. 57-66. Consejo General de Colegios Oficiales de Psicólogos Madrid, España.

Murillo, J.F. (2011). *Hacer de la Educación un ambito basado en Evidencias Científicas*. Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación, Vol. 9, No.3

Oosterhof, A. (1994). *Classroom applications of educational measurement* (2nd Ed.). New York, NY: Macmillan College Publishing Company.

Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura UNESCO (2014). *Campos de educación y capacitación 2013 de la CINE (ISCED-F 2013)*. Montreal: UNESCO-UIS.

Pedrosa, I., Suárez-Álvarez, J.& García-Cueto. E. (2014). *Evidencias sobre la validez de contenido: avances teóricos y métodos para su estimación*. Universidad de Oviedo, Acción Psicológica, junio 2014, vol. 10, n.o 2, 3-20. ISSN: 1578-908X

- Popham, W. J. (1980). *Problemas y técnicas de la evaluación educativa*. Anaya, Madrid.
- Portillo Sáenz, A. (2011). *La Educación Superior en Honduras 1941-2005*. Universidad Nacional Autónoma de Honduras, Departamento de Comercio Internacional. Tegucigalpa: Universidad Nacional Autónoma de Honduras. Recuperado el 14 de abril de 2017, de www.guspepper.net/versión%20final-3.pdf
- Puerta, L., & Marín, M. (2015). *Análisis d Validez de Contenido de un instrumento de transferencia de Tecnología Universidad-Industria de Baja California, México*. XX Congreso de Internacional de Contaduría, Administración e Informática
- Rigney, J. W. (1978). *Estrategias de aprendizaje: una perspectiva teórica*. Nueva York.: Academic Press.
- Rivas, F. y Alcantud, F. (1988). *La evaluación criterial en la educación primaria*. Madrid, España: MEC.
- Sans Martín, A. (2008). *La Evaluación de los Aprendizajes: Construcción de instrumentos*. Barcelona, ICE y Ediciones OCTAEDRO, S.L, Universidad de Barcelona.
- Salgado, R. (14 de febrero de 2017). *Exámenes de Admisión y Calidad Educativa*. El Heraldo. Obtenido de <http://www.elheraldo.hn/opinion/713447-368/ex%C3%A1menes-de-admisi%C3%B3n-y-calidad-educativa>.
- Secretaría de Educación (2002). *Currículo Nacional Básico*. Honduras.
- Secretaría de Educación. (2003). *Diseño Curricular Nacional Básico*, Tercer Ciclo. Honduras.
- Secretaría de Educación de Honduras. (2014). *Informe Nacional de Rendimiento Académico 2014*. Tegucigalpa: Proyecto Mejorando el Impacto al Desempeño Estudiantil en Honduras (MIDEH).
- Secretaría de Educación de Honduras. (2015). *Informe Nacional de Rendimiento Académico 2015*. Tegucigalpa: Proyecto Mejorando el Impacto al Desempeño Estudiantil en Honduras (MIDEH).
- Sireci, S. G. (1998a). *The construct of content validity*. Social Indicators Research, 45, 83-117.

Sireci, S. G. (1998b). *Gathering and analyzing content validity data*. Educational Assessment, 5, 299-321.

Spady, W.G. (1994^a). *Choosing Outcomes of Significance*. Educational leadership, 51(6), 18-22.

Stanley, R.P. (1971). *Teoría y aplicación de particiones planas. Parte 2*. Studies in Applied Mathematics, Vol. L. No.3. Massachusetts Institute of Technology.

Tejedor, F.J (1990). *Indicadores de la calidad docente para la evaluación del profesor universitario*. Revista Española de Pedagogía, Madrid, n.186, pp. 259-279.

Tirado, F., Backhoff, E., Larrazolo, N. y Rosas, M, (1997). *Validez predictiva del Examen de Habilidades y Conocimientos Básicos (EXHCOBA)*. Revista Mexicana de Investigación Educativa. 11, (3), 67-84.

The Council for Higher Education Accreditation. (2006). *Accreditation. and Accountability: A CHEA Special Report*.

Trost, G. (1992). *Principios y prácticas en la selección para la admisión a la Educación Superior*. XVIII Conferencia Anual de la Asociación Internacional de Exámenes, (pág. 10). Dublín.

Thurstone, LL (1938). *Habilidades mentales primarias*. Chicago: University of Chicago Press.

Thurstone, LL, y Thurstone, TG (1941). *Estudios factoriales de inteligencia*. Chicago: University of Chicago Press.

Tyler, R.W. (1949). *Basic principales of curriculum and instruction*. New York.

UNESCO (2007) *Situación Educativa de America Latina y el Caribe: garantizando la Educación de la calidad para todos.EPT-PRELAC*, Santiago de Chile, UNESCO.

UNESCO (2010), *Informe sobre las ciencias sociales en el mundo, 2010: las brechas del conocimiento*. Foro Consultivo Científico y Tecnológico (Mexico) [1], UNESCO. Director-General, 2009-2017 (Bokova, I.G.). writer of foreword [5452]

Universidad Nacional Autonoma de Honduras. (s.f.). Dirección de Sistema de Admisiones. Recuperado el 15 de Abril de 2017, de <http://www.admisiones.unah.edu.hn/Portals/0/AFICHE%202017%20ADMISION%20proceso%20I.pdf>

Universidad Nacional Autónoma de Honduras. (2015). *Normas Académicas de la Universidad Nacional Autónoma de Honduras* (Primera ed., Vol. Publicaciones de la Reforma universitaria no. 6). Tegucigalpa, D.C., Honduras: Universidad Nacional Autónoma de Honduras.

Universidad Pedagógica Nacional Francisco Morazán . (s.f.). UPNFM. Obtenido de UPNFM: <http://web.upnfm.edu.hn/admision/index.php/competencias-a-evaluar>.

Users' Guide, E. (2005). *Brussels: Directorate- General for Education and Culture*.

VanLehn, K. (1996). Cognitive skill acquisition. *Annual Review of Psychology*, 47, 513-539.

Vargas, M., Máñez, A., Cavazos, J, Cervantes, L. (2016). *Validez de Contenido de Un Instrumento de Medición para Medir el liderazgo transformacional*. Revista Global de Negocios Vol. 4, No. 1, 2016, pp. 35-45

Wiburg, K.M. (1995). *An historical perspective on instructional desing: is it time to exchange Skinner's teaching machine for Dewey's toolbox*. The first international conference on Computer support for collabortive learning (p.391).

Zenisky,A., Hambleton. R., y Robin, F. (2004). *DIF Detection and Interpretation in Large-Scale Science Assessments: Informing Item Writing Practices*. Educational Assessment, p. 61-78

Declaración de Ética de la Investigación

El presente trabajo de investigación nacional titulado “Prueba de Selección Universitaria para aspirantes a la Licenciatura en Antropología, Ciudad Universitaria, UNAH, 2018” ha sido diseñada cumpliendo con el proceso de construcción de prueba bajo la Teoría Clásica de los Test y a las normas éticas para investigación de medición.

En vista de lo anterior, yo, **KEYLA YANELY RIVERA REYES**, estudiante egresada de la Maestría en Psicometría y Evaluación Educativa con número de cuenta PSE100123, declaro bajo juramento que he desarrollado esta investigación siguiendo las instrucciones brindadas, desde el planteamiento del problema y recolección de la información, hasta el análisis de datos y elaboración de la tesis.

En tal sentido la información contenida en el presente documento es producto de mi trabajo personal, apegándome a la legislación sobre propiedad intelectual, sin haber incurrido en falsificación de la información o cualquier tipo de fraude, por lo cual me someto a las normas disciplinarias establecidas en la Facultad de Ciencias Sociales, UNAH.

Keyla Yanely Rivera Reyes

Anexos

Instrumentos para las Etapas de Construcción de Prueba

Anexo 1. Manual de capacitación de docentes en el Taller de Redacción de Ítems



Anexo 2. Acta de Compromiso de Confidencialidad de los docentes de Educación Media

**MAESTRÍA EN
PSICOMETRÍA Y
EVALUACIÓN EDUCATIVA**

**UNAH**
UNIVERSIDAD NACIONAL
AUTÓNOMA DE HONDURAS

ACTA DE COMPROMISO DE CONFIDENCIALIDAD

PROGRAMA DE CONSTRUCCIÓN DE PRUEBAS DE SELECCIÓN UNIVERSITARIA UNAH

PRUEBA _____

MAESTRÍA EN PSICOMETRÍA Y EVALUACIÓN EDUCATIVA

TEGUCIGALPA M.D.C

Yo, _____, identificado(a) con cédula de ciudadanía número _____ de _____ me comprometo a cumplir con las labores asignadas en el proceso de construcción de las pruebas de selección universitaria, con los más altos estándares de competencia ética e integridad profesional, teniendo debida consideración a la naturaleza y propósito del trabajo desarrollado. También me comprometo a no revelar directa o indirectamente a ninguna persona, durante la vigencia de este trabajo, ni después de su terminación, ninguna información que hubiera obtenido durante el mismo y que no sea de dominio público.

Firma

Fecha

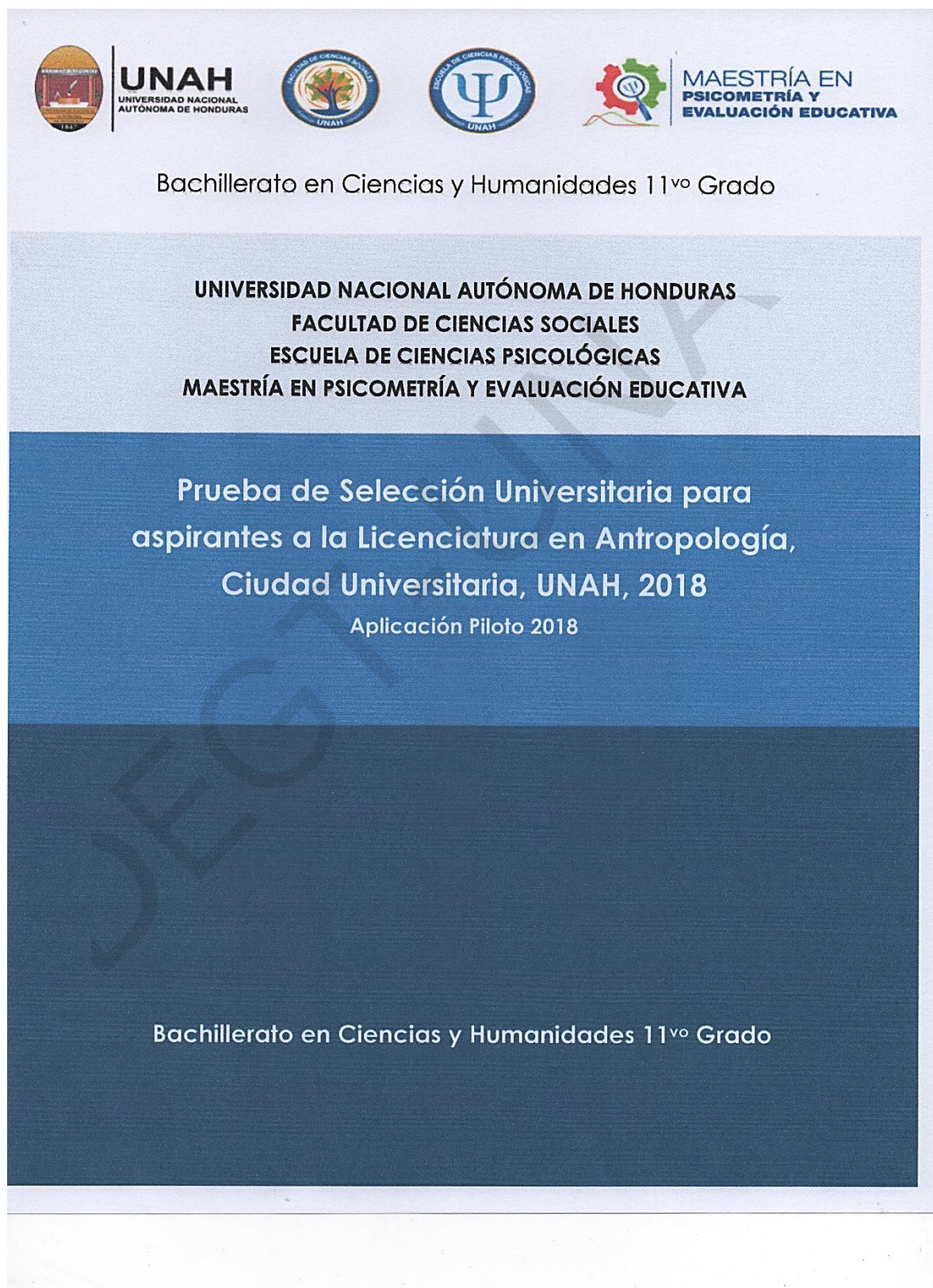
Anexo 3. Plantilla de redacción de ítems utilizada en el Taller

UNAH		FACULTAD DE CIENCIAS SOCIALES		FACULTAD DE CIENCIAS SOCIALES		MAESTRÍA EN PSICOMETRÍA Y EVALUACIÓN EDUCATIVA	
PLANTILLA DE REDACCIÓN DE ÍTEMS DE ANTROPOLOGÍA							
INFORMACIÓN GENERAL							
Espacio curricular: Antropología							
Unidad:							
Contenido:							
Expectativa de logro:							
Nivel cognitivo: (Taxonomía de Bloom)							
Reconocer	☐	Analizar	☐				
Comprender	☐	Evaluar	☐				
Aplicar	☐	Crear	☐				
Universidad Nacional Autónoma de Honduras, Tegucigalpa. (UNAH-CU)							
Carrera de Antropología de la Facultad de Ciencias Sociales.							
Docente:				Fecha:			
ÍTEM No. 1							
Clave: ☐							

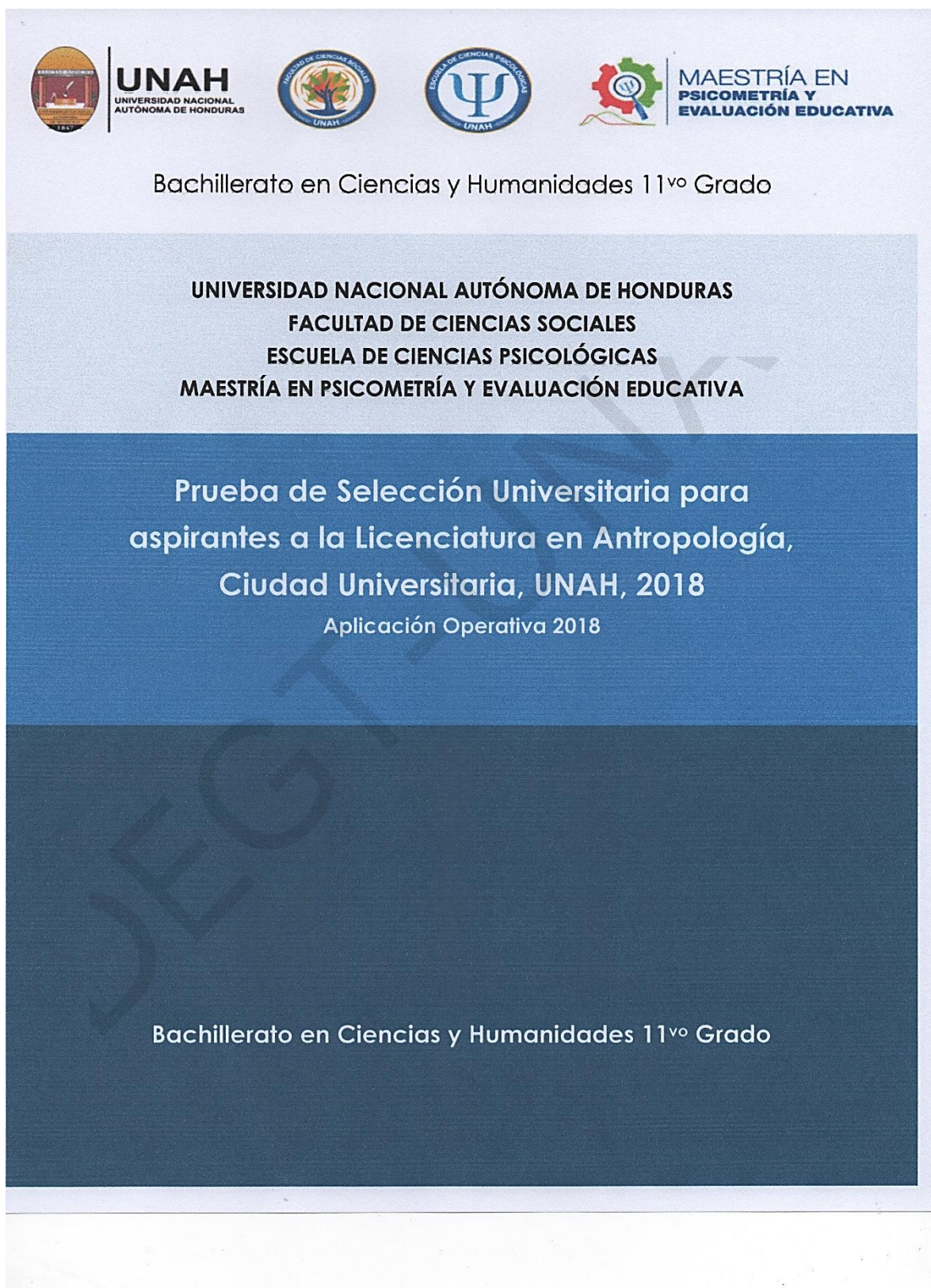
Anexo 4. Manual de aplicación de la Prueba



Anexo 5. Aplicación piloto de la prueba de selección universitaria



Anexo 6. Aplicación piloto de la prueba de selección universitaria



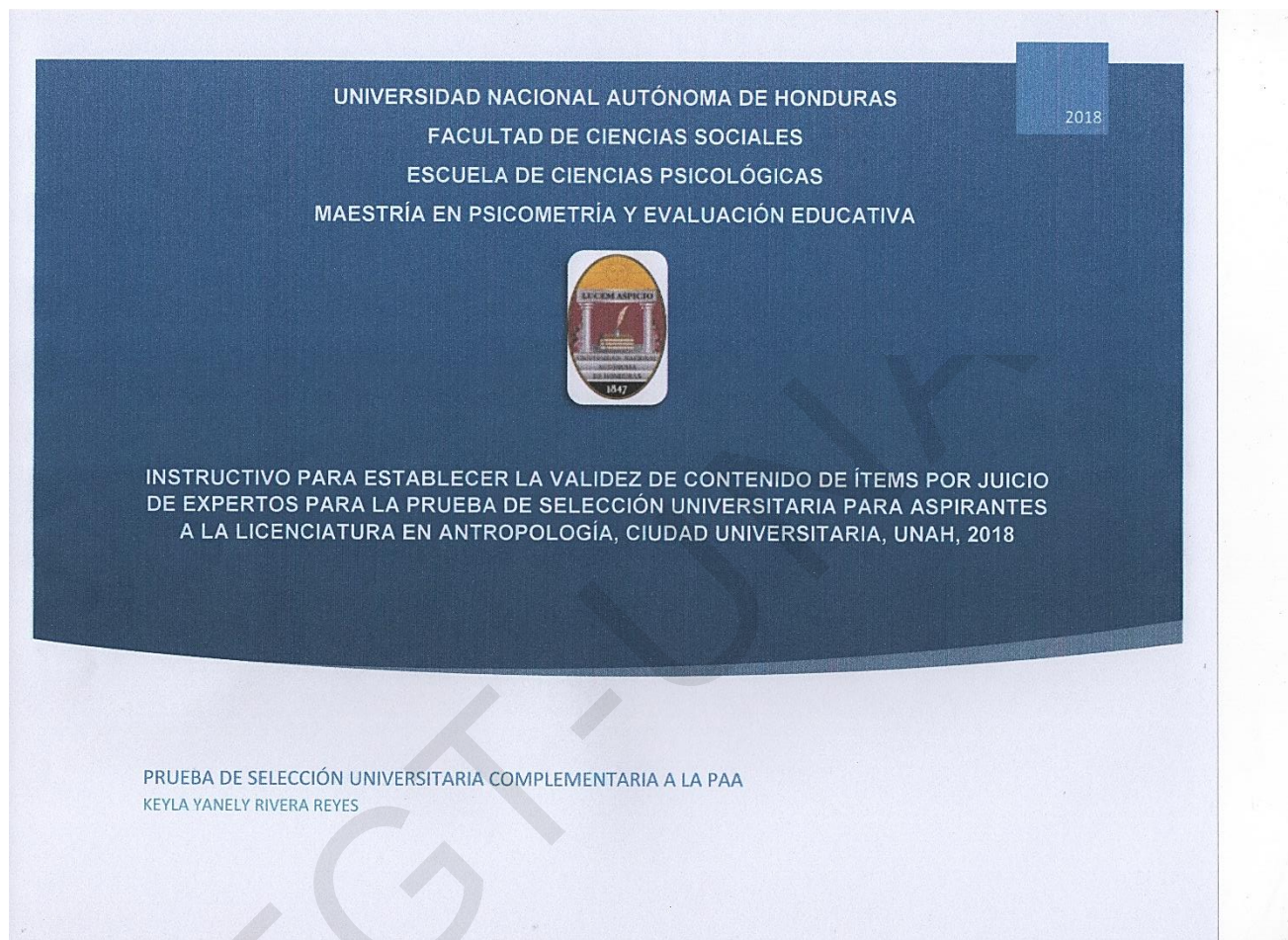
Anexo 7. Consentimiento informado de participación en el estudio**Universidad Nacional Autónoma de Honduras****Prueba de Selección Universitaria para aspirantes a la Licenciatura en
Antropología, Ciudad Universitaria, UNAH, 2018****INSTRUCCIONES**

1. Esta prueba operativa consta de sesenta (60) preguntas.
2. Esta es una prueba de conocimientos basada en el programa curricular de Educación Media en los espacios de Antropología y Fundamentos de Investigación Social.
3. Cada pregunta tiene cuatro (4) opciones, señaladas con las letras a, b, c y d, **una sola de las cuales es la respuesta correcta.**
4. Dispone de una (1) hora y veinte (20) minutos para responderla.
5. Complete todos los datos solicitados, de acuerdo con las instrucciones contenidas en el folleto de la prueba.
6. Rellene completamente el círculo de la opción que eligió, procurando no desbordar sus márgenes.
7. **Cuide su folleto.** No lo doble. No la manipule innecesariamente. Escriba en ella **solamente** los datos solicitados y las respuestas.
8. Escriba correctamente todos los datos en el folleto ÉSTOS SON DE SU EXCLUSIVA RESPONSABILIDAD. Cualquier omisión o error en ellos impedirá que se procesen adecuadamente sus resultados.
9. ES **OBLIGATORIO** DEVOLVER **ÍNTEGRAMENTE** ESTE FOLLETO ANTES DE ABANDONAR EL AULA DE CLASES.

DECLARACIÓN: Estoy en conocimiento de la seguridad con la que se rige el proceso de admisión en la Universidad Nacional Autónoma de Honduras, soy consciente de la importancia de este tipo de investigación con fines exclusivamente académicos, por lo tanto estoy de acuerdo en ser partícipe contestando este instrumento y me comprometo a no reproducir, sustraer, almacenar o transmitir, por internet o cualquier otro medio, el contenido de este folleto o alguna de sus partes.

Firma

Anexo 8. Instructivo para la validación de contenido por juicio de expertos



Anexo 9. Constancia de confidencialidad de los docentes de Antropología UNAH

CONSTANCIA DE VALIDACIÓN DE CONTENIDO DE ÍTEMS DE SELECCIÓN MÚLTIPLE CON RESPUESTA ÚNICA PARA PRUEBA DE SELECCIÓN UNIVERSITARIA PARA ASPIRANTES A ESTUDIAR ANTROPOLOGÍA EN UNAH-CU.

Quien suscribe, _____, con identidad N° _____, de profesión _____ con Grado de _____, ejerciendo actualmente como _____, en la Universidad Nacional Autónoma de Honduras (UNAH), por medio de la presente hago constar que he revisado con fines de Validación el contenido de los ítems o preguntas que formarán parte de la prueba de selección universitaria para aspirantes a estudiar Antropología en UNAH-CU.

Luego de hacer las observaciones pertinentes, firmo la presente en Ciudad Universitaria, Tegucigalpa M.D.C, el día ____ del mes ____ de 2018.

Nombre y firma

CAMPO DE EDUCACIÓN DETALLADO: ANTROPOLOGÍA
CLASIFICACIÓN INTERNACIONAL NORMALIZADA DE LA EDUCACIÓN (CINE)

4

Anexo 10. Ficha para evaluar los ítems en la Validación por expertos


UNAH




FICHA PARA VALIDACIÓN DE ÍTEMS

Instrucciones: Por favor, lea detenidamente cada uno de los enunciados y de respuesta de cada ítem. Utilice este formato para indicar su grado de acuerdo o desacuerdo con cada ítem que se presenta, marcando con una equis (x) en el espacio correspondiente según la siguiente escala: 1) Deficiente, 2) Regular, 3) Bueno y 4) Muy Bueno.

Si desea plantear alguna sugerencia para enriquecer el ítem, utilice el espacio correspondiente a observaciones, ubicado al margen derecho del formato.

Expectativa de Logro	Ítems	Criterios																Observaciones			
		Veracidad				Claridad				Coherencia				Relevancia							
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4				
Caracterizar las principales divisiones de la Antropología.	1																				
	2																				
	3																				
	4																				
	5																				
	6																				
	7																				
	8																				
	9																				
	10																				
Conocer las teorías acerca del origen de los seres humanos.	1																				
	2																				
	3																				
	4																				
	5																				
	6																				
	7																				
	8																				
	9																				
	10																				
Explicar el proceso de hominización y socialización humana.	1																				
	2																				
	3																				
	4																				
	5																				
	6																				
	7																				
	8																				
	9																				
	10																				
Reconocer la existencia de diversos tipos de organización económica y social.	1																				
	2																				
	3																				
	4																				
	5																				
	6																				

 UNAH UNIVERSIDAD NACIONAL AUTÓNOMA DE HONDURAS		FICHA PARA VALIDACIÓN DE ÍTEMS						 MAESTRÍA EN PSICOMETRÍA Y EVALUACIÓN EDUCATIVA	
	7								
	8								
	9								
	10								
	1								
	2								
	3								
	4								
	5								
	6								
Conocer las teorías sobre la llegada de los primeros habitantes a América.	7								
	8								
	9								
	10								
	1								
	2								
	3								
	4								
	5								
	6								
Diferenciar los Pueblos indígenas y afrodescendientes de Honduras.	7								
	8								
	9								
	10								
	1								
	2								
	3								
	4								
	5								
	6								

Anexo 11. Guía para moderar el grupo focal

