

UNIVERSIDAD NACIONAL AUTONOMA DE HONDURAS
FACULTAD DE CIENCIAS SOCIALES
ESCUELA DE CIENCIAS PSICOLOGICAS
MAESTRÍA EN PSICOMETRÍA Y EVALUACIÓN EDUCATIVA



TESIS

Diseño de una prueba para la evaluación del desempeño académico del área español para los estudiantes de 11° grado del Bachillerato en Ciencias y Humanidades, en Tegucigalpa 2017-2018.

Tesis para optar al grado de Magister en Psicometría y Evaluación Educativa

Presentada por

NANCI LISBETH CASTRO COLINDRES

Asesor metodológico

GERMAN MONCADA GODOY, PH.D.

Asesor temático

DENNIS CÁCERES VALLADARES, PH.D.

Tegucigalpa, M.D.C., Honduras
Octubre, 2020

Tegucigalpa, M.D.C, 06 de Octubre de 2020

SEÑORES

UNIVERSIDAD NACIONAL AUTÓNOMA DE HONDURAS

COORDINADORES MAESTRÍA EN PSICOMETRÍA Y EVALUACIÓN EDUCATIVA

Estimados señores:

De la manera más atenta, por este medio, autorizo a la Universidad Nacional Autónoma de Honduras la publicación en su REVISTA DIGITAL, WEB, CD Y CUALQUIER OTRO MEDIO, además de su uso para fines académicos de la sociedad hondureña mi trabajo de tesis titulado: **“Diseño de una prueba de desempeño académico en el área de español para los estudiantes del 11° grado del Bachillerato en Ciencias y Humanidad en Tegucigalpa 2017-2018”**, el cual he presentado como parte de los requisitos para optar por la Maestría en Psicometría y Evaluación Educativa.

Atentamente,

Nanci Lisbeth Castro

AUTORIDADES UNIVERSITARIAS

Doctor Francisco José Herrera Alvarado
Rector interino

Máster Belinda Flores de Mendoza
Vicerrectora Académica

Abogada Jessica Patricia Sánchez Medina
Secretaria General

Doctora Leonarda del Carmen Andino
Directora de Docencia

Doctor Armando Euceda
Director del Sistema de Estudios de Postgrado

Máster Carmen Julia Fajardo
Decana de la Facultad de Ciencias Sociales

Máster María José Irías
Coordinadora General de Postgrado
Facultad de Ciencias Sociales

Máster Lina María Mendoza
Coordinadora Académica de la Maestría en Psicometría y Evaluación Educativa

Resumen

El objetivo de esta investigación consistió en el diseño de un instrumento para la medición válida y confiable de las habilidades adquiridas en el área de español (lectura) por los estudiantes de 11° grado, del Bachillerato en Ciencias y Humanidades, la muestra participante está compuesta por centros educativos de educación media del distrito escolar n° 9, del municipio de Distrito central, departamento de Francisco Morazán, Honduras.

El instrumento se diseñó bajo una perspectiva metodológica enmarcada en las fases generales para el desarrollo de pruebas estandarizadas de evaluación nacional. La muestra final estuvo conformada por 441 estudiantes. La prueba en su versión operativa cuenta con 50 ítems, que miden el bloque de lectura. La calidad métrica de los elementos fue probada a través del modelo lineal clásico en el marco de la Teoría Clásica de los Test.

El rendimiento académico de la población participante, se analizó bajo la aplicación de tres métricas: niveles de desempeño, puntuación estandarizada o scale score, y porcentaje de respuestas correctas, esta última para ofrecer información porcentual en cuanto a cada bloque y componente del Currículo Nacional Básico evaluado. Los resultados también incluyen un análisis comparativo entre las categorías de colectivos según el sexo de los estudiantes y de acuerdo al tipo de administración de los centros educativos de donde provenían los examinados.

Dedicatoria

A Dios, por darme la salud y los medios para poder llevar a cabo la culminación de este proyecto académico, a todos mis seres queridos, familiares y amigos que me han motivado a cumplir mis metas profesionales.

A mi padre, Jose Castro Mendoza, que, aunque ya no me acompaña físicamente, su ejemplo, de disciplina, constancia, y esfuerzo permanecen de forma imperecedera en mi vida. A mi madre, por su apoyo incondicional en cada una de mis metas propuestas.

Finalmente, dedico este logro a mi hijo, Alejandro Josué Fuentes Castro y mi esposo Darwin Rolando Dávila, quienes me brindan incondicionalmente su apoyo, confianza y cariño para lograr mis anhelos.

Agradecimiento

A Dios, por brindarme salud y trabajo, para ser parte de este destacado programa de Maestría.

A mi hijo y esposo por el sacrificio y apoyo en cada una de las metas propuestas en mi vida.

A mis familiares y amigos que siempre me brindaron palabras con denuedo, motivándome a la propuesta de metas académicas, que han dejado experiencias maravillosas.

Agradezco a los asesores PhD. German Edgardo Moncada y PhD. Dennis Fernando Cáceres quienes a partir de su experiencia, asesoramiento y paciencia han hecho posible alcanzar esta meta.

Un agradecimiento especial a la coordinación de la maestría dirigida por la M. Sc Lina Mendoza, quien siempre motivo la culminación de este proyecto de investigación.

A todo el equipo de docentes nacionales e internacionales altamente capacitados, quienes fueron mis docentes en este programa de post grado y han enriquecido mis conocimientos respecto a la disciplina de la psicometría y la evaluación educativa.

Siglas y Acrónimos

BCH:	Bachillerato en Ciencias y Humanidades
BTP:	Bachillerato Técnico Profesional
CNB:	Currículo Nacional Básico
CGEM:	Centros Gubernamentales de Educación Media
CNGEM:	Centros No Gubernamentales de Educación Media
DCNB:	Diseño Curricular Nacional Básico
IEA:	Asociación Internacional para la Evaluación del Logro Educativos
MIDEH:	Proyecto Mejorando el Impacto del Desempeño Estudiantil de Honduras
OCDE:	Organización para la Cooperación y el Desarrollo Económicos
OREAL:	Oficina Regional de Educación para América Latina y el Caribe
PEEP:	Proyecto de Eficiencia de la Educación Primaria
PIRLS:	Programa Internacional del Progreso en Comprensión Lectora
PISA:	Programa para la Evaluación Internacional de Alumnos de la OCDE
PREAL:	Programa de Promoción de la Reforma Educativa de América Latina y el Caribe
SACE:	Sistema de Administración de Centros Educativos
SACMEQ:	Consortio de África Meridional y Oriental para la Supervisión de la Calidad de la Educación.
SE:	Secretaría de Educación de Honduras
SERCE:	Segundo Estudio Regional de Evaluación de la Calidad Educativa
TERCE:	Tercer Estudio Regional Comparativo y Explicativo
TIMSS:	Programa Internacional de Tendencias en Matemáticas y Ciencias
UMCE:	Unidad de Medición de la Calidad de la Educación
USAID:	Agencia de los Estados Unidos para el Desarrollo Internacional
USINIEH:	Unidad del Sistema Nacional de Información Educativa de Honduras

CONTENIDO

Introducción	10
CAPITULO 1. PLANTEAMIENTO DEL PROBLEMA	11
1.1. Situación Problema: construcción del objeto de estudio	11
1.2. Objetivos de la Investigación.....	13
1.2.1 Objetivo General.....	13
1.2.2. Objetivos Específicos	13
1.3 Justificación	14
CAPÍTULO 2. MARCO TEÓRICO	15
2.1. Fundamentos teóricos de la evaluación educativa	15
2.1.1. El concepto de evaluación de los aprendizajes.....	15
2.1.2 Evaluación educativa y pruebas estandarizadas	18
2.1.3 Rendimiento academico y estandares educativos.....	21
2.2 La evaluación de los aprendizajes en el área curricular español.....	22
2.2.1 La lectura como elemento primordial del aprendizaje.....	22
2.2.2 La neurodidactica en el aprendizaje de la lectura.....	24
2.2.3 Aportes de la ciencia del lenguaje para la consideración de la calidad de la educación y su evaluación.	25
2.2.4 La Evaluación del Área curricular Español: Una aproximación evolutiva en Honduras.	27
2.2.5 Relación de los Estándares educativos y las pruebas de logros	28
2.3 La psicometría en el contexto de la evaluación educativa	30
2.3.1 Aproximación histórica y definición de la psicometría	30
2.3.2. Teorias para la medicion psicometrica: Teoría Clásica de los Test, Teoría respuesta al ítem.	32
2.3.3 Análisis de los datos de una evaluación del rendimiento academico	38
2.4. Instrumentos de medida para la evaluación de los aprendizajes	44
2.4.1 Instrumentos normativos y criteriosles	45
2.4.2 Instrumentos paramétricos y no paramétricos.....	47
2.4.3 Instrumentos según el formato de sus ítems: respuesta abierta o cerrada.....	48
2.5 Fases generales para el diseño de una evaluación nacional de rendimiento académico.....	50
CAPÍTULO 3. MARCO CONTEXTUAL.....	62
3.1. Contexto educativo y social de la evaluación de los aprendizajes en Honduras 62	
3.1.1 Indicadores educativos y logro de los aprendizajes.....	62
3.1.2 Elementos determinantes en el sistema de evaluación nacional de Honduras	64
3.2 Experiencia de la evaluación de los aprendizajes en Honduras	66
3.2.1 Evaluación de los aprendizajes en el nivel educativo de educación básica	66
3.2.2 Evaluación de los aprendizajes en Educación Media: Pruebas Pre-Universitarias	69

3.2.3 Prueba de Aptitud Académica (PAA) UNAH.....	70
3.2.4 Experiencias de evaluación iexterna en Honduras	71
3.3. Ámbitos del nivel de educación media	74
3.3.1 Antecedentes y conceptualización del nivel medio.	74
3.3.2 Estructura curricular del undécimo grado.	75
3.3.3 Descripción y propósitos generales del campo curricular español en el 11° grado del Bachillerato en Ciencias y Humanidades.	77
3.3.4 Competencias cognitivas y estándares de contenido del área curricular español del 11° grado del BCH.	78
CAPÍTULO 4. MARCO METODOLÓGICO	79
4.1 Enfoque o tipo de estudio	79
4.2 Definición de las variables e hipótesis	81
4.2.1 Variable principal del estudio.....	81
4.2.2 Hipótesis.....	83
4.3 Población y Muestra	84
4.3.1 Caracterización de la población.....	84
4.3.2 Diseño de la Muestra	85
4.4 Instrumentos	87
4.4.1 Descripción del diseño de la prueba piloto	88
4.4.2 Diseño de la prueba operativa	89
4.5 Procedimiento	90
4.6 Plan de análisis	99
CAPITULO 5. RESULTADOS DE LA INVESTIGACION	104
5.1 Resultados diseño del marco operativo de la prueba.....	105
a. Resultados de la administración de la prueba piloto	107
5.2.1 Confiabilidad y validez: prueba piloto.....	107
5.2.2 Propiedades psicométricas del instrumento pilotaje: Dificultad y Discriminacion.....	111
5.2.3 Funcionamiento Diferencial de los ítems : metodo Mantel –Haenszel generalizado.....	117
5.3 Resultados diseño y administración de la prueba operativa	119
5.3.1. Confiabilidad y Validez de la prueba operativa	120
5.3.2 Estudio descriptivo y distribución normal de los datos de la prueba operativa	121
5.3.3 Comprobacion de hipotesis del estudio	123
5.3.4 Propiedades Psicometricas de la prueba operativa: Indices de discriminacion y dificultad de los items.....	128
5.4 Resultados de desempeño académico: prueba español(lectura) 11° grado BCH.	135
CAPITULO 6. DISCUSION DE LOS RESULTADOS	147
6.1 Hallazgos principales	147

Diseño del instrumento	147
Descripción del desempeño académico de los estudiantes participantes en el estudio	149
6.2 Contraste de los resultados de la prueba de fin de grado	152
6.3 Implicaciones teóricas y prácticas de los resultados	153
CONCLUSIONES.....	155
RECOMENDACIONES.....	156
BIBLIOGRAFIA	157
ANEXOS.....	171

INDICE DE TABLAS

Tabla 1 Diferencias entre la Teoria Clasica de los Test y la Teoria Respuesta al Item	37
Tabla 2. Caracteristicas de los instrumentos normativos y criteriosales	47
Tabla 3 Tasa de cobertura neta y bruta por nivel educativo en Honduras (2018)	62
Tabla 4 Operacionalizacion de la Variable de estudio rendimiento academico de los estudiantes de 11mo grado del BCH en el área español	81
Tabla 5 Distribución de la muestra en la administración de la prueba piloto	85
Tabla 6 Distribución de la muestra en la administración de la prueba operativa	87
Tabla 7 Proceso de diseño de la prueba de fin de grado de español.....	90
Tabla 8 Resultados de ítems redactados del área español 11mo grado BCH	95
Tabla 9 Métricas niveles de desempeño y escala de puntuación estandarizada	103
Tabla 10 Marco operativo de la prueba para la medición del desempeño académico de español para los estudiantes de 11mo grado de BCH	105
Tabla 11Tabla de especificaciones técnicas para el diseño de los ítems	106
Tabla 12 Índice de fiabilidad de la prueba piloto	107
Tabla 13 Índices de fiabilidad y homogeneidad de los ítems prueba “A”	108
Tabla 14 Índices de fiabilidad y homogeneidad de los ítems prueba “B”	109
Tabla 15 Parámetros psicométricos prueba piloto forma "A"	111
Tabla 16 Parámetros psicométricos prueba piloto forma "B"	114
Tabla 17 Indice de confiabilidad prueba operativa español 11mo grado	120
Tabla 18 Prueba de normalidad de los datos de la prueba operativa	121
Tabla 19 Estadísticos descriptivos según los puntajes totales de la prueba operativa.....	122

Tabla 20 Prueba comparación de medias según variable sexo: T de Student	123
Tabla 21 Descriptivos puntaje total de la prueba operativa por sexo	124
Tabla 22 Prueba comparacion de medias segun tipo de administracion de los centros educativos. "T student"	126
Tabla 23 Descriptivos prueba "T Student"	126
Tabla 24 Parámetros psicométricos prueba operativa	128
Tabla 25 Funcionamiento Diferencial de los Items (DIF) Prueba Operativa	132
Tabla 26 Puntuaciones directas y escala de puntuacion estandarizada (Scale Score) Prueba Operativa Español 11° grado BCH.....	135
Tabla 27 Rendimiento Promedio porcentual por componente Español 11mo grado 2018	145
Tabla 28 Rendimiento promedio porcentual por estandar de desempeño en Español (Lectura) 11mo grado BCH 2018	146

INDICE DE GRAFICOS

Grafico 1:Niveles de desercion y repitencia en Honduras (2016)	63
Grafico 2 Porcentaje de desercion escolar en Honduras (2014-2018).....	64
Grafico 3 Dificultad de los items de la sub prueba piloto forma "A"	113
Grafico 4 Indice de Discriminacion de los items de la prueba piloto forma "A"	114
Grafico 5 Dificultad de los ítems de la prueba piloto forma "B".....	116
Grafico 6 Indice de discriminacion prueba piloto forma "B"	117
Grafico 7 Funcionamiento Diferencial de los items, Prueba piloto forma "A"	118
Grafico 8 Funcionamiento Diferencial de los Items prueba piloto forma "B"	119
Grafico 9 Histograma normalidad de los datos prueba operativa Español 11mo grado BCH	123
Grafico 10 Gráfico comparación de medias segun sexo de los estudiantes prueba operativa 11° grado BCH, 2018.....	125
Grafico 11 Comparacion de medias variable tipo de administracion de los centros educativos	128
Grafico 12Indice de dificultad de los items de la prueba operativa 11° grado, BCH 2018	131

Grafico 13 Analisis del Funcionamiento Diferencial de los ítems prueba operativa segun Tipo de Administracion de los centros educativos	134
Grafico 14 Funcionamiento Diferencial de los ítems según sexo de los estudiantes	135
Grafico 15 Escala de puntuacion Estandarizada (100-500) Español-Lectura 11mo grado BCH 2018	137
Grafico 16 Escala de Puntuacion estandarizada (100 a 500) por centro educativo Español (Lectura) 11mo BCH 2018	138
Grafico 17 Porcentaje de estudiantes por nivel de desempeño Español (Lectura) 11mo grado BCH 2018.	139
Grafico 18 Nivel de desempeño por centro educativo Español Lectura 11mo grado BCH 2018	140
Grafico 19 Escala de puntuacion estandarizada (100 a 500) Español Lectura 11mo BCH 2018, segun tipo de administracion de los centros Educativos (Datos agregados)	141
Grafico 20 Porcentaje de estudiantes por nivel de desempeño Español Lectura 11mo BCH 2018, segun el tipo de administracion de los centros educativos	142
Grafico 21 Escala de puntuación Estandarizada (100 a 500) según Sexo, Español-Lectura 11mo grado BCH 2018 (Datos agregados).....	143
Grafico 22 Porcentaje por nivel de desempeño segun sexo Español (Lectura) 11mo grado BCH 2018	144
Grafico 23 Rendimiento Promedio Porcentual por Bloque y componente Español (Lectura) 11mo grado BCH 2018	145

INDICE DE ILUSTRACIONES

Ilustración 1 Ejemplo de Curva Caracteristica de un Item (CCI)	37
Ilustración 2 Pasos para una evaluacion nacional.....	51
Ilustración 3: Pasos para una evaluación nacional.....	51
Ilustración 4 Taxonomia de Bloom, version revisada (2000).....	55
Ilustración 5 Estructura curricular del Bachillerato en Ciencias y Humanidades	76
Ilustración 6 Tematica taller de evaluación y redacción de items	93
Ilustración 7 Formula Coeficiente alfa de Cronbach	100

INTRODUCCIÓN

El objetivo principal de esta investigación, es el diseño de una prueba estandarizada para la evaluación del rendimiento académico en el espacio curricular de español, dirigida a estudiantes de undécimo grado de la modalidad de Bachillerato en Ciencias y Humanidades. El capítulo uno, aborda el objeto de interés de la investigación, detallando los objetivos, y la justificación en cuanto a la importancia y la utilidad metodológica del estudio. El segundo capítulo brinda un recorrido teórico desde los conceptos básicos de la evaluación educativa, y fundamentos teóricos de la psicometría, hasta las teorías que guían el proceso de construcción y análisis de los test, haciendo especial referencia en el desarrollo de pruebas estandarizadas para la evaluación del rendimiento académico. Además, en este mismo apartado se establece la importancia de la medición de las competencias en la lecto escritura debido a su implicancia en el aprendizaje y en el éxito académico, como una vía trascendental para el acceso al conocimiento, la información y la cultura en general.

El tercer apartado da a conocer algunos antecedentes y la situación actual acerca de diferentes experiencias a nivel nacional y regional en el uso de las pruebas estandarizadas para la medición de las competencias de la lecto escritura, misma que es la base contextual de esta investigación. En la cuarta parte de este documento, se hace referencia a la metodología para la realización del estudio, haciendo mención de los pasos para el diseño y administración de los instrumentos, así como el plan de análisis de los datos recolectados. Finaliza el documento, con los capítulos de resultados obtenidos y la discusión de los mismos, exponiendo los principales hallazgos; sus implicaciones teóricas y prácticas, cerrando con las conclusiones y recomendaciones.

CAPITULO 1. PLANTEAMIENTO DEL PROBLEMA

1.1. Situación Problema: construcción del objeto de estudio

La escritura es considerada una actividad que satisface diferentes necesidades en el ser humano, incluyendo el conocimiento del mundo, de sí mismo y de los demás; lo que conlleva a preservar y transmitir todo tipo de información en diferentes contextos diariamente, convirtiéndose en una herramienta eficaz para las personas ofreciendo la posibilidad de expresar su interioridad, desarrollar la creatividad, comunicarse de diferentes maneras; en fin es gracias a la escritura que un país puede construir su historia y su herencia común (OREALC & UNESCO, 2015).

Comprender esta importancia ha inspirado a personas de todas las áreas para estudiar la escritura como un fenómeno cultural, en lo general, y como una habilidad humana, en lo particular. En consecuencia, durante estos últimos años, el problema de la comprensión lectora ha sido estudiada ampliamente, por su implicancia en el aprendizaje y en el éxito académico (UNESCO & LLECE, 2016).

Así mismo, en el marco de los objetivos de Desarrollo Sostenible (ODS), la importancia de la adquisición de las competencia en el área de la lectoescritura es señalada como una meta fundamental del objetivo 4, “Educación de calidad”, el cual propone que los niños, niñas y adolescentes, al final de la enseñanza primaria y/o secundaria; alcancen los resultados de aprendizaje relevantes y efectivos en lectura y matemáticas, disminuyendo las brechas encontradas en cuanto al sexo de los estudiantes (Naciones Unidas, 2018).

El desarrollo de sistemas de evaluación en América Latina en las dos últimas décadas, ha contribuido al estudio de la evaluación de la competencia lectora, como parte de los esfuerzos para modernizar y mejorar los sistemas educativos, con un significativo apoyo de diferentes organismos internacionales; y con la implementación de sistemas estandarizados de evaluación de los aprendizajes a gran escala, ganando un papel substancial en la discusión general sobre la calidad educativa (Ferrer & Fiszbein, 2015).

Sin embargo, aunque las habilidades de la lectoescritura resulta ser uno de los aprendizajes más importantes, puesto que los conocimientos de las distintas áreas curriculares se articulan de forma lingüística y simbólica; y aun teniendo en cuenta los esfuerzos que se han realizado

para mejorar la aprehensión de esta competencia, los resultados de aprendizaje en lectura y escritura en Honduras y a nivel regional, no logran los niveles esperados (UNESCO & LLECE, 2016).

Las evaluaciones de los aprendizajes en el campo nacional e internacional son cada vez más frecuentes, la Secretaría de Educación a partir del año 2007, en alianzas con las agencias y cooperantes externos en el ámbito educativo, evalúan el logro de aprendizaje de los estudiantes del nivel de educación básica en las áreas de español y matemáticas, lo que ha permitido contar con evidencias de los niveles de aprendizaje logrados en estos grados.

En el área de Español(lectura), para el 2007 solamente el 25% de los estudiantes lograban ubicarse en el nivel de desempeño satisfactorio y avanzado, mientras que en el 2017; la tendencia ha sido hacia la mejora; un total de 52% de estudiantes se agrupan en estos niveles además se ha reducido la cantidad de estudiantes en los niveles más bajos de desempeño, no obstante aunque los datos muestran una leve mejoría se sigue estando en valores que no son satisfactorios, pues a medida que se avanza de grado, los puntajes para los grados de 8° y 9° grado se acercan al nivel de desempeño insatisfactorio (Secretaría de Educación & MIDEH, 2017).

Por otra parte, los resultados de las pruebas internacionales que se han desarrollado en el país en el área de lectura; como ser el estudio internacional de progreso en comprensión lectora PIRLSS, y el reciente estudio del programa Internacional para la evaluación de estudiantes PISA-D, dirigido hacia países en desarrollo, indican que los estudiantes hondureños se agrupan en los niveles más bajos de desempeño, respecto al resto de países donde se aplicaron las pruebas (OECD, 2018; Asociación Internacional para la Evaluación del Rendimiento Educativo,2011).

En el nivel de Educación Media en las modalidades científico humanista (BCH) o bachillerato técnico profesional (BTP), aun no existen procesos de evaluación de competencias en el área de español, provocando un vacío de información en cuanto al conocimiento de los logros de aprendizaje de los educandos en este nivel educativo. En consideración a lo señalado, este estudio plantea el diseño de un instrumento de evaluación del rendimiento académico de los estudiantes de undécimo (11°) grado del Bachillerato en Ciencias y Humanidades(BCH) en el espacio curricular de español. A fin de recabar

información válida y confiable acerca del alcance de las competencias planteadas en este espacio curricular. Debido a la importancia que ha cobrado en los últimos años la necesidad del logro de competencias en la lectoescritura para el éxito escolar y social de los estudiantes, y además de los resultados reportados a nivel nacional, es que se considera que este estudio puede significar un aporte valioso en tanto que pretende diseñar y analizar las características psicométricas de una prueba estandarizada que permita evaluar las habilidades obtenidas en el espacio curricular de español (lectura), por los estudiantes de 11° grado de la modalidad científico humanista.

1.2. Objetivos de la Investigación

1.2.1 Objetivo General

Desarrollar una prueba de fin de grado, válida y confiable para la evaluación del desempeño académico de los estudiantes de 11° grado del Bachillerato en Ciencias y Humanidades, en el campo curricular de español(lectura).

1.2.2. Objetivos Específicos

1. Identificar el marco operativo de la prueba de fin de grado del espacio curricular de español para la medición del desempeño académico de los estudiantes.
2. Establecer las propiedades psicométricas de los ítems de la prueba piloto, bajo los parámetros aceptables de la Teoría Clásica de los Test.
3. Diseñar una prueba operativa que cumpla las propiedades de confiabilidad y validez.
4. Determinar el desempeño académico en español(lectura) alcanzado por los estudiantes de 11° grado del Bachillerato en Ciencias y Humanidades del distrito educativo n° 9, del municipio de Distrito Central.

1.3 Justificación

La Ley Fundamental de Educación de la república de Honduras, establece el derecho fundamental a la educación inherente a toda persona para la adquisición del conocimiento el pro del desarrollo de su personalidad y capacidades (República de Honduras, 2012). Bajo esta perspectiva la adquisición de las habilidades de lectoescritura se ha convertido en una competencia fundamental en el desarrollo del ser humano, en el plano social, personal y cultural, dada la compleja red de intercambio en que intervienen las personas en la sociedad actual (Gomez Lopez & Monroy Romero, 2009).

Goikoetxea (2012), sostiene que la ausencia de las habilidades en la lectoescritura podría conllevar a presentar dificultades generales o específicas que impactan en el aprendizaje y en la participación de los estudiantes en su vida escolar y social. Como ya se ha mencionado. Es por ello que evaluar el nivel de competencia adquirido, cobra importancia a fin de la toma de decisiones informadas frente a los resultados con fines de mejora.

El producto de esta investigación responde al problema de la ausencia de instrumentos que miden el desempeño académico en el nivel medio, específicamente en el 11° grado en el área de español (lectura). Respecto a la utilidad metodológica, el contar con una prueba que comprenda los parámetros psicométricos aceptables para la validez de sus resultados, se convierte en una referencia como ejercicio de evaluación nacional del desempeño académico en el nivel de educación media, para la identificación de los logros y desafíos que presentan actualmente los estudiantes en el aprendizaje del curricular de español(lectura).

CAPÍTULO 2. MARCO TEÓRICO

El contenido de este capítulo tiene el propósito de exponer la teoría relacionada con esta investigación, para lo que se ha organizado en tres apartados, se inicia con los fundamentos teóricos de la evaluación de los aprendizajes, seguidamente la psicometría y la medición de los aprendizajes, para finalmente detallar aspectos importantes acerca de la evaluación de las habilidades en lectura y escritura a nivel regional y nacional.

2.1. Fundamentos teóricos de la evaluación educativa

2.1.1. El concepto de evaluación de los aprendizajes

La evaluación por sí misma es una actividad inherente a la vida de los seres humanos es un proceso natural permanente en muchos aspectos cotidianos con diferentes fines; en el ámbito profesional como la medicina, la justicia, el periodismo, la construcción, la seguridad y sin duda en el ámbito de la educación sucede con un carácter más científico, sistemático, y reflexivo, orientado a la toma de decisiones (Alcaraz Salarirche, 2015).

Tradicionalmente , la evaluación ha ocupado un lugar destacado en el que hacer educativo, como un instrumento para verificar el logro académico de los estudiantes, para retroalimentar su aprendizaje y para certificar los conocimientos adquiridos (Popham W. , 2002), por ello, todos los docentes evalúan a sus alumnos, práctica que es fundamental en los procesos de enseñanza aprendizaje. Sin embargo, lo novedoso en la actualidad es que la diversidad y formas en cómo se desarrolla la evaluación y hacia quienes va dirigida; por ejemplo, hoy en día se desarrolla la evaluación de alumnos, de la práctica docente, del currículum, de las instituciones o del sistema educativo en su conjunto. Es decir, la evaluación educativa ha ampliado considerablemente sus fronteras, por lo que no se limita al desempeño académico de los estudiantes, ni a las evaluaciones que realizan los profesores en su práctica pedagógica cotidiana (UNESCO, 2018).

La evaluación educativa se puede entender de diversas maneras, dependiendo de las necesidades, propósitos u objetivos de la institución educativa, tales como: el control y la medición, el enjuiciamiento de la validez del objetivo, la rendición de cuentas y desde esta perspectiva se puede determinar en qué situaciones educativas es pertinente realizar una valoración, una medición o la combinación de ambas concepciones (Mora Vargas, 2004).

Históricamente el interés por la evaluación educativa se remonta a tiempos muy antiguos. Algunos autores lo sitúan en el 2000 a.C y es desde entonces hasta la actualidad, el concepto de evaluación ha ido modificándose a través de diferentes conceptualizaciones y características que han dado lugar a una serie de diversas concepciones. La base de las teorías de la evaluación del aprendizaje de los estudiantes, es un elemento tan antiguo y omnipresente en las instituciones escolares que indudablemente puede considerarse esencial su implementación para la medición de la calidad de todo sistema educativo en América Latina y el mundo (Martínez Rizo, evaluación educativa y Pruebas estandarizadas. elementos para enriquecer el debate , 2011).

La evaluación puede definirse como un proceso permanente y sistemático de recolección de evidencias acerca de qué tanto los estudiantes comprenden los saberes objeto de enseñanza, debiendo promover en los estudiantes reflexiones novedosas e interesantes que les ayuden a profundizar en los contenidos trabajados en el aula y a establecer conexiones entre estos saberes y otros; además de permitir identificar problemáticas puntuales de comprensión que tienen los evaluados y, por ende, debe poder orientar acerca de los ajustes que requiere el proceso de formación. Debe ofrecer a los evaluados, además, la oportunidad para reconocer por sí mismos sus avances y sus dificultades en el proceso de aprendizaje, y debe servir para que otras audiencias (padres de familia y comunidad en general) puedan ser informados acerca de la eficacia del proceso educativo (LLECE, 2009, p.5).

Bajo este escenario, la evaluación de los aprendizajes cumple con diferentes funciones desde diagnóstica, como formativa y reguladora que coadyuva de manera eficaz, en mejorar la calidad de los sistemas educativos, y aunque ciertamente medir los conocimientos de los estudiantes no resulta por sí mismo una medida suficiente para explicar el rendimiento académico constituye una condición necesaria para la toma de decisiones en materia de política educativa de un país.

Los resultados de la evaluación educativa resulta una información valiosa en cuanto al establecimiento de objetivos cuantitativos, cualitativos de planes educacionales, así como para que los tomadores de decisión puedan evaluar diferentes alternativas para la asignación de recursos, combinación de insumos y/o tecnologías pedagógicas y asignar recursos y esfuerzos para lograr los objetivos definidos en un programa curricular (Arancibia V. , 1997).

Otro autores señalan que la evaluación de los aprendizajes puede entenderse como aquella acción dentro del proceso de enseñanza y aprendizaje donde las y los docentes de manera particular, o los organismos privados y públicos habilitados para ello determinan a través de pruebas relativamente objetivas, el nivel de logro en términos de adquisición de habilidades o competencias por parte de los estudiantes, una vez obtenidos los resultados la evaluación de los aprendizajes converge en las acciones didácticas y pedagógicas progresivas que los agentes educativos realizan en el aula y fuera de la institución educativa (Prado Delgado & Medina Rivilla , 2014).

Como se ha reflejado en los párrafos anteriores la tarea de evaluar consiste en recoger información sobre qué y cómo aprende cada estudiante, cuáles son sus progresos al participar en situaciones de enseñanza, de modo que se pueda junto con ellos detectar logros, pero también incomprensiones y dificultades para tomar medidas necesarias y superarlas; pero al mismo tiempo, debe entenderse como un mecanismo de revisión y renovación constante del proceso pedagógico.

La evaluación constituye una orientación fundamental para la toma de decisiones didácticas. Esta provee información a los docentes para impulsar el aprendizaje y ajustar la enseñanza; por lo tanto, articula los procesos de enseñanza aprendizaje así como la reflexión pedagógica, exigiendo un análisis constante y permanente de sus resultados, además de la revisión de los principios que la orientan a fin de promover un impacto significativo en la práctica de la enseñanza en el aula de clases (Vergara Reyes, 2011).

El interés de la sociedad por mejorar la calidad de la educación para que los estudiantes puedan construir los conocimientos y desarrollar las habilidades y actitudes significativas y relevantes para su vida presente y futura, ubica las prácticas de enseñanza y, de manera especial, a las evaluativas en un lugar central. Sin embargo, es necesario develar los múltiples factores que convergen y que las condicionan, tales como los contenidos disciplinarios, los recursos didácticos, las relaciones interpersonales, el apoyo de los padres, las características socioculturales del entorno y, de manera crítica, las concepciones que sustentan los profesores, dada su gravitación en casi todas las actividades que se realizan en el aula.

Si bien es importante conocer los resultados del aprendizaje de los estudiantes es preciso entenderlos, por una parte, como una oportunidad para conocer sus dificultades y debilidades,

ayudarles a superar los problemas detectados y habilitarlos para continuar sus estudios posteriores. Por otra, es preciso que los profesores los consideren como una fuente de relevante información para reorientar y transformar sus prácticas en post del desarrollo de sus facultades.

La primera conclusión general que se postula a partir de este somero análisis acerca del origen y conceptualización de la evaluación de los aprendizajes consiste en sostener que evaluar el aprendizaje es un proceso consustancial al desarrollo del proceso docente educativo, para comprobar el grado de cumplimiento de los objetivos formulados en los planes y programas de estudio, mediante la valoración de los conocimientos y habilidades que los estudiantes adquieren y desarrollan; y que constituye a su vez, una vía para la retroalimentación y la regulación de dicho proceso para mejorar la calidad de la educación.

Por otro parte, el propósito de cualquier evaluación define su estructura, sus contenidos y sus usos y, determinando así sus alcances y limitaciones. Las evaluaciones de aprendizaje o logro educativo se pueden clasificar en dos grandes categorías: 1) las que diseña el docente para utilizar en su salón de clase, con el objetivo de retroalimentar y calificar a sus estudiantes, y 2) las que desarrollan grupos de especialistas, que se basan en la literatura científica (con marcos de referencia teóricos y metodológicos rigurosos), que tienen como propósito evaluar de manera objetiva, confiable y válida lo que los estudiantes han aprendido en un dominio escolar determinado, independientemente del contexto en que ha ocurrido su aprendizaje (Popham, 2001a; 2002). A continuación, una aproximación de estas últimas.

2.1.2 Evaluación educativa y pruebas estandarizadas

Es importante para este estudio desarrollar algunas ideas alrededor de este enunciado, en virtud de que su fundamentación teórica reside en la evaluación del aprendizaje de un campo curricular a través de una prueba estandarizada. La estandarización de pruebas para valorar los sistemas educativos ha ido en ascenso desde que el discurso de las competencias comenzó a protagonizar la agenda educativa en los países europeos. Los autores de diferentes programas de evaluación de los aprendizajes de los estudiantes colocan a las pruebas estandarizadas como un modelo de eficiencia en educación, definiendo este tipo de evaluación como un proceso de recogida de información útil y valiosa para emprender procesos de mejora (Fernandez Navas, Alcaraz Salarirche, & Sola Fernandez, 2017).

En este sentido la importancia de la evaluación de los aprendizajes, a través de pruebas estandarizadas, se hizo presente en diferentes partes del mundo, surgiendo tanto organismos nacionales como internacionales para medir lo que los estudiantes son capaces de aprender al término de ciertos grados de la educación básica o al cumplir una determinada edad. La evaluación estandarizada de los aprendizajes bajo diversas formas y modalidades ha sido un campo en constante desarrollo de tal manera que actualmente se cuenta con sistemas de tal grado de desarrollo y sofisticación que han convertido a la evaluación educativa en una disciplina, una industria y una profesión (Arzola Franco , 2017).

Estas evaluaciones han tenido dos grandes propósitos: 1) conocer la eficacia de los países en materia educativa y 2) hacer recomendaciones de política pública para mejorar la calidad y equidad de la educación en los países participantes. Entre las organizaciones que destacan en estos esfuerzos evaluativos se encuentran: la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO) y la Organización para la Cooperación y Desarrollo Económico (OCDE).

Esta última organización ha destacado por su proyecto PISA, que impactó al mundo desde su primera aplicación en el año 1997 y que ha logrado interesar en la actualidad a más de 70 países, siendo un referente en este tipo de evaluación estandarizada, dada la necesidad de obtener datos comparables entre países sobre el rendimiento académico de los estudiantes, el cual tiene por objetivo medir hasta qué punto el alumnado de 15 años, hacia el final de la educación obligatoria, ha adquirido conocimientos y destrezas clave que son esenciales para la plena participación en las sociedades modernas (OCDE;, 2017).

En la historia de la educación, destaca el surgimiento de la evaluación estandarizada, cuyo propósito y formato son diferentes a las que los docentes realizan en el aula. Estas evaluaciones, también conocidas como objetivas o de gran escala, rebasan el ámbito del aula para proporcionar resultados que sean confiables, válidos y comparables entre distintas poblaciones de estudiantes, a través de instrumentos analizados con base a modelos matemáticos, en el marco de teorías de medición específicas (Backhoff Escudero, 2018).

Características de las evaluaciones estandarizadas de aprendizaje

En los países del continente americano, las evaluaciones estandarizadas se empezaron a utilizar como instrumentos para la selección de personas a diversas instituciones:

principalmente a las educativas y a las fuerzas armadas, en el caso de Estados Unidos. Así mismo, estos instrumentos se han utilizado para certificar las competencias de algunas profesiones y otorgar licencia o permisos para ejercerlas, tales como la medicina. En el área educativa, desde mediados del siglo pasado, las evaluaciones de aprendizaje estandarizadas se empezaron a usar para comparar la calidad educativa de los países, Bloom (1969) citado por (Martinez Rizo & Blanco, La evaluación educativa en México: experiencias, avances y desafíos , 2001). Y, actualmente, una gran cantidad de naciones en el mundo han creado instituciones para la evaluación de la calidad de la educación midiendo además diversos componentes de sus sistemas educativos a forma de rendición de cuentas a la sociedad en general.

Entre las características más sobresalientes de las pruebas estandarizadas

- Se diseñan de tal manera que las preguntas, las condiciones para su administración, los procedimientos de calificación y la manera de interpretar los resultados son uniformes, consistentes y comparables de una evaluación a otra.
- No necesariamente son pruebas de alto impacto, de tiempo limitado o pruebas cuyo formato de respuesta es la opción múltiple. Las preguntas pueden ser simples o complejas y no se limitan a medir el logro educativo.
- Están diseñadas para administrarse a grandes grupos de personas, como es el caso de las pruebas de admisión a las universidades y las evaluaciones de aprendizaje que se realizan para evaluar la calidad educativa de un país.
- Su desarrollo requiere de personal especializado y capacitado en el desarrollo de instrumentos de evaluación, entre los que destacan: psicólogos expertos en medición y psicometría, especialistas en currículo y docentes de las asignaturas y grados escolares que se evalúan.
- Deben de contar con evidencias que garanticen la validez y confiabilidad de sus resultados.

Como se ha reflejado en párrafos anteriores las evaluaciones estandarizadas tienen un siglo de vida y se han utilizado para una diversidad de propósitos educativos. Una de sus bondades es la de poder administrarlas simultáneamente en poblaciones muy grandes de estudiantes. Para ello, se ha tenido que utilizar el formato de selección de respuesta, donde sólo una de las opciones es la correcta. Esto permite que los dispositivos ópticos o

electrónicos puedan leer y calificar las respuestas de los estudiantes de manera automática. Sin embargo, este formato de selección de respuestas impone varias limitaciones que se deberán superar (Backhoff Escudero, Vázquez Lira, Contreras Roldán, Caballero Meneses, & Rodríguez Jiménez, 2017).

Las pruebas estandarizadas están diseñadas para permitir una comparación confiable de los resultados entre todas las personas examinadas, porque todos responden la misma prueba. Sin embargo, es importante hacer notar que, a menudo, los individuos no han tenido las mismas oportunidades para aprender y adquirir las competencias que evalúa una prueba estandarizada, por lo que a lo largo de la implementación de los sistemas de evaluación, se han desarrollado estudios de factores asociados al aprendizaje simultáneos a la evaluación cognitiva, en los mismos se ha incluido la participación de los actores del sistema educativo, desde los padres familia, como los docentes, personal directivo de centros educativos así como las autoridades y gestores de la educación.

Las evaluaciones estandarizadas ayudan a hacer visibles un conjunto de aspectos centrales de la labor educativa, aportan información sobre: en qué medida los alumnos están aprendiendo lo que se espera de ellos al finalizar ciertos grados o niveles; cuál es el grado de equidad o inequidad en el logro de dichos aprendizajes; cómo evolucionan a lo largo de los años, tanto los niveles de logro como la equidad en el acceso al conocimiento por parte de los diversos grupos sociales (Popham W. , 2009).

2.1.3 Rendimiento académico y estándares educativos

Resulta importante destacar la complejidad en la conceptualización de la variable rendimiento académico; ante la disyuntiva, se acentúa que la perspectiva del estudio subyace en definirlo como un fenómeno en el que confluyen muchos factores , donde probablemente una de las variables más empleadas o consideradas por los docentes e investigadores para aproximarse a este son: las calificaciones escolares; comprendiendo el resultado de las mismas como el nivel de conocimientos demostrado en un área o materia comparado con la norma de edad y nivel académico”, razón de ello que existan estudios que pretendan calcular algunos índices de fiabilidad y validez de éste criterio (Jimenez Moreno, 2016, pág. 103).

El Currículo Nacional Básico es el instrumento normativo que establece las capacidades competencias, conceptos, destrezas, habilidades y actitudes que debe lograr todo sujeto del

sistema educativo nacional en los niveles, ciclos y modalidades que rectora la Secretaría de Educación de Honduras (Secretaría de Educación, 2004).

Estándares de desempeño: la transformación más significativa de la evaluación nacional de logros de aprendizaje

La incorporación de *estándares de desempeño* a partir de la entrega de resultados de la medición de 2007 es probablemente la transformación más significativa en el área de evaluación educativa en los últimos años. La introducción de estándares de desempeño permite complementar la interpretación relativa de los resultados, basada fundamentalmente en comparaciones con puntajes previos o con grupos de referencia con una interpretación de resultados que se contrasta con niveles absolutos de logro competencia claramente descritos e ilustrados, pasando así de un énfasis en puntaje promedio a un énfasis en los aprendizajes que evidencian o no los alumnos y alumnas en la prueba.

2.2 La evaluación de los aprendizajes en el área curricular español

2.2.1 La lectura como elemento primordial del aprendizaje.

Leer comprensivamente y escribir con coherencia son actividades imprescindibles para participar en la mayoría de las interacciones sociales, los ciudadanos actuales requieren de saber utilizar diferentes soportes de la lectura tanto como el papel o una pantalla (UNESCO, 2014). El propósito inherente de la lectura es la comprensión. Esta supone la capacidad del lector de disponer y utilizar diferentes estrategias que le ayuden a comprender distintos textos, intenciones textuales, objetivos de lectura y a resolver múltiples situaciones para hacer propio un determinado contenido. Se resalta que la comprensión es un proceso de carácter estratégico (Muñoz & Ocaña, 2017).

Asegurar la comprensión precisa de la puesta en marcha de diferentes acciones que permitan a quien lee activar sus conocimientos previos para construir ideas sobre el contenido, organizar la información relevante, detenerse cuando lo requiera e incluso regresar y reflexionar; lo que implica espacios para desarrollar una lectura individual sin eliminar la posibilidad de concurrencias de diálogo y discusión. En este orden de ideas, la lectura en contextos académicos ha cobrado mayor importancia y está llamada a participar cada vez más activamente de los procesos de aprendizaje continuo y aprendizaje para la vida.

La competencia lectora ha sido siempre un instrumento primordial en el aprendizaje escolar. Varias razones avalan esta afirmación. En primer lugar, los conocimientos de las distintas áreas y materias se articulan fundamentalmente de forma lingüística y simbólica; en segundo lugar, una gran parte de los conocimientos se obtienen a través de los textos escritos; por último, las diferencias observadas en lectura son, en gran medida, predictores del futuro desarrollo educativo de los alumnos y alumnas (Dorronsoro, 2005, p.8).

La lectura como bloque de contenido en el DCNB se concibe como una tarea prioritaria de la escuela. Desde su inicio en el Primer Ciclo de Educación Básica, la lectura supone una apropiación de significados a través de la manipulación de textos completos, coherentes, bien conectados y relacionados con un contexto comunicativo determinado. La lectura no debe agotarse en una instancia de lectura literal, sino que supone procesos de inferencia, valoración y crítica, que configuran el desarrollo de estrategias de comprensión lectora (DCNB, p. 35).

Para evaluar esta importante competencia académica, se consideran tres características fundamentales con respecto a los estándares de lectura: 1. Comprensión de las ideas del texto. 2. Comprensión de vocabulario dentro del contexto. 3. Comprensión de unidades estructurales del texto. En correspondencia con lo planteado en el DCNB, la evaluación aquí analizada, refleja énfasis en la comprensión lectora, reconociendo que toda lectura implica una búsqueda de significado. En este sentido, el compromiso de la escuela es formar participantes competentes de la lectura y de la escritura con el fin de lograr que las personas (niños, niñas, jóvenes y adultos) conciban y experimenten el acto de leer como parte fundamental de la vida que hace posible aprender, informarse, divertirse e imaginar.

Esta concepción interactiva es la que subyace en las evaluaciones internacionales de la lectura que se están aplicando actualmente y, desde esta perspectiva, el propósito de la lectura adquiere una relevancia especial tanto en la enseñanza como en la evaluación de la competencia lectora. PIRLS (Progress in International Reading Literacy Study), evaluación internacional de la capacidad lectora aplicada a niños y niñas de 9-10 años, define la competencia lectora como:

La habilidad para comprender y utilizar las formas lingüísticas requeridas por la sociedad y/o valoradas por el individuo. Los lectores de corta edad son capaces de construir

significados a partir de una variedad de textos. Leen para aprender, para participar en las comunidades de lectores del ámbito escolar, y para disfrute personal (Mullis, 2006, p. 5).

Nuestro sistema educativo trata de responder a estos nuevos retos y define objetivos y establece planes para su consecución. Una de las aspiraciones es que los alumnos adquieran aquellas competencias necesarias para formarse en la sociedad del conocimiento, entre las que se encuentra la competencia lectora. De ahí que, en el contexto internacional actual, la lectura constituya un objetivo prioritario tanto en la Educación Básica como en la Secundaria. Las pruebas de rendimiento académico en lectura son tan valiosas, porque muestran cómo estamos y a dónde debemos orientar los objetivos de enseñanza.

El Congreso Nacional de la República de Honduras aprobó el Decreto No. 218-2012 de la Ley de Fomento para la Lectura y el Libro, la cual en algunos Considerandos establece: a) Que la lectura es una herramienta extraordinaria de trabajo intelectual ya que pone en acción las funciones mentales agilizando la inteligencia; b) Que la lectura potencia la formación estética y educa la sensibilidad estimulando las buenas emociones y los buenos sentimientos. (La Gaceta, 2013) Normativa que pone en evidencia la importancia que actualmente se le asigna a la lectura para el desarrollo intelectual tanto a nivel internacional, como en el local hondureño.

2.2.2 La neurodidáctica en el aprendizaje de la lectura

La Neurociencia es una de las disciplinas biomédicas de gran importancia en la actualidad, la cual indica cómo la actividad del cerebro tiene relación con la conducta y el aprendizaje. Los procesos de aprendizaje y la experiencia propiamente dicha van modelando el cerebro que se mantiene a través de incontables sinapsis; estos procesos son los encargados de que vayan desapareciendo las conexiones poco utilizadas y que tomen fuerza las que son más activas. Si bien las asociaciones entre neuronas se deciden, sobre todo, en los primeros quince años de vida, y hasta esa edad se va configurando el diagrama de las células nerviosas, las redes neuronales dispondrán todavía de cierta plasticidad (Vargas Castro, Jara Castro, Lozada Meza, & Dume Villacis, 2019)

Desde el ámbito científico surge el aporte de las Neurociencias que, en su conjunto neurobiología, neurofisiología, neurodesarrollo, neurociencia cognitiva, neurolingüística, neuropsicología, neuropsiquiatría, psicofarmacología, etc., representan un potencial

alternativo para los problemas de la educación, Cajal (2006), citado por (Moncada & Lopez, 2014). La neurociencia puede proporcionar fundamento teórico científico que ayude a mejorar y justificar la práctica pedagógica de docentes partiendo de la idea de conocer cómo aprende y cómo funciona el cerebro puede mejorar la práctica pedagógica y las experiencias de aprendizaje (Vargas Castro, et.al., 2019).

En esta línea, la neurodidáctica de la lengua y la literatura promueve la inclusión de actividades que fomenten el aprendizaje lingüístico en un entorno motivador, que estimule, a su vez, la reciprocidad y la comunicación basada en el intercambio natural en un ambiente lúdico que propicie la sociabilidad, la creatividad y el equilibrio socio-emocional del alumno; aspectos que favorecerán la comunicación equilibrada entre las personas, ya que mediante ello se establecerán las bases de las futuras relaciones personales, pues la comunicación es un proceso de negociación entre sujetos que pertenecen a sociedades y culturas diferentes, que favorece el conocimiento (OEI, 2018).

La neuroeducación, es un campo nuevo y abierto de esperanza en el área del aprendizaje y trata sobre cómo enseñar de modo cada vez más fructífero. Las aportaciones de la neuroeducación llevarán a conocer qué herramientas puede proveer la neurociencia, que de modo práctico sirvan para enseñar de forma más eficiente tanto en la escuela como en la enseñanza media o en la universidad, y en todo proceso de enseñanza, sea general o especializada; dichas herramientas deben servir para detectar problemas neurológicos y psicológicos, que impidan o interfieran en los niños la tarea de aprender con facilidad en el colegio; además estas herramientas deben servir para formar mejores ciudadanos críticos, con un equilibrio entre cognición y emoción (Mora T., 2013).

2.2.3 Aportes de la ciencia del lenguaje para la consideración de la calidad de la educación y su evaluación.

En buena medida los conocimientos que adquiere un estudiante le llegan a través de la lectura. Durante el proceso de enseñanza aprendizaje, desde la primaria hasta la educación postgraduada, se necesita leer una variedad de textos para apropiarse de diferentes conocimientos y la importancia del hecho, no sólo radica en los contenidos, sino en la cantidad, estilo y propósitos de la lectura.

Con frecuencia se considera que los alumnos saben leer, porque pueden visualizar los signos y repetirlos oralmente, o bien porque tienen la capacidad para decodificar un texto escrito. Sin embargo, la decodificación no es comprensión y esto es el resultado de un primer nivel de lectura con el cual no debería conformarse el lector (Huerta, 2009, p.2).

La OCDE, señala que el concepto de comprensión lectora retomada por muchos países, es un concepto mucho más amplio que la noción tradicional de la capacidad de leer y escribir (alfabetización), en este sentido, señala la OCDE, que la formación lectora de los individuos para la efectiva participación en la sociedad moderna requiere de la habilidad para decodificar el texto, interpretar el significado de las palabras y estructuras gramaticales, así como, construir el significado.

La comprensión lectora involucra, por tanto, la habilidad de comprender e interpretar una amplia variedad de tipos de texto y así como de dar sentido a lo leído al relacionarlo con los contextos en que aparecen. El proyecto “Estudio Internacional de Progreso en Comprensión Lectora” (PIRLS) siglas en inglés, (2006) perteneciente a la Asociación Internacional para la Evaluación del Rendimiento Educativo (IEA) siglas en inglés, evalúa las capacidades lectoras de los estudiantes y afirma que la lectura tiene como finalidad entender cómo es el mundo y cómo ha sido y por qué las cosas funcionan de la manera en que lo hacen, y que los lectores pueden ir más allá de la adquisición de información y utilizarla para practicar el razonamiento y la acción (Citado en Pérez, 2007, p. 28).

La lectura es un concepto importante dentro del proceso de la comprensión lectora, Echevarría (2006), comenta que en la actualidad se conceptualiza a ésta como un proceso basado en el texto, de naturaleza interactiva, con propósitos específicos, y que depende tanto del texto como de la persona que lo lee. La lectura es un proceso complejo y coordinado que incluye operaciones preceptuales o regular, lingüístico y conceptual, y los lectores a su vez también representan los conceptos y los hechos que se describen en el texto (Bello, 2006, p.9).

2.2.4 La Evaluación del Área curricular Español: Una aproximación evolutiva en Honduras.

En noviembre 2002, fue aprobada la propuesta: Plan de Educación para Todos y Todas Iniciativa de Apoyo Ágil Honduras, 2003-2015. Este plan señaló como metas finales para el año 2015, que todos los niños y las niñas de Honduras hubiesen completado seis grados de Educación Básica a la edad de 12 años, con un rendimiento académico de 70% en las áreas de Español y Matemáticas, así como la universalización de la Educación Prebásica. Y estas metas se convirtieron en las guías orientadoras de la acción estatal en el ámbito educativo nacional y de la Cooperación Internacional vinculada al área de educación operando en el país.

Para alcanzar estas metas, los países cooperantes han apoyado a Honduras en el sector educativo durante todo este período a través de la Mesa Redonda de Cooperantes Externos en Educación (MERECE), y desde el año 2003 alineando su cooperación con el Plan EFA – FTI (2003-2015), mediante las modalidades de Proyectos Bilaterales y/o Fondo Común En Honduras, en el año 2015 se aplicaron por cuarto año consecutivo pruebas para Matemáticas y Español a una muestra de centros educativos que ofertan educación básica en todo el país.

Conocer el nivel de desarrollo de las habilidades adquiridas en los campos del conocimiento de matemáticas y de comprensión lectora es cada vez más indispensable, 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 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 (Benavidez Ormaza, 2010).

En Honduras, el alineamiento que existe entre el currículo nacional, libros de texto y otros materiales educativos ha hecho posible el desarrollo e implementación de un sistema de evaluación basado en estándares de clase mundial. Los materiales de evaluación sirven para medir el nivel académico estudiantil al inicio del año escolar (pruebas diagnósticas), mensualmente (pruebas formativas) y al finalizar el año escolar (pruebas sumativas).

Acciones como la determinación de estándares de desempeño de forma conceptual (descriptores de niveles de desempeño) y de forma operativa (puntos de corte); se han presentado reportes de los resultados de las pruebas a nivel nacional, regional y escolar. Estos resultados se presentaron en función de la cantidad de estudiantes en cada nivel de desempeño. Se observó cierto nivel de resistencia respecto al cambio de reportar porcentaje correcto, pero con actividades de promoción en los medios y mediante talleres desarrollados a nivel nacional existe una buena perspectiva que el sistema es aceptado (Lansdale, Vukmirovic , & Estrada, 2012)

En el año 2016 se aplicaron por quinto año consecutivo pruebas para estudiantes de primero a noveno grado en las áreas curriculares de Matemáticas y Español a una muestra de centros educativos que ofertan educación básica a nivel nacional. Este tipo de muestra tiene representatividad para todo el país (424 centros que ofertan primero y segundo ciclo de educación básica y 225 que ofertan el tercer ciclo, para un total de 649 centros y 80,357 estudiantes de los 18 departamentos de la división político administrativa hondureña. (Secretaria de Educación & MIDEH, 2016).

El tipo de pruebas aplicadas y el nivel de muestra seleccionada permiten presentar resultados para cada Bloque y Componente del contenido curricular de Matemáticas y de Español para cada grado, en diferentes niveles de análisis: Nacional, centro educativo, grado y sección. Sin embargo, el presente informe enfatiza que esta valiosa información por sí misma no genera mejoras en el sistema educativo. Es una condición necesaria pero no suficiente. Solamente si se elaboran e implementan los Planes de Mejoras basados en los resultados obtenidos en cada nivel de gestión (direcciones departamentales, municipales, centros educativos y aulas), se puede esperar un impacto de la evaluación en la mejora de los aprendizajes.

2.2.5 Relación de los Estándares educativos y las pruebas de logros

Los estándares de desempeño establecen los niveles de aprendizaje logrado por los estudiantes. Las evaluaciones alineadas a los estándares sirven para determinar el nivel de dominio de los estándares por los estudiantes. Definir Estándares de Desempeño es un procedimiento para conceptualizar los niveles de desempeño que se usarán para evaluar los aprendizajes: se trata de decidir acerca del número y el propósito de los niveles de

desempeño, escoger los nombres de los niveles de desempeño (Etiquetas de Niveles de Desempeño) y desarrollar descripciones generales y específicas de los niveles de desempeño (Descriptores de Niveles de Desempeño)

Los estándares definen las expectativas académicas de los estudiantes: los estándares de contenido definen lo que los estudiantes deben saber y saber hacer La evaluación de los aprendizajes y su importancia en la mejora de la calidad de la educación. La importancia de la evaluación de los aprendizajes ha sido un tema de interés colectivo, sin embargo, muchos autores manifiestan que la importancia radica en informar a partir de los resultados obtenidos, y por otro lado permite dar la pauta para orientar la toma de decisiones hacia la búsqueda de la mejora de la calidad de la educación. Es así, que coinciden en que el fin primordial de las evaluaciones de los aprendizajes es la “utilización de los resultados”, siendo este el concepto que predominó en las primeras etapas del desarrollo histórico de la evaluación (ONE, 2009).

El objetivo de la evaluación del aprendizaje, como actividad genérica, es valorar el aprendizaje en su proceso y resultados. Las finalidades o fines marcan los propósitos que signan esa evaluación. Las funciones se refieren al papel que desempeña para la sociedad, para la institución, para el proceso de enseñanza-aprendizaje, para los individuos implicados en éste (Pérez, 2001).

La evaluación de aprendizajes tiene un potencial único para mejorar el desempeño de los alumnos. Su importancia es tal que puede ser el eje de toda la enseñanza, pues a partir de ella se podría definir qué, cómo y cuándo se enseña. No obstante, para que la evaluación explote su potencial se requiere que esté estrechamente ligada a los procesos de enseñanza y aprendizaje. De esa forma, los diferentes actores que intervienen en el quehacer educativo (docentes, alumnos, padres de familia y la escuela como institución) se verán beneficiados (García, Aguilera, Pérez y Muñoz, 2011).

Sin embargo, si nos preguntamos por qué se evalúa, la cuestión remite inmediatamente a la intención de quien construye el objeto de evaluación (conocer las respuestas de los alumnos a un conjunto de situaciones problemáticas, por ejemplo). Si, en cambio, el interrogante está referido a para qué se evalúa, la preocupación refiere a los efectos de la evaluación sobre la acción (que se hará con los resultados). En el primer caso, los propósitos o intenciones del evaluador dependen de cual sea su concepción sobre la evaluación. Las

intenciones más reconocidas son las de medir, apreciar y comprender, que se corresponden con sus respectivas “filosofías” subyacentes (ONE, 2009).

Cabe destacar que, si la educación es vista como un medio por el cual una sociedad se puede transformar y lograr cambios significativos, entonces resulta imprescindible la existencia de un sistema que rinda cuentas de la manera en la que está funcionando, si realmente los estándares educativos se están cumpliendo al 100%, asimismo saber qué tan importante resultan las evaluaciones que se están realizando en el sistema educativo. Es por ello que la implementación de una evaluación continua de los aprendizajes permitiría tener mayor control a las autoridades educativas, sobre lo que el estudiante está adquiriendo durante el proceso, y si realmente está siendo preparado para dar respuesta a las necesidades de una sociedad cambiante (Cháves, 2015).

2.3 La psicometría en el contexto de la evaluación educativa

2.3.1 Aproximación histórica y definición de la psicometría

La evaluación educativa ha sido producto de dos disciplinas que han convergido históricamente; la psicología y la educación. Desde hace más de cien años, la psicología se ha interesado en evaluar ciertos atributos de los individuos: su personalidad, su inteligencia y sus capacidades cognitivas. Con este interés nació la psicometría, campo disciplinario cuyo propósito es medir cuantitativamente las características de los individuos. Con nuevas herramientas estadísticas fue posible diseñar y construir diversos exámenes estandarizados para medir el logro académico de los estudiantes (Anastasi & Urbina, 1997).

El discurso acerca de las teorías de la psicometría y la medición proceden desde el siglo XVIII por Laplace y Gauss (1777-1846). La aplicación de estas ideas al campo educativo comenzó desde el XIX, distinguiéndose países como Alemania, Inglaterra, los Estados Unidos y, en menor medida, Francia y las regiones francófonas de Suiza y Bélgica. A partir de los primeros años del siglo XX se desarrolla la metodología de la teoría clásica de los test (TCT) a partir de los conceptos de la confiabilidad y el modelo estadístico de las puntuaciones, con las nociones de puntaje verdadero, error de medida y confiabilidad de la prueba. Los trabajos pioneros fueron los del inglés Charles Spearman, entre 1904 y 1913, y la obra del norteamericano Edward L. Thorndike *An introduction to theory of mental and social measurement*, publicada en Nueva York, también en 1904 (Martínez Arias M. , 1995).

Visto someramente los orígenes de la psicometría es preciso definir formalmente su concepto, entre la diversidad de literatura entre distintos autores, su definición puede ser determinada a través de tres aproximaciones, la primera aborda la psicometría a partir de los instrumentos que utiliza, definiéndola como la disciplina encargada de desarrollar los fundamentos para la construcción y administración de los test (Martinez Arias & Hernandez LLoreda, 2006).

Una segunda corriente determina, el concepto de psicometría a partir del objeto sobre el cual aplica su análisis o estudio, de forma que la explica como la disciplina científica orientada a la evaluación de los fenómenos psicológicos, ajustada al significado etimológico de la palabra psicometría, que en griego se traduce *psique*: “alma”, “aliento o “intelecto” y *metrón* que significa proceso de medición, dentro de esta perspectiva la función de esta disciplina es desarrollar medidas científicas en la áreas de personalidad y las capacidades, de manera que la científicidad se convierte en sinónimo de estandarización, indicando el uso de los test en el contexto científico, bajo un estatus métrico de las puntuaciones obtenidas como resultado del proceso de medición psicológica, en este sentido la psicometría es el conjunto de métodos, técnicas y teorías implicadas en la medida de las variables psicológicas teniendo en cuenta las propiedades métricas exigibles para esa medida (Muñiz J. , 1996).

Una tercera aproximación incorpora un tercer elemento vinculada a la finalidad a la que sirve, y en este sentido la psicometría contribuye al desarrollo de la evaluación psicológica proporcionando, teorías, métodos, técnicas , que permite describir, clasificar, diagnosticar los fenómenos psicológicos u objetos de medida, es decir la finalidad que orienta el desarrollo y la administración de los test se enlaza con la de contribuir a la recolección de información que permitan la toma de decisiones (Fernandez Ballesteros, 2000).

En atención al enfoque planteado en los párrafos anteriores una definición que incorpora los elementos descritos anteriormente define a la psicometría dentro del marco de la evaluación y la intervención psicológica:

La psicometría es una rama de la psicología que, mediante teorías, métodos y técnicas vinculados al desarrollo y la administración de test, se ocupa de la medida indirecta de los fenómenos psicológicos con el objetivo de hacer descripciones, clasificaciones, diagnósticos, explicaciones o predicciones que

permitan orientar una acción o tomar decisiones sobre el comportamiento de las personas en el ejercicio profesional de la psicología (Meneses, y otros, 2013, p. 38).

Partiendo de esta definición que integra las tres aproximaciones referidas, la psicometría en el área educativa; interviene en una variedad de procesos académicos desde el diagnóstico de problemas de aprendizaje, hasta la medición de habilidades, conocimientos, calidad de los sistemas educativos, así mismo en el campo de la orientación profesional contribuye a diagnosticar habilidades académicas específicas, como los test vocacionales, sin dejar de mencionar las bondades que esta rama de la psicología nos refiere para la toma de decisiones informadas.

2.3.2. Teorías para la medición psicometrica: Teoría Clásica de los Test, Teoría respuesta al ítem.

Al hablar de teorías para la medición psicométrica, paralelamente se habla de los test como instrumentos de medida. El origen de los test, se remonta desde los emperadores chinos 300 años antes de Cristo, donde empleaban pruebas para evaluar los oficiales, que iban a entrar en servicio. Pero fue Binet y Simón (1905) quienes establecieron en una nueva escala tareas cognoscitivas para evaluar aspectos como el juicio, la comprensión y el razonamiento. Seguidamente (Terman,1916) desarrolla la revisión de la escala de Binet, 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 (Muñiz J. , 2010).

La escala de Binet, da inicio al uso de métricas o escalas individuales en la medición, que llega hasta estos días. En el año 1917, las teorías de los tests reciben un gran impulso con los tests colectivos Alfa y Beta, que surgieron de la necesidad del ejército estadounidense en reclutar rápidamente soldados para la primera guerra mundial. El test Alfa fue administrado a la población general y el Beta dirigido a las personas analfabetas o que no dominaban el idioma inglés. El uso de las pruebas dio los resultados esperados; y una vez culminada la guerra, las empresas y otras organizaciones gubernamentales adoptaran la práctica del uso de los tests como herramientas útiles para distintos propósitos (Muñiz J. , 2010).

Hay dos grandes enfoques o teorías para la construcción y análisis de los test en su conjunto, estas son la Teoría Clásica de los Tests (TCT) y el enfoque de la Teoría de Respuesta a los Ítems (TRI) (Arregui, Wolfe, & Ravela, 2008). A continuación, se detalla en un sentido laxo las generalidades, haciendo hincapié en las propiedades psicométricas que cada una de estas examina, profundizando en la Teoría Clásica de los Test, misma que es el enfoque con el cual se analizarán los resultados de esta investigación.

Teoría Clásica de los Test:

La Teoría Clásica de los Test, fue formulada en 1904 por Charles Spearman. Se le conoce también como modelo lineal clásico, reformulado por Lord y Novick (1968) donde incluye tres conceptos fundamentales: la puntuación verdadera, la puntuación empírica y el error de medida, abriendo paso al nuevo enfoque de la Teoría Respuesta al Ítem que será puntualizado más adelante. El enfoque clásico es el predominante en la construcción de los test, el cual trata de asegurar que el instrumento que se utiliza mida con precisión, y con poco error (Muñiz J. , Teoría Clásica de los Tests, 2000).

La teoría clásica de los tests, a través de Spearman, propone un modelo lineal clásico, el cual consiste en asumir que la puntuación que una persona obtienen en un test, que se denomina puntuación empírica (X), está formada por dos componentes que son la puntuación verdadera (V) más el error de medición (e), siendo este último ocasionado por factores externos en el proceso de medición; la formulación de este modelo lineal corresponde a la siguiente: $X = V + e$, con esto se quiere decir que el puntaje observado en una medición (X), es igual al puntaje verdadero (V), más el error (e) (Borsboom, 2005).

Esta teoría presenta tres supuestos básicos, (Muñiz J. , 2010): El puntaje verdadero es igual a la esperanza matemática o valor esperado de las puntuaciones observadas $V = E(X)$: 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 administrase 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.

La correlación entre el puntaje verdadero en un test y el error en ese test es igual a cero. Es decir, no existe relación entre ambos $r(v, e) = 0$: en el segundo supuesto 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.

La correlación entre los errores dados en dos test diferentes es igual a 0. es decir, los errores son independientes $r(e_j, e_k) = 0$. 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.

La ecuación de la Teoría Clásica de los Test (TCT), significa que mientras menor sea el término de error, mayor será la semejanza entre el puntaje observado y el verdadero. Lo importante en términos de la confiabilidad de las puntuaciones de un test, es estimar cuan bien representan las puntuaciones observadas a las verdaderas; buscando estimar la relación que existe entre el puntaje verdadero y el puntaje observado lo que indica que mientras más fuerte sea la relación lineal entre ambos, mejor será la representación del puntaje verdadero mediante el puntaje observado.

La teoría clásica de los test, dentro de sus bondades y limitaciones ha generado algunos estudios sobre todo en función del tratamiento en cuanto error de medida; numerosas investigaciones se han desarrollado por diversos autores a fin de estimar los distintos componentes del error de medida. De todos estos intentos sobresale por su enfoque sistemático, 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 (Webb & Shavelson, 2005).

Otro desarrollo psicométrico surgido en el marco clásico ha sido el de los Tests Referidos al Criterio (TRC). Como se ha mencionado en el apartado anterior estos son utilizados en el ámbito educativo y en la evaluación en contextos laborales. Los indicadores psicométricos clásicos desarrollados a partir del modelo lineal clásico no se adaptaban bien a la filosofía de

construcción de estos nuevos tests, por lo que se ha desarrollado todo un conjunto de tecnología psicométrica específica para complementar el cálculo de la fiabilidad y validez, así como para establecer los puntos de corte que determinan si una persona domina o no el criterio evaluado (Berk, 1985).

Las principales propiedades medidas en el marco del modelo lineal clásico, son el índice de dificultad (ID) de un ítem (p-valué), la índice discriminación (D) y/o correlación ítem-test, la valoración del sesgo y el coeficiente de fiabilidad de Spearman, dentro de los cuales existen algunas limitaciones, por ejemplo, la dificultad de los ítems, o los coeficientes de fiabilidad dependen en gran medida del tipo de muestra utilizada para calcularlos. Este tipo de problemas encuentran una solución adecuada dentro del marco de la Teoría Respuesta al Ítem (TRI). Con la anterior declaración no se quiere invalidar las bondades de la TCT, al contrario, se puede afirmar que la TRI, constituyen un excelente complemento que, dependiendo de las características de la medición, ofrece soluciones a problemas no muy bien resueltos en el modelo lineal clásico; en conclusión, ambas teorías son utilizadas perfectamente en la construcción y análisis de los test. A continuación, se ofrece los fundamentos principales sobre los cuales subyace la Teoría Respuesta al Ítem.

Teoría de Respuesta al Ítem (TRI)

Esta teoría surgió 40 años después del inicio de la TCT, como se indicaba en los párrafos anteriores ha contribuido resolviendo algunas limitaciones que presenta la TCT, la TRI se basa en la estadística de los ítems en lugar de los puntajes del test, resultando con ello una herramienta que permite medir un rasgo latente de una persona. La utilidad de este modelo en el campo educativo radica en que postula que la probabilidad que un estudiante de una respuesta correcta depende del nivel de habilidad del examinado y del análisis de tres parámetros: dificultad y discriminación de los ítems si como de un tercer elemento que no examina la TCT y que corresponde al parámetro de pseudo-adivinanza (Vukmirovic, 2016)

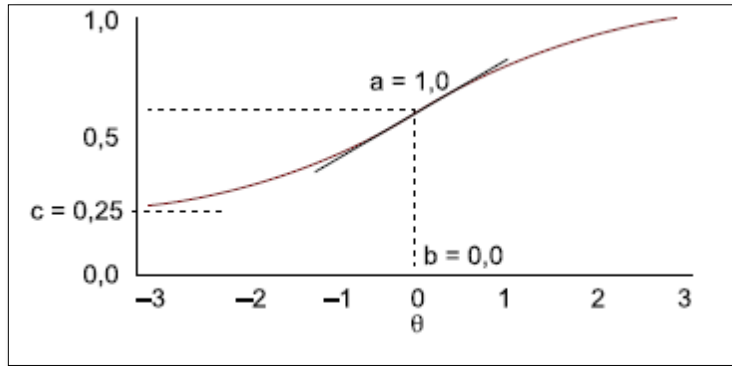
En cualquier contexto de medida existe una variable de interés subyacente que, en el ámbito educativo, se refieren a los conocimientos, convirtiéndose en variables cognitivas que pueden también ser psicológicas, como la inteligencia, todas estas variables deben tener como soporte un constructo teórico. Estas variables subyacentes son llamadas en la

nomenclatura de TRI “rasgos latentes” o “habilidades”. Entonces, la TRI es una herramienta que nos permite cuantificar un rasgo latente de una persona (Debera & Nalbarte, 2006).

De acuerdo con Torres Fernández y Artavia Medrano (2010), citado por (Moya, Pérez Gómez, & Ruiz Cordovéz, 2016) la TRI establece un enfoque metodológico que se apoya en modelos que asocian la probabilidad de respuesta de los estudiantes examinados con la medida de sus habilidades en el tema evaluado y se caracterizan por un reforzamiento en el principio de confiabilidad de los instrumentos de evaluación, el cual permite la consistencia de los resultados con independencia de los individuos que la responden por su parte la TCT, establece la noción de habilidad por medio del llamado puntaje verdadero que se define como "el valor esperado a partir de la destreza observada en la prueba en cuestión" (Hambleton, Swaminathan, & Rogers, 1991).

Luego de conceptualizar la Teoría Respuesta al ítem para entender sus fundamentos es importante conocer las principales características como una alternativa a la teoría clásica de los test, estas son: 1. Las características de los reactivos no dependen del grupo del cual fueron obtenidas. 2. Los puntajes que describen la habilidad del examinado no dependen del test en su conjunto. 3. El modelo se expresa al nivel del reactivo más que al nivel del test. 4. El modelo no requiere de pruebas paralelas para determinar el índice de confiabilidad y finalmente 5. Provee una medida de la precisión de cada índice de habilidad.

La TRI, se basa en el cálculo, para cada ítem de una serie de parámetros, que asumen un modelo matemático para la medición del rasgo latente a partir de tres parámetros: la discriminación del ítem, su dificultad y el acierto al azar. Los dos principios básicos que de acuerdo con Hambleton, et. al., (1991), sustentan esta teoría son la ejecución de un examinado en una prueba pueden ser predichos por un conjunto de rasgos latentes y habilidades; y la relación entre las respuestas de los examinados a los reactivos y el conjunto de rasgos que subyacen a la respuesta ante el reactivo, pueden describirse por una función llamada curva característica del ítem (CCI). La siguiente figura muestra un ejemplo de la CCI.

Ilustración 1 Ejemplo de Curva Característica de un Ítem (CCI)

Fuente: Tomado de (Meneses, y otros, 2013, p.251)

A partir de la curva de características de los ítems (CCI) se miden los tres parámetros de la TRI. El parámetro identificado con la letra a) mide la discriminación del ítem, que indica que cuanto mayor sea la pendiente, mayor será la discriminación. El parámetro con la letra b, mide la dificultad a partir del eje X, correspondiente a la probabilidad de acierto, cuanto mayor sea el parámetro b, más difícil será el ítem, finalmente el parámetro c, mide el nivel de la respuesta al azar y se le conoce también como índice de pseudo-azar. En la ilustración 1, corresponde al valor de X, que se intercepta en el eje Y, se interpreta como la probabilidad de acertar cuando el conocimiento del ítem es nulo por parte de quienes lo responden. Estos cálculos requieren de programas software específicos y de personal calificado para desarrollar los análisis requeridos.

Tabla 1 Diferencias entre la Teoría Clásica de los Test y la Teoría Respuesta al Ítem

Aspectos	Teoría Clásica	Teoría de Respuesta a los Ítems
Modelo	Lineal	No Lineal
Asunciones	Débiles (fáciles de cumplir por los datos)	Fuertes (difíciles de cumplir por los datos)
Invarianza de las mediciones	No	Sí
Invarianza de las propiedades del test	No	Sí
Escala de las puntuaciones	Entre cero y la puntuación máxima en el test	Entre $-\infty$ y $+\infty$
Énfasis	Test	Ítem
Relación Ítem-Test	Sin especificar	Curva Característica del Ítem
Descripción de los ítems	Índices de Dificultad y de	Parámetros a, b, c

Discriminación		
Errores de medida	Error típico de medida común para toda la muestra	Función de Información (varía según el nivel de aptitud)
Tamaño Muestral	Puede funcionar con muestras entre 200 y 500 sujetos aproximadamente	Se recomiendan más de 500 sujetos, aunque depende del modelo

Fuente: Tomado de (Muñiz J. , Teoría de los Test: Teoría Clásica y Teoría respuesta al ítem, 2010, p.64)

2.3.3 Análisis de los datos de una evaluación del rendimiento académico

Para el análisis de los ítems de un test, como se ha demostrado anteriormente, se distinguen dos grandes teorías, la primera basada en el modelo lineal clásico a partir de la TCT, y el análisis a través de la CCI, que propone la TRI. A continuación, se profundiza acerca de las propiedades psicométricas que miden la calidad de un test a partir de la Teoría Clásica de los Test, siendo esta el fundamento científico sobre la cual se presentaran los resultados de esta investigación.

Propiedades psicométricas de los ítems en el marco de la Teoría Clásica de los Test (TCT)

La evaluación de la calidad métrica de los elementos de un test, describe las características estadísticas de estos y de su contribución a la medición del constructo. Los ítems de un test por ejemplo deben ser capaces de discriminar entre los examinados de alto rendimiento frente a los que no lo tienen. El análisis de los ítems se lleva a cabo en dos fases: en el estudio piloto y en el diseño de la prueba en su versión final. para comprobar la calidad psicométrica de los ítems se deben estimar los parámetros de dificultad (p-value), discriminación, así como el análisis de los distractores. Para la valoración del sesgo en la medida se emplea el análisis del Funcionamiento diferencial de los ítems (DIF). Así mismo estas características psicométricas contribuyen a las fiabilidad y validez de los resultados. (Barbero Garcia, 2006).

Índice de dificultad: se define como el coeficiente entre el número de individuos que aciertan el ítem y el número total de individuos que responden el ítem. Según Morales (2007) el índice de dificultad se interpreta de la siguiente manera: • 0,0-0,15 se considera muy difícil • 0,15-0,4 se considera difícil • 0,4-0,6 se considera moderado • 0,6-0,85 se considera fácil • 0,85-1,0 se considera muy fácil. Esta relación muestra que la dificultad como la facilidad de un ítem o reactivo, son conceptos excluyentes y complementarios con respecto a la unidad.

En un grupo de examinados, una pregunta será más fácil si ella es acertada por un mayor número de sujetos, y por tanto será más difícil cuanto más sujetos la fallen (Hurtado Mondoñedo, 2018). La dificultad de un ítem se relaciona con su contenido, si existe complejidad en los contenidos, en correspondencia los elementos desarrollados serán difíciles o lo contrario; generalmente las pruebas miden contenidos variados, de forma que el índice de dificultad debe ir correspondiente al contenido que se pretende medir (Meneses, y otros, 2013).

Índice de discriminación: se define como la capacidad que tiene un ítem, para distinguir entre los examinados de buen rendimiento en relación a los que no demuestran la habilidad requerida. También se define como la correlación entre las puntuaciones de los sujetos en el ítem y sus puntuaciones. Conocer cuáles alumnos tiene un mejor aprendizaje que otros es uno de los principales objetivos de las evaluaciones en el ámbito educativo, por consiguiente, el poder de discriminación de un ítem es importante para la calidad de un examen.

El análisis de esta propiedad se analiza de la siguiente forma: se divide a los estudiantes con mejores puntuaciones y con las peores puntuaciones, y se espera que un ítem “bueno” sea respondido correctamente en su mayoría por aquellos estudiantes con un buen desempeño, y que aquellos que tuvieron un bajo desempeño fallen en mayor proporción el ítem. Esta fórmula se conoce como “Índice de discriminación basado en grupo extremos”, donde se toma el cuartil inferior y el cuartil superior como grupos extremos (García-Cueto, 2005). La fórmula para calcular este índice es el siguiente: $D = P^+ - P^-$

Donde P^+ es la proporción de acertantes al ítem del grupo superior, y P^- es la proporción de acertantes del grupo inferior. Ambas proporciones se obtienen del número de personas que aciertan el ítem entre el número de personas que forman cada cuartil.

Para ayudar a interpretar las puntuaciones del índice de discriminación, Ebel (1965) propone las siguientes clasificaciones:

Valores	Interpretación
Igual o mayor que 0,40	El ítem discrimina muy bien
Entre 0,30 y 0,39	El ítem discrimina bien
Entre 0,20 y 0,29	El ítem discrimina poco

Entre 0,10 y 0,19 Ítem límite. Se debe mejorar

Menor de 0,10 El ítem carece de utilidad para discriminar.

Coeficiente de correlación biserial puntual: Es la correlación entre las puntuaciones de los sujetos en el ítem y las puntuaciones de los sujetos en toda la prueba.

Calidad de los instrumentos: confiabilidad y validez

Otro importante aspecto que surge al analizar un instrumento es su calidad técnica. Cuanto mejor es la calidad de un instrumento, más útil será, más confianza se le puede tener a los puntajes obtenidos y mayor será la confianza para tomar decisiones a partir de estos resultados, por lo que es imperativo usar instrumentos de alta calidad al hacer evaluaciones (Center for Assessment and Research, James Madison University, 2014). Los dos principales elementos que dan cuenta de la calidad de un instrumento son su validez y su confiabilidad.

Validez de un instrumento

El concepto de validez ha ido sufriendo transformaciones a lo largo del tiempo. Tradicionalmente, la validez de un instrumento se ha entendido como hasta qué punto el instrumento efectivamente mide lo que sus autores declaran que mide (Cohen,, Manion, & Morrison, 2007). Sin embargo, más recientemente los especialistas en evaluación han considerado que la validez no es una propiedad fija e inherente del instrumento, sino que es un juicio, basado en evidencia, sobre qué tan apropiadas son las inferencias realizadas o acciones implementadas a partir de los puntajes de una prueba en un determinado contexto (Salvia & Ysseldyke, 2004; Cohen y Swerdlik, 2009).

Entonces, la validación puede ser vista como el desarrollo de un argumento de validez sólido para los usos propuestos de los puntajes de un instrumento y su relevancia para el uso propuesto (Joint Committee on Standards for Educational and Psychological Testing of the American Educational Research Association, American Psychological Association and National Council on Measurement in Education, 1999)⁸. Definida de esta manera, la validez no es una propiedad inherente al instrumento, sino que se relaciona con el objetivo de evaluación.

Tipos de validez:

Validez de contenido: En los tests no se pueden evaluar todos los conocimientos de los estudiantes, sino solo una muestra de ellos, por lo tanto, es muy importante que esta sea una muestra adecuada del área de aprendizaje que interesa evaluar. Si esto se logra, aumenta nuestra posibilidad de hacer inferencias válidas sobre los logros de aprendizaje en un cierto dominio (Darr, 2005). Este tipo de evidencia requiere mirar el contenido del instrumento para analizar la relación con el constructo que se quiere medir (Joint Committee on Standards for Educational and Psychological Testing, 1999; Cohen y Swerdlik, 2009).

Para saber qué es lo que un test mide, no basta con guiarse por el nombre, es fundamental mirar los ítems que lo componen (Center for Assessment and Research, James Madison University, 2014). Se puede analizar cada ítem en relación con el dominio, o pedir la opinión de expertos sobre qué tan adecuadamente estos abordan el dominio (Joint Committee on Standards for Educational and Psychological Testing, 1999). Si un instrumento es bueno, tendrá ítems que evalúen diferentes aspectos del tema evaluado, y expertos en el área, que no están familiarizados de antemano con los ítems, estarán de acuerdo en qué evalúa cada ítem. Existen dos riesgos que deben ser evitados. Uno es la sub-representación del constructo, es decir, que elementos importantes del constructo que se quiere evaluar no estén siendo evaluados. El otro es la varianza relacionada con constructos que son irrelevantes para lo que se está midiendo, por ejemplo, en una prueba de lectura, el conocimiento previo del tema o la respuesta emocional frente al texto, o en un test de matemática, la velocidad de lectura o el vocabulario (Joint Committee on Standards for Educational and Psychological Testing, 1999).

Validez de constructo: Como parte de la valoración psicométrica de la prueba, se desarrolla la aplicación del Análisis Factorial como uno de los métodos más significativos para evaluar la validez de constructo. Este tiene como uno de sus objetivos proporcionar a la prueba un fundamento científico. La validez de constructo es definida como el grado en que una prueba mide y evalúa una característica psicológica, demostrando las construcciones teóricas que explican en cierta medida el resultado encontrado (Brown, 1993; Muñiz, 1994; Anastasi y Urbina, 1998).

Cuando se investiga la validez de constructo, por lo general se desea saber qué propiedad psicológica puede explicar la varianza de la prueba. El investigador se pregunta: ¿qué construcción teórica mide la prueba?, ¿qué factores o constructos dan cuenta de la varianza en el desempeño de la prueba? La validez de constructo adquiere gran importancia en la Psicometría e investigación científica porque vincula las nociones y prácticas psicométricas con las nociones teóricas (Kerlinger, 2000). No se trata sólo de validar el instrumento sino aproximarse a la teoría que fundamenta el instrumento. El aspecto significativo acerca de la validez de constructo es el estar focalizado en la teoría, los constructos teóricos y la investigación científica empírica que implica la prueba de relaciones hipotetizadas entre las variables. La validación de constructo requiere una teoría sustantiva para definir el constructo que va a ser medido, así como una teoría de la medición para proporcionar la medida (Tapia & Luna, 2010)

Análisis Factorial: El análisis factorial es una técnica estadística multivariada que se incorpora a la metodología cuantitativa que involucra variables latentes. Estas variables no observables, denominadas frecuentemente constructos, son variables que no pueden medirse de manera directa: se estiman a través de variables manifiestas (observadas). Ejemplos de variables latentes podrían ser la inteligencia, el nivel de ansiedad, el nivel socioeconómico, el capital cultural, el grado de satisfacción con un producto o el nivel de razonamiento verbal. Variables observadas podrían ser la respuesta a un reactivo de un examen, el número de aciertos en un examen, la intensidad con que se lanzó una pelota, el número de computadoras en una vivienda, etcétera.

En el análisis factorial se asume que la variable latente es continua: los individuos pueden ordenarse de mayor a menor nivel del atributo bajo estudio. El objetivo primordial de esta herramienta es estudiar la estructura de correlación entre un grupo de variables medidas, asumiendo que la asociación entre las variables puede ser explicada por una o más variables latentes, que en el caso del análisis factorial se les reconoce como factores. Dicho de otra manera, la correlación entre el grupo de variables se explica por la presencia de los factores subyacentes a ellas. En el caso de que esta estructura de correlación pueda explicarse a través de un solo factor, estaremos ante un modelo unifactorial; por el contrario, si necesitamos más de un factor para explicar estas correlaciones, utilizaremos un modelo multifactorial. En este último caso, se espera que las variables que componen cada uno de

estos factores estén fuertemente correlacionado, y con correlaciones débiles con las variables que componen el resto de los factores.

Confiabilidad de un instrumento

La confiabilidad significa la consistencia entre los puntajes de un test obtenido por los mismos individuos en distintas ocasiones o entre diferentes conjuntos de ítems equivalentes (APA, 1999). Se refiere a la consistencia con que el instrumento mide, o visto de otro modo, al grado de error presente en la medida, en teoría, un instrumento perfectamente confiable mide siempre de la misma manera (Cohen y Swerdlik, 2009). Un test confiable entrega resultados consistentes a lo largo del tiempo. Por ejemplo, estudiantes con el mismo nivel de comprensión lectora que toman un test de comprensión lectora tendrán puntajes similares o idénticos, sin importar cuándo lo tomen, asumiendo que su nivel de comprensión lectora no ha variado. Se define también como la precisión con que el test mide lo que mide, en una población determinada y en las condiciones normales de aplicación.

Para determinar la confiabilidad, al igual que la validez, debe entenderse en contextos y propósitos evaluativos específicos. Sin embargo, ya que la confiabilidad remite a cuánta variación es esperable entre una medición y otra, se entiende de una manera más estrechamente estadística que la validez. Desde la teoría clásica, los enfoques más típicos para analizar la confiabilidad son: coeficientes derivados de la administración de formas paralelas en sesiones independientes, coeficientes obtenidos por la administración del mismo instrumento en ocasiones separadas (conocido también como “test re-test”, o “coeficiente de estabilidad”) y coeficientes basados en la relación entre puntajes derivados de ítems individuales o subtests dentro de un test (Cohen y Swerdlik, 2009).

Dimensiones de la confiabilidad:

- Estabilidad: si se pretende evaluar en qué grado el puntaje de un individuo en un test está libre de errores de medición causados por cambios personales aleatorios en el examinado y por cambios en las condiciones de administración.
- Consistencia interna: si se intenta conocer en qué medida la elección de la muestra de ítems que componen la prueba resulta una fuente de error en la medición.

- Confiabilidad inter-examinadores: Si se desea estimar en qué grado la medición de un rasgo a través de un instrumento es independiente de la subjetividad del evaluador

Es importante señalar que la Fiabilidad o confiabilidad de un instrumento ayuda a discriminar a los estudiantes que han estudiado y quienes no, ósea que la debe tener una capacidad de discriminar. La validez de criterio es una propiedad externa del instrumento es una forma externa de evaluarse se necesita comprobar que los resultados del instrumento realmente corresponden a lo estamos evaluando.

El coeficiente de confiabilidad más usado en teoría clásica es el Alpha de Cronbach, que pertenece a esta última categoría. Alpha se desarrolló en 1951 para entregar una medida de la de la consistencia interna de un test o una escala, es decir, identificar cuánto los ítems miden el mismo concepto, por lo tanto, si un test tiene varias escalas puede ser más apropiado usar alpha en forma separada para cada escala. Si los ítems se correlacionan entre sí, al valor de alpha aumenta. Pero este valor puede aumentar también por la cantidad de ítems (Webb, Shavelson & and Haertel, 2006).

Sus valores posibles se mueven entre 0 y 1. En general un alpha de .7 o más se considera aceptable (Institute for Digital Research and Education, UCLA, n.d.), por ejemplo, para la evaluación de programas (Center for Assessment and Research, James Madison University, 2014), pero si los resultados tendrán consecuencias individuales es mejor obtener valores superiores a .8 (Webb, Shavelson & and Haertel, 2006).

2.4. Instrumentos de medida para la evaluación de los aprendizajes

Desde un punto de vista métrico, los instrumentos que se utilicen para la evaluación del rendimiento académico deben reunir condiciones tales como ser objetivos, claros, comprensibles por las partes involucradas, preferiblemente cuantitativos, fiables y válidos. En el apartado siguiente se amplía con un poco más de detalle los pasos que deben seguirse para desarrollar instrumentos de evaluación que cumplan estas condiciones. No obstante, conviene aclarar desde el principio que un instrumento métricamente adecuado es condición necesaria, pero no suficiente para llevar a cabo una evaluación exitosa (Muñiz, Cueto Garcia, Vicente Martinez, Moreno, & Fidalgo, 2005) .

Existen diversas maneras de clasificar los instrumentos de evaluación de aprendizajes, una forma de clasificar los sistemas de evaluación es según las consecuencias en función de sus resultados. De forma que, existen los sistemas de bajas implicancias, que generan información para fines formativos, de decisiones “blandas”; y los de altas implicancias, que usan los resultados de pruebas para diversos fines con consecuencias para los involucrados, tales como incentivos para los docentes, definir la promoción educativa de los estudiantes o bien informar a la población acerca del rendimiento académico de los estudiantes en relación al logro de habilidades contempladas en los diversos programas de estudio (Covacevich, 2014).

Así mismo los instrumentos se pueden clasificar según las características técnicas en cuanto a su desarrollo y abordaje en la interpretación de los resultados, es decir si la medida del aprendizaje es directa a través de pruebas de conocimiento o indirecta, basado en la percepción de estudiantes, padres de familia, docentes (Covacevich, 2014). A continuación, se dan a conocer sucintamente las pruebas de aprendizaje que son una medida directa del aprendizaje, clasificadas como instrumentos normativos y criterios, paramétricos y no paramétricos finalizando con los de respuesta cerrada versus respuesta abierta.

2.4.1 Instrumentos normativos y criterios

Las evaluaciones estandarizadas de los aprendizajes pueden proporcionar dos tipos de resultados de acuerdo a los tipos de instrumentos utilizados, de acuerdo a sus características existen instrumentos referidos a la norma y los referidos a criterios. Los primeros sirven para comparar los resultados de un individuo con respecto a una población de referencia, con la cual se normalizan estadísticamente las puntuaciones. Así, los resultados de una persona pueden presentarse en puntuaciones Z, en una escala predefinida, en puntuaciones percentilares o, bien, en niveles de. Por su parte, los resultados de una evaluación referidos a un criterio hablan de la cantidad o proporción de competencias que un estudiante domina, del total de competencias evaluadas. Los resultados se presentan, por lo general, de dos maneras: porcentaje de respuestas correctas y si se cuenta o no con el nivel requerido para ejecutar una tarea (Backhoff Escudero, 2018).

La diferencia entre evaluación normativa y evaluación criterial es introducida por Glaser en 1963, autor que utiliza por primera vez la expresión “evaluación criterial” Popham, (1983)

citado por (Heredia Manrique, 2009). Estos dos tipos de estrategias de evaluación hacen referencia explícita al rendimiento del estudiante, cada una dando su propia interpretación del mismo, con una perspectiva muy técnica y operativa. La característica fundamental de la evaluación criterial es apreciar el logro de los objetivos por parte de cada alumno, se fundamenta en la precisión de una serie de aspectos: universo de medida, objetivos de aprendizaje, elementos de medida, punto de corte, que tienen como finalidad, saber si cada alumno ha adquirido el conocimiento mínimo del área curricular. La evaluación criterial técnicamente es viable, aunque costosa, como modalidad de evaluación para países en vías de desarrollo, pero sin lugar a duda la información generada a raíz de estos instrumentos es sumamente valiosa para la toma de decisiones en materia de política educativa.

La evaluación normativa por su parte, tiene como objetivo ver la posición relativa de cada alumno respecto a los demás y cumple su función cuando con ella se sabe quién es mejor (o peor) que quién. Así, trata de identificar el talento del alumno, de manera que se puedan establecer comparaciones y luego hacer una selección entre los alumnos, Popham (1983): citado por (Alcaraz Díaz, 2003). La definición técnica de evaluación referida a la norma refiere que es un test basado en pautas de normalidad destinado a determinar la posición de un sujeto examinado en relación con el rendimiento de un grupo de otros sujetos que haya hecho ese mismo test. Ése será el grupo normativo respecto al cual se compara el rendimiento de quien haga dicho test. Los test basados en pautas de normalidad son excelentes instrumentos para la finalidad con que fueron creados, discriminar entre los examinados en una situación de *numerus clausus* en la que sólo hay posibilidades de admisión para un número limitado de individuos (Popham W. , Evaluación basada en criterios, 1983).

La evaluación normativa tiene como referencia las realizaciones de los alumnos del mismo grupo, procurando la medición de las diferencias individuales de rendimiento, bajo un modelo de curva normal cuyo fin es clasificar los estudiantes. La media del grupo se utiliza como criterio de superación o no (punto de corte) y será variable, más alta o más baja, en función de las capacidades demostradas por el grupo de alumnos: siempre habrá un 50% de alumnos máster por encima de la media y un 50% de alumnos no-máster por debajo de la media. Así, el mayor o menor rendimiento individual se disgregará dentro del rendimiento del grupo, el cual, puede ser muy variable (Gomez Arceo, 1990).

Tabla 2. Características de los instrumentos normativos y criterios

Característica	Test referido a norma	Test referido a criterio
Objeto que se evalúa	Mide una variable psicológica o rasgo.	Mide un conjunto de conocimientos o competencias (criterios).
Análisis de la fiabilidad	A partir de las diferencias entre los sujetos.	A partir del dominio que presenta el individuo sobre el tema.
Objetivo de evaluación	Posición relativa de un individuo respecto al resto del grupo.	Grado de conocimiento que presentan los individuos en el dominio.
Ámbito de aplicación	Personalidad, actitudes, etc.	Educación, ámbito laboral evaluación de programas, etc.

Fuente: (Meneses, y otros, 2013. p.113)

La tabla 2, resume los elementos básicos que contempla cada uno de estos test y sobre todo considerando que, aunque existen diferencias marcadas, en la práctica distinguir los procesos de cada test, no siempre se muestra de forma muy perceptible. Para finalizar este pequeño apartado se puntualiza que mientras los test referidos a la norma (TRN) tienen como objetivo conocer si la puntuación obtenida por el sujeto “A” es superior a la del sujeto “B” cuando se valora X, los test referidos al criterio (TRC), plantean si ambos sujetos tienen la habilidad para resolver X (Meneses, y otros, 2013).

2.4.2 Instrumentos paramétricos y no paramétricos

Estos test utilizan métodos diferentes, debido a los objetivos que tienen cada uno, sin embargo, se complementan en el análisis de sus resultados. Las pruebas paramétricas permiten representar a la población general, por ejemplo, a la de un determinado país, o de un cierto grupo dentro de un país. Generalmente los test paramétricos han sido piloteados y estandarizados en la población completa y ofrecen muchos datos sobre su muestreo, validez y confiabilidad. Los tests no paramétricos, en cambio, han sido diseñados para un grupo específico y no consideran a la población general (Cohen, Manion, & Morrison, 2007).

Las ventajas de los tests paramétricos son que permiten comparar sub-grupos con la población general, por ejemplo, los resultados de una escuela específica contra el puntaje promedio nacional, realizar análisis sofisticados de los datos, y hacer inferencias. Los

instrumentos no paramétricos, en cambio, son particularmente útiles para muestras pequeñas y específicas, por ejemplo, un cierto curso de una cierta escuela, y pueden ser diseñadas específicamente para lo que se desea medir en esa situación. Además, ya que los análisis que se pueden hacer necesariamente son menos sofisticados, son más fáciles de realizar por cualquier usuario (Cohen, Manion y Morrison 2000).

Un ejemplo de un test paramétrico sería una prueba nacional censal, en que los resultados obtenidos por cada escuela se pueden comparar con los obtenidos por el promedio de las escuelas del país, y de uno no paramétrico, una prueba diseñada por una escuela para diagnosticar los conocimientos en matemática de sus estudiantes al inicio del año escolar.

2.4.3 Instrumentos según el formato de sus ítems: respuesta abierta o cerrada

Un test, resulta un conjunto intencionalmente diseñado de ítems; a través de cuyas respuestas se logra inferir el grado de habilidad de quienes son evaluados, en relación con el objeto de evaluación. En este sentido, una prueba constituye la agregación sinérgica y articulada de ítems para la comprobación de habilidades acerca de las personas que la responde, en función de la eficacia con que lo hace (Pardo Adames & Rocha Gaona, 2009). En tal sentido, los instrumentos que se usan para medir aprendizaje tradicionalmente se dividen en aquellos que son de respuesta abierta y los de respuesta cerrada. Estos últimos implican escoger alternativas dentro de una serie de opciones provistas; el ejemplo más conocido son las pruebas estandarizadas de ítems de selección múltiple. Los instrumentos de respuesta construida, en cambio, son las evaluaciones en que se debe crear la respuesta, tales como ensayos, reportes, u exámenes orales (Center for Assessment and Research, James Madison University, 2014).

A cada formato se le asocia la capacidad para provocar en los examinados procesos cognitivos particulares; no obstante, más que el formato en sí mismo, lo importante es la combinación de la formulación de la tarea de evaluación, la naturaleza del objeto evaluado y las características de la población lo que determina que un formato de ítems resulte más apropiado para evaluar unos procesos cognitivos que otros.

Los instrumentos de respuesta abierta, se componen generalmente de tres elementos básicos: un contexto que delimita las condiciones para responder, una tarea de evaluación que indica lo que se espera que el evaluado produzca y además unos criterios que delimitan

las condiciones esperables del producto. Los tipos de ítems utilizados en estos instrumentos corresponden a los ítems de tipo completación, y de construcción de ensayos cortos, largos, de ejecución, y productos escritos. Su revisión es compleja y requiere de un equipo de profesionales especializados para su revisión, en especial en aquellos que evalúan la expresión oral, donde resulta más exigente incluso que la evaluación de la producción escrita (Pardo Adames & Rocha Gaona, 2009).

Por otra parte, los instrumentos de respuesta cerrada, contienen algunas ventajas desde su aplicación y revisión, entre estos se mencionan los instrumentos con ítems de selección múltiple, selección alterna, dependientes de contexto y también aquellos elementos llamados falso y verdadero. Sobre estos se puntualiza acerca de las ventajas de los instrumentos de selección múltiple pues debido a que se pueden administrar de forma rápida a grandes números de estudiantes abordando muchos temas, hoy resultan un tipo de ítems frecuentemente utilizados en las evaluaciones nacionales para el rendimiento académico a nivel nacional e internacional.

No obstante, este tipo de elementos ofrecen menos riqueza, menos profundidad y tienden a enfocarse en la capacidad de recordar y otras habilidades cognitivas de orden inferior. Las ventajas de los de respuesta construida son que proveen una información más profunda de los que los estudiantes saben y pueden hacer, pero como se indicaba anteriormente son caros porque deben ser puntuados por jueces entrenados especialmente para ello, lo que además toma bastante tiempo, y en general requieren tiempos de aplicación más largos; además, si la puntuación no es efectuada correctamente y de la misma manera por todos los correctores, la comparación de los puntajes se ve afectada, lo que no ocurren en el caso de los de respuesta cerrada. (Center for Assessment and Research, James Madison University, 2014).

Finalmente, existen algunas pruebas que combinan ítems tanto de respuesta abierta como cerrada. Por ejemplo, una prueba puede estar conformado principalmente por ítems de selección múltiple, pero complementar la información recogida con una o dos preguntas abiertas, que requieran que el estudiante redacte un texto y exponga argumentos, lo que permite evaluar habilidades que no se pueden recoger de la misma manera con solo ítems de selección múltiple, para ello se emplea una rúbrica y una metodología en su revisión, que logre de una forma estándar determinar la habilidad de los examinados .

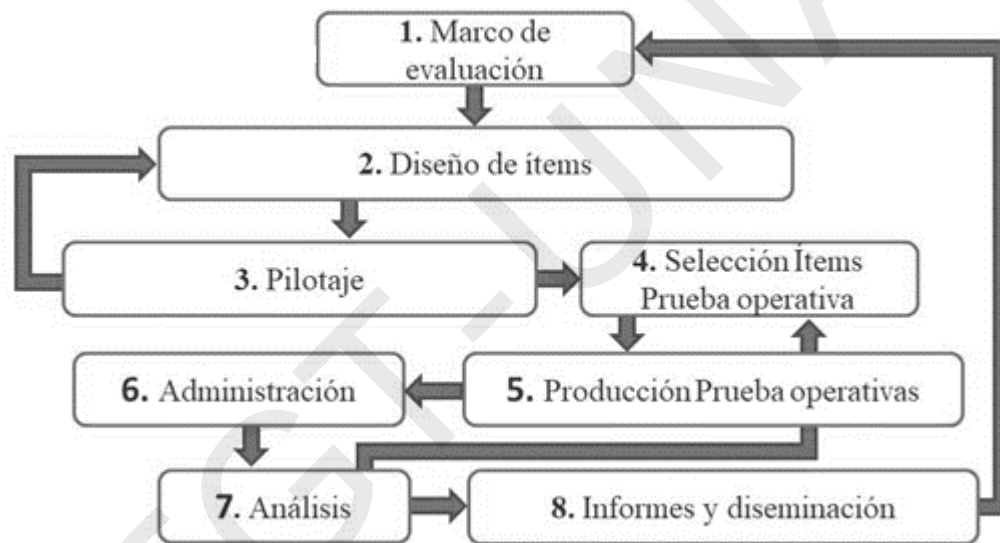
2.5 Fases generales para el diseño de una evaluación nacional de rendimiento académico

Dada la incuestionable importancia de la construcción, adaptación y uso de test, resulta más que adecuado reconsiderar periódicamente el proceso y secuenciación de las tareas que se dan entorno a estas actividades, persiguiendo con esto mejorar la labor que supone la medida de lo psicológico a través de un test (Perez & Carretero-Dios, 2005) . El proceso de construcción de un instrumento de medición debe ser definido objetivamente, debe tener principios teóricos y métricos para su confiabilidad y validez. Un diseño mal construido genera información incompleta, datos no precisos, generando información nada confiable.

Todo proceso de construcción de un instrumento de medida comienza por una justificación detallada y precisa de cuáles son las causas que motivan su construcción. Asimismo, hay que delimitar con claridad cuál es la variable objeto de medición, cuál va a ser el contexto de aplicación o circunstancias en el que se va administrar el instrumento de evaluación, el tipo de aplicación (individual, colectiva), el formato de aplicación (lápiz y papel, informática), y qué decisiones se van a tomar a partir de las puntuaciones (Muñiz & Fonseca Pedrero, 2008).

Existen algunos principios básicos que deben tenerse en cuenta en todo ejercicio de evaluación educativa, estos principios corresponden al proceso general para la construcción de un test. De la calidad en el diseño de los instrumentos depende la confiabilidad y validez en cuanto a los resultados. Diversas literaturas al respecto enmarcan pasos en común para este tipo de instrumentos, sin embargo, para este trabajo de investigación se ha considerado una revisión reciente, desarrollada por Prue Anderson y George Morgan (2016), en la serie de libros, del grupo Educación del Banco Mundial: Evaluaciones nacionales del rendimiento académico, así como otra literatura homologa a la propuesta de esta obra en mención.

Ilustración 2 Pasos para una evaluación nacional



Fuente: Elaboración propia, adaptado documento: Paradigmas y tendencias en medición y evaluación educativa (Fast & Vukmirovic , 2016)

En la ilustración 2, se recoge de forma esquemática, los principales pasos o fases que responden a la construcción de un instrumento para la evaluación del rendimiento académico y a continuación se describe cada una de ellas:

1. Diseño del marco de evaluación:

Inicialmente el proceso de construcción de un instrumento de medida comprende una justificación detallada y precisa de cuáles son las causas que motivan su construcción, incluyendo una definición de lo que se está evaluando, determinar las características de las diferentes tareas que se emplearan y por supuesto estipular cuáles serán las bases sobre las cuales se desarrollaran la interpretación de sus resultados Kirsch (2001), Messick (1987), citado por (Anderson & Morgan, Evaluaciones de rendimiento academico., 2016).

Esta etapa comprende la elaboración de un marco conceptual (Test Framework), mismo que tiene como función principal asegurar coherencia y consistencia en la construcción de una evaluación, definiendo desde la estructura organizadora de la evaluación, hasta el detalle del contenido a evaluar y la población objetivo. El muestreo probabilístico siempre es preferible al no probabilístico, para la estimación del tamaño muestral requerido para un determinado error de medida ha de acudir a los textos especializados, o consultar los expertos en la tecnología de muestreo.

El diseño del marco de evaluación abarca los siguientes aspectos:

- Detalles precisos de la población meta que se quiere evaluar.
- Explicación precisa del propósito de la evaluación.
- Detalles del contenido a evaluar, extensión de la prueba y proporción de los ítems
- Cualquier información contextual que permita entender el motivo y forma de la evaluación.
- Tabla de estándares de contenido que representa un análisis del contenido a evaluar.

Una vez determinado el marco conceptual de la prueba, se elaboran dos documentos que guían el desarrollo, análisis y reportes de una prueba, estos corresponden al marco operativo de la variable de medida y la tabla de especificaciones técnicas del instrumento de medida (Fast & Vukmirovic , 2016).

Marco operativo de la prueba: el objetivo esencial de este documento es la definición operativa, semántica y sintáctica de la variable medida, así como las facetas o dimensiones

que la componen (*American Educational Research Association et al., 1999; Carretero-Dios y Pérez, 2005; Lord y Novick, 1968; Wilson, 2005*).

El constructo a evaluar debe ser definido en términos operativos, para que pueda ser medido de forma empírica. Esta fase requiere de personal especializado en el área del currículo que será objeto de medida. Lo anterior se traduce, por un lado, en delimitar la variable objeto de medición, y considerar todas las dimensiones relevantes de la misma, y por otro, identificar con claridad los niveles de procesamiento cognitivo más representativos de la variable de medición (Muñiz & Fonseca Pedrero, 2008).

El marco operativo brinda una relación de los ítems que conformaran la prueba operativa, señalando con detalle información como ser: la proporción de ítems dentro de la prueba en su versión final y según cada área del currículo (bloques, componentes, estándares); así mismo deberá pormenorizar la cantidad de ítems que abordan distintas habilidades del proceso cognitivo Bloom (2000), además mostrar la proporción de ítems de opción múltiple y de respuesta abierta correspondientemente (Fast & Vukmirovic(2016) Anderson & Morgan, 2016).

Tristan y Pedraza (2017) en su artículo *La Objetividad en las Pruebas Estandarizadas* mencionan que la definición del objeto de medida es el punto de partida para el proceso de evaluación porque es el que permite hacer apreciaciones cualitativas o cuantitativas respecto de los atributos inherentes del objeto, propiciando profundizar en sus características y funciones. La definición del objeto a medir debería ser una de las responsabilidades y preocupaciones de los diseñadores de pruebas y cuestionarios, porque sin una definición objetiva, es altamente probable que se obtengan resultados y conclusiones poco fieles del objeto que se está midiendo.

En este caso, se trata de un instrumento para la medición de habilidades cognitivas adquiridas en un año en el campo del conocimiento español y de no hacer una adecuada definición puede verse el enorme riesgo de generar un instrumento que proporcione información deficiente en términos de objetividad. Una vez delimitados el propósito de la evaluación y la definición operativa de la variable que interesa medir se debe llevar a cabo determinadas especificaciones relacionadas con el instrumento de medida.

Los requerimientos de administración del instrumento de medida se refieren a la cual va a ser el soporte administración o el tipo de aplicación que se va a realizar, y cuando y en qué lugar se va aplicar el instrumento de medida con sus requerimientos cognitivos, en relación con los ítems que se debe especificar el tipo, el número, la longitud, el contenido y el orden (disposición) de los mismos, así como el formato de respuesta o el tipo de alternativas que se van a utilizar.

La tabla de especificaciones técnicas describe la forma detallada y precisa aspectos concernientes al número de ítems que serán diseñados para el ensamblaje de la versión final. Una tabla de especificaciones para el diseño de ítems representaría entonces una extensión del marco operativo, en función que sería un riesgo el diseñar únicamente el número de ítems que precisa la versión final. La información de la tabla de especificaciones consiste en la multiplicación de los ítems por 3 o 4, según el nivel de confianza se tenga de la calidad de los ítems a pilotear y el tiempo y recursos disponibles para el diseño y revisión de los mismos (Fast & Vukmirovic ,(2016), Pardo Adames & Rocha Gaona, (2009).

2. Construcción de los ítems

La construcción de los ítems constituye una de las etapas más decisivos dentro del proceso de construcción del instrumento de medida (Downing, 2006; Schmeiser y Welch, 2006). Los ítems son la materia prima, los ladrillos, a partir de la cual se forma un instrumento de evaluación, por lo que una construcción deficiente de los mismos, incidirá en las propiedades métricas finales del instrumento de medida y en las inferencias que se extraigan a partir del mismo.

En el momento que se cuenta con un constructo claramente delimitado en cuanto a sus facetas o componentes operativos, e igualmente ha sido establecida la red de relaciones esperadas tanto para el constructo en general como para cada una de sus facetas, puede emprenderse la tarea de construcción de ítems. (Carretero-Dios 2005). Los ítems se definen como la formulación de un enunciado de forma clara, sencilla y mide el logro del aprendizaje de los estudiantes.

Para redactar reactivos se debe tomar en cuenta los siguientes criterios: el contenido de los reactivos deberá reflejar los programas de estudios, la dosificación de contenidos y estándares, deberán evaluar contenidos según el contexto cultural y lingüístico de los

estudiantes; haciendo uso de un lenguaje sencillo y de acuerdo al grado que se evalúa. Se debe evitar las expresiones técnicas que dificulten la comprensión de la pregunta o proposición.

En los ítems de selección múltiple, se debe considerar el nivel de procesamiento cognitivo que mide cada estándar; diseñar ítems desde las dos perspectivas, tanto de contenido curricular como de nivel cognitivo con el objetivo de que el instrumento pueda discriminar entre alumnos de distintos niveles de capacidad o habilidad, desde los más flojos (nivel insuficiente) hasta los más capacitados (nivel avanzado) (Muñiz, Cueto Garcia, Vicente Martinez, Moreno, & Fidalgo, 2005). Cada vez se vuelve más ineludible en el desarrollo de pruebas de evaluación nacional, el comprobar que los alumnos sean capaces de demostrar destrezas cognitivamente más complejas y no solamente de conocimiento y memorización de datos y hechos y procesos sencillos. Hay dos modelos que se puede utilizar para la calificación de nivel cognitivo: Bloom y Webb, la ilustración 3, muestra las habilidades propuestas por Benjamin Bloom (1956), revisada posteriormente (2000).

Ilustración 4 Taxonomía de Bloom, version revisada (2000)



Fuente: tomado de (Gual, 2010)

La Taxonomía de Bloom resulta una herramienta clave para los docentes en la etapa del diseño de ítems para instrumentos de evaluación. En el marco operativo de una prueba se deben proponer reactivos desde las dos perspectivas, tanto de contenido curricular como de nivel cognitivo, con el objetivo de mostrar la gama de capacidades de los alumnos que se contempla evaluar.

Del mismo modo, los ítems deben ser comprensibles para la población a la cual va dirigido el instrumento de medida, evitándose en todo momento un lenguaje ofensivo y/o discriminatorio. Ítems con una redacción defectuosa o excesivamente vagos van a incrementar el porcentaje de varianza explicada debido a factores espurios o irrelevantes. Los ítems de respuesta cerrada, en especial los de selección múltiple son el formato más utilizado en los procesos de evaluación estandarizada, está constituido por dos partes esenciales: un enunciado, que contiene la pregunta o tarea de evaluación y algunas (entre 3 y 5) opciones que plantean posibles respuestas al enunciado, entre las cuales sólo una es la respuesta correcta. La coherencia y la concordancia gramatical entre el enunciado y las opciones son características que los desarrolladores de pruebas deben cuidar con esmero (Pardo Adames & Rocha Gaona, 2009).

Los ítems pueden de selección múltiple, pueden estar acompañados de un material informativo que se presenta para delimitar la tarea de evaluación. A esta información se le denomina contexto y puede ser un texto, una gráfica, una tabla, etc., en especial en el área de español, lectura y escritura se utilizan ítems dependientes. El evaluado se remite a la información que ofrece el contexto y a la pregunta implícita en el enunciado para pensar y elegir la respuesta válida entre las opciones. En algunos ítems, el contexto se funde con el enunciado y no se reconoce fácilmente como una parte distinta de este. En otras ocasiones, el contexto está dado por el saber previo que se supone tienen los evaluados acerca del objeto de evaluación, gracias al proceso formativo en el que están inmersos y, por tanto, no se hace explícito.

La principal ventaja de los ítems de selección múltiple, es que pueden abordar una diversidad de logros, estableciendo diferencias sutiles en el conocimiento y comprensión de los alumnos. Sin embargo una de las limitaciones es que se requiere amplia experiencia para desarrollar ítems de buena calidad y otra situación es el elemento de que pueden ser respondidos al azar, limitando la medición de la habilidad que pretende evaluar el constructo (Pardo Adames & Rocha Gaona, 2009).

3. Pilotaje de los ítems

Las pruebas previas o pruebas piloto resultan un elemento esencial del diseño de pruebas. Estas pruebas se administran a alumnos que tengan las mismas características que los que van a participar en la prueba final. Se deben incluir instituciones educativas de diferentes tamaños, en diferentes áreas geográficas y con estudiantes de diferentes niveles socioeconómicos. En cuanto a la administración de las pruebas pilotos lo recomendable, es que debe llevarse a cabo con un año de antelación, en la misma época del año que la prueba final será aplicada (Anderson & Morgan, 2016).

Los estudiantes que participen en la prueba piloto no deben ser parte de la muestra en la versión final. Si la prueba operativa se va a administrar a un grupo de muestra aleatorio seleccionado estadísticamente, el grupo de muestra final se debe seleccionar antes de elegir las escuelas para la prueba previa. Por otra parte; una buena práctica es realizar una prueba previa con un número de ítems dos o tres veces mayor que el necesario para la prueba final. En cuanto a la duración de la prueba piloto, esta debe tener una duración similar a la de la prueba final.

La finalidad de cualquier estudio piloto es examinar el funcionamiento general del instrumento de medida, si cumple con las propiedades métricas en una muestra de participantes con características semejantes a la población objeto de interés. Esta fase es de suma importancia ya que permite detectar, evitar y corregir posibles errores, así como llevar a cabo una primera comprobación del funcionamiento del instrumento de evaluación en el contexto aplicado (Muñiz & Fonseca Pedrero, 2008).

Una vez tabulados los datos, se procede a los análisis de la calidad métrica de los ítems. En función de criterios sustantivos y estadísticos algunos ítems son descartados mientras que otros pueden ser modificados. Es importante sistematizar una estrategia que deje constancia de qué ítems fueron eliminados o modificados; dependiendo de ello se puede considerar considera la incorporación de nuevos ítems. Todas las actividades deben ir destinadas a seleccionar los ítems con las mejores características métricas para lograr maximizar las propiedades finales del instrumento de evaluación. Finalmente, se debe construir una nueva versión del instrumento de medida que es revisada de nuevo por el grupo de expertos y que será la que en última instancia se administre en el estudio final de campo.

4. Selección de los ítems de la prueba operativa

La selección de ítems de la prueba previa para la prueba final depende en gran medida, del marco de evaluación, en especial de las especificaciones técnicas. La segunda condición para la selección de los ítems de la prueba en su versión final es las propiedades de medición de los ítems. Prue Anderson y George Morgan (2016), proponen los siguientes criterios para la selección de los ítems:

- El ítem se ajusta a las especificaciones técnicas.
- El porcentaje de alumnos que responde correctamente al ítem oscila entre el 40 y el 80 por ciento.
- El ítem tiene una tasa baja de respuestas faltantes.
- El índice de discriminación (correlación entre el puntaje del ítem y el puntaje total de la prueba) es superior a 0,2.
- La inclusión de este ítem mejora la confiabilidad de la prueba.
- El sesgo del ítem está dentro de límites aceptables para los grupos de alumnos principales.
- La correlación punto biserial de la clave es positiva y superior a 0,2.
- Todos los distractores son verosímiles (es decir, han sido seleccionados por al menos el cinco por ciento de los alumnos) y tienen una correlación punto biserial cero o negativa.

En general, las decisiones respecto de la inclusión o exclusión de ítems deben tener en cuenta las especificaciones técnicas y las áreas del plan de estudio que la prueba pretende evaluar. Los ítems con valores estadísticos por debajo de lo aceptable según el modelo estadístico de análisis deben ser revisados y de ser posible ser ensayados nuevamente. A pesar de ello, cuando ítems con estadísticas deficientes abordan contenidos significativos del marco operativo de la prueba y de no existir ítems disponibles, puede ser necesario incluirlos en la versión final de la prueba.

5. Ensamblaje de formas: producción de la prueba operativa

Una vez seleccionados los ítems con mejores propiedades psicométricas se procede al diseño de la prueba operativa. La organización de la información del instrumento, así como el manual de administración del mismo, debe tener claridad y precisión. Se recomienda que

el diseño de la prueba piloto y operativa sean iguales o similares. La prueba final, debe conformarse de ítems fáciles al principio para dar confianza a los examinados, continuando con ítems que presenten mayor grado de dificultad, de forma que los alumnos no decidan abandonar culminar la prueba, por resolver ítems difíciles. Los ítems relacionados o dependientes de un párrafo, mapa, grafico etc, deben estar juntos sin importar su nivel de dificultad.

En esta fase se compone, se edita y se lleva a imprimir la primera versión del instrumento de medida, además de construir la base de datos con la claves de corrección, normalmente las pruebas se imprimen en ambas caras de las hoja, con una calidad de papel adecuada, para asegurar que los ítems impresos en una cara no interfieran en la presentación de los ítems impresos en el lado opuesto, en cuanto a las ilustraciones, gráficos, se recomienda el uso de escala de grises, aunque este último depende del presupuesto para la impresión, el tamaño de la fuente también debe ser el adecuado (Greaney & Kellaghan, 2016).

6. Administración de la prueba operativa

Desde la aplicación de la prueba piloto se han debido definir las condiciones de administración convenientes y necesarias para los propósitos de evaluación, generalmente a través de un manual de administración de la prueba, concertado previamente con los involucrados en la aplicación de la prueba definitiva. Lo anterior a fin de garantizar que las condiciones para responder la prueba sean estandarizadas para la población objetivo; esperando que todos los evaluados reciban las mismas instrucciones antes de la aplicación; que ninguno de ellos haya tenido conocimiento previo del contenido de la prueba; o bien tenga la oportunidad de buscar o de recibir ayuda extra de fuente alguna para responder la prueba (Pardo Adames & Rocha Gaona, 2009).

Las condiciones de la aplicación, tales como la disposición de materiales (cuadernillos de prueba, hoja de respuestas, instrucciones, lápiz, borrador, etc.), la acomodación de las sillas o pupitres, así como la iluminación, la ventilación y el tiempo disponible para responder deben ser iguales para todos los evaluados y, finalmente, que cuando las circunstancias lo ameriten, se tomen las medidas oportunas para que personas de la población objetivo con alguna limitación física reciban el apoyo requerido para afrontar el proceso de evaluación en condiciones equivalentes al resto de la población.

7. Análisis de los resultados

Inicialmente el equipo destinado para la corrección de los ítems, en caso de hacerlo de forma manual, debe estar debidamente capacitado, máximo para aquellos ítems de respuesta abierta. La representatividad y generalizabilidad de los resultados depende en gran medida de que la muestra elegida sea realmente representativa de la población objetivo de estudio. Elegir una muestra pertinente en cuanto a representatividad y tamaño es esencial, si se falla en esto todo lo demás podría resultar un esfuerzo nulo.

Seguidamente se procede al estudio de las propiedades métricas del mismo: análisis de los ítems, estudio de la dimensionalidad del constructo, estimación de la fiabilidad, obtención de evidencias de validez y construcción de baremos (Pardo Adames & Rocha Gaona, 2009). En general las estadísticas clásicas de los ítems: el índice de discriminación, dificultad, se pueden interpretar directamente, tal y como se ha detallado en el apartado de análisis de los datos de una evaluación estandarizada.

El análisis de los datos a través de la TRI, suele ser más intuitiva con software más complejos, los parámetros a , b y c de este modelo, no se deben interpretar de forma aislada, porque aunque cada uno describe un comportamiento específico de los ítems, la relación entre las respuestas a los ítems y la prueba en general, es el resultado de las interacciones entre los tres parámetros y los niveles de competencia de los estudiantes (Shiel & Cartwright, 2016).

8. Informes y diseminación

La última fase en la elaboración de instrumentos para la medición del aprendizaje, corresponde a la presentación de los resultados, estos dependen en primer lugar de los objetivos de la evaluación por lo que debe considerarse los aspectos en cuanto al contexto político en el que se lleva a cabo la evaluación nacional y sus implicaciones para el uso de los resultados de la misma. Por otra parte, determinar una forma amigable para la presentación de resultados para población no especializada, será imperioso, esto para el cumplimiento la cuestión de la rendición de cuentas, que es la forma más usual en muchas administraciones gubernamentales y a la cual se han relacionado estrechamente las actividades de evaluación nacional, y además considerando que el uso de los resultados determina su utilidad y trascendencia. La información en este sentido debe mostrar también

la calidad de los instrumentos utilizados, de forma que los datos recabados en cuanto al aprendizaje de los estudiantes (conocimiento, habilidades, actitudes y hábitos que los estudiantes adquirieron como resultado de su escolarización) sean de crucial importancia para el uso de los resultados con el fin de mejorar el aprendizaje (Kellaghan, Greaney, & Scott Murray, 2016).

Presentados los resultados no quiere decir que el proceso de validación concluya aquí, posteriores estudios deberán seguir recogiendo evidencias de validez que permitan tomar decisiones fundadas a partir de las puntuaciones de los individuos. (Muñiz J. 2008). Alcanzar la objetividad en el proceso de evaluación junto con la validez y la confiabilidad permite disponer de pruebas mejor diseñadas, más robustas, donde las perfeccionadas herramientas de medición facultan tomar decisiones en beneficio de los individuos y de la sociedad en su conjunto (Tristán López & Pedraza Corpus, 2017).

Calidad de los instrumentos

La validez de constructo: Es dimensionar los conceptos que queremos medir. Es una demostración empírica de que una prueba está midiendo la construcción que pretende medir. La validez en general se refiere a la exactitud con que pueden hacerse medidas significativas y adecuadas con un test. La validez del constructo se define como un proceso a través del cual se acumula evidencia empírica, con el objetivo de establecer la pertinencia en la medición de un concepto o constructo teórico, esto a partir de inferencias o interpretaciones que se elaboran con base en las puntuaciones obtenidas en una prueba. La validez de constructo es entonces el concepto unificador que integra las consideraciones de validez de contenido y de criterio en un marco común para probar hipótesis acerca de relaciones teóricamente relevantes.

5. Versión definitiva, informe final y manual del instrumento de medida: Se procede a la elaboración de la versión definitiva del instrumento de medida, se envía un informe de resultados a las partes interesadas (alumnos, profesores, departamentos, etc.) y se elabora el manual del mismo que permita su utilización a otras personas o instituciones interesadas. El manual de la prueba debe de recoger con todo detalle todas las características relevantes de la prueba.

CAPÍTULO 3. MARCO CONTEXTUAL

3.1. Contexto educativo y social de la evaluación de los aprendizajes en Honduras

3.1.1 Indicadores educativos y logro de los aprendizajes

El bajo rendimiento escolar o académico, es el resultado de múltiples factores y causas, entre los que se encuentran: los de carácter individual con relación a su estado de salud, el contexto familiar, la situación socioeconómica y otros asociados al propio sistema educativo (Guerrero, Lucero , Segura, & Tovar Cuevas, 2013,p.656). Durante la última década en Honduras se ha desarrollado evaluación externa de los aprendizajes en los espacios curriculares de español y matemáticas en los tres ciclos de educación básica (1° a 9° grado), lo que ha traído consigo disponibilidad periódica de datos sobre los logros en rendimiento académico que se han alcanzado. Los resultados muestran que ha habido mejoras en Español, no obstante en matemáticas, los desafíos se agrandan en los grados superiores de 7°, 8° y 9° grado (FEREMA, 2017). En el marco de estos resultados confluyen indicadores educativos como la cobertura, eficiencia, calidad, equidad, gestión y financiamiento, todo lo anterior es determinante en los centros educativos donde se produce todo el desarrollo intelectual, académico y técnico para lograr el aprendizaje (SE & MIDEH, 2017).

Bajo esta perspectiva, un indicador básico de eficacia del sistema educativo es que los niños, niñas y jóvenes asistan regularmente a clases en la edad oportuna, para medir este indicador se utiliza la tasa de cobertura neta y bruta. En Honduras los datos muestran avances modestos, sin embargo, en los niveles de tercer ciclo de educación básica y educación media existen brechas importantes. en la tabla 3, se muestra los porcentajes alcanzados en cobertura hasta el año 2018.

Tabla 3 Tasa de cobertura neta y bruta por nivel educativo en Honduras (2018)

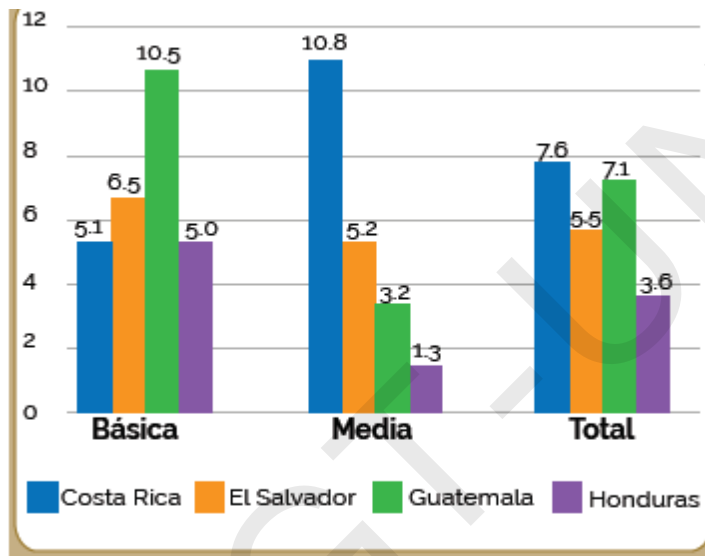
Nivel Educativo	2015 (%)	2016 (%)	2017 (%)	2018 (%)
Cobertura Bruta 3-5 años (Educ. Pre-Básica)	39.5	40.7	42.8	44.3
Cobertura Bruta 6-11 años (Educ. Básica. I-II Ciclo o 1ero-6to Grado)	97.9	96.9	95.3	94.1

Cobertura Bruta 12-14 años (Educ. Básica. III Ciclo o 7mo-9no Grado)	68.8	68.6	68.4	68.2
Cobertura Bruta 15-17 años (Educ. Media)	42.8	44.3	44.6	45.7

Fuente: USINIEH (2018)

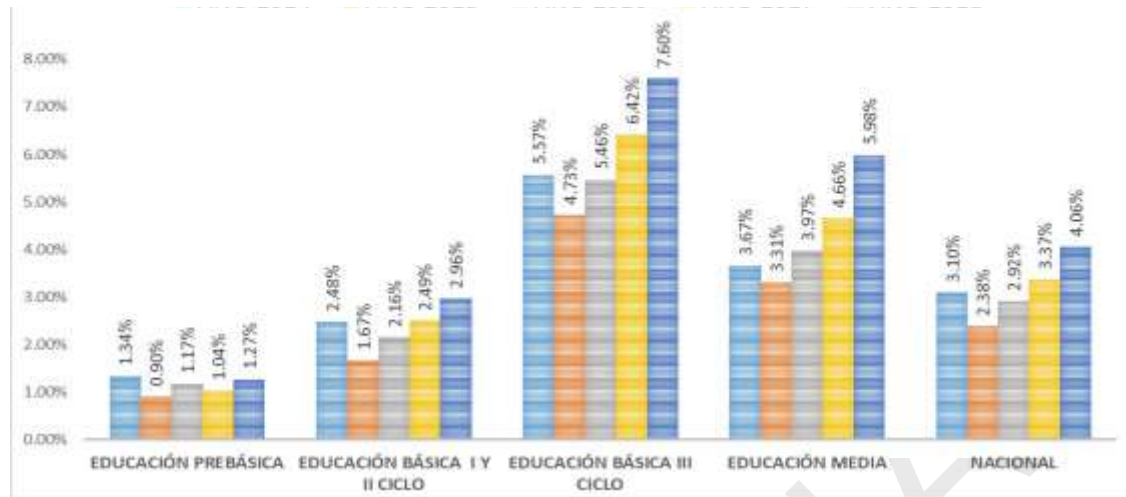
Otro indicador importante asociado al logro de los aprendizajes, es el correspondiente a la repitencia y la deserción en los estudiantes. En Honduras, frente a los países de Centroamérica, el índice de repitencia presenta la tasa más baja. El grafico 1, permite apreciar con mayor precisión.

Grafico 1:Niveles de desercion y repitencia en Honduras (2016)



Fuente: Informe de rendimiento académico (2017), tomado de Quinto informe de Estado de la región de Centroamérica (2016). p.363

La deserción escolar por su parte, en la cual inciden a su vez diferentes factores, es un indicador que forma parte de las dimensiones básicas del área de eficiencia del sistema educativo, al igual que en el indicador de repitencia, la tasa de deserción en Honduras es baja. El grafico 2, compara un periodo de cinco años, en los cuales en el resultado nacional no se supera el 5%.

Grafico 2 Porcentaje de desercion escolar en Honduras (2014-2018)

Fuente: PEI 2018-2022

3.1.2 Elementos determinantes en el sistema de evaluación nacional de Honduras

Conocer el nivel de desarrollo de las habilidades adquiridas en los campos del conocimiento de matemáticas y de comprensión lectora es cada vez más indispensable, 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 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 (Benavidez Ormazza, 2010).

En Honduras, el alineamiento que existe entre el currículo nacional, libros de texto y otros materiales educativos ha hecho posible el desarrollo e implementación de un sistema de evaluación basado en estándares de clase mundial. Los materiales de evaluación sirven para medir el nivel académico estudiantil al inicio del año escolar (pruebas diagnósticas), mensualmente (pruebas formativas) y al finalizar el año escolar (pruebas sumativas).

Acciones como la determinación de estándares de desempeño de forma conceptual (descriptores de niveles de desempeño) y de forma operativa (puntos de corte); se han presentado reportes de los resultados de las pruebas a nivel nacional, regional y escolar. Estos resultados se presentaron en función de la cantidad de estudiantes en cada nivel de desempeño. Se observó cierto nivel de resistencia respecto al cambio de reportar porcentaje

correcto, pero con actividades de promoción en los medios y mediante talleres desarrollados a nivel nacional existe una buena perspectiva que el sistema es aceptado (Lansdale, Vukmirovic , & Estrada, 2012).

En el año 2016 se aplicaron por quinto año consecutivo pruebas para estudiantes de primero a noveno grado en las áreas curriculares de Matemáticas y español a una muestra de centros educativos que ofertan educación básica a nivel nacional. Este tipo de muestra tiene representatividad para todo el país (424 centros que ofertan primero y segundo ciclo de educación básica y 225 que ofertan el tercer ciclo, para un total de 649 centros y 80,357 estudiantes de los 18 departamentos de la división político administrativa hondureña. (Secretaría de Educación & MIDEH, 2017).

El tipo de pruebas aplicadas y el nivel de muestra seleccionada permiten presentar resultados para cada Bloque y Componente del contenido curricular de Matemáticas y de español para cada grado, en diferentes niveles de análisis: Nacional, centro educativo, grado y sección. Sin embargo, el presente informe enfatiza que esta valiosa información por sí misma no genera mejoras en el sistema educativo. Es una condición necesaria pero no suficiente. Solamente si se elaboran e implementan los Planes de Mejoras basados en los resultados obtenidos en cada nivel de gestión (direcciones departamentales, municipales, centros educativos y aulas), se puede esperar un impacto de la evaluación en la mejora de los aprendizajes.

En español los resultados comparativos por departamento indican que Ocotepeque, Copán, Valle, Atlántida y Comayagua, mostraron los mejores niveles de aprendizaje en el año 2015. Con excepción de Valle, los otros cuatro departamentos ya aparecían como los de mejores resultados en la evaluación del año 2010.El año 2015 marcó una fecha importante para la educación hondureña, en tanto señaló el final del período de la iniciativa Plan EFA-FTI Honduras 2003-2015. En este marco general puede plantearse que la educación hondureña ha venido avanzando en la última década, con cambios importantes en el último lustro, pero mantiene pendientes muchas tareas importantes aún. En este sentido se ha señalado que “El sistema educativo hondureño no está brindando una educación de calidad que promueva crecimiento económico a largo plazo, apoye a diversos sectores de la población a salir de la pobreza y a la vez, reduzca las importantes inequidades existentes” (Luque, 2014: 37). De manera que, pese a los significativos avances de los últimos años, los

principales indicadores educativos del país siguen estando entre los más atrasados de la región.

Sin embargo, varios factores tales como una década de amplio apoyo de la cooperación internacional a la educación hondureña y una significativa mejora en la gobernabilidad del sistema en los cuatro últimos años (pasando desde poco más de 100 días de clases en el año 2003, hasta 200 días en la actualidad), parecen haber tenido algún efecto en relación con los niveles de aprendizaje medidos por pruebas estandarizadas. Para el año 2014, los resultados nacionales de evaluación externa estandarizada promediaron 70% en Español y 57% en Matemáticas (SE, 2015, p. 45-47). Esta mejora fue refrendada a finales de ese mismo año por los resultados del Tercer Estudio Regional Comparativo y Explicativo TERCE de la UNESCO (que aplicó las pruebas en el año 2013), en el cual Honduras obtuvo el 10º lugar de quince países de Latinoamérica en español y Matemáticas de tercer grado y los lugares 12º y 11º en sexto grado, los mejores resultados en pruebas estandarizadas internacionales hasta la fecha (UNESCO, 2014).

3.2 Experiencia de la evaluación de los aprendizajes en Honduras

3.2.1 Evaluación de los aprendizajes en el nivel educativo de educación básica

En Honduras han existido varios esfuerzos por evaluar la calidad de la educación, y en particular la evaluación de los aprendizajes, aspecto fundamental cuando se evalúa la calidad educativa. Uno de los esfuerzos pioneros estuvo enmarcado en el Proyecto de Eficiencia de la Educación Primaria (PEEP) durante el período 1986-1995, el cual pretendía desde una perspectiva integral, mejorar el costo-efectividad, la eficiencia interna y la calidad de la educación del sistema de escuelas primarias públicas.

Uno de los componentes de este proyecto era el de evaluación del aprendizaje, el cual implicaba una innovación significativa en el sistema educativo hondureño, ya que incluía la elaboración y aplicación de pruebas estandarizadas para evaluar el rendimiento escolar de los alumnos a nivel nacional. (SE, 2009). Este fue el primer gran esfuerzo nacional para la evaluación del rendimiento académico a través de pruebas estandarizadas, ya que durante 1990-1994 se estuvieron aplicando las pruebas estandarizadas para los primeros cinco grados en las cuatro asignaturas básicas (Matemáticas, Español, Ciencias Naturales y Estudios Sociales) y evaluando las escuelas de 10 municipios.

Unidad de Medición de la Calidad Educativa (UMCE): Una vez finalizado el Proyecto PEEP, se creó la Unidad de Medición de la Calidad de la Educación (UMCE), financiada por el Banco Mundial y administrada técnicamente por la Universidad Pedagógica Nacional Francisco Morazán. A partir del año 1997 y hasta el año 2004 se aplicaron pruebas estandarizadas de Español y Matemáticas (en el 2004 también en Ciencias Naturales), en una muestra de escuelas a nivel nacional de los 18 departamentos (con énfasis para el 3er y 6to grado), de manera paralela se elaboraron estudios sobre los Factores Asociados al rendimiento académico. Estas pruebas se realizaron teniendo como referencia la estructura curricular plasmada en los “Rendimientos Básicos” e Indicadores de Evaluación establecidos para los seis grados en las cuatro asignaturas básicas, elaborados previamente en el PEEP. (SE, 2009).

Según la UMCE, en el año 2002, se suscribe un convenio entre la Secretaría de Educación y la Universidad Pedagógica Nacional Francisco Morazán, donde la UMCE tenía como misión principal realizar mediciones del rendimiento académico cada dos años, a una muestra representativa de alumnos de Tercero y Sexto Grados en las asignaturas de Español, Matemáticas y Ciencias Naturales y la investigación de los factores asociados (Banco Mundial, 2001). En el año 2007, la UMCE pasa a convertirse en una Unidad Técnica del Instituto de Investigación y Evaluación Educativas y Sociales (INIEES), creado por la UPNFM en ese mismo año. De igual forma, se firmó el Convenio Interinstitucional entre la SE y la UPNFM, donde se establece que la UMCE es la unidad responsable de realizar la Evaluación Externa a los Centros Educativos de todo el país.

En febrero del año 2010, en Honduras se formula la visión de país 2010-2038, a través del Decreto 286-2009, publicado en La Gaceta N° 32129 y que contempla la Ley para el establecimiento de una Visión de País y la adopción de un plan de Nación para Honduras. En el plan de Nación 2010-2022 en el lineamiento estratégico No.4, se deja contemplado la: Educación como medio de emancipación social, entre los indicadores para este lineamiento se encuentran cuatro, cuyo seguimiento corresponde a la UMCE-UPNFM. (UMCE, 2017).

Proyecto Mejorando el Impacto al Desempeño Estudiantil Hondureño (MIDEH): a partir de finales del año 2004, hasta el 2017, la evaluación de los aprendizajes es impulsada por la Secretaría de Educación con el apoyo de la cooperación externa a través del proyecto Mejorando el Impacto al Desempeño Estudiantil de Honduras, este proyecto ha contribuido

al diseño de estándares nacionales educativos de matemáticas y español, además de programaciones mensuales en matemáticas y español, y un sistema amplio de evaluación formativa y sumativa a través de la elaboración de Pruebas Formativas Mensuales y las Pruebas de Fin de Grado (SE, 2009).

Durante la última década los resultados de la evaluación externa de los aprendizajes en Español y Matemáticas habían estado estancados en valores muy bajos cercanos al 40%, como promedio nacional, sin embargo, la introducción de un nuevo currículo de educación básica, a partir del 2004, mediante un sistema de materiales educativos alineados a dicho currículo y que sirven de apoyo a la labor docente en el aula (incluyendo Libros de Texto, Guías didácticas para el docente, Estándares Educativos, Programaciones Educativas Mensuales, Estándares de Desempeño, Pruebas Formativas Mensuales y Cuadernos de Trabajo para alumnos) parece haber cambiado esa tendencia.

El desarrollo e implementación de este sistema de materiales educativos ha sido posible con el apoyo de la cooperación internacional como parte de la iniciativa: Education for All (EFA) que cuenta con apoyo directo de la cooperación canadiense, española, europea, Sueca y Alemana, y de proyectos específicos como PROMETAM de JICA y MIDEH de USAID, por lo que puede considerarse como evidencia de una intervención exitosa en un sistema educativo nacional. (Hernández, 2009). Además de los problemas de cobertura, el registro educativo hondureño manifiesta agudos problemas en torno a los niveles de aprendizaje que alcanzan los estudiantes. La Secretaría de Educación a través de la Unidad de Medición de la Calidad Educativa UMCE ha venido evaluando una muestra nacional de alumnos de escuelas primarias en los últimos diez años y los resultados han sido estables: Media de entre 30 y 40% tanto en Español como en Matemáticas en estudiantes de diferentes grados. El informe del 2005 indica que menos del 20% de los alumnos aprueban los exámenes aplicados a nivel nacional en Español, Matemáticas y Ciencias Naturales. (UPNFM, 2011).

En noviembre del 2015, por cuarto año consecutivo en el país, la Secretaría de Educación por medio de la Dirección General de Evaluación y Currículo (DGCE), coordinó la evaluación estandarizada de fin de grado en las asignaturas de Matemáticas y Español en los tres ciclos de educación básica, de primero a noveno grado. En los tres años anteriores se aplicó a una muestra de representación nacional, pero en el 2015, gracias al generoso apoyo de la cooperación internacional acreditada en el país, se aplicó a una muestra con

representatividad hasta el nivel municipal. (Secretaria Educacion Honduras & MIDEH, 2015).

La calidad de los aprendizajes que se obtienen en los centros educativos sigue siendo el talón de Aquiles del sistema educativo nacional. Aunque se ha invertido mucho en los elementos claves que contribuyen a mejorar la calidad de la educación (formación de maestros, textos y otras ayudas pedagógicas), todavía queda mucho que avanzar en materia de eficiencia de la educación por no hablar de la pertinencia de los aprendizajes. Las pruebas realizadas por la UMCE a lo largo de varios años mostraron la baja calidad de los aprendizajes que generan los procesos educativos del aula.

3.2.2 Evaluación de los aprendizajes en Educación Media: Pruebas Pre-Universitarias

La Secretaría de Educación de Honduras, como parte de los procesos de evaluación estandarizada en el nivel de educación media, aplica la prueba “Pre Universitaria” dirigida a los estudiantes que están por egresar; y es coordinada por la Dirección General de currículo y evaluación, teniendo como fundamento legal para la aplicación de la misma, el acuerdo 0487-SE-2014 del 01 de abril de 2014 el cual establece: “Aplicar dos Evaluaciones externas de implicancia directa a todos los alumnos de último año del nivel de educación media como requisito obligatorio para la graduación y certificación de este nivel; esta evaluación tendrá el nombre de “Preuniversitaria” (Secretaria de Educacion , Informe de Resultados del Rendimiento Académico de los egresados de Educación Media, 2014).

Este mismo informe refiere que la prueba preuniversitaria se aplicó por primera vez, en noviembre del año 2014, teniendo como propósito la obtención de información válida y confiable del rendimiento académico que logran los estudiantes que egresan de educación media, de la zona rural como de la zona urbana, de las instituciones gubernamentales y no gubernamentales; respecto a cuatro áreas consideradas básicas: habilidad lingüística, matemáticas, conocimiento científico en ciencias naturales y ciencias sociales. Los ejercicios que integran la prueba son de “opción múltiple”, en cada uno de ellos se ofrecen cuatro posibilidades de respuesta anteceditas por una letra, entre las cuales se debe seleccionar la respuesta correcta encerrándola en un círculo.

Valorar el resultado de los aprendizajes con que egresan los educandos del sistema educativo nacional, así como el resultado de la adecuación de las prácticas docentes a las

necesidades educativas que nacen del contexto y de lo previsto en los planes y programas que conforman el currículo de educación media, buscan prevenir y disminuir la cantidad de egresados con bajo rendimiento académico (Secretaría de Educación; 2015).

Finalmente, esta información establece a manera de diagnóstico, datos, que describen el nivel de aprendizaje que logran los educandos que egresan del Sistema Educativo Nacional, lo que permite la toma de decisiones informadas y consensuadas por parte de las autoridades del nivel central, departamental y de los centros educativos.

3.2.3 Prueba de Aptitud Académica (PAA) UNAH

En Honduras la Universidad Nacional Autónoma de Honduras, desde el viene aplicando a los futuros estudiantes la Prueba de Aptitud Académica (PAA), realizando hasta este día 32 procesos de admisión, lo que representa un total 364,385 aplicaciones de pruebas donde han participado un total de 295,270 (personas), de ese total 238,579 aspirantes realizaron la PAA solo una vez (1), de los cuales fueron admitidos un total de 194,908 (82%)

Un total de 44,267 aspirantes realizaron la PAA 2 veces siendo admitido 28,466 (64%). Unos 12,424 se sometieron a la PAA 3 veces de los cuales fueron admitidos 4,582 (37%) quedando solamente 7,842 que han realizado la prueba en las 3 oportunidades y en ninguna lograron alcanzar el puntaje mínimo de 700 puntos para ser admitidos. A manera de resumen de los 295,270 estudiantes aspirantes que han realizado la prueba una, dos o tres veces en estos 10 años, un total de 227,956 (77%) han sido admitidos a la UNAH, 59,472 aspirantes, es decir el 20% todavía tienen una o dos oportunidades de volver a realizar la Prueba para ser admitidos y solo 7,842 (3%) no alcanzaron el puntaje mínimo de 700 y ya no pueden volver a realizar la prueba para ser admitidos a la UNAH.

El análisis estadístico de los 10 años de aplicación, muestra que la mayor cantidad de aspirantes que se presentan a las pruebas provienen del sector público, sin embargo, los quienes presentan levemente mejores resultados de las Pruebas de Aptitud Académica son los aspirantes de las instituciones privadas (UNAH & Dirección de Sistema de Admisión, 2016).

3.2.4 Experiencias de evaluación iexterna en Honduras

Estudio Internacional de Progreso en Comprensión Lectora PIRLS / Estudio Internacional de Tendencias en Matemáticas y Ciencias TIMSS

En el año 2011, Honduras decide participar en los estudios PIRLS (Progress in International Reading Literacy Study) y TIMSS (Trends in Mathematics and Science Study) por ser los más consolidados y los de mayor seguimiento a nivel internacional. PIRLS, en español significa “Estudio Internacional de Progreso en Comprensión Lectora”, y evalúa la comprensión lectora en alumnos de cuarto grado de Educación Primaria en un momento importante de su aprendizaje: “En ese momento, los alumnos suelen haber aprendido a leer y ahora están ya leyendo para aprender” (Mullis, I. V. S. et ál., 2009, p. 8).

PIRLS 2011 es el cuarto estudio que la IEA realiza sobre el aprendizaje de la lectura desde 1991, en el que tuvo lugar el primero de ellos (Reading Literacy Study). Ya con el nombre de PIRLS y el actual diseño, el estudio se ha llevado a cabo en 2001, 2006 y 2011, consolidando así una secuencia de ciclos de cinco años.

El estudio TIMSS, “Estudio Internacional de Tendencias en Matemáticas y Ciencias” en español, evalúa el rendimiento en estas dos materias de los alumnos de cuarto grado de Educación Primaria y segundo de Secundaria. Los inicios de TIMSS se remontan a 1964, como prueba solo de matemáticas que, en los años ochenta, continuaría con estudios separados de matemáticas y de ciencias, contabilizados como “segundos estudios”. En 1995, con el nombre de Third International Mathematics and Science Study (TIMSS), se llevó a cabo el primer estudio conjunto de ambas materias y dio comienzo la serie en ciclos de cuatro años (1995, 1999, 2003, 2007 y 2011).

Es importante destacar que en 2011 han coincidido ambos estudios por lo que los países han tenido la ocasión de aplicar los estudios PIRLS Y TIMSS a una misma muestra de alumnos y hacer así análisis comparativos entre los rendimientos del alumnado en estas tres materias.

Revisando los resultados según (IEA, 2011), En la prueba de TIMSS de matemáticas para 4to grado, Honduras obtuvo un puntaje promedio de 396 solo por encima de Yemen que obtuvo 348, ubicándose por debajo de la media en esta escala que es de 500, en matemáticas para 8vo. Grado igualmente el puntaje promedio internacional es de 500, Honduras obtuvo

338 que está por debajo de la media. Para la prueba de Ciencias en 4to. Grado nuestro país obtuvo 432 ubicándose por encima de países como Bostwana y Yemen con 367 y 345 respectivamente y finalmente para 8vo. Grado se obtuvo 369 igualmente ubicándose por debajo de la media internacional.

Cabe aclarar que algunos países, la dirección de PIRLS y TIMSS recomendó aplicar la prueba en sexto grado de Primaria. En el caso de Botswana, Honduras y Sudáfrica para PIRLS y Botswana, Honduras y Yemen para TIMSS, las pruebas se han realizado en sexto curso de Educación Primaria, por considerarlas más apropiadas para este nivel en el caso de los países indicados (Instituto Nacional de evaluación Educativa, 2011).

3.1.4.2 Tercer Estudio Regional Comparativo y Explicativo TERCE

El Tercer Estudio Regional Comparativo y Explicativo (TERCE) tiene entre sus principales objetivos entregar información relevante que contribuya a la mejora la calidad de los sistemas de educación de la región. El TERCE provee a los países de evidencia empírica determinante como un aporte para la puesta en marcha de políticas educativas que permitan que las niñas y niños de estos países puedan tener acceso a educación de calidad. Para ello, el estudio enfatiza su trabajo no solo en la medición de los aprendizajes, sino también en el análisis de los factores que explican diferencias de logro académico entre alumnos y entre escuelas.

El TERCE evalúa logros de aprendizaje en las disciplinas de lenguaje (lectura y escritura) y matemática en tercer y sexto grados de educación primaria y, además, ciencias naturales en sexto grado. En este tercer estudio participaron 16 sistemas educativos, entre ellos, 15 países más un estado sub-nacional, como ser: Argentina, Brasil, Chile, Colombia, Costa Rica, Ecuador, Guatemala, Honduras, México, Nicaragua, Panamá, Paraguay, Perú, República Dominicana, Uruguay, y Estado de Nuevo León (México).

Los resultados y las recomendaciones del estudio de factores asociados del TERCE se organizaron en tres categorías:

- a) Características de los estudiantes y sus familias; las evidencias muestran que el desempeño académico de los estudiantes está directamente influenciado por sus antecedentes escolares, las prácticas educativas en el hogar y las características socioeconómicas, demográficas y culturales de sus familias.

b) Características del docente, prácticas pedagógicas y recursos en el aula; los hallazgos demuestran que el docente y las prácticas en el aula son unas de las principales variables que afectan el rendimiento escolar. Entre los aspectos importantes están la formación inicial y continua de los docentes, la motivación, las prácticas pedagógicas y de los recursos disponibles en el aula que tienen una relación con los resultados, y

c) Características de las escuelas, ya que es donde se lleva a cabo el proceso de enseñanza aprendizaje. Entre los principales factores escolares que se asocian al aprendizaje se encuentra: La desigualdad en los resultados académicos entre escuelas y al interior de estas. Las desigualdades socioeconómicas explican la mayor parte de las brechas de aprendizaje entre escuelas, pero no las diferencias de logro al interior de cada establecimiento educativo. Otro factor que incide es la violencia en el entorno de la escuela afecta negativamente los aprendizajes. Recursos escolares, el estudio señala que los recursos materiales de las escuelas (infraestructura, instalaciones y servicios) influyen en los resultados académicos de los alumnos y por último el cumplimiento de la jornada escolar completa es otra variable que pertenece a la esfera de los recursos de las escuelas. (Organización de la Naciones Unidas para la Educación, la Ciencia y la Cultura., 2015).

3.3. Ámbitos del nivel de educación media

3.3.1 Antecedentes y conceptualización del nivel medio.

En el acápite anterior relacionado con la estructura de los niveles de educación formal, se dio a conocer a efectos de este estudio nuestro contexto en la educación media mencionando las modalidades que actualmente existen en el sistema. En función de ubicar el objeto de estudio de esta investigación referida al diseño de una prueba para la evaluación de logros obtenidos en el área de Español en el décimo grado de educación media, es necesario desarrollar brevemente una cronología y conceptualización del mismo.

El Currículo Nacional Básico (2004), define la modalidad científico humanista como la oferta educativa que brindara los conocimientos básicos para la continuación de los estudios en el nivel Superior, en ese momento, esa modalidad era llamada Bachillerato en Ciencias y Letras, pero tras el proceso de la segunda reforma educativa mediante la Ley Fundamental de Educación se estructuró en el año 2012 un nuevo diseño curricular de la Educación Media, teniendo en cuenta la rápida evolución de la ciencia y la tecnología y de las condiciones cambiantes del mundo y sus repercusiones en la realidad nacional, así mismo, las exigencias de la vida democrática y ciudadana, razón por la cual en el año 2014, después de múltiples propuestas y revisiones por parte del equipo técnico de la Secretaría de Educación mediante acuerdo N° 0112-SE-2014, el congreso nacional aprueba en forma permanente el funcionamiento de los planes de estudio de la carrera en Bachillerato en Ciencias y Humanidades, la cual tiene una duración de dos años para los centros gubernamentales y no gubernamentales, aboliendo el funcionamiento del Bachillerato en Ciencias y Letras (Gaceta, 2014).

Ampliando el tema de los antecedentes, se puede mencionar que las iniciativas que se relacionan con el actual Bachillerato en Ciencias y Humanidades, se remontan a la década de los 60, cuando la educación media se estructuraba en dos ciclos: Ciclo común de Cultura general que comprendía de primero a tercer curso y el ciclo diversificado, el cual tenía una duración de dos años para los llamados anteriormente Bachilleratos en Ciencias y Letras y tres años para el diversificado de carreras como Perito Mercantil, Secretaria, Bachiller en administración de empresas etc. Posteriormente un convenio entre Honduras y un consorcio de universidades del Estado de Florida Estados Unidos, trajo consigo acciones de reforma a

la Educación Media, ejecutadas a través de la Secretaria de Educación, lamentablemente el intento no logro grandes avances en el asunto.

Es a partir de 1990 se promueven otras iniciativas que aportan lineamientos teóricos y metodológicos que son tomados para llevar a cabo un Curriculum del Bachillerato en Ciencias y Humanidades, derogando al Bachillerato en Ciencias y Letras vigente, entre estos el hecho que la Educación Media tiene como fundamento la Educación Pre básica y la Educación Básica con una duración entre ambos niveles de al menos diez años (10), un año para la primera y nueve para la segunda, además que a través del currículo del Bachillerato en Ciencias y Humanidades se comparte el planteamiento de que durante estos diez años se logre la formación integral del educando para que desarrolle diversas capacidades cognitivas, sociales y afectivas, orientada a una cultura general que lo capacite par el ingreso a la Universidad y la inserción al mercado laboral.

De acuerdo con (Solís M. A., 2004) el Bachillerato Científico Humanista es básicamente similar al Bachillerato Técnico Profesional ya que ambos se relacionan con el enfoque basado en competencias, la diferencia radica en la metodología y contenido para la construcción de las mismas. En ese sentido las competencias científico humanistas se encuentran influenciadas por las relaciones de la vida universitaria y el mundo de la sociedad. En las interacciones de la vida universitaria a los estudiantes se les brindan las competencias básicas indispensables para el aprovechamiento de la educación superior o universitaria. El mundo de la sociedad ayuda a tener un panorama general de la vida laboral, profesional y de convivencia en sociedad.

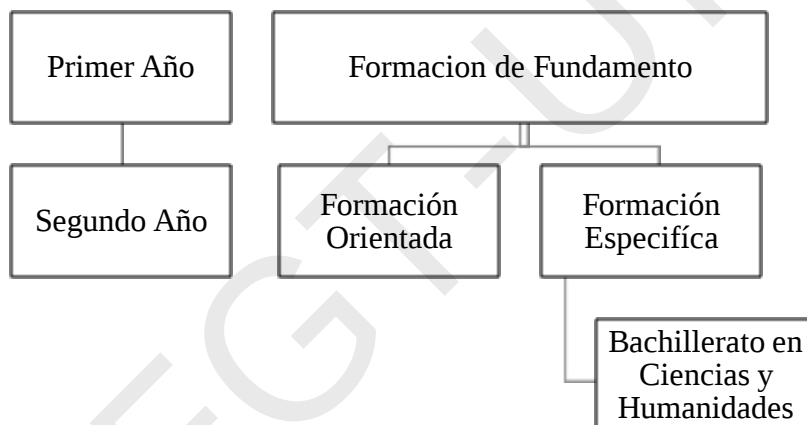
3.3.2 Estructura curricular del undécimo grado.

Desde el punto de vista curricular, la educación media integra un cuerpo de campos de conocimiento que constituyen una base común con el propósito de preparar a los estudiantes para la transformación del país y formar ciudadanos responsables y productivos. La curricular de este nivel corresponde a la formulación de aspectos estructurales y de organización del currículo en consonancia a las exigencias de la situación educativa, de la realidad que atraviesa el país, del contexto global, de las innovaciones en materia de tecnología aplicada a los campos del conocimiento y específicamente del sistema de educación media, en el cual implementa este Bachillerato en el conocimiento científico y de las Humanidades.

La Formación de fundamento: proporciona un conjunto de competencias básicas que son comunes a las modalidades de Bachillerato Técnico Profesional y Bachillerato en Ciencias y Humanidades, nucleadas alrededor de áreas y espacios curriculares de carácter obligatorio. La educación es considerada un asunto de interés social y es por ello que la incorporación de veedores de diferentes sectores demanda cada vez una rendición de los procesos educativos cada vez más transparentes

Para ejemplo de la estructura curricular en la modalidad del Bachillerato en Ciencias y Humanidades, se presenta el *cuadro 1* que detalla bloques de formación, áreas curriculares y sus correspondientes campos del conocimiento, intencionadamente organizados en un plan de estudios capaz de asegurar a los estudiantes una sólida formación en los campos científico, tecnológico y social, capacitándolo para continuar estudios de nivel superior y desenvolverse en el ambiente laboral. (Secretaría de Educación, 2013, págs. 13-14).

Ilustración 5 Estructura curricular del Bachillerato en Ciencias y Humanidades



Fuente: Planes y programas de estudio primer año, bachillerato en ciencias y humanidades (2014)

Formación de fundamento: proporciona un conjunto de competencias básicas que son comunes a las modalidades de Bachillerato Técnico Profesional y Bachillerato en Ciencias y Humanidades, nucleadas alrededor de áreas y espacios curriculares de carácter obligatorio. El décimo grado que es el grado a evaluar en este estudio y es también llamada “formación de fundamento”.

Formación orientada: constituida por aquellas competencias de espacios curriculares que guían hacia la formación específica en los campos laborales y académicos. Tanto en el Bachillerato Técnico Profesional como en el Bachillerato en Ciencias y Humanidades

cumplen con la doble función de profundizar en las competencias fundamentales que se iniciaron en el primer año de estudios de la modalidad, así como la de aportar aquellas competencias orientativas correspondientes a cada ámbito (Científico / Humanista) de estudios.

Formación específica: aporta las competencias propias que habilitan al egresado para continuar estudios en el nivel superior. Las competencias correspondientes a este bloque, por tanto, deben responder a requerimientos académicos establecidos por las instituciones de Educación Superior. (Secretaría de Educación, 2013, págs. 13-14)

3.3.3 Descripción y propósitos generales del campo curricular español en el 11° grado del Bachillerato en Ciencias y Humanidades.

El área de comunicación orienta el desarrollo del pensamiento razonador y crítico, comprende el lenguaje oral y escrito, la lectura, la conciencia lingüística y la expresión literaria y artística; además, integra la interpretación artística del lenguaje estructurado como factor de la transmisión cultural de valores y de recreación del espíritu como producto de un bien cultural. Integra el aprendizaje y conocimiento de un idioma extranjero, desarrolla las habilidades y destrezas de hablar, escuchar, escribir y leer para interactuar con libertad en un mundo globalizado (Secretaría de educación, 2014).

En el Bachillerato en Ciencias y Humanidades el desarrollo dichas competencias se encuentran plasmadas en el área de comunicación, a través de los espacios curriculares de lengua y literatura en el primer semestre, y lenguaje-pensamiento crítico en el segundo semestre. El lenguaje juega un importante papel en la vida de los educandos, ya que es un valioso medio en el desarrollo integral, tanto en sí mismo como por ser puerta de entrada, en el conocimiento científico y tecnológico, a la cultura del trabajo y a la vida productiva, pues está íntimamente ligado al carácter esencial de la persona humana y de la sociedad y al rápido aprendizaje en las condiciones de constantes cambios (Secretaría de educación, 2014).

Objetivos del área de comunicación: procura el desarrollo de las capacidades de comprensión y expresión de todos los mensajes que hacen posible el intercambio comunicativo entre las personas. Su meta es posibilitar la adquisición de las competencias comunicativas de los educandos, de su capacidad de conocer y producir enunciados adecuados en diferentes contextos educativos. Los campos del conocimiento asociados en

esta área son: literatura general y científica, la lengua en el contexto mundial, la semiología, lingüística, filología y lengua extranjera (Secretaría de educación, 2014).

3.3.4 Competencias cognitivas y estándares de contenido del área curricular español del 11° grado del BCH.

El espacio curricular Lengua y Literatura está desarrollado bajo el enfoque comunicativo organizado en base a competencias que los estudiantes de Bachillerato en Ciencias y Humanidades deben evidenciar durante el primer semestre. El programa profundiza en el análisis de textos propios de la orientación académica, interacciones orales o escritas, como condición previa para el logro de las competencias. El programa sigue pautas del constructivismo centrándose en las funciones, saberes, y estrategias comunicativas que garantizan la formación lingüística y pragmática del estudiante.

Las competencias generales que pretende lograr el programa de Lengua y Literatura para el 11° grado de BCH, son las siguientes:

- Aplicar los fundamentos de la comunicación en el ámbito universitario y personal, en base a las demandas de la sociedad del conocimiento.
- Producir textos escritos propios de su área profesional para establecer mejores niveles de comunicación.
- Interpretar textos escritos que correspondan a su quehacer académico y personal.
- Interpretar mensajes orales que contribuyan a la eficiencia en su área de desempeño.

En el segundo semestre, el plan de estudios y programas curriculares para el Bachillerato en Ciencias y Humanidades, contempla la asignatura de Lenguaje y Pensamiento Crítico, orientado a propiciar el conocimiento de las habilidades de lectura, redacción y expresión oral, además pretende desarrollar las habilidades de búsqueda y selección de información relevante, conocer las características y lineamientos de un reporte de lectura, ejercitar su elaboración, identificando ideas principales, secundarias, globales e inferenciales.

El propósito general de este espacio curricular es lograr que el estudiante adquiera el pensamiento crítico a través del mejoramiento de la escritura, lectura, expresión oral, estructura y funciones de la lengua desde una perspectiva científico literario. Las competencias generales que pretende lograr son:

- Desarrollar la capacidad para la comunicación efectiva, escrita, y oral para el crecimiento personal y la interacción social.
- Desarrollar el pensamiento sistemático analítico, dialógico y crítico mediante el perfeccionamiento de la comprensión y análisis de la lectura, la redacción y la comunicación oral.
- Propiciar la toma de conciencia acerca de la importancia de la lengua en la interpretación de la realidad, la intervención en el ámbito personal y profesional.
- Fortalecer la meta cognición y la autorregulación a través de la metodología del espacio curricular con el fin de reforzar la motivación intrínseca y extrínseca para favorecer la autonomía de los aprendizajes.
- Desarrollar la conciencia ética a partir de las actividades y las relaciones de comunicación.
- Propiciar la transferencia de las competencias del lenguaje y pensamientos a otros ámbitos de la vida.

CAPÍTULO 4. MARCO METODOLÓGICO

En este apartado se detalla la metodología que sustenta el diseño de la prueba para la evaluación del rendimiento académico del área español de los estudiantes del undécimo grado del Bachillerato en Ciencias y Humanidades del departamento de Francisco Morazán., finaliza el capítulo definiendo de forma pormenorizada los parámetros y procedimientos aplicados para obtener los resultados de rendimiento académico.

4.1 Enfoque o tipo de estudio

Este estudio ha sido desarrollado bajo un enfoque metodológico cuantitativo, el cual supone un conjunto de procesos secuenciales y probatorios a través de la recolección de los datos que se fundamentan en la medición de variables o conceptos comprendidos en las hipótesis con procedimientos e instrumentos estandarizados y aceptados por una comunidad científica (Hernández Sampieri, Fernandez Collado, & Baptista Lucio, 2014).

El alcance de la investigación es descriptivo teniendo como principal objetivo especificar las propiedades psicométricas de los reactivos y del instrumento en sí mismo; el cual, una vez desarrollado el proceso de la administración, consecuentemente permitió categorizar el

desempeño académico de los estudiantes y compararlo según el sexo y tipo de administración de los centros educativos participantes.

El diseño es transversal conducente a un solo momento de recolección de datos sobre la base de un marco conceptual y operativo de la prueba; para luego analizar minuciosamente los resultados, a fin de extraer generalizaciones significativas que contribuyan a describir la variable de rendimiento académico con la aplicación de modelos estadísticos, y métricas para la medición del desempeño académico.

Así mismo, la finalidad de este estudio es de carácter práctico, permitiendo establecer en este sentido un conocimiento acerca de lo que han aprendido los estudiantes, cuanto han aprendido y las diferencias de los grupos respecto al aprendizaje.

4.2 Definición de las variables e hipótesis

4.2.1 Variable principal del estudio

Siguiendo el paradigma de investigación de este estudio, las variables que se utilizaron son cuantitativas-continuas, definiendo como variable principal de este estudio el rendimiento académico de los estudiantes de Undécimo grado del Bachillerato en Ciencias y Humanidades, en el área curricular de español, dicha variable, se han subdividido dimensiones y sub dimensiones conformadas por los bloques y componentes curriculares a evaluar; los indicadores se convierten en los indicadores. Ver tabla 4.

Tabla 4 Operacionalización de la Variable de estudio rendimiento académico de los estudiantes de 11mo grado del BCH en el área español

Variable	Definición conceptual	Definición Operacional	Dimensiones	Sub dimensiones	Indicadores
Rendimiento Académico en el área español de los estudiantes de Undécimo grado del Bachillerato en Ciencias y Humanidades	El Nivel de conocimientos demostrado en un área o materia comparado con la norma de edad y nivel académico	Medición del rendimiento académico través de la administración de una prueba estandarizada	Lectura	Tipos de texto	Leen y utilizan con diversos propósitos textos narrativos, descriptivos, expositivos y persuasivos, tanto literarios como funcionales e informativos, de varias fuentes como libros, medios masivos de comunicación e internet, incluyendo contenido lingüístico y gráfico. Demuestran comprensión de las ideas globales, principales, secundarias e inferenciales de un texto leído, incluyendo la interpretación de gráficos e iconos

Vocabulario	Identifican el contexto, la intención comunicativa, el punto de vista y el estilo de los textos narrativos, descriptivos, expositivos y persuasivos Interpretan información explícita e implícita de textos diversos para formular planteamientos en sentido crítico. Analizan diferentes puntos de vista y evalúan rasgos científicos, sociológicos y culturales de los textos. Comprenden, aprovechan y evalúan críticamente imágenes y textos, iconos verbales y de informática. Reconocen y entienden en textos leídos una variedad de palabras conocidas. Interpretan palabras desconocidas utilizando diversas estrategias. Interpretan en textos leídos el significado de términos técnicos, lenguaje no-estándar y palabras con múltiples significados. Reconocen e interpretan el lenguaje figurado, la intencionalidad del autor y otros aspectos semánticos de las palabras.
-------------	--

Fuente: Elaboración propia. Adaptado del Currículo Nacional Básico (2004).

4.2.2 Hipótesis

Tras el recorrido teórico realizado en este estudio, se ha planteado que en el análisis del rendimiento académico a nivel nacional e internacional surgen diferencias respecto al aprendizaje de los estudiantes, reflejados en los resultados. La comparación es uno de los procedimientos más utilizados en la investigación evaluativa, si bien los datos obtenidos no son absolutos por sí mismos, se vuelve necesario comparar entre diferentes categorías de colectivos y poder interpretar mejor los resultados de la evaluación estandarizada (Secretaría de Educación & MIDEH, 2017, pág. 54).

Un análisis que es habitual en una evaluación del rendimiento académico, es la comparación de los resultados de los estudiantes a partir del tipo de administración de los centros educativos de los cuales provienen. En la experiencia nacional e internacional se ha reflejado que los resultados de los estudiantes procedentes de centros educativos no gubernamentales son superiores a los obtenidos por los estudiantes de centros educativos oficiales, incluso en un reciente estudio internacional, Honduras presenta la brecha más amplia entre escuelas públicas y privadas de la región, donde el 91% de los estudiantes evaluados de centros educativos oficiales presentaron bajos desempeños, comparado con el 69% de centros educativos no gubernamentales (OCDE & PISA-D, 2018).

Así mismo al hacer comparaciones a partir del sexo de los estudiantes se encuentra que los niveles de aprendizaje de las mujeres en el campo curricular de lecto escritura es superior al de los hombres, marcando como tendencia que las niñas una vez que logran el acceso a la escuela y a la oportunidad de aprender, permanecen y culminan con mayor éxito sus estudios (UNESCO UIS, 2017).

Finalmente, como se ha señalado en apartados anteriores, y tomando en cuenta el principal objetivo de esta investigación referido al diseño de un instrumento para la evaluación del rendimiento académico, es importante acentuar que las cualidades psicométricas de validez y confiabilidad configuran los atributos fundamentales para el diseño, administración, interpretación y uso de los resultados de las pruebas estandarizadas (Tristan López & Pedraza Corpus, 2017). Desde las perspectivas anteriores, las hipótesis de este estudio se delimitan en el marco del objetivo 4, referente a la descripción del desempeño de los estudiantes, estableciéndose de la siguiente forma:

- Los estudiantes de sexo femenino obtienen un mayor puntaje total de la prueba respecto a los estudiantes de sexo masculino.
- Los estudiantes de institutos no gubernamentales obtienen un mayor puntaje de respuestas correctas en la prueba de español, en relación a los estudiantes de instituciones oficiales.

4.3 Población y Muestra

4.3.1 Caracterización de la población

La población objeto de este estudio fueron los estudiantes de educación media de undécimo grado del Bachillerato en Ciencias y Humanidades, que comprenden las edades entre los dieciséis (16) años a los (18) años. Se contempló en el marco muestral los centros educativos oficiales y no gubernamentales que ofertan el Bachillerato en Ciencias y Humanidades del departamento de Francisco Morazán, específicamente del municipio de Distrito central.

Una particularidad de la Educación Media en Honduras, es referente a las debilidades estructurales que se reflejan en el bajo rendimiento de los educandos en aquellos aspectos exigidos por el Nivel de Educación Superior, tal y como se explicó en los apartados anteriores, los egresados de Educación Media, en sus diferentes modalidades reflejan vacíos conceptuales, procedimentales y actitudinales en su formación, que se constituyen en carencias que afectan su desempeño laboral al insertarse en el mercado del trabajo y en lo académico al continuar estudios de educación superior (Secretaria de Educacion , Planes y Programas Bachilleratos en Ciencias y Humanidades , 2013).

En el sistema educativo de Honduras el Bachillerato en Ciencias y Humanidades es el espacio académico que permite a los estudiantes alcanzar la madurez necesaria para la realización de estudios de nivel universitario. En su estructura se complementa la formación en las ciencias de la naturaleza, del hombre y de la sociedad con el estudio de las manifestaciones socio culturales que le faciliten un buen comportamiento ciudadano así como el aprecio de la diversidad presente en la humanidad; formación que una vez adquirida supone que el estudiante estará en condiciones de seleccionar y desempeñarse con éxito en alguna de las carreras que ofrece el nivel de educación superior (Secretaria de Educacion , Planes y Programas Bachilleratos en Ciencias y Humanidades , 2013, p.10).

4.3.2 Diseño de la Muestra

Tipo de Muestreo

Para el desarrollo de esta investigación se utilizaron dos tipos de muestreo, debido a la característica del estudio referido al diseño de un instrumento que requirió de dos etapas de administración de instrumentos, la primera para la aplicación del pilotaje de reactivos y la segunda para la recopilación de datos para el análisis del rendimiento académico a través de la administración de la prueba operativa.

Diseño y tamaño de la muestra en la prueba piloto

Las pruebas previas o pruebas piloto son un elemento esencial del desarrollo de una prueba para la evaluación del rendimiento académico, estas deben ser administradas a alumnos(as) que tengan las mismas características que los que van a participar en la prueba final. La etapa del pilotaje conlleva a un tipo de muestreo no probabilístico, por conveniencia, que permitió seleccionar aquellos casos accesibles que acepten ser incluidos. Esto, fundamentado en la conveniente accesibilidad y proximidad de los sujetos para el investigador (Otzen & Manterola, 2017).

Esta primera etapa se consideró criterios como la accesibilidad de los centros para ser partícipes del estudio y además tomando en cuenta el tipo de administración de estos. Finalmente, para la aplicación de esta prueba se tomó 263 estudiantes de tres centros educativos que incluye dos centros educativos del municipio de Distrito central y un centro educativo del municipio de Talanga.

Tabla 5 Distribución de la muestra en la administración de la prueba piloto

Tipo de administración	Ubicación	Nombre del centro Educativo	F	M	T	N° Secciones
Oficial	Distrito Central	Instituto España Jesús Milla Selva	97	89	186	6
Oficial	Talanga	Instituto Tecnológico Brassavola Digbyana	23	24	47	2
No gubernamental	Distrito Central	Instituto Tecnológico Taular	0	30	30	1
Total			120	143	263	8

Fuente: Elaboración propia.

Diseño y tamaño de la muestra en la prueba operativa

Para la prueba operativa, la selección de los centros educativos se realizó mediante un muestreo probabilístico aleatorio simple, polietapico consistente en dos etapas, en la primera utiliza las unidades de muestreo de mayor tamaño que se llaman unidades de muestreo primarios, PSUs primary sampling units, mientras que en la etapa final se utilizan las unidades de muestreo más pequeño denominadas unidades de muestreo secundarias, SSUs secondary sampling units (Mandeville, 2010). De manera que, dado a que la población muestra que se precisa para este estudio, es de centros educativos de educación media del departamento de Francisco Morazán, en específico del municipio de Distrito central, el cual a su vez está dividido en quince (15) Distrito educativos, se consideraron criterios como ser distritos con centros educativos de educación media gubernamentales y no gubernamentales que ofertaran el Bachillerato en Ciencias y Humanidades.

Considerando los criterios anteriores la técnica utilizada para seleccionar las unidades de análisis del estudio, fue un muestreo aleatorio simple, donde mediante dos etapas se seleccionaron la población final. En la primera etapa: se seleccionó aleatoriamente el distrito educativo que formarían parte de este estudio donde fueron excluidos los distritos educativos, 2, 3 y 12 donde los centros educativos de educación media no ofertan bachillerato en ciencias y humanidades, el distrito seleccionado de forma aleatoria fue el distrito número 9, ubicado en la ciudad de Comayagüela.

En un segundo momento se realizó respectivamente el procedimiento aleatorio simple, en las unidades de muestreo secundarias determinadas a través de las secciones de 11° grado de cada Centro de educación media del distrito educativo n° 9, seguidamente aplicando un método estadístico de aleatoriedad a través del programa Excel, se procedió a la elección de los Centros Educativos de educación Media participantes en una forma aleatoria al azar.

A cada centro educativo se le concedió un número aleatorio determinando que los números aleatorios más altos fueran parte del estudio de manera que todos los centros educativos tenían la misma probabilidad de ser elegidos. Los centros educativos resultantes como parte de la muestra de estudio finalmente fueron tres centros oficiales y tres centros no gubernamentales; los oficiales fueron: CEMG Jesús Aguilar Paz, CEMG Abelardo R. Fortín,

CEMG Mixto Hibueras y los no gubernamentales CEMNG Renacimiento, CEMNG San Francisco y el CEMNG Inmaculada Concepción.

Tabla 6 Distribución de la muestra en la administración de la prueba operativa

Tipo de administración	Ubicación	Nombre del centro Educativo	F	M	T	N° Secciones
Oficial	Distrito Central	CEMG Jesús Aguilar Paz	23	15	38	1
Oficial	Distrito Central	CEMG Mixto Hibueras	138	36	174	6
Oficial	Distrito Central	CEMG Abelardo Fortín	12	9	21	1
No gubernamental	Distrito Central	CEMNG Renacimiento	40	36	76	3
No gubernamental	Distrito Central	CEMNG Inmaculada Concepción	19	14	33	1
No gubernamental	Distrito Central	CEMNG San Francisco	56	43	99	4
	Total		288	153	441	16

Fuente: Elaboración propia

La muestra para el presente estudio lo constituyen 441 estudiantes del undécimo grado del bachillerato en Ciencias y Humanidades, 288 mujeres y 153 hombres de tres centros educativos oficiales y tres no gubernamentales.

4.4 Instrumentos

Para determinar el logro de los aprendizajes de los estudiantes y realizar las comparaciones entre grupos muestrales, se utilizó una prueba estandarizada, que supone una serie de pasos y/o actividades en común, que van desde el diseño de un marco de evaluación, proceso de redacción y revisión de ítems, administración de pruebas de pilotaje, entre otros hasta llegar a la producción de la prueba operativa para la recolección de datos y redacción de informe final (Greaney & Kellaghan, 2016).

En las siguientes líneas se aborda el diseño de los instrumentos utilizados para la recolección de datos dando a conocer en forma general las especificaciones de contenidos y procesos cognitivos que están a la base de los ítems que componen la prueba.

4.4.1 Descripción del diseño de la prueba piloto

Las pruebas previas o pruebas piloto son un elemento esencial del desarrollo de pruebas estandarizadas, estas se administran a estudiantes que tengan las mismas características que los que van a participar en la prueba final (Anderson & Morgan , 2016).

En el diseño del instrumento de pilotaje de los ítems, se estructuró a través de dos sub pruebas, forma “A” y forma “B” con 43 ítems cada una, vinculadas entre sí, ensambladas con 35 ítems únicos y ocho ítems ancla. Esta vinculación permitió determinar la dificultad global de los ítems, sin importar en la sub prueba que fueron presentados.

De forma que, finalmente 86 reactivos fueron llevados al proceso de pilotaje, siendo estos seleccionados atendiendo el marco de evaluación, que corresponde a las especificaciones técnicas o mapa de la prueba, diseñado por docentes especialistas curriculares del área español; dicho mapa o esquema se basa en los documentos que forman parte del Diseño Curricular Nacional Básico (DCNB), que comprende el plan de estudio y programas curriculares, así como las programaciones del undécimo grado del Bachillerato en Ciencias y Humanidades .

Los ítems fueron de tipo selección múltiple, donde en cada reactivo al estudiante se le brindó cuatro posibles opciones de respuesta de las cuales debe seleccionar la correcta. Este tipo de ítems es ampliamente utilizado en pruebas estandarizadas debido a que reportan información del constructo a evaluar en forma efectiva y eficiente, asegurando la validez, confiabilidad y objetividad del instrumento en su totalidad (Downing & Haladyna, 2006).

El formato de los ítems de acuerdo a la temática declarada en el bloque de lectura fue de reactivos dependientes de un contexto, donde se presenta un párrafo o texto, por lo general, de mediana extensión, un afiche, una figura o una tabla de datos, donde a partir de ese contexto se derivan regularmente entre cuatro y ocho ítems de selección múltiple, orientados a recabar el dominio puntual a evaluar (Pardo Adames & Rocha Gaona, 2009).

El bloque que se evaluó fue el de lectura, y los componentes fueron tipología textual y vocabulario; los reactivos comprendieron los niveles de procesamiento cognitivo de conocimiento, comprensión y aplicación, según la taxonomía de Benjamín Bloom (1956), que examina diferentes miradas a este dominio categorizando y ordenando habilidades de pensamiento y objetivos. Los estudiantes indicaron sus respuestas, rellenando un circulo, en

la letra que estimaron correspondiente a la respuesta correcta; donde en los criterios de revisión, la máxima puntuación es: puntuación 1, para la respuesta opción correcta y ninguna puntuación: puntuación 0, para otras respuestas.

4.4.2 Diseño de la prueba operativa

La prueba estandarizada para la evaluación de desempeño académico del espacio curricular español en el undécimo grado del bachillerato en ciencias y humanidades, se estructuró con 50 ítems de selección múltiple. La decisión respecto a la inclusión o exclusión de ítems en la prueba operativa, consideró criterios tales como, si el ítem se ajustaba a las especificaciones técnicas o mapa de la prueba, y considerando las propiedades psicométricas de medición de los ítems.

Los parámetros psicométricos se basaron en los propuestos por la Teoría Clásica de los Test (TCT), de forma que se consideró aquellos ítems, que observaron un P-Value o dificultad clásica en el rango de 0.3 a 9, además que el índice de discriminación(correlación entre el puntaje del ítem y el puntaje total de la prueba) de estos; presentaran valores arriba de 0.3, otro parámetro estimado para la selección de los reactivos fue el análisis de opciones y distribución de puntajes, donde se tomó en cuenta aquellos ítems que estadísticamente sus opciones fueron plausibles, siendo la correlación punto biserial de cada opción correcta, positiva y las opciones incorrectas registraran correlaciones negativas, una última condición para la selección de los ítems de la prueba operativa consistió en la observación del Funcionamiento Diferencial del ítem (DIF) o sesgo, incluyendo aquellos que estaban dentro de los límites aceptables. Esta información se incluyó en una plantilla (ítem card), para facilitar el trabajo de la selección.

La cobertura curricular de la prueba final comprendió diez estándares de desempeño del bloque de lectura, declarados en el campo del conocimiento o asignaturas de lengua y literatura en el primer semestre y la asignatura de lenguaje y pensamiento crítico en el segundo semestre, ambas partes de las programaciones del undécimo grado del Bachillerato en Ciencias y Humanidades.

Los reactivos se construyeron considerando los niveles de procesamiento cognitivo de conocimiento, comprensión y aplicación, donde el 56% de los reactivos fueron de nivel cognitivo de comprensión; ya que los docentes especialistas del área, en el diseño de la tabla

de especificaciones de la prueba; afirmaron que las habilidades que los estándares sugieren que los estudiantes logren en este grado, se relacionan con la aprehensión de habilidades de orden superior que requieren pensamiento sistémico o bien pensamiento crítico por parte de los educandos.

4.5 Procedimiento

Resulta conveniente destacar que las operaciones fundamentales para la elaboración de una prueba estandarizada para la evaluación del rendimiento académico son el diseño de las especificaciones técnicas, el análisis de los reactivos, la determinación de estándares o puntos de corte, y la determinación de la confiabilidad y la validez (Leiva Barajas, 2011). Por consiguiente, a continuación, se describe el detalle de las actividades que resultaron para el diseño de la prueba de español y la producción de sus resultados.

Para el diseño de la prueba estandarizada se establecieron ocho fases con procedimientos básicos y productos que dieran lugar al logro del objetivo principal de esta investigación. Resulta importante mencionar que estas fases se determinaron considerando lineamientos y estándares para instrumentos de evaluación educativa utilizados en nuestro país, que según el Informe Nacional de desempeño académico (2017) estos adoptan una metodología sólida que han demostrado su validez para los propósitos que se persiguen.

La tabla 2. Muestra el proceso para la implementación de este estudio enlistando las principales etapas y productos alcanzados en cada una de estas, en base al modelo de construcción de pruebas criteriosales alineadas al Currículo Nacional Básico.

Tabla 7 Proceso de diseño de la prueba de fin de grado de español

Etapas	Productos	Participantes en la investigación
I. Preparación del marco de evaluación y selección de la población objetivo	-Marco muestral -Marco conceptual de la prueba.	-Investigador
II. Planificación del taller de evaluación educativa y redacción de ítems	-Plan del taller de evaluación educativa. -Convocatoria a docentes especialistas.	- Autoridades secretaria de Educación/DDE Francisco Morazán.
III. Desarrollo del taller de evaluación educativa y redacción de ítems	-Tabla de especificaciones técnicas de la prueba.	-Docentes especialistas del área español.

	-Redacción de 215 ítems del área español 11mo grado BCH.	
	-Mapa de la prueba pilotaje.	
	-Ensamblaje de la prueba de pilotaje.	
	-Revisión del cuadernillo, impresión y codificación de las pruebas pilotos.	-Docentes especialistas del área español.
IV. Pilotaje de los ítems	-Manual de aplicación de la prueba.	-Docentes aplicadores de las pruebas piloto.
	-Capacitación a aplicadores de la prueba	
	-Implementación de la prueba piloto estructurada con 78 reactivos.	-Población escolar seleccionada para el pilotaje.
	-Calificación manual de los ítems.	
	-Análisis psicométrico de los ítems a través del programa SPSS.	
V. Selección de los ítems Prueba operativa	-Análisis de la fiabilidad de la prueba piloto.	-Investigador
	-Elaboradas 78 Ítems Cards (tarjetas de ítems) para la revisión y selección de los reactivos de la prueba operativa.	-Docentes especialistas del área español.
	-Mapa de la prueba operativa.	-Docentes especialistas del área español.
	-Ensamblaje de la prueba operativa.	
	-Revisión del cuadernillo, impresión y codificación del instrumento.	-Estudiantes de la carrera de Psicología UNAH como aplicadores de las pruebas estandarizadas.
VI. Producción de la prueba operativa o definitiva	-Capacitación a aplicadores de la prueba operativa.	
	-Implementación de la prueba operativa estructurada con 50 reactivos.	-Población escolar seleccionada para el estudio de rendimiento académico.
	-Calificación manual de los ítems.	
	-Base de datos Excel depurada.	
	-Análisis de parámetros psicométricos de los ítems.	
VII. Análisis de los resultados de la prueba de español 11mo grado BCH	-Análisis de fiabilidad de la prueba operativa.	Investigador
	-Análisis según métricas de niveles de desempeño, escala de puntuación estandarizada y porcentaje de respuestas correctas.	
VIII. Presentación de los resultados de desempeño académico de los estudiantes de 11mo grado BCH, en el área español.	-Descripción de los resultados de desempeño académico de los centros educativos participantes en el estudio.	Investigador

Fuente: elaboración propia

I. Desarrollo del marco de evaluación y selección de la población objetivo:

El marco de evaluación de la prueba, brinda un esquema general para guiar el desarrollo del instrumento, dicho marco contribuye a generar una idea clara del constructo que se pretende evaluar (Vukmirovic & Fast, Paradigmas y tendencias en medicion y evaluación educativa, 2017).. Este paso tuvo como resultado el documento que corresponde al punto de partida que guía el desarrollo de una prueba de evaluación nacional, se trata del marco conceptual (Test Framework) que establece aspectos generales de la evaluación, como ser los detalles del propósito de la evaluación, la población meta a evaluar, y la tabla de especificaciones técnicas de la prueba.

De lo anterior, además se derivó la construcción del marco muestral de la prueba, detallado en el apartado anterior, es importante señalar que los centros educativos que participaron tanto en la prueba de pilotaje como en la prueba final, se excluyeron al momento de hacer la selección de los docentes especialistas que se encargaron de la construcción, revisión y depuración de los ítems de la prueba.

II. Plan del taller de evaluación educativa y redacción de ítems

Considerando que la calidad de una prueba depende en gran medida de la claridad con la que esta cumple con su objetivo y la precisión con la cual los reactivos coinciden con las especificaciones técnicas trazadas (Anderson & Morgan , 2016); se determinó llevar a cabo un taller de evaluación educativa y redacción de ítems con docentes especialistas curriculares en el área de español, para diseñar y validar la tabla de especificaciones técnicas de la prueba y desarrollar el proceso de redacción de los ítems de selección múltiple, para la construcción del instrumento.

El taller fue desarrollado con la asesoría de la Universidad Nacional Autónoma de Honduras (U.N.A.H.) a través de la Maestría en Psicometría y Evaluación Educativa como una actividad de vinculación Universidad con la Secretaría de Educación, Sub-Secretaria de Asuntos Técnicos Pedagógicos, Dirección General de Currículo y Evaluación, Dirección General de Desarrollo Profesional Docente y Dirección Departamental de Educación de Francisco Morazán.

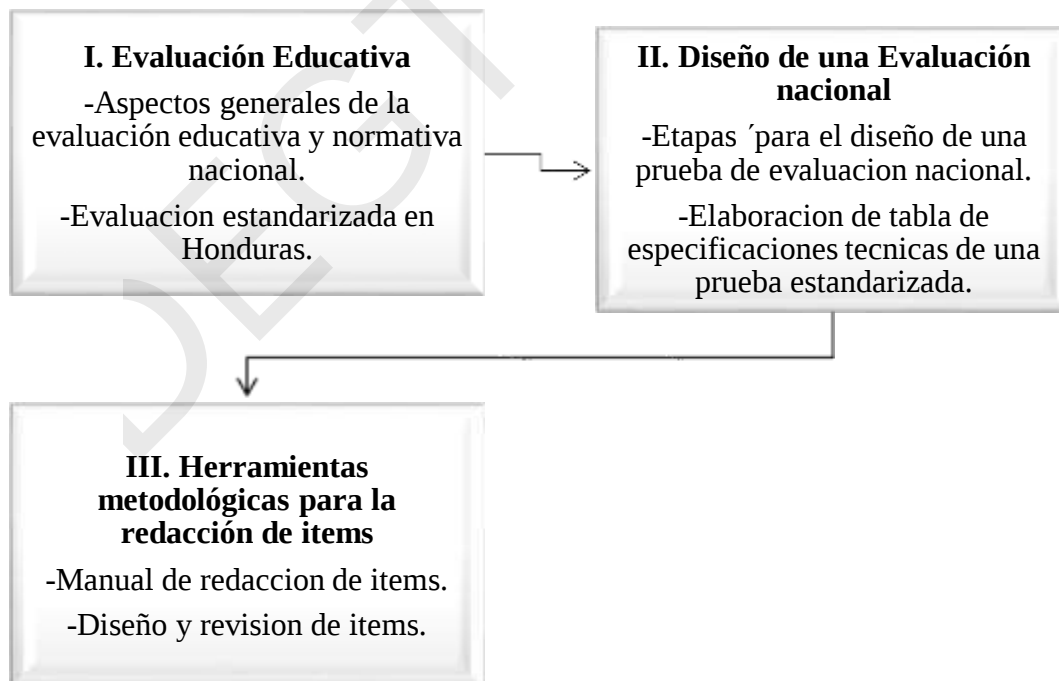
Una vez llevadas a cabo las gestiones correspondientes para la organización del Taller; con la autorización de la Dirección Departamental de Educación se realizó la convocatoria

del equipo de docentes para la redacción de ítems. Los documentos preparados para los docentes convocados al taller, consistió en dos manuales referidos a la Evaluación Educativa y a la redacción de ítems, elaborado por el equipo de pasantes de la Maestría en Psicometría y Evaluación Educativa.

III. Desarrollo del taller de evaluación educativa y redacción de ítems:

El Taller de Evaluación Educativa y redacción de ítems, se llevó a cabo en el lugar que ocupa el Centro Regional de Formación Profesional Centro Su Oriente(CRFPCSO), con la participación de 80 docentes de las especialidades de español, matemáticas, ciencias naturales y ciencias sociales del nivel básico y medio. El contenido del taller desarrolló una secuencia didáctica detallada en un guion metodológico, enmarcado en el fortalecimiento de las capacidades técnicas de los(as) docentes en cuanto a la evaluación de logros de aprendizajes, teniendo especial interés en ofrecer a los educadores principios metodológicos para la construcción de reactivos de alta calidad para la conformación de instrumentos de evaluación de logros de aprendizajes de los diferentes campos del conocimiento. La figura ,sintetiza el contenido desarrollado en el Taller.

Ilustración 6 Tematica taller de evaluación y redacción de ítems



Fuente: Elaboración propia.

Lo anterior condujo a lograr los objetivos principales del taller, referidos al diseño y validación de la tabla de especificaciones técnicas de la prueba y la conformación de un banco de ítems de opción múltiple, para la evaluación del rendimiento académico del área español en el undécimo grado del BCH, aunado a estos objetivos el taller persiguió la finalidad de demostrar a través de fundamentos teóricos y científicos el papel de la evaluación estandarizada como una herramienta potencialmente valiosa para mejorar la calidad de la educación.

La tabla de especificaciones técnicas fue diseñada por el equipo de docentes luego de haber compartido el proceso que la literatura de elaboración de pruebas, señala acerca de esta etapa. Los(as) docentes analizaron la tabla de estándares propuesto en las programaciones del DCNB para las asignaturas de Lengua y literatura I y Lenguaje y Literatura II, las cuales se encuentran articuladas a las competencias que establecen los planes y programas de estudio para el 11° grado de BCH, en los espacios curriculares de Lengua y Literatura y Lenguaje y pensamiento crítico. Lo anterior para llegar al consenso en términos de que debe evaluarse. Cumpliendo la norma para la construcción de ítems de una prueba de rendimiento nacional, se acordó seleccionar los estándares dirigidos a evaluar las habilidades centrales. Debido a las limitaciones de tiempo y recursos, razón por la cual igualmente se estableció que la prueba evaluaría únicamente el bloque de lectura, considerando que para los ítems del bloque de escritura se requiere un método holístico para su revisión que implica la contratación de jueces y la construcción de rubricas para la asignación de puntuaciones.

Finalmente, el diseño del marco operativo de la prueba, determinado por los docentes especialistas determinaron que la extensión esta, constaría de diez estándares (10) del Bloque de Lectura, que incluye los componentes de tipos de texto y vocabulario, con una proporción de cincuenta (50) ítems, que abordaron distintas habilidades del proceso cognitivo, que se detalla en el apartado de resultados. Para el diseño de los ítems, cada docente tenía el manual de diseño de ítems, discutido previamente y una copia de la tabla de especificaciones, que es una extensión del marco operativo que refleja 4 veces el número de ítems necesario para la prueba operativa, de manera que, en total por nivel cognitivo, se diseñaron 24 ítems de conocimiento, 116 de comprensión y 60 de aplicación, que hicieron un total de 200 reactivos.

Los ítems se diseñaron en una plantilla (ver anexo), y la metodología consistió en la distribución a los docentes de forma equitativa de los estándares por componente, se

determinó un tiempo para la redacción de cada reactivo para luego proceder a una plenaria que tenía como propósito revisar la objetividad y el cumplimiento de los criterios que debe cumplir cada ítem, según el manual de redacción, dado a conocer en el taller. En la plantilla de ítems se evaluaron elementos como ser, si estos se encontraban apegados a las especificaciones técnicas, si eran claros, relevantes, si establecían una proposición interesante y justa para los estudiantes, finalmente los docentes especialistas daban sus opiniones acerca de cada reactivo en los momentos de la plenaria y se tomaban como aprobado o rechazado para ser mejorado. La tabla 4 esquematiza con más detalle el número de ítems diseñados.

Tabla 8 Resultados de ítems redactados del área español 11mo grado BCH

Bloque	Componente	Estándar	Ítems Diseñados	Ítems Aceptados	Ítems Rechazados
Lectura	1. Tipos de texto	1.1	20	16	4
		1.2	20	12	8
		1.3	20	14	6
		1.4	20	18	2
		1.5	20	8	12
		1.6	28	13	15
	2. Vocabulario	2.1	12	6	6
		2.2	16	9	7
		2.3	24	12	12
		2.4	20	17	3
			200	125	75

Fuente: Elaboración propia.

Lo anterior condujo a lograr los productos principales del taller, referidos al diseño y validación de la tabla de especificaciones técnicas de la prueba y la conformación de un banco de ítems de opción múltiple, conformado por 24 ítems de conocimiento, 116 de comprensión y 60 de aplicación, que hicieron un total de 200 reactivos. Aunado a estos objetivos el taller

persiguió la finalidad de demostrar a través de fundamentos teóricos y científicos el papel de la evaluación como una herramienta potencialmente valiosa para mejorar la calidad de la educación.

IV. Pilotaje de los ítems

Una vez concertada la tabla de especificaciones de la prueba y revisados cada uno de los ítems, se estructuró el mapa de la prueba piloto, conformada por dos sub pruebas previas; forma “A” y forma “B”, para el ensamblaje de las mismas los ítems fueron codificados, y las pruebas se vincularon compartiendo ocho ítems en ambas pruebas, que se situaron en la parte inicial y central de los formularios, lo que permitió medir la consistencia de las estadísticas de los ítems. En esta fase también se consideró la diagramación y el diseño de los ítems que son elementos cruciales para la claridad y el atractivo de la prueba. Los ítems se diseñaron en forma de recuadro cada texto correspondiente a la pregunta, de manera que la lectura se mostrara agradable a los estudiantes.

Para iniciar la validación de los ítems, una vez ensambladas las formas, estas fueron revisadas por un equipo de coordinadores académicos con vasta experiencia en el campo de español, a fin de establecer a esta instancia un tipo de validez de criterio, no así estadística. Fue preciso asegurar una de las facetas de la validez, que tiene relación con el grado de representatividad de los estándares de español en la prueba, a través de especialistas que no hayan sido los redactores de los ítems, con el fin de que este equipo determinara si los ítems de las pruebas brindaban una cobertura adecuada del campo de español.

La administración de la prueba piloto se llevó a cabo en dos institutos del municipio de Distrito Central, y uno del municipio de Talanga, detallado anteriormente en el marco muestral de este mismo apartado, la administración se hizo bajo condiciones estandarizadas de aplicación, se realizó una capacitación a los docentes seleccionados como aplicadores de la prueba, el documento que orientó este procedimiento fue el manual de aplicación de la prueba (ver anexo 4). Los estudiantes resolvieron la prueba de forma individual haciendo uso de lápiz grafito. La corrección de la prueba y el ingreso de datos de la prueba piloto se hizo de forma manual, en el programa del paquete de office Excel,

se realizaron procedimientos de doble revisión para garantizar la consistencia y confiabilidad de la base de datos. El procesamiento y análisis de datos se realizó a través de los paquetes estadísticos Statistical Package for the Social Sciences (SPSS).

V. Selección de los ítems Prueba operativa:

Para la selección de los ítems de la prueba, se consideraron los parámetros psicométricos propuestos por la Teoría Clásica de los Test (TCT). Las estadísticas de los reactivos se describieron individualmente en una plantilla o tarjeta de ítems (ver anexo N°5) esta plantilla contiene datos generales del ítems como ser enunciado del ítem, bloque, componente y estándar al que pertenece, función que desempeño, año de aplicación, y la cantidad de la muestra que lo contesto, además contiene los estadísticos psicométricos, como ser: índice de dificultad (P-Value), el índice de discriminación, análisis de opciones así como la imagen y la descripción de la categoría del DIF.

De forma que se consideró aquellos ítems, que observaron un P-Value o dificultad clásica en el rango de 0.3 a 9, además que el índice de discriminación(correlación entre el puntaje del ítem y el puntaje total de la prueba) de estos; presentaran valores arriba de 0.3, otro parámetro estimado para la selección de los reactivos fue el análisis de opciones y distribución de puntajes tomando en cuenta aquellos ítems que estadísticamente sus distractores resultaron verosímiles con no menos del 5%, y donde aquellos la correlación punto biserial de cada opción correcta, se observaran valores positivos y las opciones incorrectas registraran correlaciones negativas, una última condición para la selección de los ítems de la prueba operativa consistió en la observación del Funcionamiento Diferencial del ítem (DIF) o sesgo, incluyendo aquellos que estaban dentro de los límites aceptables. Completada esta revisión de las características psicométricas, se seleccionaron los ítems que mejor estadísticas obtuvieron para ser incluidos en la prueba operativa. De esto este proceso, en base a la tabla de especificaciones técnicas de la prueba se seleccionaron cincuenta (50) reactivos.

VI. Producción y administración de la prueba operativa o definitiva

Esta etapa consiste previamente en la elaboración del mapa de la prueba operativa (ver anexo 6), para el ensamblaje de la misma y su diagramación. Se verificaron detalles de impresión en cuanto a la organización de la información sobre el alumno en la portada,

además del formato y distribución de los ítems, se corrigieron errores en el uso de mayúsculas, puntuación, y ortografía, ilustraciones y tamaño de la fuente utilizada.

Se imprimieron 450 cuadernillos en papel tamaño carta, cada prueba estaba identificada con un código efectos del control en el empaque y distribución para el momento de la aplicación. Previo a la implementación de la prueba definitiva, a través de la Dirección Departamental de Educación de Francisco Morazán, se realizó la convocatoria (ver anexo 7) a los estudiantes de los seis centros educativos seleccionados mediante oficio circular. Como aplicadores de la prueba definitiva, participaron veinticuatro (24) estudiantes de la carrera de Psicología de la Universidad Nacional Autónoma de Honduras, a quienes preliminarmente se les socializó el Manual de aplicación de la prueba.

Considerando los hallazgos en la implementación de la prueba piloto, se estimó un tiempo de dos (2) horas para que los estudiantes dieran respuesta a los ítems de la prueba definitiva. Se procuró condiciones estándares para la aplicación tales como la disposición de materiales (Cuadernillos, lápices, borradores) la acomodación de las sillas, la iluminación, la ventilación y el tiempo disponible para responder fueran iguales para todos los evaluados, así mismo se indicó a los aplicadores que si las circunstancias lo ameritaran, se tomaran las medidas necesarias para que estudiantes de la población objetivo con alguna limitación física reciban soporte adecuado para afrontar el proceso de evaluación en condiciones equivalentes al resto de la población.

VII. Análisis de los resultados de la prueba de español 11mo grado BCH

La calificación de los ítems se hizo de forma manual, donde en los criterios de revisión, la máxima puntuación fue 1, para la respuesta opción correcta y puntuación 0, para otras respuestas, se asignó el valor 9, para los ítems que no fueron contestados y para aquellos que presentaban dos opciones de respuesta.

Los datos recopilados recibieron inicialmente un tratamiento estadístico descriptivo con tablas y graficas analizados mediante el programa Statistical Package for the Social Sciences (SPSS). Posteriormente mediante el modelo lineal clásico de la TCT, se determinó el índice de fiabilidad, dificultad, discriminación, y los valores del Funcionamiento Diferencial de los Ítems (DIF). Finalmente, en la transformación e

interpretación de las puntuaciones, se consideraron tres métricas utilizadas en evaluaciones nacionales e internacionales correspondiente a la métrica de Niveles de desempeño, Escala de puntuación estandarizada o Scale Score y el porcentaje de respuestas correctas.

VIII. Presentación de los resultados de desempeño académico de los estudiantes de 11mo grado BCH, en el área español.

Los resultados se presentaron a través de gráficas y tablas, que describen los niveles de desempeño académico de la población objetivo de este estudio. El nivel de análisis permitió exponer un análisis detallado de los niveles de aprendizaje mostrados por los estudiantes del undécimo grado del Bachillerato en Ciencias y Humanidades, en el área de español.

4.6 Plan de análisis

El análisis de los datos se realizó en dos momentos, en la aplicación de la prueba piloto y la prueba final. Los programas utilizados para analizar los datos fueron SPSS Statistics y el programa Excel del paquete de Office, estos procedimientos estadísticos se basaron en la Teoría Clásica de los Test (TCT). El procedimiento para ambos momentos inicialmente consistió en explorar la base de datos para controlar la presencia de errores o valores fuera de rango, se generaron descriptivos de centralización y dispersión. Para comprobar la naturaleza de los datos, e indicar si la distribución es simétrica, sesgada o agrupada se realizó la prueba de Kolmogorov Smirnov. A continuación, se detallan el tratamiento de los datos en cada etapa, y que, para efectos de los objetivos y comprobación de hipótesis establecidos en este estudio, fueron aplicados.

Análisis de los datos prueba piloto

La finalidad del tratamiento de los datos de la prueba piloto fue obtener información estadística de los ítems que ayudare a seleccionar aquellos con mejores parámetros para el diseño de la prueba final. Habiendo obtenido las estadísticas generales, consecutivamente se analizo la confiabilidad de cada sub prueba (Forma “A” y “B”) a través del método de covarianza de los ítems utilizando como estadístico el coeficiente Alfa de Cronbach, este permitió verificar la consistencia interna de los puntajes del instrumento, examinando la

correlación elemento-total corregida de ambas pruebas, que equivale al coeficiente de homogeneidad corregida.

La fórmula del coeficiente de alfa de Cronbach utilizada fue:

Ilustración 7 Formula Coeficiente alfa de Cronbach

$$\alpha = \frac{k}{k-1} \left(\frac{St^2 - \sum s^2}{St^2} \right)$$

donde (α) es igual al coeficiente alfa de Cronbach, (k) representa el número de ítems de la prueba, ($\sum s^2$) corresponde a la sumatoria de la varianza de cada ítem y (St^2) constituye la varianza del total de las puntuaciones del test. Esta se aplicó a través del software estadístico SPSS.

Para determinar la validez de la prueba, se determinó sustraer evidencias de validez de contenido, pues debido al carácter unidimensional de la prueba al medir un solo bloque de lectura se desestimó comprobar la validez de constructo que implica un análisis factorial exploratorio o confirmatorio. La validez de contenido implica realizar interpretaciones y usos adecuados de los resultados o de la información obtenida de la prueba, una de sus facetas es el grado de representatividad del contenido de la prueba en lo que se refiere al currículo a evaluar (Anderson & Morgan, Evaluaciones de rendimiento académico., 2016). A diferencia de otros tipos de validez, esta no se representa de manera estadística con índice, sino que por lo general se estima de manera objetiva o intersubjetiva a través del método “juicios de expertos”, este se conformó a través de tres docentes de institutos del Distrito Central que no fueron seleccionados en la muestra para la prueba pilotaje ni en su forma operativa. En este procedimiento se utilizaron el marco operativo, la tabla de especificaciones técnicas y una versión preliminar de las sub pruebas piloto impresas.

Una vez determinada la confiabilidad y validez de la prueba, los siguientes pasos en el tratamiento de los datos de las pruebas pilotos consistió en analizar los indicadores de calidad de los reactivos a través de los parámetros de discriminación, dificultad y la valoración de sesgo o DIF. El índice de dificultad (Valor P) de los ítems se calculó mediante los procesos estadísticos del programa SPS. Este equivale a la proporción de los y las estudiantes que

contestan cada ítem correctamente, sus valores fluctúan entre 0 y 1, los cercanos a 1 indican una baja dificultad y próximos a 0 demuestran dificultad máxima.

El índice de discriminación calculado en la prueba fue la correlación ítems-test, biserial-puntual; este cálculo permite cuantificar la relación entre una variable binaria (el acertar o no el ítem) y una variable de escala, la puntuación total obtenida en la prueba. La correlación punto-biserial es un tipo específico de la correlación usado para ítems dicotómicos (ítems de opción múltiple). Una alta correlación ítem-puntaje total indica que los estudiantes que contestaron un ítem correctamente, también recibieron altos puntajes en la prueba (y viceversa). Los valores arriba de 0.3 son deseables, aquellos valores bajo 0.25 deben ser investigados.

Finalmente se realizó el análisis del funcionamiento Diferencial de los ítems (DIF), que implica el cálculo de las diferencias inesperadas en el desempeño en un ítem por medio del procedimiento matemático del método Mantel –Haenszel. Se analizó dos tipos de DIF, por sexo, tomando como grupo focal las mujeres, y por tipo de administración de los centros educativos identificando como grupo focal los estudiantes de institutos no gubernamentales.

Análisis de los datos de la prueba operativa

Las propiedades psicométricas de los ítems y del instrumento en su conjunto, fueron analizadas atendiendo especialmente a la fiabilidad y validez de las medidas obtenidas a partir del modelo lineal clásico que brinda la teoría clásica de los test (TCT) empleando el software SPSS. Inicialmente se obtuvo las estadísticas descriptivas del instrumento, realizando las primeras observaciones acerca de la normalidad de los datos, mediante la prueba de Kolmogorov Smirnov, que consiste en un procedimiento de “bondad de ajuste” que permite verificar el grado de concordancia existente entre la distribución del conjunto de datos y una distribución teórica específica de las puntuaciones que obtuvieron los estudiantes en la prueba.

Para determinar la confiabilidad de los datos se recurrió al método de consistencia interna a través del coeficiente Alfa (α) de Cronbach (1951), que se deriva del cálculo de las varianzas de cada ítem y de las varianzas de las puntuaciones totales del instrumento. Por otra parte, la validez de contenido se determinó mediante el análisis realizado por docentes expertos en el área de español acerca de la tabla de especificaciones técnicas y la

representatividad de los ítems que evalúan los estándares en el instrumento, así como la calidad en el contenido, redacción y relevancia de los reactivos. La comprobación de hipótesis, se determinó realizando un procedimiento de comparación de medias que establece la prueba “T de Student”, representado en tablas y gráficos de cajas el análisis de las diferencias respecto a las puntuaciones directas de los estudiantes según las variables de sexo y tipo de administración de centros educativos.

Finalmente, la descripción del rendimiento académico de los (as) estudiantes de undécimo grado del BCH, se presentó a través de tres métricas, que han sido utilizadas en la experiencia nacional respecto a la evaluación del desempeño académico de los estudiantes de educación básica, están corresponden a la métrica escala de puntuación estandarizada o scale score (100 a 500), métrica niveles de desempeño, anclada a la escala estandarizada(insatisfactorio (100-199), debe mejorar(200-299), satisfactorio (300-399), y avanzado(400-500) por último la métrica porcentaje de respuestas correctas (SE & MIDEH, Informe Nacional de Desempeño Académico, 2017).

Para obtener la medición de las métricas escala de puntuación estandarizada (Scale score), y niveles de desempeño, se realizó un procedimiento de transformación de las puntuaciones directas a través del programa estadístico SPSS y Excel. Los pasos se resumen en lo siguiente:

1. Las frecuencias absolutas (el número real de preguntas que contesto correctamente el estudiante), se transformaron en puntuaciones estandarizadas Z, este cálculo precisó utilizar el valor de la media y la desviación típica del puntaje total.

$$Z = \frac{x - \bar{x}}{s}$$

X= Puntaje en la prueba

$\bar{x}$ = Media en la prueba

S= Desviación estándar

Las puntuaciones z, que resultaron con valores negativos, significan el número de desviaciones típicas que se encuentran por debajo de la media obtenida en la prueba.

2. Una vez obtenidas las puntuaciones “z”, la métrica de puntuación estandarizada o Scale Score, convirtió los puntajes brutos de los estudiantes (el número real de preguntas que contestó correctamente), en una puntuación de escala común (100 a 500). La fórmula aplicada fue:

$$\text{Escala puntuación estandarizada} = a \times Z_i + b$$

a=desviación estándar

Z_i=puntaje z

B=media

Se utilizó una media teórica de 300 y una desviación estándar de 50, de acuerdo al procedimiento de la experiencia nacional en desarrollo de esta escala, misma que se encuentra en correspondencia a la métrica de niveles de desempeño que distribuye a los estudiantes en distintas categorías según rangos de puntuaciones, de forma que el siguiente paso fue delimitar los puntos de corte.

3. Los niveles de desempeño definidos para interpretar los resultados toman como base teórica las categorías ya establecidas en el sistema de evaluación nacional de Honduras, alineados a los estándares evaluados en la prueba. Cada nivel tiene su descripción general y están identificados en una escala de colores. Este modelo de presentación de resultados contribuye en hacer más comprensible y relevante la información aportada por la prueba. Los niveles de desempeño definidos son: Avanzado, Satisfactorio, Debe Mejorar e Insatisfactorio. La tabla 5 presenta la categorización general por niveles de desempeño, anclada a la escala de puntuación estandarizada utilizada en este estudio.

Tabla 9 Métricas niveles de desempeño y escala de puntuación estandarizada

Nivel de desempeño	Descriptores generales	Escala de puntuación estandarizada (100-500)
	Avanzado	Los estudiantes en este nivel demuestran un desempeño excepcional en los temas evaluados.
	Satisfactorio	Los estudiantes en este nivel han cumplido con el objetivo educativo, tienen los conocimientos, habilidades y destrezas necesarias para tener éxito en el siguiente grado.

	Debe Mejorar	Los estudiantes en este nivel tienen conocimientos mínimos de los temas evaluados, sus conocimientos, habilidades y destrezas son suficientes para justificar su avance al siguiente grado.	200 a 299
	Insatisfactorio	Los estudiantes en este nivel no tienen los conocimientos mínimos de los temas evaluados. Sus conocimientos, habilidades y destrezas no son suficientes para justificar su avance al siguiente grado.	100 a 199

Fuente: Elaboración propia, adaptada del informe nacional de rendimiento académico 2017, secretaria de educación de Honduras y Proyecto de Medición del Desempeño Académico (MIDEH).

Culmina el análisis con la presentación de los resultados según la métrica de porcentaje promedio de respuestas correctas, convirtiendo los resultados en datos porcentuales, exponiendo con mayor detalle las fortalezas y debilidades encontradas por bloque y estándar evaluado; cabe señalar que, aunque esta forma de análisis y comunicación de resultados presenta limitaciones técnicas para hacer análisis comparativos entre pruebas y a través del tiempo (SE & MIDEH, 2017, pág. 33)., sin embargo debido a la naturaleza de este estudio de corte transversal, ha sido utilizada por la facilidad para comunicar la información obtenida a públicos no especializados en evaluación, dada a su semejanza con las calificaciones escolares.

CAPITULO 5. RESULTADOS DE LA INVESTIGACION

Los resultados han sido estructurados de acuerdo al plan de análisis, detallado en el apartado anterior, y en correspondencia a los objetivos de este estudio han sido organizados en tres partes. Inicialmente se da a conocer el resultado del diseño del marco operativo que oriento la estructura de la prueba, seguidamente se presentan los análisis realizados acerca de las evidencias de confiabilidad y validez del instrumento, a través del proceso de pilotaje de los ítems, para finalmente mostrar un estudio descriptivo acerca de los resultados de desempeño académico obtenido por los estudiantes de undécimo grado de BCH en el área español, que incluye la comprobación de hipótesis del estudio así como el reporte de los resultados de aprendizaje a través de las métricas de niveles de desempeño académico, escala de puntuación estandarizada y porcentaje de respuestas correctas.

5.1 Resultados diseño del marco operativo de la prueba

A lo largo de este estudio, se explicaron los pasos establecidos para el diseño de una prueba que mida el desempeño académico, que van desde el momento que se decide que estándares evaluar, hasta la producción del informe con sus resultados. Los objetivos propuestos por la naturaleza de esta investigación, consistió primeramente en determinar un marco de evaluación que oriente la construcción del instrumento, donde el producto principal fue obtener el marco operacional de la prueba y la tabla de especificaciones técnicas, misma que es el insumo principal para el diseño de los ítems y estructura de la prueba.

A través del taller de evaluación educativa y redacción de ítems, los especialistas del área español determinaron que el instrumento debería estar conformada por 50 reactivos de opción múltiple, evaluando el bloque de lectura, y los componentes correspondientes a tipos de texto y vocabulario. La tabla 3, muestra el marco operativo de la prueba indicando con detalle, el número de estándares que evalúa el instrumento, así como la distribución de los ítems por nivel de procesamiento cognitivo según el modelo definido por Bloom (2000).

Tabla 10 Marco operativo de la prueba para la medición del desempeño académico de español para los estudiantes de 11mo grado de BCH

Bloque	Componente	Estándar	Ítems por nivel cognitivo			Total general
			A	B	C	
Lectura	1. Tipos de texto	Leen y utilizan con diversos propósitos textos narrativos, descriptivos, expositivos y persuasivos, tanto literarios como funcionales e informativos, de varias fuentes.	1	2	2	5
		Demuestran comprensión de las ideas globales, principales, secundarias e inferenciales de un texto leído, y la interpretación de gráficos e iconos.	0	4	1	5
		Identifican el contexto, la intención comunicativa, el punto de vista y el estilo de los textos narrativos, descriptivos, expositivos y persuasivos.	0	2	3	5
		Interpretan información explícita e implícita de textos diversos para formular planteamientos en sentido crítico.	0	3	2	5
		Analizan diferentes puntos de vista y evalúan rasgos científicos, sociológicos y culturales de los textos.	0	4	1	5

		Comprenden, aprovechan y evalúan críticamente imágenes y textos, iconos verbales y de informática.	2	4	1	7
	1.6					
	2.1	Reconocen y entienden en textos leídos una variedad de palabras conocidas.	1	1	1	3
	2.2	Interpretan palabras desconocidas utilizando diversas estrategias	1	3	0	4
2.						
Vocabulario	2.3	Interpretan en textos leídos el significado de términos técnicos, lenguaje no-estándar.	0	5	1	6
	2.4	Reconocen e interpretan el lenguaje figurado y otros aspectos semánticos de las palabras	1	1	3	5
Total de ítems por nivel cognitivo			6	29	15	50
Porcentaje Nivel Cognitivo			12%	58%	30%	100%

Fuente. Elaboración propia

Nota: Niveles cognitivos, A: Conocimiento, B: Comprensión, C. Aplicación

En el mismo orden, como parte de los resultados en el diseño del marco de evaluación, se presenta la información general de la tabla de especificaciones técnicas de la prueba (ver anexo 7), documento que orientó la construcción de los ítems por parte de los docentes especialistas. La tabla 7, detalla de forma puntualizada el número total de reactivos diseñados por componente, estándar y nivel cognitivo. La propuesta fue la redacción de 24 ítems de conocimiento, 116 de comprensión y 60 de aplicación, lo que hizo un total de 200 reactivos.

Tabla 11 Tabla de especificaciones técnicas para el diseño de los ítems

Bloque	Componente	Nivel Cognitivo			Total
		A	B	C	
Lectura	Tipos de texto	12	76	40	128
	Vocabulario	12	40	20	72
	Total	24	116	60	200

Fuente. Elaboración propia. Resumen de la tabla de especificaciones técnicas de la prueba operativa que representa una extensión del marco operativo de la prueba.

Nota: Niveles cognitivos, A: Conocimiento, B: Comprensión, C. Aplicación

a. Resultados de la administración de la prueba piloto

A continuación, se presentan los resultados de la aplicación de la prueba piloto, que obedece al objetivo concerniente a establecer los parámetros psicométricos de confiabilidad y validez de contenido de la prueba; así como el índice de discriminación y dificultad de los ítems, y el funcionamiento diferencial (DIF) de los reactivos, teniendo como grupos de interés el sexo de los estudiantes y el tipo de administración de los centros educativos parte de la muestra. La prueba piloto se estructuró en dos sub pruebas o formas “A” y “B”. En esta primera recolección de datos, la población objetivo participante fue de 235 estudiantes del undécimo grado del BCH, de dos centros de educación media oficiales y uno no gubernamental.

5.2.1 Confiabilidad y validez: prueba piloto

La confiabilidad de las puntuaciones obtenidas, se estimó por medio de la prueba estadística Alfa (α) de Cronbach obteniéndose un valor de .98 en la forma “A” y .97 en la forma “B”; ninguno de los ítems contempló correlaciones negativas, lo que indica que los puntajes totales de los ítems en relación al puntaje total del test reflejaron índices de homogeneidad aceptables que demuestra validez como consistencia interna de los ítems.

Tabla 12 Índice de fiabilidad de la prueba piloto

Forma	Estadístico Fiabilidad	Alfa de Cronbach basada en los elementos tipificados	N de elementos
A	.976	.975	43
B	.972	.970	43

Fuente: Resultados del análisis de datos de las sub pruebas piloto “A” y “B” Español 11° grado BCH 2018

Siguiendo el análisis de confiabilidad de la prueba piloto, la tabla 13, permite observar la consistencia en las puntuaciones de los ítems de la sub prueba “A”, dando evidencia que los ítems están midiendo un mismo constructo estando positivamente correlacionados; los valores obtenidos en la correlación elemento-total corregida, oscilaron entre .414 y .893, y el índice de fiabilidad de cada ítem si se elimina el elemento, se mantuvo constante con .975; lo que demuestra un índice de homogeneidad en las puntuaciones en la categoría aceptable.

De manera que al no observar puntuaciones negativas en los resultados de la correlación elemento total corregida ≤ 3 , no se hizo necesario la eliminación de los reactivos para mejorar el índice de fiabilidad.

Tabla 13 Índices de fiabilidad y homogeneidad de los ítems prueba “A”

Elementos	Correlación elemento-total corregida	Alfa de Cronbach si se elimina el elemento
ESP11018_111BOA01	.490	.976
ESP11018_111AOA02	.793	.975
ESP11018_111COA03	.893	.975
ESP11018_111BOA04	.473	.976
ESP11018_112BAA05	.432	.976
ESP11018_112COA06	.812	.975
ESP11018_112BOA07	.691	.975
ESP11018_112BOA08	.801	.975
ESP11018_113BAA09	.818	.975
ESP11018_113COA10	.792	.975
ESP11018_113BOA11	.716	.975
ESP11018_113BOA12	.693	.975
ESP11018_113BOA13	.580	.976
ESP11018_114BOA14	.692	.975
ESP11018_114BAA15	.717	.975
ESP11018_114BAA16	.855	.975
ESP11018_114BAA17	.892	.975
ESP11018_115BOA18	.517	.976
ESP11018_115BOA19	.681	.975
ESP11018_121BOA20	.834	.975
ESP11018_115BOA21	.890	.975
ESP11018_115COA22	.608	.976
ESP11018_121AOA23	.698	.975
ESP11018_116AAA24	.691	.975
ESP11018_116BOA25	.749	.975
ESP11018_116AOA26	.739	.975
ESP11018_116BOA27	.830	.975
ESP11018_116BOA28	.770	.975
ESP11018_116AOA29	.495	.976
ESP11018_121AAA30	.848	.975
ESP11018_122BAA31	.877	.975
ESP11018_124BOA32	.888	.975

ESP11018_124BOA33	.836	.975
ESP11018_122BAA34	.768	.975
ESP11018_122AAA35	.855	.975
ESP11018_122BAA36	.414	.976
ESP11018_123CAA37	.551	.976
ESP11018_123COA38	.787	.975
ESP11018_123BOA39	.654	.975
ESP11018_123BOA40	.841	.975
ESP11018_124BOA41	.834	.975
ESP11018_124BOA42	.755	.975
ESP11018_124COA43	.509	.976

Fuente: Elaboración propia. Resultados análisis de datos SPSS, sub prueba piloto Forma “A” (2018)

En la tabla 14, se muestran los valores observados en cuanto a la consistencia y uniformidad en la medición de los ítems de la sub prueba “B”; el análisis incluye los índices de homogeneidad donde se aprecian cuatro reactivos con valores bajos en la correlación-elemento total corregida, estos se encuentran en las posiciones (2), (19), (20) y (34), los primeros tres elementos corresponden al componente de tipología textual y el último a vocabulario. La correlación elemento total corregida de estos elementos, mostro puntuaciones de .313 y .393, dando lugar a excluir la utilidad de dichos ítems para la construcción de la escala, logrando una confiabilidad de .973, hecho que no fue estadísticamente significativo.

Tabla 14 Índices de fiabilidad y homogeneidad de los ítems prueba “B”

Elementos	Media de la escala si se elimina el elemento	Varianza de la escala si se elimina el elemento	Correlación elemento- total corregida	Alfa de Cronbach si se elimina el elemento
ESP11018_111BOB01	11.40	155.138	.462	.972
ESP11018_111AOB02	11.54	159.369	.339	.972
ESP11018_111COB03	11.37	156.227	.440	.972
ESP11018_111BOB04	11.45	156.556	.478	.972
ESP11018_112BAB05	11.22	151.301	.898	.971
ESP11018_112COB06	11.24	152.550	.815	.971
ESP11018_112BOB07	11.26	152.424	.798	.971
ESP11018_112BOB08	11.31	153.070	.881	.971
ESP11018_113BAB09	11.32	154.423	.666	.972
ESP11018_113COB10	11.30	153.280	.796	.971

ESP11018_113BOB11	11.40	156.583	.706	.972
ESP11018_113BOB12	11.40	155.566	.684	.972
ESP11018_113BOB13	11.40	156.703	.715	.972
ESP11018_114BOB14	11.45	156.864	.641	.972
ESP11018_114BAB15	11.34	154.919	.538	.972
ESP11018_114BAB16	11.36	154.453	.591	.972
ESP11018_114BAB17	11.20	151.297	.812	.971
ESP11018_115BOB18	11.40	156.087	.834	.972
ESP11018_115BOB19	11.51	159.191	.342	.972
ESP11018_121BOB20	11.57	160.683	.313	.973
ESP11018_115BOB21	11.25	151.990	.889	.971
ESP11018_121BOB22	11.38	156.407	.626	.972
ESP11018_115AOB23	11.40	156.096	.673	.972
ESP11018_116AAB24	11.38	154.844	.817	.972
ESP11018_116AOB25	11.35	154.177	.802	.971
ESP11018_116COB26	11.30	154.408	.767	.972
ESP11018_116BOB27	11.47	158.404	.806	.972
ESP11018_116BOB28	11.46	157.343	.606	.972
ESP11018_116BOB29	11.36	154.846	.854	.972
ESP11018_121AAB30	11.23	151.676	.791	.971
ESP11018_122BAB31	11.25	152.016	.897	.971
ESP11018_124BOB32	11.30	152.878	.882	.971
ESP11018_124BOB33	11.20	150.996	.947	.971
ESP11018_122BAB34	11.42	156.074	.397	.973
ESP11018_122AAB35	11.31	154.078	.914	.972
ESP11018_122BAB36	11.45	157.633	.665	.972
ESP11018_123CAB37	11.47	157.891	.527	.972
ESP11018_123COB38	11.30	153.143	.868	.971
ESP11018_123BOB39	11.39	155.000	.901	.972
ESP11018_124COB40	11.46	156.745	.969	.972
ESP11018_124BOB41	11.32	153.876	.894	.971
ESP11018_123BAB42	11.30	152.870	.753	.971
ESP11018_124COB43	11.28	152.852	.822	.971

Fuente: Elaboración propia. Resultados análisis de datos SPSS, sub prueba piloto Forma “B” (2018)

Validez de contenido

La evidencia de la validez de contenido a través de juicios de expertos, tuvo como propósito determinar la relación existente entre los ítems que componen el test y lo que este

pretende evaluar, prestando atención tanto a la relevancia como a la representatividad de los ítems. Los docentes participantes en esta etapa determinaron que la prueba presentaba una cobertura adecuada para brindar evidencias suficientes del rendimiento académico de los estudiantes en el área español, coincidiendo en el contenido y redacción de los ítems, así como en la extensión de la prueba, tipo de formato y diagramación de la misma.

Sus aportes fueron concluyentes para determinar algunos cambios en las pruebas pilotos consistentes en una mejor distribución de los ítems de la forma “A” donde en opinión de los expertos, la prueba comprendía textos más amplios que la prueba “B”, lo que podría resultar un desbalance en términos de dificultad de la prueba.

5.2.2 Propiedades psicométricas del instrumento pilotaje: Dificultad y Discriminación

Respecto a la determinación de la calidad estadística de los ítems, la tabla 15, muestra el resumen de los estadísticos de los ítems de la forma “A”. se detallan a continuación las estadísticas obtenidas en los valores de dificultad (p-value), discriminación, así como el funcionamiento diferencial de los ítems, donde de este último se dedica un apartado para detallar con mayor precisión los resultados.

Tabla 15 Parámetros psicométricos prueba piloto forma "A"

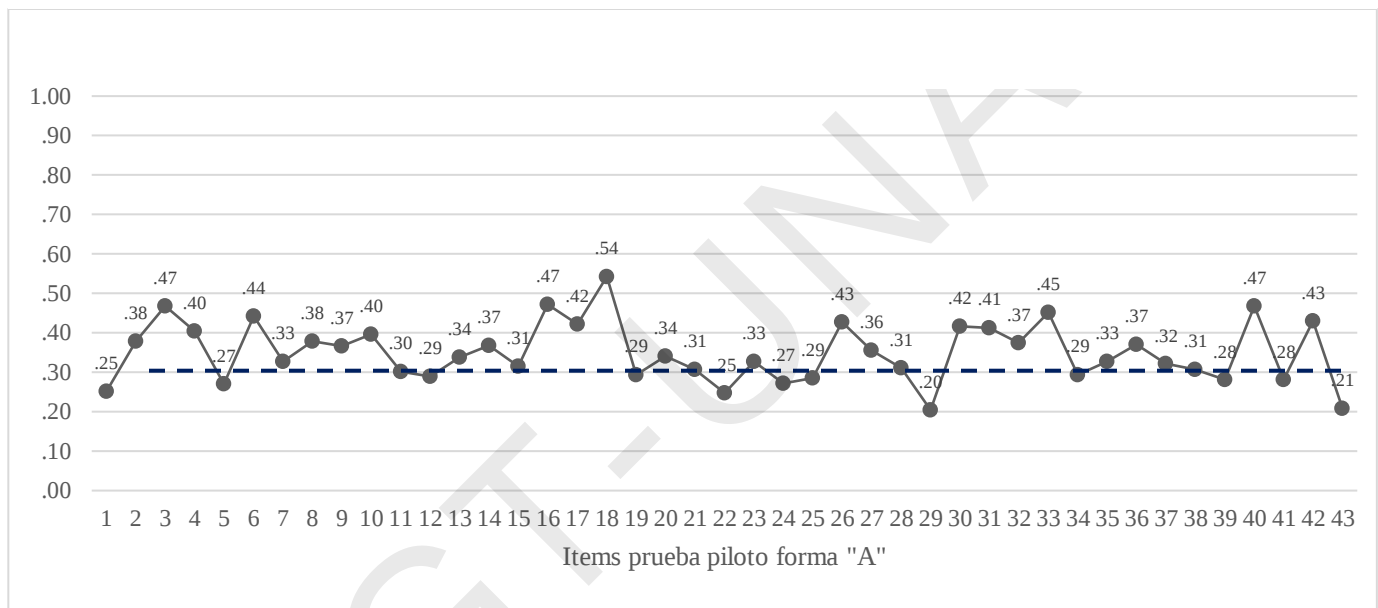
Ítems	Dificultad	Discriminación Correlación Ítem- puntaje total	DIF Sexo	Categoría	DIF CE	Categoría
1	.25	.58	.05	A	.08	B
2	.38	.77	.06	B	.10	B
3	.47	.87	.07	B	.00	A
4	.40	.47	.04	A	.01	A
5	.27	.43	.02	A	.16	C
6	.44	.85	.02	A	.10	B
7	.33	.73	.04	A	.05	A
8	.38	.74	.04	A	.07	B
9	.37	.75	.04	A	.00	A
10	.40	.81	.00	A	.06	B
11	.30	.68	.02	A	.13	C
12	.29	.67	.01	A	.05	A
13	.34	.56	.03	A	.17	C
14	.37	.62	.03	A	.23	C
15	.31	.67	.05	A	.00	A

16	.47	.66	.04	A	.02	A
17	.42	.87	.08	B	.06	B
18	.54	.67	.00	A	.01	A
19	.29	.65	.02	A	.20	C
20	.34	.73	.03	A	.11	C
21	.31	.63	.01	A	.01	A
22	.25	.61	.03	A	.12	C
23	.33	.71	.02	A	.20	C
24	.27	.66	.06	B	.13	C
25	.29	.65	.06	B	.17	C
26	.43	.64	.13	C	.05	A
27	.36	.63	.02	A	.04	A
28	.31	.67	.06	B	.07	B
29	.20	.52	.02	A	.04	A
30	.42	.82	.04	A	.03	A
31	.41	.78	.09	B	.38	C
32	.37	.82	.01	A	.12	C
33	.45	.89	.06	B	.20	C
34	.29	.67	.05	A	.03	A
35	.33	.66	.01	A	.04	A
36	.37	.45	.06	B	.02	A
37	.32	.54	.05	A	.06	B
38	.31	.73	.00	A	.16	C
39	.28	.63	.11	C	.02	A
40	.47	.89	.06	B	.00	A
41	.28	.64	.06	B	.04	A
42	.43	.86	.02	A	.02	A
43	.21	.53	.02	A	.10	B

Fuente: Elaboración propia. Análisis de datos programa SPSS, prueba piloto forma "A"

La tercera columna de la tabla 11, muestra los estadísticos del valor P-Value o dificultad de los ítems, estos oscilaron entre .20 y .54, ubicándose en la categoría de dificultad media y alta. Un número total de 18 ítems presentan estadísticas con mayor dificultad en la sub prueba “A”, mientras que 21 presentan dificultad categoría aceptable y 4 elementos dificultad media. Cabe destacar que el valor P-Value de un ítem debe relacionarse con el concepto que recoge, es decir que la complejidad de las habilidades evaluadas en esta prueba es correspondiente a los valores de dificultad.

Grafico 3 Dificultad de los ítems de la sub prueba piloto forma "A"



Fuente. Elaboración propia, análisis de resultados prueba piloto forma “A”.

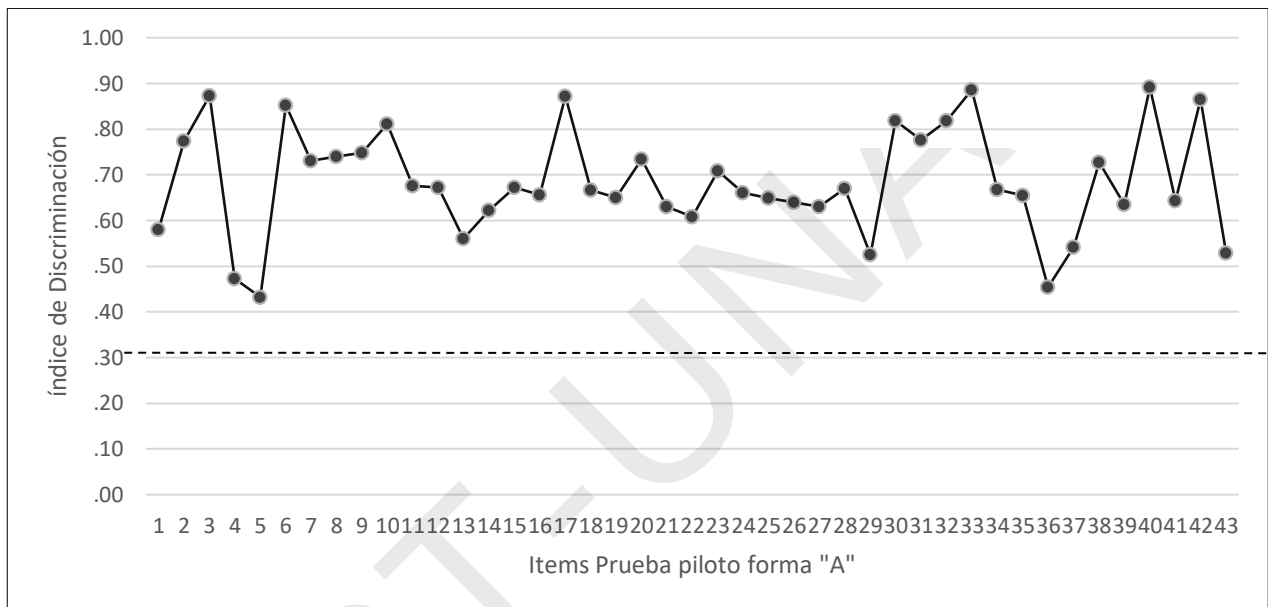
El gráfico 3, presenta con mayor simplicidad los ítems que se ubicaron con un p-valúe por debajo 0.3, se observa un total de 7 ítems del componente tipos de texto y 6 ítems del componente vocabulario con valores menores 0.3, que los ubica en categoría de dificultad alta. El ítem con menor dificultad fue el ítem 18 correspondiente al componente tipos de texto.

En cuanto a los valores de la discriminación o índice de correlación ítem-puntaje total, observados en la siguiente columna de la tabla 10, permite visualizar puntuaciones positivas para este parámetro psicométrico, con puntajes entre .43 y .89, lo que indica según los puntos de corte de los valores de correlación, que el total de los ítems de la sub prueba “A” se

encuentran en la categoría de alta discriminación, mostrando un aceptable grado de acuerdo existente, entre los puntajes individuales de los ítems y el puntaje total de la prueba.

La figura 4, permite observar lo acotado anteriormente, el total de los 43 ítems de la prueba piloto “A”, revela una alta discriminación, lo que a pesar de que los valores de la dificultad fueron altos, los hacen elegibles para conformar la prueba operativa.

Grafico 4 Índice de Discriminación de los ítems de la prueba piloto forma "A"



Fuente: prueba piloto Español-Lectura, 11mo grado BCH, forma “A” (2018)

Nota: Los valores $\geq .30$, indican una discriminación aceptable. Los 43 ítems de la sub prueba piloto “A”. se categorizan con alta discriminación

En el mismo orden de análisis se procedió para los ítems de la sub prueba piloto forma “B”. La tabla 16, muestra los resultados obtenidos en cada uno de los parámetros psicométricos.

Tabla 16 Parámetros psicométricos prueba piloto forma "B"

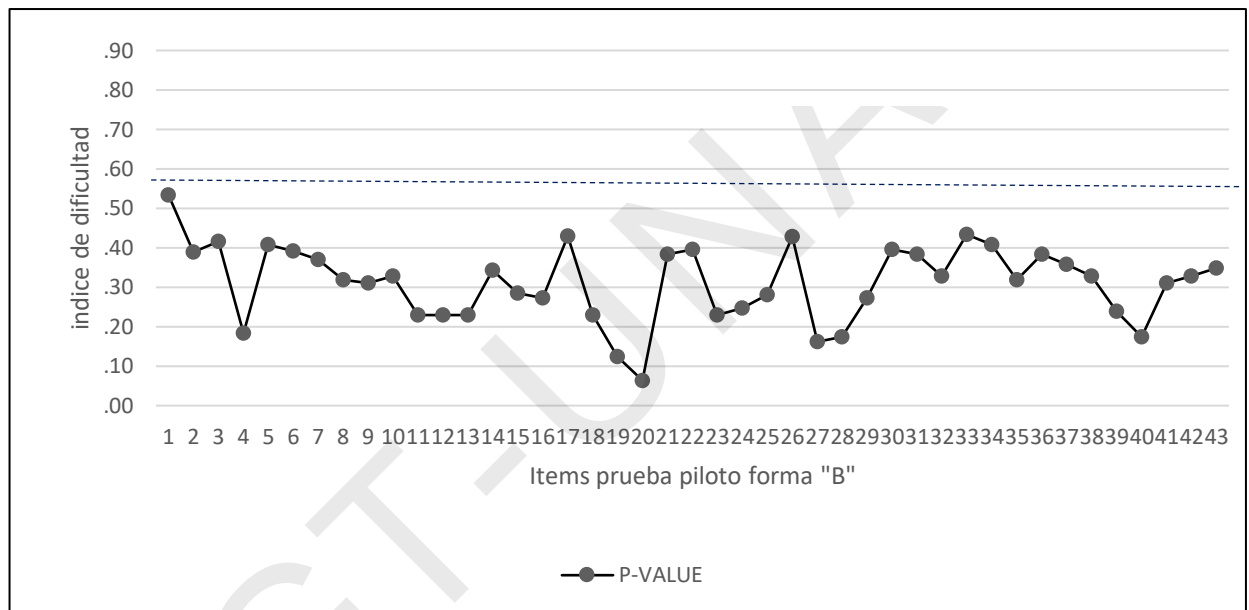
Ítems prueba piloto	Índice de Dificultad	Índice Discriminación Correlación Ítem-puntaje total	DIF Sexo	Categoría	DIF tipo de ADMON.	Categoría
1	.53	.66	.01	A	.11	C
2	.39	.50	.13	C	.30	C
3	.42	.54	.05	A	.07	B
4	.18	.58	.01	A	.08	B

5	.41	.89	.00	A	.03	A
6	.39	.79	.01	A	.16	C
7	.37	.81	.00	A	.02	A
8	.32	.78	.03	A	.08	B
9	.31	.67	.01	A	.11	C
10	.33	.76	.01	A	.07	B
11	.23	.53	.03	A	.02	A
12	.23	.62	.04	A	.12	C
13	.23	.51	.07	B	.06	B
14	.34	.55	.03	A	.13	C
15	.29	.64	.05	A	.06	B
16	.27	.69	.06	B	.02	A
17	.43	.88	.08	B	.05	A
18	.23	.57	.01	A	.03	A
19	.12	.36	.06	B	.01	A
20	.06	.25	.01	A	.05	A
21	.38	.84	.03	A	.00	A
22	.40	.63	.03	A	.19	C
23	.23	.57	.18	C	.13	C
24	.25	.68	.10	B	.16	C
25	.28	.71	.00	A	.06	B
26	.43	.66	.04	A	.22	C
27	.16	.41	.03	A	.02	A
28	.17	.51	.02	A	.11	C
29	.27	.65	.03	A	.04	A
30	.40	.86	.05	A	.01	A
31	.38	.84	.04	A	.00	A
32	.33	.79	.04	A	.06	B
33	.43	.91	.02	A	.05	A
34	.41	.80	.04	A	.14	C
35	.32	.69	.03	A	.02	A
36	.38	.47	.04	A	.14	C
37	.36	.47	.06	B	.05	A
38	.33	.77	.01	A	.15	C
39	.24	.67	.04	A	.06	B
40	.17	.57	.09	B	.24	C
41	.31	.71	.12	C	.16	C
42	.33	.79	.01	A	.04	A
43	.35	.78	.08	B	.04	A

Fuente: Análisis de datos SPSS sub prueba piloto “B”.

El índice de dificultad de los ítems de la sub prueba “B”, observo valores entre .06 y .53, donde 26 ítems se ubican en una dificultad media, y 17 reactivos se categorizaron en el nivel de dificultad alta. La figura 5, muestra en el eje (Y) el índice de dificultad y en el eje (X) la posición de los ítems. Un total de 19 ítems observaron una dificultad entre .31 y .40, mientras que 11 ítems se registraron con valores de dificultad entre .40 y .54.

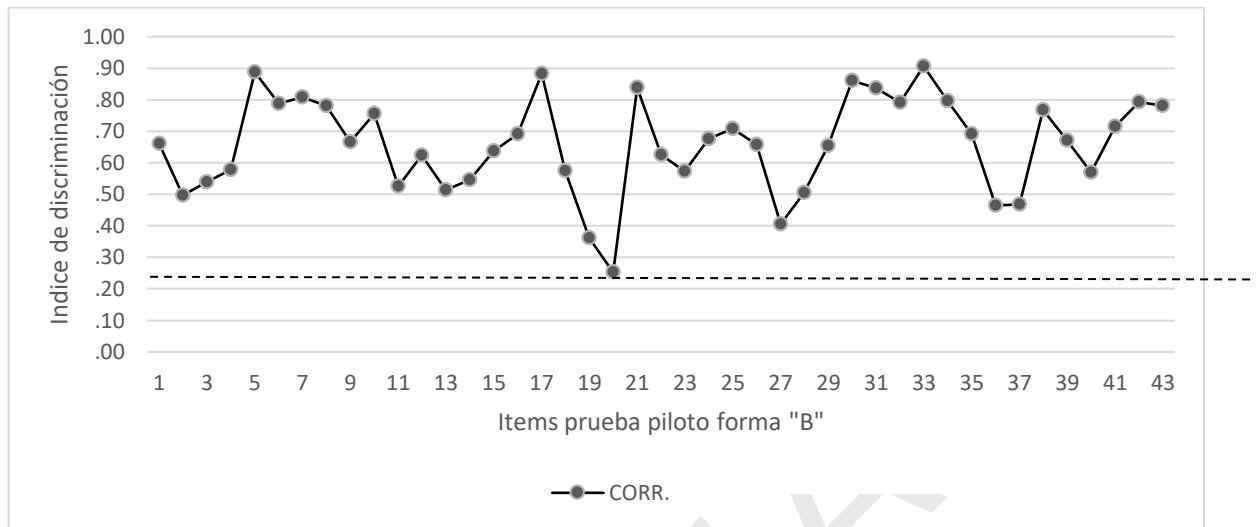
Grafico 5 Dificultad de los ítems de la prueba piloto forma "B"



Fuente: Elaboración propia: Análisis de datos prueba piloto, forma “B”.

Como se logra apreciar en la sub prueba forma “B”, se encontraron los ítems con mayor dificultad respecto a la sub prueba forma “A”. Los elementos que resultaron estadísticamente con alta dificultad fueron los ítems de posición 19 y 20, con valores de 0.12 y 0.06, haciéndolos descartables para conformar la prueba operativa, siendo que los resultados en cuanto a la correlación también los ubicaban en una categoría de baja discriminación.

En cuanto a la capacidad de discriminación de los reactivos se obtuvieron valores entre .25 y .91; lo que dio como resultado un total de 41 ítems con una correlación aceptable. En la figura 4, se muestra el índice de discriminación los cuales para ambas pruebas mantuvieron valores aceptables, todos los valores son positivos de forma que revelan discriminación entre .25 y .91.

Gráfico 6 Índice de discriminación prueba piloto forma "B"

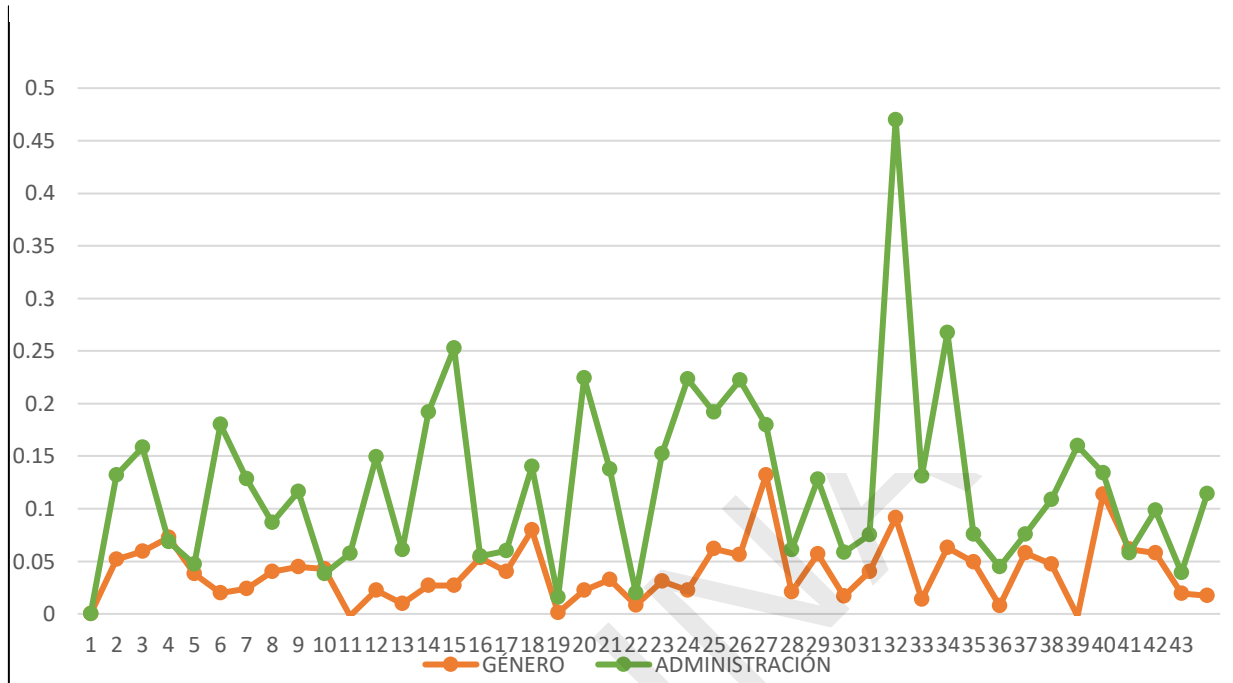
Fuente: Elaboración propia. Análisis de datos prueba piloto español 11° grado BCH, forma "B" (2018)

Las medidas de discriminación de los reactivos de la sub prueba "B", indican que tan adecuadamente separa o diferencia un reactivo entre los estudiantes que tienen un alto rendimiento en el test, respecto a los que lo tienen bajo. Ningún elemento se ubicó bajo la categoría de inaceptable (0-0.20), sino que aun aquellos reactivos que puntuaron con alta dificultad, resultaron con índice de discriminación alta.

5.2.3 Funcionamiento Diferencial de los ítems : metodo Mantel –Haenszel generalizado

Como punto de partida para el análisis de sesgo de los ítems y aseguramiento de la validez de las puntuaciones, se realizó el análisis de las diferencias inesperadas en el desempeño de los ítems (DIF), a través del método Mantel-Haenszel, identificando dos sub grupos de interés, hombres-mujeres, tomando como grupo referencia las mujeres y focal los hombres.

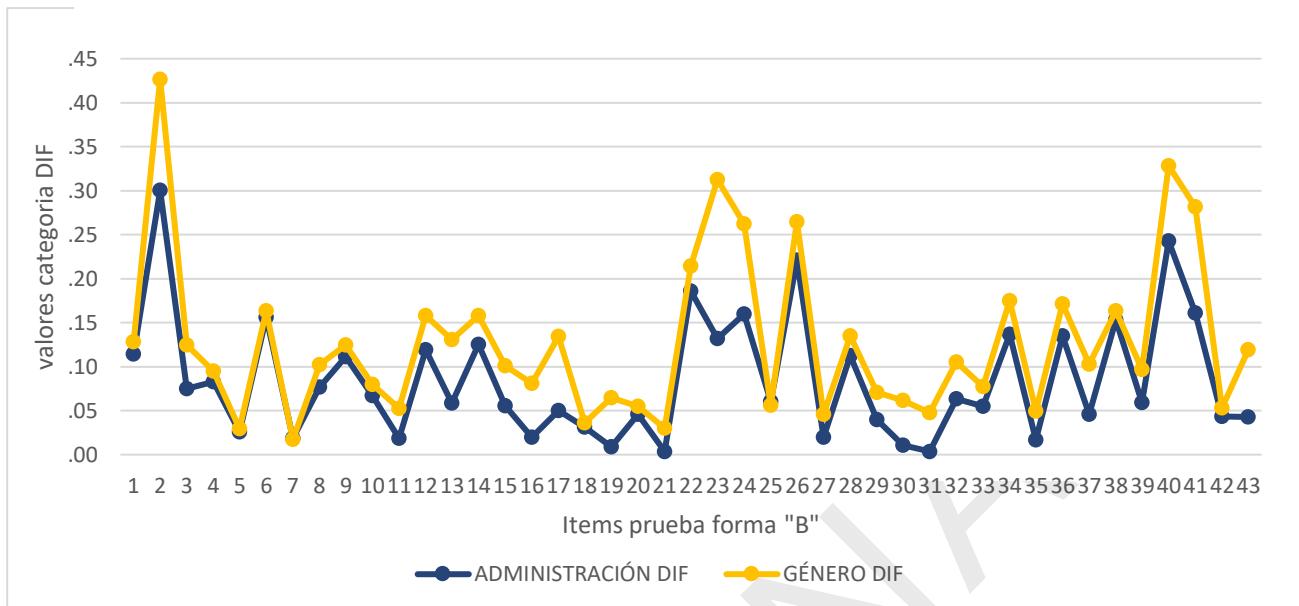
El otro conjunto de datos analizado se realizó segmentando la población de estudio de acuerdo al tipo de administración de los centros educativos, teniendo como grupo focal los centros educativos de educación media oficiales y como referencia los centros educativos no gubernamentales. El estadístico MH generalizado, presentado con anterioridad en la tabla 16, clasifica los ítems en tres grupos o categorías A: no hay grupo favorecido (0 - 0.05), B (0.05 -1): uno de los grupos es ligeramente favorecido y grupo C(+1.00), uno de los grupos se ve muy favorecido. La figura 7, permite representar con mayor claridad la variabilidad de las puntuaciones de este estadístico.

Grafico 7 Funcionamiento Diferencial de los ítems, Prueba piloto forma "A"

Fuente: Elaboración propia. Análisis de datos prueba piloto de español undécimo grado BCH forma "A"

En la prueba piloto de español forma "A", el estadístico MH generalizado, según el sexo de los estudiantes dio como resultado 2 ítems en el grupo "C", dando como supuesto que estos favorecen estadísticamente al grupo de referencia que son las mujeres. Por otra parte, en cuanto al DIF por tipo de administración de los centros educativos 14 elementos estadísticamente estarían favoreciendo a los estudiantes de institutos no gubernamentales. Al realizar la revisión en contenido y redacción de los ítems encontrados con DIF mencionados anteriormente, no se encuentra mayor evidencia de sesgo en los mismos, sin embargo, cuatro de estos ítems considerando que los índices en dificultad y discriminación presentan categorías inaceptables, se excluyeron para ser parte de la prueba operativa.

La figura 5, representa las puntuaciones del DIF, de los ítems de la sub prueba piloto "B", donde se permite apreciar que estadísticamente se encontraron diferencias en el desempeño de los ítems analizados según la administración de los centros educativos, ya que 16 elementos, se ubicaron en la categoría "C" (Alto), lo que significaría que estos ítems están siendo mejor contestados por los estudiantes de los institutos no gubernamentales.

Grafico 8 Funcionamiento Diferencial de los Items prueba piloto forma "B"

Fuente: Elaboración propia. Prueba piloto de español forma "B" (2018)

Para determinar el sesgo de estos elementos, se revisó los índices discriminación y dificultad constatando que 12 ítems se encuentran en categoría aceptable y al revisar la pertinencia y coherencia del contenido, no se establece evidencia suficiente para determinar su calificación como ítems sesgados. Por otra parte 4 elementos correspondientes a las posiciones 12, 24, 28 y 40 de la sub prueba "B", analizando la estadística en parámetros en dificultad, discriminación, y que además sugieren que favorecen mediamente a las mujeres (B), se determinó excluirlos de la prueba final, aunque en la revisión de su contenido no se demuestre fehacientemente la presencia de sesgo en los mismos.

En ambas sub pruebas piloto formas "A" y "B", se hizo una revisión exhaustiva acerca de la posibilidad de sesgo en los ítems específicamente en el análisis de los resultados por tipo de centro educativo, lo que permitió concluir que, si bien este estadístico es un punto de partida para el estudio de sesgo de los ítems, en este caso, no resulto suficiente evidencia debido a que dichos elementos presentan los parámetros de dificultad y correlación aceptables, además de un contenido, redacción y nivel cognitivo adecuado.

5.3 Resultados diseño y administración de la prueba operativa

Una vez seleccionados los ítems con mejores estadísticas y de acuerdo a la cobertura curricular propuesta por los especialistas del área; se estructuró la prueba operativa, para

seguidamente llevar a cabo el proceso de aplicación de la misma, con ello dando respuesta al objetivo número tres de esta investigación, referente al diseño de una prueba que asegurase validez y confiabilidad en la medición del desempeño académico en el área de español-lectura. La prueba se aplicó en 3 centros educativos oficiales y 3 no gubernamentales, participando 288 mujeres y 153 hombres, que hizo un total de 441 estudiantes del undécimo grado del Bachillerato en Ciencias y Humanidades del Distrito educativo N° 9.

Este apartado inicia con las pruebas de confiabilidad y validez del instrumento, para luego mostrar un estudio descriptivo de los resultados obtenidos que incluye las pruebas de normalidad de los datos; consecutivamente se detalla el comportamiento de los ítems en la prueba operativa a través de los parámetros psicométricos de discriminación, dificultad, así como el análisis de sesgo a través del DIF. Continúa la presentación de resultados con la comprobación de hipótesis del estudio a través de las pruebas T Student; finalizando el capítulo, con el análisis del rendimiento académico mostrados por los estudiantes en el área de español a través de las métricas escala estandarizada (100-500) anclada a los niveles de desempeño, concluyendo los análisis con el porcentaje de respuestas correctas.

5.3.1. Confiabilidad y Validez de la prueba operativa

La confiabilidad del instrumento se analizó a través del estadístico alfa de Cronbach (α), obteniéndose un alfa de .97 el cual refleja la consistencia interna de la prueba a partir de la variación entre sus 50 ítems respecto a la varianza total del test. ver tabla 17.

Tabla 17 Índice de confiabilidad prueba operativa español 11mo grado

Estadístico Fiabilidad	Alfa de Cronbach basada en los elementos tipificados	N de elementos
.971	.972	50

Fuente: Resultados prueba operativa de español 11mo grado BCH (2018)

Obtener una puntuación alta en el coeficiente de Cronbach, no hizo necesario la reducción de ítems, para efectos de mejorar el índice de confiabilidad según el método de consistencia interna.

La validez de contenido del instrumento se determinó a través de la definición del marco operativo de la prueba operativa, que mostro de acuerdo a los especialistas del área de español una cobertura curricular optima, en cuanto a la pertinencia del contenido de los ítems para la medición del desempeño académico en este campo del conocimiento. Este tipo de validez no tiene un coeficiente estadístico fijo, sino que este criterio depende de la objetividad del instrumento, así como de la muestra seleccionada.

5.3.2 Estudio descriptivo y distribución normal de los datos de la prueba operativa

Pruebas de normalidad de los datos

A través del programa SPSS, se realizó la prueba de bondad de ajuste de Kolmogorov Smirnov, para determinar la normalidad en la distribución de los datos. En la tabla 14 se muestra el valor del estadígrafo (Z de Kolmogorov-Smirnov), lo que dio por resultado 1.789, con un valor p (Sig. Asintót. bilateral) de 0,03.

Tabla 18 Prueba de normalidad de los datos de la prueba operativa

Prueba de Kolmogorov-Smirnov para una muestra		
Parámetros normales ^{a,b}	Media	29.11
	Desviación típica	6.956
Diferencias más extremas	Absoluta	.085
	Positiva	.048
	Negativa	-.085
Z de Kolmogorov-Smirnov		1.789
Sig. asintót. (bilateral)		.003

Fuente: Análisis de datos programa SPSS

Nota: a. la distribución de contraste es normal. b. se han calculado a partir de los datos

Como el valor (p) fue menor al valor de significancia de 0,05, se rechaza la hipótesis nula (H_0) y se concluye que los datos no presentan suficiente evidencia para afirmar que siguen una distribución normal. Para ampliar y visualizar el análisis de la normalidad de los datos, la tabla 14, permite observar los descriptivos obtenidos en las medidas de tendencia central y dispersión, tales como la media, mediana, varianza, así como el valor de la asimetría y curtosis, los que reflejan datos importantes de acotar para establecer con mayor precisión el comportamiento o variabilidad de los datos recabados.

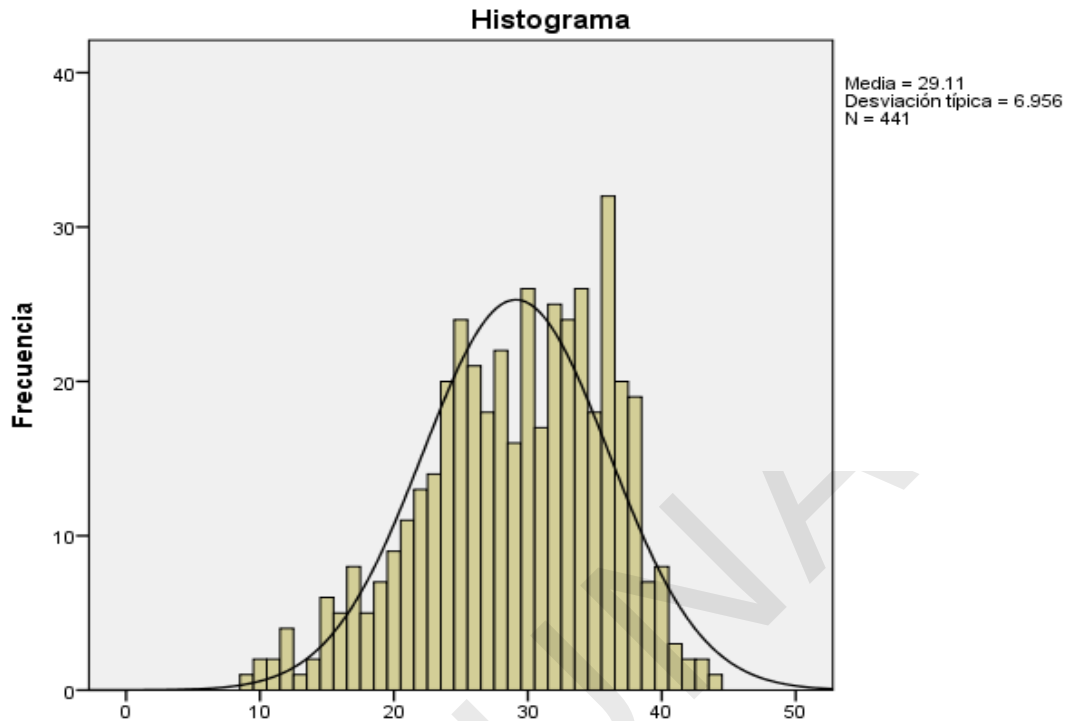
Tabla 19 Estadísticos descriptivos según los puntajes totales de la prueba operativa

Medidas	Valores
N validos	441
N perdidos	0
Media	29.11
Error típ. de la media	.331
Mediana	30.00
Moda	36
Desv. típ.	6.956
Varianza	48.384
Asimetría	-.472
Error típ. de asimetría	.116
Curtosis	-.282
Error típ. de curtosis	.232
Rango	35
Puntaje mínimo	9
Puntaje máximo	44

Fuente: Resultados prueba de español-lectura, 11mo grado BCH (2018)

La estadística descriptiva de los datos permite observar que la media de estos fue de 29, lo que indica que en promedio los estudiantes contestaban 29 preguntas correctas de la prueba, mientras que la mediana de los datos registro un valor de 30, que significa que la mitad de los 441 estudiantes evaluados contestaron entre nueve (puntaje mínimo) y treinta ítems acertadamente. La medida del rango de los datos fue de 35, lo que supone una alta dispersión de los datos.

El valor de la asimetría resulto negativo $-.472$, de forma que siendo menor que cero, se confirma que la distribución de los datos es asimétricamente negativa y los datos se tienden a reunir más en la parte derecha de la media, lo que se puede comprobar al observar el valor de la moda que es 36; en cuanto al coeficiente de la curtosis, con una medida $-.282$ se evidencia que hubo una ligera concentración de los datos, por lo que se clasifica como una curtosis platicúrtica. El histograma del gráfico 3, representa visualmente la distribución de los datos.

Grafico 9 Histograma normalidad de los datos prueba operativa Español 11mo grado BCH

Fuente: Elaboración propia. Análisis de datos prueba de español-lectura 11mo grado BCH

5.3.3 Comprobación de hipótesis del estudio

Las hipótesis planteadas se comprobaron a través de la comparación de medias por medio de la prueba “T de Student”: Comparación de medias según el sexo de los estudiantes

H_0 = La media del puntaje total de los estudiantes de sexo femenino es menor que la media del sexo masculino.

H_1 = La media del puntaje total de los estudiantes de sexo femenino es mayor que la media del sexo masculino.

Tabla 20 Prueba comparación de medias según variable sexo: T de Student

	Prueba de Levene		Prueba T para la igualdad de medias						
	F	Sig.	t	gl	Sig. (bilateral)	Diferencia de medias	Error típ. de la diferencia	95% Intervalo de confianza para la diferencia	
								Inferior	Superior
Se han asumido varianzas iguales	0.083	0.774	-0.954	439	0.341	-0.664	0.696	-2.032	0.704

No se han asumido varianzas iguales	-0.962	316.886	0.337	-0.664	0.691	-2.023	0.695
-------------------------------------	--------	---------	-------	--------	-------	--------	-------

Fuente: Elaboración propia. Análisis de datos prueba operativa

La tabla 20, muestra los resultados más importante en la fila superior “se asumido varianzas iguales”, donde la prueba de Levene con una valor ($p = 0,774$), permite asumir la homogeneidad de varianzas en las muestras al ser mayor que el valor alfa de 0.05, seguidamente el estadístico de la t de Student observa una medida de -9.54 (con 43 grados de libertad) y un valor “p” asociado (significancia bilateral) de 0.341, mayor al valor alfa de 0,05, lo que significa que no se han encontrado diferencias estadísticamente significativas por lo que se acepta la hipótesis nula y se asume que no existe diferencias significativa entre la media del puntaje total de los estudiantes de sexo femenino en relación al sexo masculino.

Se muestra los descriptivos que señalan que la media obtenida por el sexo femenino fue 28.88 y para el sexo masculino fue 29.54, teniendo una diferencia de 0.66 lo que reafirma lo señalado en la tabla de la comparación de medias “T Student”.

Tabla 21 Descriptivos puntaje total de la prueba operativa por sexo

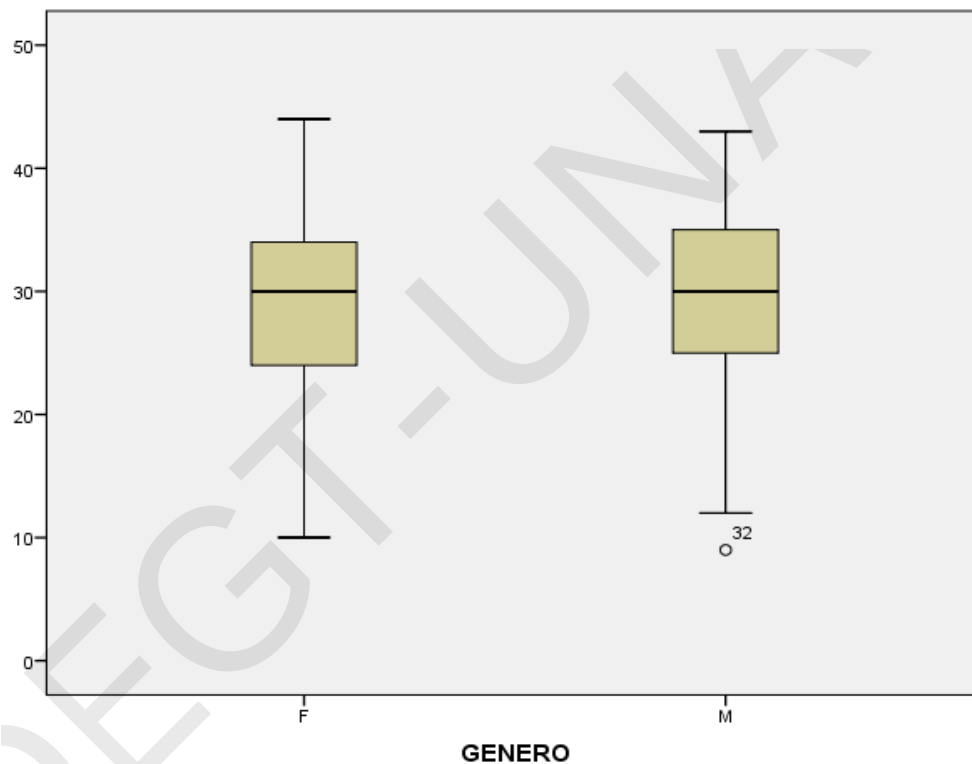
	Sexo	Estadística	Sesgo	Típ. Error	Intervalo de confianza al 95%	
					Inferior	Superior
F	Media	28.88	.00	.21	28.45	29.30
	N	288		10	269	307
	Desv. típ.	7.016	-.006	.117	6.775	7.240
	Curtosis	-.282	.006	.091	-.442	-.082
	Asimetría	-.461	-.001	.052	-.568	-.368
	Mínimo	10				
	Máximo	44				
	Mediana	30.00	-.10	.30	29.00	30.00
	% del total de N	65.3%				
M	Media	29.54	-.01	.40	28.72	30.28
	N	153		10	134	172
	Desv. típ.	6.842	-.009	.236	6.374	7.302
	Curtosis	-.263	.010	.192	-.593	.192
	Asimetría	-.493	.002	.105	-.697	-.276
	Mínimo	9				
	Máximo	43				

Mediana	30.00	.12	.86	28.00	31.00
% del total de N	34.7%				

Fuente: Elaboración propia. Análisis de datos de la prueba Español-Lectura (2018) 11mo grado, BCH.
*F: Femenino, M: Masculino.

Para consolidar la hipótesis, se observa en el gráfico de cajas, la distribución de las puntuaciones para ambos sexos, donde el valor de la media del sexo femenino y masculino, no muestra una diferencia significativa.

Grafico 10 Gráfico comparación de medias segun sexo de los estudiantes prueba operativa 11° grado BCH, 2018



Fuente: Elaboración propia. Resultados análisis de los datos prueba Español 11° grado BCH 2018

Comparación de medias por tipo de administración de los centros educativos

Hipótesis 2

H_0 = La media del puntaje total de los estudiantes de institutos no gubernamentales es menor que la media de los estudiantes de institutos gubernamentales.

H_1 = La media del puntaje total de los estudiantes de institutos no gubernamentales es mayor que la media de los estudiantes de institutos gubernamentales.

Tabla 22 Prueba comparacion de medias segun tipo de administracion de los centros educativos. "T student"

	Prueba de Levene para la igualdad de varianzas		Prueba T para la igualdad de medias						
	F	Sig.	t	gl	Sig. (bilateral)	Diferencia de medias	Error típ. de la diferencia	95% Intervalo de confianza para la diferencia	
								Inferior	Superior
Se han asumido varianzas iguales	3.119	0.078	-4.797	439	0	-3.106	0.648	-4.379	-1.834
No se han asumido varianzas iguales			-4.826	438.9	0	-3.106	0.644	-4.371	-1.841

Fuente: Elaboración propia. Resultados prueba operativa Español-Lectura (2018) 11mo grado BCH

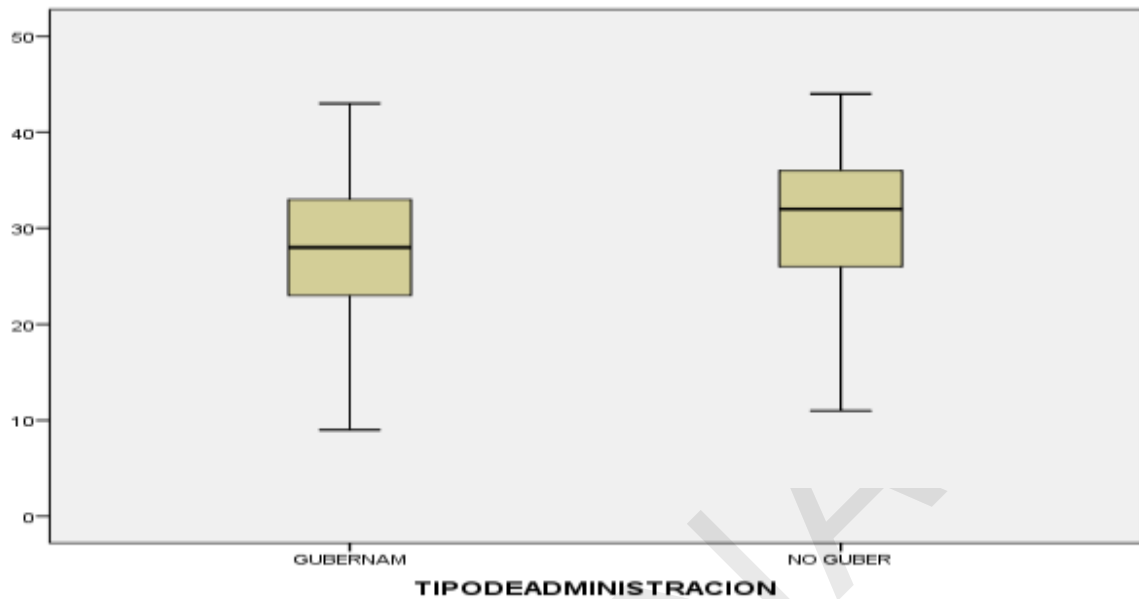
De la misma forma para el tipo de Administración de centros educativos, inicialmente se observa que la prueba de Levene ($p = 0,78$), presenta homogeneidad de varianzas, siendo mayor que el valor alfa de 0.05; en el caso del estadístico de la t de Student, tiene una medida de -4.797 con un valor “p” de significancia bilateral de 0,00 lo que indica que siendo igual que 0, se rechaza la hipótesis nula, estadísticamente existe una diferencia mínima entre la media de la variable puntaje total de los centros educativos gubernamentales respecto a los no gubernamentales. En la tabla 23, los estadísticos se observan de manera más puntual, la media en los institutos gubernamentales fue de 27.64 mientras que en los no gubernamentales se registró un valor de 30.75, lo que se puede apreciar mejor en el gráfico de cajas.

Tabla 23 Descriptivos prueba "T Student"

		Bootstrap ^a				
TIPODEADMINISTRACION	Statistic	Sesgo	Error	Típ.	Intervalo de confianza al 95%	
					Inferior	Superior
GUBERNAMENTAL	Media	27.64	.00	.00	27.64	27.64

NO GUBERNAMENTAL	N	233	0	0	233	233
	Desv. típ.	7.118	.000	.000	7.118	7.118
	Curtosis	-.470	.000	.000	-.470	-.470
	Asimetría	-.341	.000	.000	-.341	-.341
	Mínimo	9				
	Máximo	43				
	Mediana	28.00	0.00	0.00	28.00	28.00
	% del total de N	52.8%				
	Media	30.75	.00	.00	30.75	30.75
	N	208	0	0	208	208
	Desv. típ.	6.398	.000	.000	6.398	6.398
	Curtosis	.046	.000	.000	.046	.046
	Asimetría	-.587	.000	.000	-.587	-.587
Total	Mínimo	11				
	Máximo	44				
	Mediana	32.00	0.00	0.00	32.00	32.00
	% del total de N	47.2%				
	Media	29.11	.00	.00	29.11	29.11
	N	441	0	0	441	441
	Desv. típ.	6.956	.000	.000	6.956	6.956
	Curtosis	-.282	.000	.000	-.282	-.282
	Asimetría	-.472	0.000	.000	-.472	-.472
	Mínimo	9				
	Máximo	44				
	Mediana	30.00	0.00	0.00	30.00	30.00
	% del total de N	100.0%				

Fuente: Elaboración propia

Grafico 11 Comparacion de medias variable tipo de administracion de los centros educativos

Fuente: Elaboración propia

5.3.4 Propiedades Psicométricas de la prueba operativa: Índices de discriminación y dificultad de los ítems

En la tabla 24, se muestra el coeficiente de discriminación punto biserial y el índice de dificultad observados en cada uno de los ítems de la prueba operativa.

Tabla 24 Parámetros psicométricos prueba operativa

Posición	Coeficiente Discriminación Correlación Biserial-Puntual	Índice Dificultad P-Valué
1	0.61	0.59
2	0.69	0.51
3	0.47	0.43
4	0.71	0.19
5	0.33	0.13
6	0.81	0.36
7	0.77	0.31
8	0.74	0.26
9	0.63	0.23
10	0.61	0.36
11	0.64	0.27

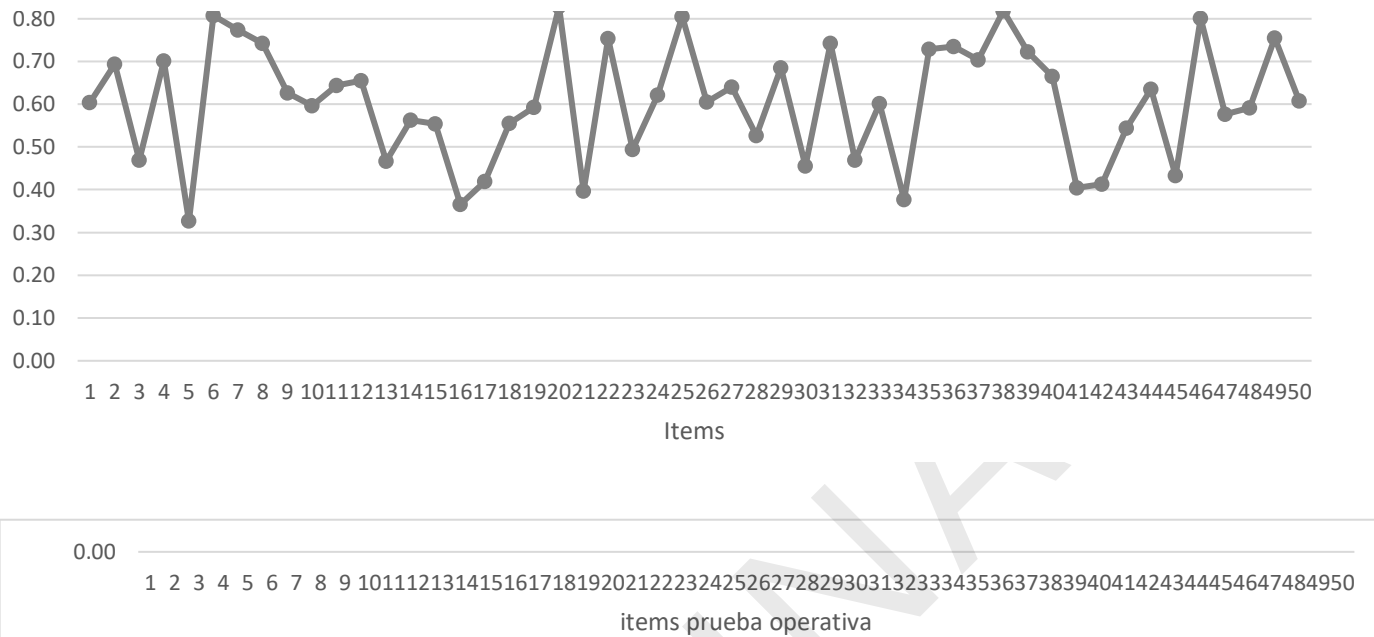
12	0.66	0.35
13	0.47	0.25
14	0.56	0.29
15	0.55	0.27
16	0.37	0.32
17	0.42	0.27
18	0.56	0.23
19	0.59	0.48
20	0.83	0.42
21	0.41	0.22
22	0.75	0.64
23	0.49	0.26
24	0.62	0.41
25	0.81	0.28
26	0.61	0.41
27	0.64	0.23
28	0.53	0.33
29	0.68	0.31
30	0.46	0.21
31	0.74	0.35
32	0.47	0.27
33	0.61	0.22
34	0.38	0.26
35	0.73	0.26
36	0.73	0.28
37	0.71	0.42
38	0.82	0.51
39	0.72	0.59
40	0.66	0.13
41	0.41	0.24
42	0.41	0.33
43	0.54	0.39
44	0.63	0.41
45	0.43	0.31
46	0.81	0.35
47	0.58	0.31
48	0.59	0.48
49	0.76	0.43
50	0.61	0.43

Fuente: Elaboración propia. Datos prueba Español Lectura 11mo grado BCH (2018)

El total de los ítems que conformaron la prueba, mostraron una correlación punto biserial positiva ≥ 3 , indicando que los estudiantes que contestaron correctamente cada ítem, también recibieron un alto puntaje en la prueba. Sin embargo, aunque siempre en la categoría aceptable, los elementos 5,16 y 34 del bloque de tipos de textos mostraron las correlaciones más bajas entre 0.33 y 0.38 con valores de dificultad bajos desde 0.13. Mientras que los ítems 6,20 y 25 del componente tipos de texto, y los reactivos 38 y 46 del componente vocabulario, son los que mejor discriminan por estar cercanos a 1, y con una categoría de dificultad aceptable.

En cuanto al índice de dificultad, los ítems 4, 5 y 30 del componente tipos de texto, y el ítem 40 del componente vocabulario se categorizan como los ítems más difíciles presentando porcentajes solamente entre el 13% y 19% de estudiantes respondieron acertadamente. El ítem con menor índice de dificultad se observó en el elemento de posición 22 correspondiente al componente tipos de texto donde el 64% de los estudiantes lo acertaron; registrando un coeficiente de discriminación de 0.75, que lo convierte en un ítem con propiedades psicométricas bastante aceptables.

Las siguientes ilustraciones muestran la distribución de los ítems según los parámetros psicométricos mencionados.

Grafico 12Índice de dificultad de los ítems de la prueba operativa 11° grado, BCH 2018

Fuente: Elaboración propia. Resultados del análisis psicométrico de los ítems Prueba Español-Lectura 11mo

5.3.5 Estudio del Funcionamiento Diferencial de los Ítems (DIF) de los ítems de la Prueba Operativa

El procedimiento utilizado para la detección del Funcionamiento Diferencial de los ítems DIF, al igual que en la prueba piloto, se realizó a través del procedimiento Mantel-Haenszel, que permitió identificar diferencias en las respuestas a los ítems, según el sexo de los estudiantes y de acuerdo al tipo de administración de los centros educativos de los cuales provenían. Los ítems se clasificaron en tres grupos o categorías A: no hay grupo favorecido, B: uno de los grupos es ligeramente favorecido y grupo C, uno de los grupos se ve muy favorecido.

En los resultados se encontró al ser ordenados de acuerdo al puntaje total de la prueba, en el caso del cálculo de DIF por tipo de administración de los centros educativos; 22 ítems se registraron en categoría (A), 21 en categoría (B) y en categoría (C) 7 reactivos. Por otra parte, los resultados de DIF de acuerdo al sexo de los estudiantes, resulto en categoría (A) 28 ítems, categoría (B) 14 ítems, y en la categoría C, se ubicaron 8 ítems. La tabla 13, detalla con más precisión.

Tabla 25 Funcionamiento Diferencial de los Items (DIF) Prueba Operativa

Posición	DIF Sexo	Categoría	DIF Administración	Categoría
1	0.05	A	-0.1	B
2	0.04	A	-0.07	B
3	0.07	B	-0.05	A
4	0.01	A	-0.05	A
5	-0.11	C	-0.03	A
6	-0.08	B	0.06	B
7	0.09	B	-0.12	C
8	-0.08	B	0.05	A
9	0.12	C	0.01	A
10	0.05	A	-0.04	A
11	0.04	A	0.02	A
12	-0.08	B	-0.1	B
13	0.13	C	0	A
14	0.07	B	0.03	A
15	-0.06	B	0.19	C
16	0.03	A	0.1	B
17	-0.02	A	0.11	C
18	-0.06	B	0.07	B
19	-0.11	C	0.04	A
20	0	A	0.05	A
21	0.03	A	0.06	B
22	-0.04	A	0.08	B
23	0	A	0.1	B
24	-0.06	B	0.09	B
25	-0.01	A	-0.03	A
26	0.04	A	-0.01	A
27	-0.01	A	0.01	A
28	0.1	B	-0.13	C
29	0.01	A	0.03	A
30	0.15	C	-0.07	B
31	-0.14	C	0.1	B
32	0.03	A	0.02	A
33	-0.03	A	-0.08	B
34	0.02	A	-0.05	A

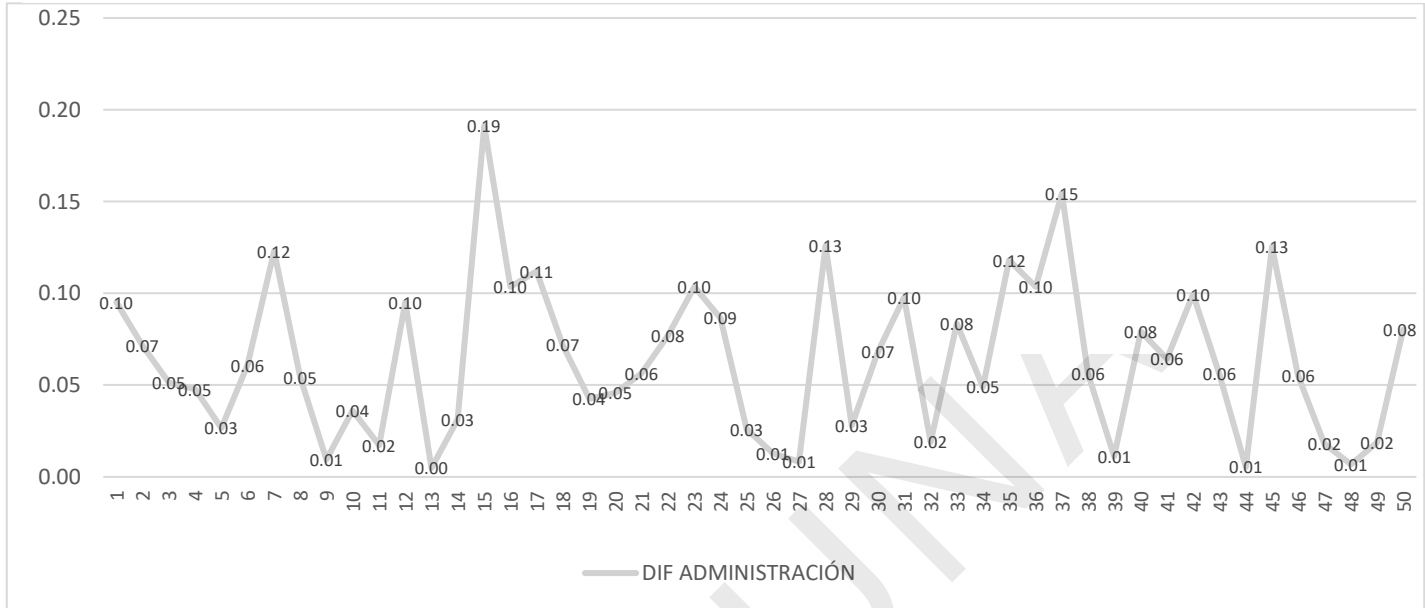
35	0.06	B	-0.12	C
36	0.03	A	-0.1	B
37	0.14	C	-0.15	C
38	-0.03	A	-0.06	B
39	0	A	-0.01	A
40	0.01	A	0.08	B
41	0	A	-0.06	B
42	-0.06	B	0.1	B
43	-0.12	C	-0.06	B
44	-0.01	A	0.01	A
45	0.03	A	-0.13	C
46	-0.07	B	0.06	B
47	0.09	B	-0.02	A
48	-0.04	A	0.01	A
49	-0.02	A	0.02	A
50	0.01	A	-0.08	B

Fuente: elaboración propia, resultados prueba Español-lectura 11mo grado BCH (2018)

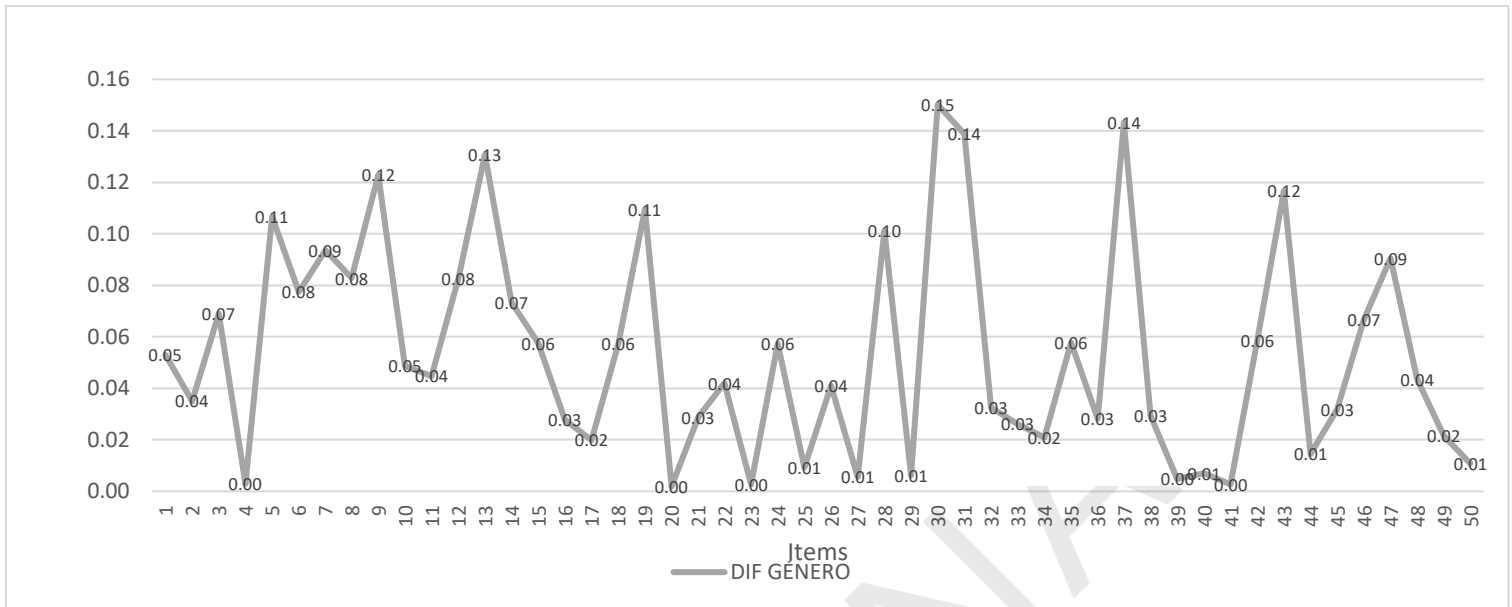
Haciendo el análisis, del estadístico Mantel-Haenszel, podemos resaltar que el ítem en la posición 37, observa categoría “C” en ambos sub grupos de estudio, lo que podría suponer que el contenido de este, favorece sistemáticamente a las mujeres y a los estudiantes de centros educativos no gubernamentales. Sin embargo, al revisar el contenido no existe suficiente evidencia para determinar un posible sesgo en dicho elemento, considerando además que tanto la correlación biserial-puntual, y el índice de dificultad son aceptables.

Las figuras siguientes, permiten visualizar la detección del DIF según el sexo de los estudiantes y tipo de administración de los centros educativos.

Grafico 13 Analisis del Funcionamiento Diferencial de los ítems prueba operativa segun Tipo de Administracion de los centros educativos



Fuente: elaboración propia, resultados prueba Español-lectura 11mo grado BCH (2018)

Grafico 14 Funcionamiento Diferencial de los ítems según sexo de los estudiantes

Fuente: Elaboración propia. Análisis de los ítems prueba operativa 11° grado BCH 2018

5.4 Resultados de desempeño académico: prueba español(lectura) 11° grado BCH.

Desde una perspectiva eminentemente metodológica, se presenta en este apartado la transformación de las puntuaciones directas obtenidas por los estudiantes parte de la muestra de este estudio. en la escala de puntuación estandarizada, permitió inicialmente identificar los niveles de desempeño alcanzados por los estudiantes. La tabla 21, detalla con precisión los valores encontrados.

Tabla 26 Puntuaciones directas y escala de puntuacion estandarizada (Scale Score) Prueba Operativa Español 11° grado BCH

Puntuación directa	Frecuencia absoluta	Porcentaje acumulado	Puntaje z	Escala puntuación estandarizada
9	1	0.2	-2.9	155
10	2	0.7	-2.8	160
11	2	1.1	-2.6	170
12	4	2	-2.5	175
13	1	2.3	-2.3	185
14	2	2.7	-2.2	190
15	6	4.1	-2	200

16	5	5.2	-1.9	205
17	8	7	-1.8	210
18	5	8.2	-1.6	220
19	7	9.8	-1.5	225
20	9	11.8	-1.3	235
21	11	14.3	-1.2	240
22	13	17.2	-1	250
23	14	20.4	-0.9	255
24	20	24.9	-0.7	265
25	24	30.4	-0.6	270
26	21	35.1	-0.5	275
27	18	39.2	-0.3	285
28	22	44.2	-0.2	290
29	16	47.8	0	300
30	26	53.7	0.1	305
31	17	57.6	0.3	315
32	25	63.3	0.4	320
33	24	68.7	0.6	330
34	26	74.6	0.7	335
35	18	78.7	0.9	345
36	32	85.9	1	350
37	20	90.5	1.1	355
38	19	94.8	1.3	365
39	7	96.4	1.4	370
40	8	98.2	1.6	380
41	3	98.9	1.7	385
42	2	99.3	1.9	395
43	2	99.8	2	400
44	1	100	2.2	410
50			3	450
Total	441			

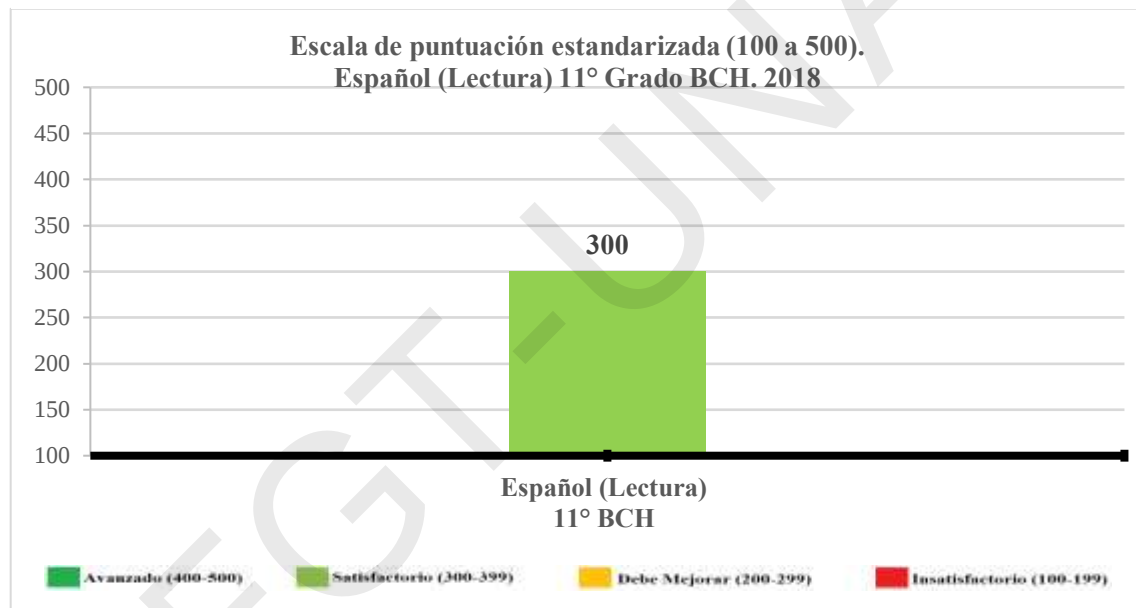
Fuente: Elaboración propia. Transformación de puntuaciones directas obtenidos en la prueba de Español 11mo grado BCH 2018-2019.

Los valores negativos de Z, de acuerdo a la frecuencia acumulada permite observar que 195 estudiantes contestaron correctamente entre 9 y 28 ítems, es decir por debajo de la media, que fue de 29, mientras que 246 participantes obtuvieron un numero de preguntas correctas

igual o mayor que la media, entre 29 y 44 preguntas acertadas, siendo este último (44) el puntaje máximo obtenido en la prueba, con la frecuencia de una persona.

La última columna muestra la métrica de puntuación estandarizada o scale score, esta escala convirtió los puntajes brutos de los estudiantes (el número real de preguntas que contestó correctamente), en una puntuación de escala común (100 a 500). A partir de los valores positivos de Z, mayores a 29, tal y como se indicó anteriormente se ubicaron 246 participantes, lo que significa que en la métrica scale score, dichos estudiantes obtuvieron un puntaje igual o mayor a 300 puntos.

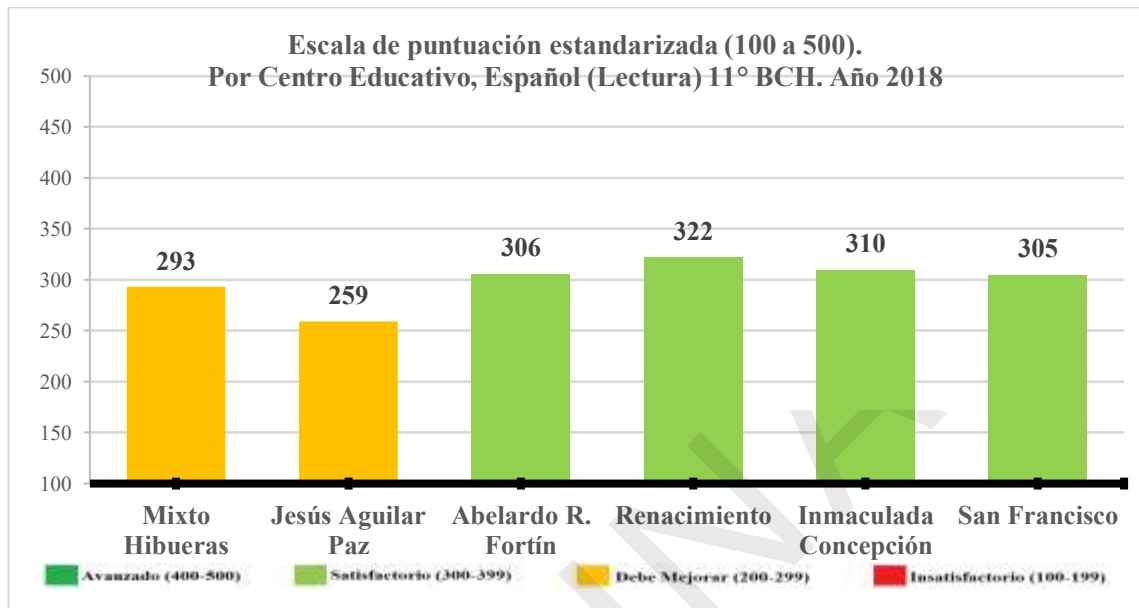
Grafico 15 Escala de puntuacion Estandarizada (100-500) Español-Lectura 11mo grado BCH 2018



Fuente: elaboración propia

En español (lectura) en la escala de puntuación estandarizada Scale Score, con datos agregados de los seis centros educativos, los estudiantes se agruparon en el nivel de desempeño satisfactorio con 300 puntos respectivamente.

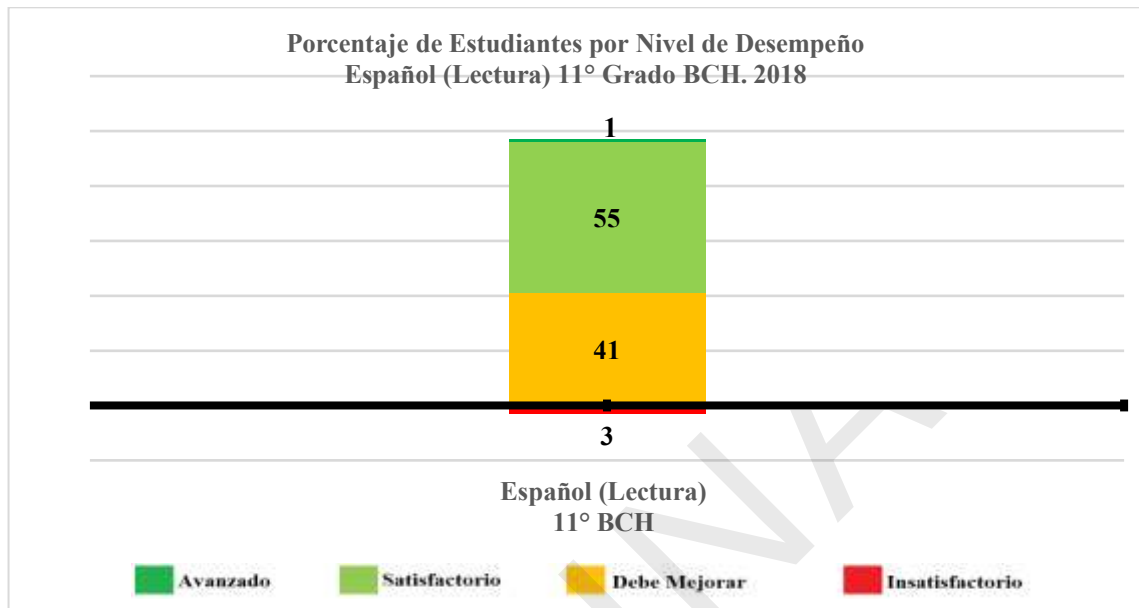
Grafico 16 Escala de Puntuacion estandarizada (100 a 500) por centro educativo Español (Lectura) 11mo BCH 2018



Fuente: elaboración propia

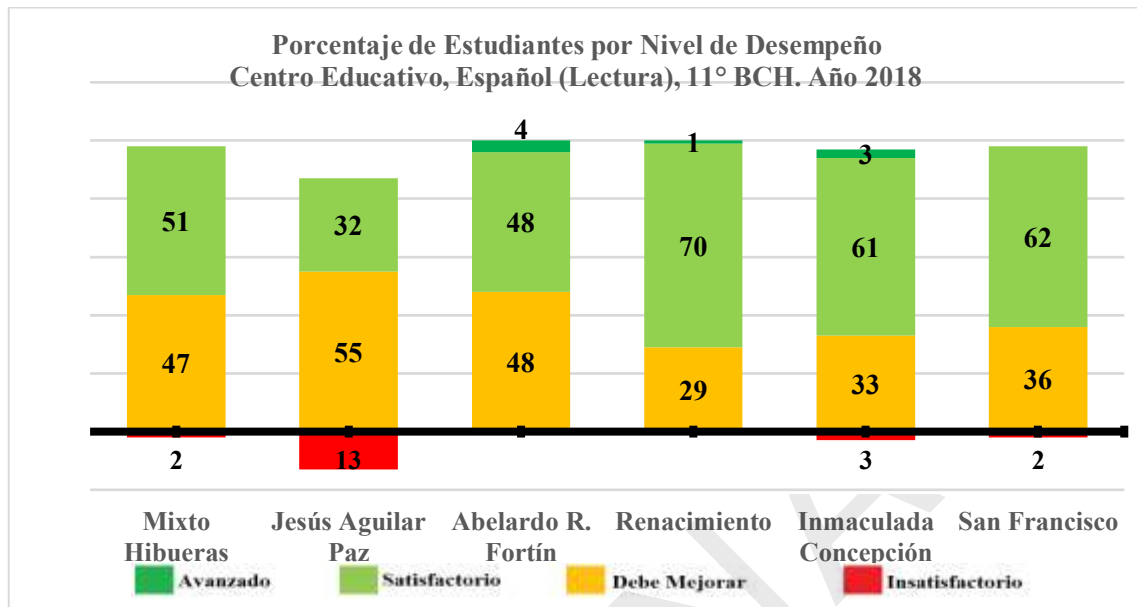
En los resultados desagregados por centro educativo según la escala de puntuación estandarizada (100 a 500), cuatro de los centros educativos parte de este estudio, se agruparon en el nivel satisfactorio, mientras que el Instituto Mixto Hibueras e Instituto Jesús Aguilar, ambos centros educativos oficiales reflejaron los resultados más bajos, en el nivel debe mejorar.

Grafico 17 Porcentaje de estudiantes por nivel de desempeño Español (Lectura) 11mo grado BCH 2018.



Fuente: Elaboración propia

Los resultados correspondientes a la métrica niveles de desempeño, en español (lectura) permite observar que más de la mitad (56%) de los estudiantes se encuentran en los niveles satisfactorio y avanzado, por otra parte, el 41% alcanzaron resultados en el nivel debe mejorar, 3% de los estudiantes evaluados están en el nivel insatisfactorio.

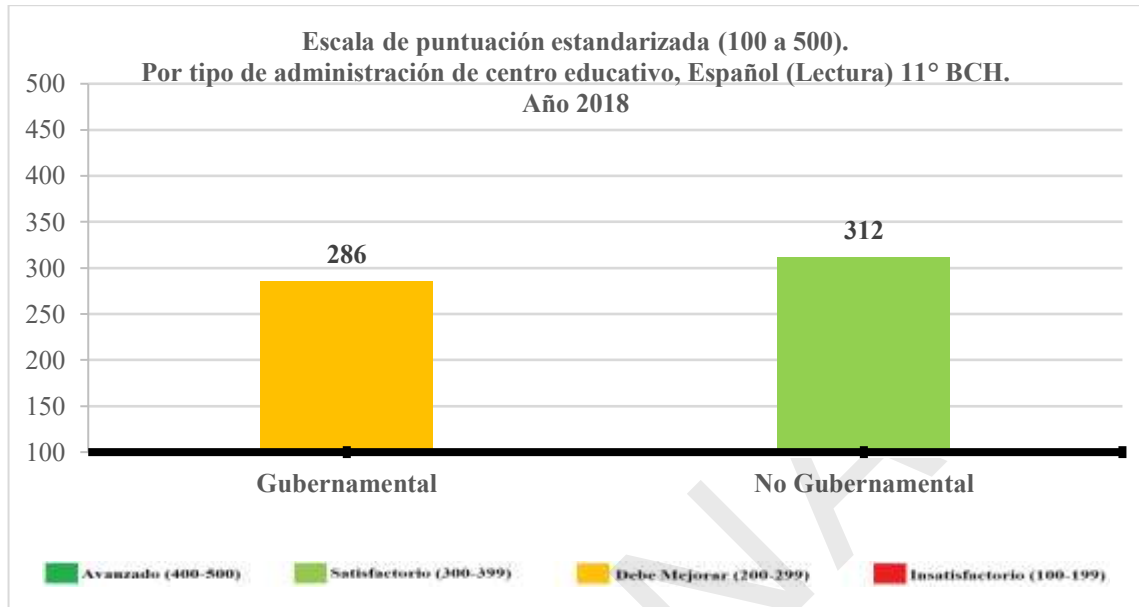
Grafico 18 Nivel de desempeño por centro educativo Español Lectura 11mo grado BCH 2018

Fuente: Elaboración propia

En la prueba Español (lectura) 11mo grado BCH, los estudiantes de los centros educativos no gubernamentales mostraron los mejores resultados, más del 60% de la población participante se ubicaron en el nivel de desempeño satisfactorio, mientras que, de los centros educativos oficiales, solamente entre el 32% y 51% de los estudiantes se agruparon en este nivel.

Por su lado, el 68% de la población participante del Centro de Educación Media Gubernamental Jesús Aguilar Paz, se agruparon en los niveles de más bajo desempeño; así mismo en los institutos oficiales Mixto Hibueras y Abelardo R. Fortín, un poco menos de la mitad de los estudiantes evaluados, como se observa en el grafico 5; entre el 47% y 49%, se concentraron en los niveles de desempeño debe mejorar e insatisfactorio.

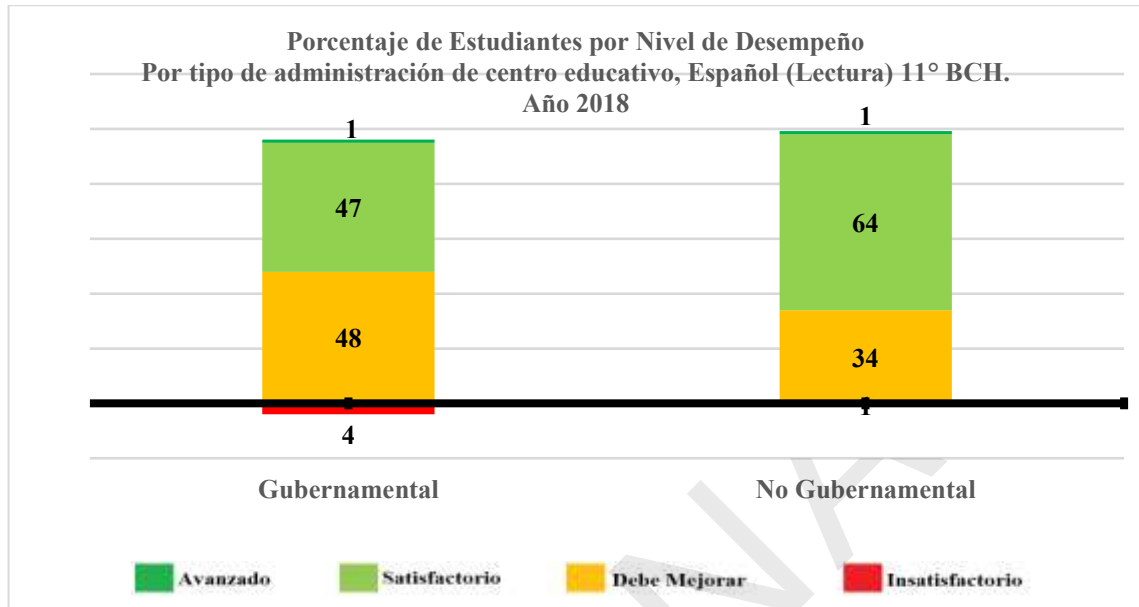
Grafico 19 Escala de puntuación estandarizada (100 a 500) Español Lectura 11mo BCH 2018, segun tipo de administracion de los centros Educativos (Datos agregados)



Fuente: Elaboración propia

Según la Escala de puntuación Estandarizada en datos agregados, el desempeño académico de los estudiantes de los centros educativos oficiales en la prueba español (lectura) 11mo grado BCH, se agruparon en el nivel de desempeño debe mejorar, siendo inferior al resultado mostrado por los estudiantes de los Centros Educativos de Educación Media no gubernamentales, quienes obtuvieron un puntaje de 312 puntos, que los ubica en el nivel de desempeño satisfactorio.

Grafico 20 Porcentaje de estudiantes por nivel de desempeño Español Lectura 11mo BCH 2018, segun el tipo de administracion de los centros educativos

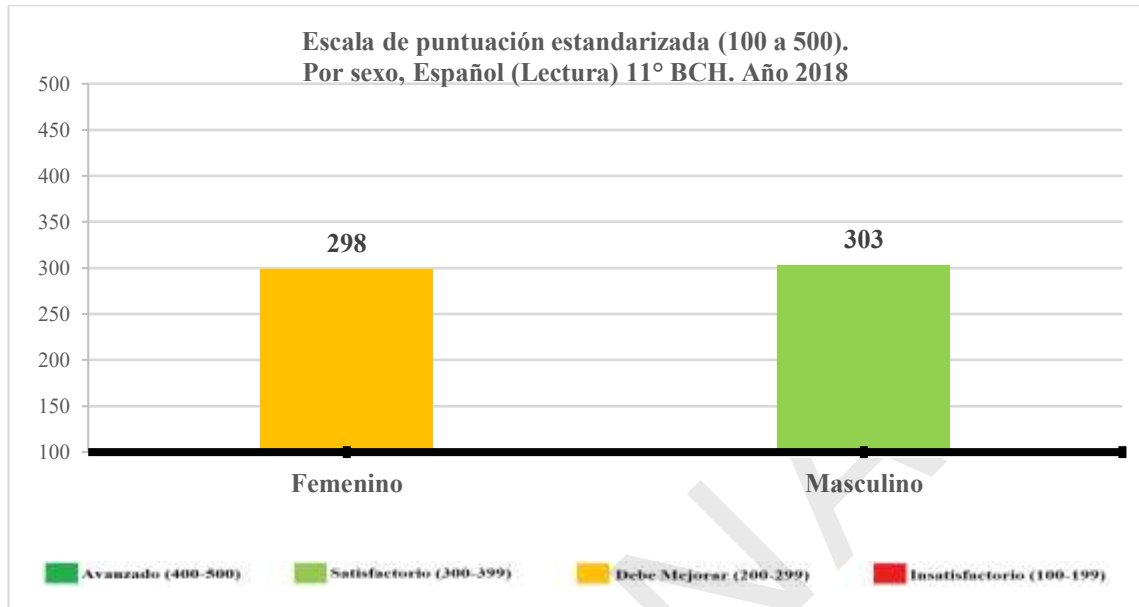


Fuente: elaboración propia

El gráfico 7, muestra el nivel de desempeño obtenido por los estudiantes de centros educativos oficiales y no gubernamentales, donde un 64% de los estudiantes de centros educativos no gubernamentales se agruparon en el nivel satisfactorio, siendo superior en comparación al nivel de desempeño de los estudiantes de centros educativos oficiales, con un 47% agrupados en este nivel de desempeño esperado.

Se observa además que el 33% de los estudiantes de institutos no gubernamentales se agrupó entre los niveles debe mejorar y únicamente el 1% en el nivel de desempeño insatisfactorio, mientras que el 48% de los estudiantes de centros educativos oficiales, que representa un poco menos de la mitad de los estudiantes evaluados, se agruparon en el nivel debe mejorar, y un 4% en el nivel más bajo (insatisfactorio).

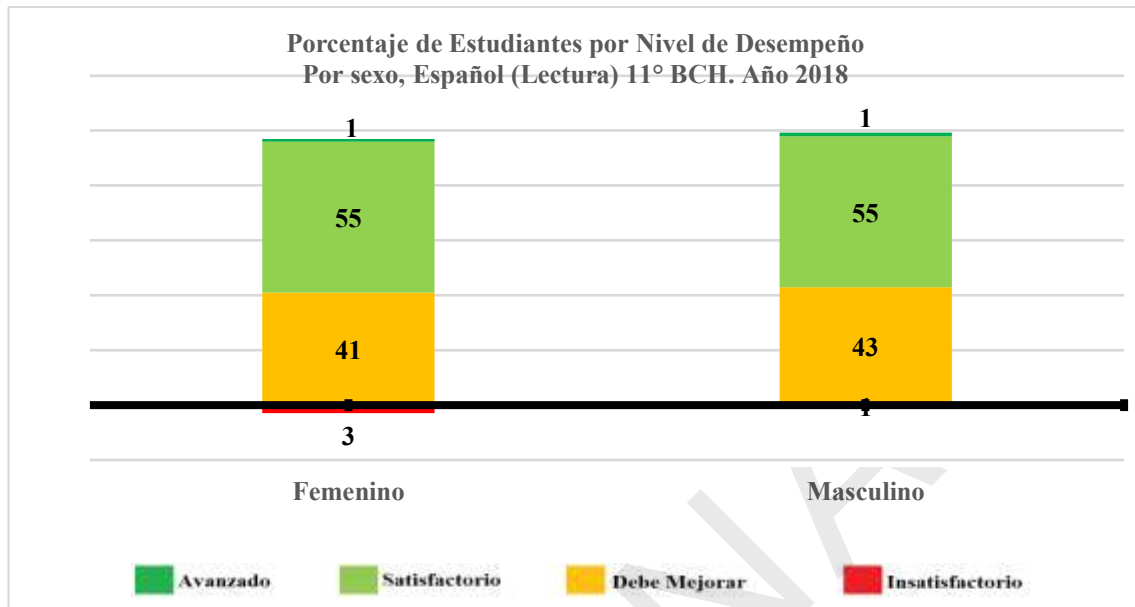
Grafico 21 Escala de puntuación Estandarizada (100 a 500) según Sexo, Español-Lectura 11mo grado BCH 2018 (Datos agregados)



Fuente: Elaboración propia

En la español (lectura), 11mo grado BCH, en la escala de puntuación estandarizada según sexo, el grafico 8, permite observar, que los estudiantes de sexo masculino obtuvieron 303 puntos, lo que indica un desempeño levemente mejor que las estudiantes de sexo femenino con una puntuación de 298 y aunque estadísticamente la diferencia fue solamente de 5 puntos, en la métrica niveles de desempeño fue significativa, porque debido a la puntuación obtenida, los hombres se ubicaron en el nivel satisfactorio y las mujeres en el nivel debe mejorar.

Grafico 22 Porcentaje por nivel de desempeño segun sexo Español (Lectura) 11mo grado BCH 2018



Fuente: elaboración propia.

El 56% de los estudiantes de sexo masculino y femenino participantes en este estudio se agruparon en los niveles satisfactorio y avanzado. En el nivel debe mejorar, se agrupó el 41% de las mujeres y el 43% de los hombres. Así mismo el grafico 9, presenta el porcentaje de estudiantes en el nivel más bajo (insatisfactorio) donde un 3% de las mujeres se ubicaron en este nivel mientras que de los hombres solamente el 1% se ubicó en este nivel respectivamente.

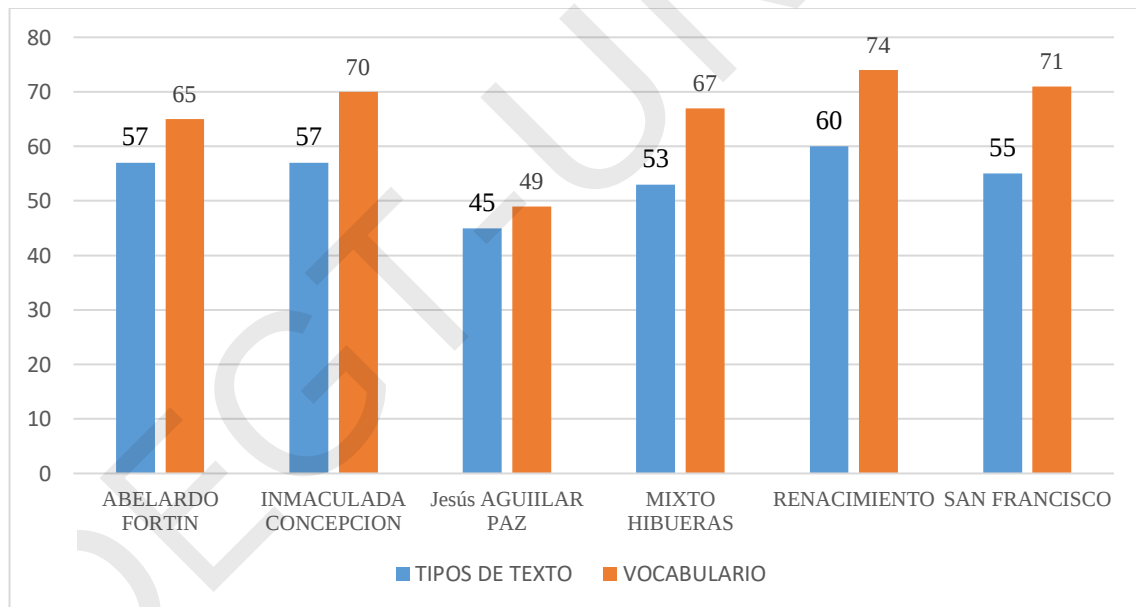
Porcentaje de respuestas correctas en relación con los bloques y componentes del área curricular español

Este nivel de detalle permite identificar en cuales bloques de contenido y componentes se presentaron las mayores fortalezas y debilidades cognitivas por parte de los estudiantes participantes en este estudio. La tabla 21, muestra el rendimiento porcentual por componente, obtenido por los estudiantes de los seis centros educativos evaluados, siendo los mejores resultados los encontrados en el componente vocabulario, aunque, hubo centros educativos, es especial de los institutos no gubernamentales, que mostraron resultados aceptables en ambos componentes.

Tabla 27 Rendimiento Promedio porcentual por componente Español 11mo grado 2018

Bloque	Componente	Centros de Educación Media Oficiales			Centros de Educación Media No Gubernamentales		
		Abelardo Fortín	Jesús Aguilar Paz	Mixto Hibueras	Inmaculada Concepción	Renacimiento	San Francisco
Lectura	Tipos de texto	59	49	55	58	62	57
	Vocabulario	61	44	60	66	69	65

Fuente: elaboración propia

Grafico 23 Rendimiento Promedio Porcentual por Bloque y componente Español (Lectura) 11mo grado BCH 2018

Fuente. Elaboración propia

El centro educativo con rendimiento menor al 50% de porcentaje de respuestas correctas en los componentes evaluados, fue el centro de educación media oficial “Jesús Aguilar Paz”, y el centro de educación media con mayor rendimiento porcentual fue el centro de educación media no gubernamental “Renacimiento” quien obtuvo resultados de 60% en el componente tipos de texto y 74% en vocabulario.

Tabla 28 Rendimiento promedio porcentual por estandar de desempeño en Español (Lectura) 11mo grado BCH 2018

Componente	Estándar	Centros Educativos Oficiales			Centros gubernamentales	Educativo	No
		Abelardo Fortín	Jesús Aguilar Paz	Mixto Hibueras	Inmaculada Concepción	Renacimiento	San Francisco
1. Tipos de texto	1.1	51%	45%	43%	52%	56%	53%
	1.2	78%	64%	68%	68%	76%	75%
	1.3	57%	48%	54%	60%	58%	51%
	1.4	61%	44%	56%	48%	57%	52%
	1.5	49%	45%	47%	53%	58%	49%
	1.6	59%	42%	55%	60%	61%	56%
2.2 Vocabulario	2.1	71%	42%	70%	88%	79%	80%
	2.2	44%	36%	46%	51%	53%	50%
	2.3	52%	43%	55%	52%	57%	62%
	2.4	55%	39%	55%	65%	63%	58%

Fuente: Elaboración propia

En la tabla 21, se indica con detalle, en cuales componentes y estándares presentaron mayor habilidades y dificultades los estudiantes que fueron evaluados a través de este estudio. Los estudiantes del Centro de Educación Media no gubernamental Renacimiento son los que mayor porcentaje de respuestas correctas mostraron con un 76% en el componente Tipos de Texto y 69% en Vocabulario. Los valores más bajos los obtuvo el Centro de Educación Media Gubernamental Jesús Aguilar Paz con resultados promedios de 49% en el componente Tipos de texto y 44% en Vocabulario.

El estándar con mayor porcentaje de respuestas correctas del componente tipos de texto, corresponde al estándar 1.2, evaluado a través de los ítems, 6,7,8,9 y 10, el cual sugiere a los estudiantes mostrar comprensión de ideas globales, principales y secundarias incluyendo gráficos e iconos. En relación con el componente vocabulario el estándar 2.1, medido a través de los elementos 24 y 27, observo el mayor porcentaje de respuestas correctas, su contenido precisaba que los estudiantes reconocieran palabras desconocidas en una variedad de textos leídos.

Por otra parte, el estándar 1.5, del componente tipos de texto, solicitaba a los estudiantes analizar diferentes puntos de vista y evaluar rasgos científicos, sociológicos y culturales de los textos a través de los ítems 22,23, 25 y 26; y tal como se demuestra en la tabla 21, estos fueron los elementos con menor acierto entre los estudiantes ya que el porcentaje de respuestas correctas osciló entre el 45% y 48%.

En el componente vocabulario, reconocer e interpretar el lenguaje figurado y la intencionalidad del autor a través de algunas palabras, resulto una habilidad difícil de mostrar por los estudiantes evaluados, el grado de acierto en los seis ítems que median este estándar (2.1) se redujo entre el 36% y 43% respectivamente, lo que los categorizo como reactivos con un p-value en la categoría difícil.

CAPITULO 6. DISCUSION DE LOS RESULTADOS

En este apartado se resumen los aspectos más relevantes de acuerdo a los objetivos planteados en esta investigación, en cuanto al diseño del instrumento y los resultados de aprendizaje alcanzados por los estudiantes. A partir de estas consideraciones en los siguientes párrafos se mencionan los principales hallazgos, el contraste de los resultados de la prueba de Español (Lectura) 11mo grado BCH, frente a estudios similares, para finalmente establecer conclusiones y recomendaciones.

6.1 Hallazgos principales

Diseño del instrumento

De conformidad con los análisis realizados en cuanto a la evidencia de la calidad del instrumento diseñado, bajo el esquema de la Teoría Clásica de los Test a través del cálculo de los parámetros de discriminación, dificultad y confiabilidad, la prueba para la evaluación

del rendimiento académico de los estudiantes de undécimo grado en el área español, dio muestra de ser una prueba confiable, valida y consistente en sus resultados.

Consistencia interna de la prueba

El análisis de fiabilidad de las puntuaciones de la prueba operativa dio por resultado un coeficiente alpha de Cronbach de .97, destacando además que ninguna correlación ítem-test resulto negativa; lo que permite concluir que cada uno de los 50 ítems que conformo la prueba contribuyeron de forma positiva a la consistencia interna de la misma.

Validez de los elementos

La validez se obtuvo mediante una cuidadosa planificación y delimitación de los estándares de desempeño enmarcados en el CNB, para el área de español en el undécimo grado del Bachillerato en Ciencias y Humanidades, lo que exigió el desarrollo del “*Taller de Evaluación Educativa y Redacción de ítems*” , hecho que permitió el estudio de las programaciones, planes de estudio, libros de texto por parte de un equipo de docentes especialistas en el área de español y literatura con amplia experiencia en la enseñanza de este campo curricular, quienes a su vez asesoraron la elaboración del marco operativo de la prueba y la tabla de especificaciones técnicas, documentos determinantes para la validez y objetividad de una prueba estandarizada para medir el desempeño académico.

Lo anterior fue determinante para lograr un nivel de detalle en los resultados porcentuales por bloque, componente y estándar, dando lugar a la identificación de las áreas del currículo en la que los estudiantes participantes de este estudio, dieron evidencia de haber adquirido las competencias o bien reflejan debilidades en la adquisición de dichas competencias.

Análisis psicométrico de los ítems

Los 50 ítems que conformaron la prueba, mostraron una correlación punto biserial alta ≥ 3 , indicando que los estudiantes que respondieron correctamente cada ítem, también recibieron un alto puntaje en la prueba, ningún reactivo se registró por debajo del valor de 0.25 con discriminación baja. De forma que, un total de 37 elementos mostraron alta correlación con valores mayores a 0.50, lo que resulta correspondiente de una prueba estadísticamente con un nivel de confiabilidad alto demostrando una idónea capacidad discriminativa de los ítems que conformaron el instrumento, lo que suma validez a la prueba.

Respecto al índice de dificultad de los ítems, el 68% (34 elementos), resultaron con una dificultad media lo que significa que más de la mitad de la prueba, presento una dificultad adecuada a los conceptos medidos. 11 ítems mostraron dificultad alta, donde solamente entre el 13% y 19% de los estudiantes contestaron correctamente los ítems, finalmente 5 elementos observaron dificultad baja, siendo contestados por el 64% de la población participante en el estudio. Es importante señalar que aun aquellos ítems que mostraron alta dificultad, observaron una correlación punto- biserial aceptable.

En síntesis, los estadísticos de los parámetros psicométricos de la prueba reflejan que el instrumento conservó un equilibrio entre la complejidad del contenido de los reactivos. con un grado de complejidad alta, media y baja, presentando parámetros psicométricos aceptables; aunque es claro que aquellos ítems que mostraron menor discriminación y una dificultad muy alta pueden ser revisados para mejorarlos o proponer otro reactivo para evaluar esta habilidad.

Descripción del desempeño académico de los estudiantes participantes en el estudio

1. En la escala de niveles de desempeño, con datos agregados los Centros Educativos de educación Media evaluados, se ubicaron en el nivel de desempeño satisfactorio con 300 puntos.

Los resultados con datos agregados en la escala de nivel de desempeño, ubica a los institutos en el nivel de desempeño satisfactorio; sin embargo hay diferencias marcadas entre los institutos oficiales y no gubernamentales en los resultados desagregados por centro educativo según la escala de puntuación estandarizada (100 a 500), los centros educativos oficiales Instituto Mixto Hibueras e Instituto Jesús Aguilar Paz con 259 puntos se ubicaron en el nivel de desempeño debe Mejorar, mientras que el Instituto Técnico Abelardo R. Fortín se ubicó en el nivel de desempeño Satisfactorio con 306 puntos. Respecto a los centros de educación media no gubernamentales Inmaculada Concepción, San Francisco y Renacimiento alcanzaron entre 305 y 322 puntos agrupándose en el nivel satisfactorio.

2. Los niveles de aprendizaje que muestran los estudiantes de centros de educación media no gubernamental, son ligeramente superiores respecto a los gubernamentales.

El nivel de desempeño satisfactorio en los institutos no gubernamentales fue de 64% mientras que en los centros educativos oficiales únicamente el 47%; además la proporción de estudiantes en la categoría de insatisfactorio es de 4% en los centros de educación media oficiales y en los no gubernamentales el 1% se ubicó en este nivel.

3. En la escala de puntuación estandarizada, los estudiantes de sexo masculino se agruparon en el nivel satisfactorio y las mujeres en el nivel de desempeño debe mejorar.

Al hacer las comparaciones en base a sexo, se observó que los niveles de desempeño de los varones fue de 303 y el de las mujeres 298, lo que deduce una leve diferencia entre el rendimiento de los varones respecto al de las mujeres; contrario a los resultados de rendimiento académico a nivel nacional y regional, que han revelado que las niñas muestran mejor desempeño en lectura que los varones, y aunque la diferencia no es verdaderamente significativa en el caso de este estudio, fue determinante para que los varones se ubicaran en el nivel satisfactorio mientras que las mujeres en el nivel debe mejorar.

4. En la escala de niveles de desempeño según el sexo de los estudiantes, el porcentaje de hombres y mujeres en el nivel satisfactorio fue de 55%.

El 55% de los estudiantes del sexo femenino y masculino se agruparon en el nivel de desempeño satisfactorio. Aunque los resultados mantienen tendencia a suponer una ligera diferencia favorable al sexo masculino, se encontró el mismo porcentaje de estudiantes en nivel satisfactorio para ambos sexos. Una diferencia no significativa estadísticamente se encontró en el nivel de desempeño insatisfactorio, resultando el 1% de los estudiantes del sexo masculino en este nivel, mientras que en el sexo femenino la proporción fue de 3% en el nivel insatisfactorio.

5. En el rendimiento promedio porcentual por bloque de contenido, componentes y estándares evaluados, los mejores resultados se encontraron en el componente vocabulario.

Este nivel de detalle permitió identificar en cuales bloques de contenido y componente se presentan fortalezas y debilidades cognitivas por parte de los

estudiantes de acuerdo a la evaluación del área curricular español (lectura). sin embargo, hubo centros educativos que reportaron buenos resultados en ambos componentes.

El centro educativo con rendimiento menor al 50% de porcentaje de respuestas correctas fue el centro de educación media gubernamental “Jesús Aguilar Paz”, y el centro de educación media con mayor rendimiento porcentual fue el centro no gubernamental “Renacimiento” quien obtuvo resultados de 60% en el componente tipos de texto y 74% en vocabulario.

De los diez estándares evaluados, seis corresponden al componente tipos de texto y cuatro a vocabulario. En el componente tipos de texto, los mejores resultados mostrados por los centros educativos evaluados fue el estándar 2, medido a través de los ítems 6,7,8,9 y 10 referente a demostrar comprensión de las ideas globales, principales, secundarias e inferenciales de un texto leído, incluyendo la interpretación de gráficos e iconos. En el componente de vocabulario los resultados más altos fueron en el estándar 1, concerniente a reconocer y entender en textos leídos una variedad de palabras conocidas, evaluado en los ítems 24 y 27 de la prueba.

Los resultados relativamente bajos se observan en el componente tipos de texto, en especial los valores del estándar 5, evaluado a través de los ítems 22, 23, 25 y 26 los cuales proponen a los estudiantes demostrar habilidades de análisis de diferentes puntos de vista y evaluar rasgos científicos, sociológicos y culturales de los textos; de los seis institutos evaluados solamente los centros educativos no gubernamentales “Inmaculada Concepción” y “Renacimiento” alcanzaron resultados de 53% y 58% de respuestas correctas respectivamente.

Algo similar ocurre en el componente vocabulario en el estándar que refiere el reconocimiento e interpretación del lenguaje figurado, la intencionalidad del autor y otros aspectos semánticos de las palabras, donde los institutos no gubernamentales obtuvieron resultados de 44%, 36% y 46%. En los resultados de acuerdo a la escala de puntuación estandarizada el puntaje menor obtenido fue de 155 y el mayor de 410.

El 55% de estudiantes se ubicaron en el nivel satisfactorio, mientras que 41% se ubicó dentro del nivel debe mejorar, 3% en el nivel insatisfactorio y el 1% en el nivel avanzado.

6.2 Contraste de los resultados de la prueba de fin de grado

Un estudio realizado por PREAL (2017), revela que en el área de Español existe un leve avance de los niveles de aprendizaje en los niños y niñas de Honduras. Por su parte la evaluación realizada por la SEDUC a través del Proyecto MIDEH, desde el año 2008, a los grados de primero a noveno grado, reflejo un moderado avance en esta área.

En los apartados anteriores se indica que en el estudio TERCE (2013), se refleja que el problema de desempeño de los estudiantes en las áreas de lectura, escritura, matemáticas y ciencias es regional. Honduras, aunque por debajo del promedio de todos los países evaluados; en los resultados de la prueba de Lectura, los estudiantes de Tercer grado de educación básica, superan ligeramente a países como Republica Dominicana, Panamá, Paraguay, Nicaragua y Guatemala (Laboratorio Latinoamericano de la Evaluacion de la calidad (LLECE), 2014).

Así mismo en el último estudio llevado a cabo de forma municipal por parte de MIDEH (2017), en el Municipio de Distrito Central, los estudiantes de 1ro a 6to grado, se encuentran en el nivel satisfactorio en Español(lectura) con 301 puntos, tendencia que disminuye en el III ciclo de educación básica, ya que de 7mo a 9no grado, los estudiantes se ubican con 290 puntos en el nivel debe mejorar, por su parte los resultados según la métrica de puntuación estandarizada y niveles de desempeño que se han descrito en este estudio, muestran que el 55% de los estudiantes se agrupan en el nivel de desempeño satisfactorio; pero se debe considerar que aunque el tamaño de la muestra es admisible para esta investigación, no es representativa para afirmar una generalización, por lo que los resultados obtenidos son solo significativas para la muestra del estudio.

Sin embargo en cuanto al rendimiento académico respecto al sexo de los estudiantes, las diferencias no fueron significativas , estadísticamente en la prueba de hipótesis, a través de la comparación de medias “T Student”, en su significancia bilateral dio por resultado 0.341, mayor al valor alfa de 0,005, lo que permite concluir que no existen diferencias estadísticamente significativas entre los resultados de puntaje total en lo referente a la media

de preguntas contestadas correctamente, donde la media de las mujeres fue de 28.88 y para el sexo masculino de 29.54.

Así mismo en las escalas de nivel de desempeño y estandarizada los estudiantes de sexo masculino y femenino muestran una diferencia no muy significativa sin embargo en la categorización por niveles de desempeño, aunque sea por una leve diferencia, ubica a los estudiantes de sexo femenino en el nivel debe mejorar con 298 puntos y los varones en satisfactorio con un puntaje de 303.

Por otra parte, los niveles de aprendizaje que muestran los estudiantes de centros de educación media no gubernamental, son ligeramente superiores respecto a los gubernamentales. El nivel de desempeño satisfactorio en los institutos no gubernamentales fue de 64% mientras que en los gubernamentales de 47%.

Estos resultados también son consecuentes a los observados en el reciente estudio PISA-D, donde en ciencia y lectura los estudiantes hondureños se agruparon en los niveles más bajos de desempeños, por debajo de la media de los países de América Latina y el Caribe y además en los resultados de desempeño entre escuelas públicas y privadas, Honduras presenta la brecha más alta entre escuelas públicas y privadas de los países de la región (OCDE, 2018)

6.3 Implicaciones teóricas y prácticas de los resultados

Tal y como se ha venido acotando en los apartados anteriores, la evaluación estandarizada es entendida como un proceso de recogida de información útil y valiosa para emprender procesos de mejora en la práctica educativa, esto en el mejor de los casos, pues contribuyen de alguna forma a identificar el progreso de logro de competencias de los estudiantes de una escuela, de una región determinada, o bien del sistema educativo de un país. (Fernandez Navas, Alcaraz Salarirche, & Sola Fernandez, 2017).

La prueba para la medición del desempeño académico del área español en el undécimo grado, logró categorizar a las seis instituciones participantes en un nivel de desempeño, sin embargo debido al objetivo de esta evaluación, misma que no implica la triangulación como metodología de investigación; la construcción de la información solo consideró un tratamiento estadístico de los resultados mostrados por los estudiantes, sin cavilar en los factores multicausales que hoy en día explican el desempeño académico de los estudiantes.

En este sentido, el enfoque de esta evaluación ha permitido comprobar y comparar el rendimiento académico de los estudiantes en el área curricular español, información que coadyuva de alguna manera en la toma de decisiones informadas y la construcción de un plan de mejora de las prácticas de enseñanza con una amplia participación de la comunidad educativa.

Un efecto positivo esperado acerca de las evaluaciones estandarizadas y que la literatura acerca de esta nos señala reiteradamente, es el uso de sus resultados para enriquecer las prácticas de enseñanza y evaluación de los docentes en el aula de clases, a fin de que esta propicie una apertura que conlleve a los maestros(as) a nuevos horizontes conceptuales en relación a la didáctica y a la forma de evaluar a los(as) estudiantes.

De igual forma, los resultados de este estudio también pueden contribuir a la toma de decisiones en cuanto a la importancia de ciertos contenidos y habilidades que los jóvenes deberían adquirir y que muchas veces son desarrollados en un sentido muy laxo o en el peor caso no son parte del marco conceptual dentro del cual los docentes de esta área organizan su enseñanza.

Finalmente, los reactivos conforman un banco de ítems con estadísticas que conllevan a suponer calidad del instrumento, bajo una propuesta de ítems con una complejidad cognitiva, que, de lograr conocer el instrumento por parte de los docentes, puede alcanzar enriquecer las rutinas de evaluación en el aula de clases.

CONCLUSIONES

1. Las propiedades psicométricas de la prueba para la evaluación del rendimiento académico del área español en el undécimo grado del Bachillerato en Ciencias Humanidades, fueron aceptables en el marco de los parámetros que establece la Teoría Clásica de los Test (TCT).
2. El análisis de los resultados del desempeño académico, agrupa a los estudiantes en el nivel de desempeño satisfactorio en la prueba de español (lectura) en los datos agregados; sin embargo, existe una brecha entre el desempeño académico de los estudiantes de instituciones gubernamentales respecto a los no gubernamentales, hecho que en estudios de evaluaciones nacionales e internacionales sigue siendo latente en Honduras.
3. Los resultados de la prueba de Español(Lectura), no demostraron una relación clara entre el desempeño académico y el sexo de los estudiantes ya que estadísticamente no se comprobó diferencias significativas entre las medias obtenidas no obstante, en el análisis de rendimiento académico en la escala de puntuación estandarizada y los niveles de desempeño, aun por una diferencia mínima, los estudiantes de sexo masculino alcanzaron el nivel satisfactorio con 303 puntos, mientras que las mujeres con una puntuación de 298, se agruparon en el nivel debe mejorar.
4. La prueba para la evaluación del desempeño académico de español midió el bloque de Lectura a través de los componentes tipología textual y vocabulario. Los resultados muestran que en el componente Tipos de texto, se presentaron los resultados más bajos en relación al componente vocabulario.
5. Una de las deficiencias más importante en relación a los resultados, es la ausencia de conocimiento en los ítems relacionados con la comprensión, análisis y evaluación de rasgos científicos, sociológicos y culturales de los textos, así como también en el reconocimiento e interpretación del lenguaje figurado, la intencionalidad del autor y otros aspectos semánticos de las palabras en un texto leído.
6. A partir de los resultados asociados a los niveles de desempeño se obtuvo información sustantiva respecto de los conocimientos y las habilidades que han desarrollado los estudiantes en el área español, esta información se puede convertir en un valioso aporte para orientar el trabajo pedagógico.

RECOMENDACIONES

1. La Secretaría de Educación debe multiplicar esfuerzos para desarrollar el proceso de establecimiento de estándares de desempeño en el nivel de educación media, de manera que permita la elaboración de ítems que respondan a las especificaciones técnicas de pruebas de desempeño académico, que midan justamente lo que este enseñando en el aula de clases.
2. La estandarización de pautas o directrices en la aplicación del instrumento son cruciales en la validación de las puntuaciones, por lo que, en nuevas aplicaciones similares a esta línea de investigación, será necesario contar con suficiente recurso humano que contribuya a promover un control de calidad en la administración de la prueba que garantice la equidad de condiciones de los evaluados.
3. Aunque la tesis referida centra su objetivo en el diseño de una prueba estandarizada que mida el desempeño académico de los estudiantes de BCH, en el área de lecto escritura, se plantea como recomendación general, el incremento de esfuerzos de todos los actores del sector educativo para articular estrategias de evaluación que profundicen en las razones de los resultados favorables o desfavorables obtenidos por los estudiantes, a fin de generar líneas de acción contextualizadas, que coadyuven a un marco de acción en el aula, que permita desde todas las aristas posibles realizar una verdadera mejora en la calidad de la educación, tal cual es, el propósito de la aplicación de las pruebas estandarizadas.

BIBLIOGRAFÍA

- Alas Solis, M., Moncada, G., Hernandez , R., & Morales, R. (2017). *Informe de Progreso Educativo de Honduras. Año 2017*. Anual, FEREMA, UPNFM, Tegucigalpa.
- Alcaraz Díaz, F. (2003). Un modelo de evaluación criterial para el área de Matemáticas en la enseñanza obligatoria, ensayos. *Revista de Estudios de la Escuela Universitaria de Magisterio de Albacete*, 225-244.
- Alcaraz Salarirche, N. (2015). Aproximacion historia de la evaluacion educativa. *Revista Iberoamericana de Evaluacion Educativa*, 8(1), 11-25.
- Amador Lopez, J. (Noviembre de 2008). Tesis Maestria Investigacion Educativa. *La Reforma Educativa en Honduras, el Currículo Nacional Basico y la actitud de los docentes del nivel de educación Basica, desde la perspectiva de la Teorica de la accion razonada*, 43-51. /
- Anastasi, A., & Urbina, S. (1997). Principales contextos del uso actual de las pruebas. En A. Anastasi, & S. Urbina, *Tests psicológicos* (págs. 475-553). Prentice Hall.
- Anderson , P., & Morgan, G. (2016). *Evaluaciones de rendimiento academico*. (Vol. Volumen 2 Desarrollo de pruebas y cuestionarios para una evaluacion nacional de rendimiento academico). (V. Greaney, & T. Kellaghan, Edits.) Washington D.C.: Banco Mundial. doi:10.1596/978-1-4648-07459-9
- Anderson, P., & Morgan , G. (2016). *Evaluaciones nacionales del rendimiento académico* (Vol. Volumen II: Desarrollo de pruebas y cuestionarios para una evaluación nacional del rendimiento académico). (V. Greaney, & T. Kellaghan, Edits.) Washington DC: Banco Mundial.
- Arancibia, J. (2015). La estandarización de la evaluación. Las pruebas nacionales e internacionales ¿medición o evaluación? *Internacional de la Educación para América Latina*, 21.

- Arancibia, V. (1997). Los Sistemas de Medicion y Evaluacion de la calidad de la educación. *Documentos de Laboratorio Latinoamericano de Evaluación de la calidad de la educación LLECE*, 1-41.
- Arguello, C. (11 de Junio de 2011). Métodos Cuantitativos de investigacion educacional. Venezuela.
- Arregui, P., Wolfe, R., & Ravela, P. (2008). Las Evaluaciones Educativas que America Latina necesita. *Revista Iberoamericana de Evaluacion Educativa*.
- Arzola Franco , D. (2017). Evaluación, pruebas estandarizadas y procesos formativos: una experiencia en centros educativos del norte de Mexico. *Educacion* , XXVI(50), 28-46.
- Asociación Internacional para la Evaluación del Rendimiento Educativo . (2011). *Progress in International Reading Literacy Study*.
- Backhoff Escudero, E. (2018). Evaluación estandarizada de logro educativo: contribuciones y retos. *Revista Digital Universitaria, UNAM*, 19(6), 1-15.
- Backhoff Escudero, E., Vázquez Lira, R., Contreras Roldán, S., Caballero Meneses, J., & Rodríguez Jiménez, J. (2017). *Cambios y tendencias de aprendizaje en México: 2000-2015*. Cuaderno de investigacion , Instituto Nacional para la Evaluacion de la educacion, Mexico.
- Banco Mundial . (7 de Abril de 2017). *Honduras Panorama general* . Obtenido de Banco Mundial : <http://www.bancomundial.org/es/country/honduras/overview>
- Barbero Garcia, M. I. (2006). Analisis de la calidad metrica de los Items. *DIALNET Metodologia de las ciencias del comportameniento UNED*.
- Barrenechea, I. (2010). *Evaluaciones Estandarizadas: Seis Reflexiones Críticas*. Arizona, Estados Unidos. Obtenido de <file:///C:/Modulo%20I%20Seminario/275019712008.pdf>
- BCH, B. (2017). *Indice Mensual de Actividad Economica IMAE*. Tegucigalpa: Departamento de Estadistica macroeconomica BCH.

- Benavidez Ormaza , V. (12 de Abril de 2010). Las evaluaciones de logros educativos y su relacion con la calidad de la educacion. *Revista Iberoamericana de Educacion*(53), 83-96.
- Berk, R. (1985). Una guía para la construcción de pruebas con criterios de referencia. (T. Haladyna, Ed.) *Revista para la medición educativa*, 22(4), 313-316.
- Borsboom, D. (2005). *Measuring the mind: Conceptual issues in moderno psychometrics*. Cambridge University Press.
- Chacón Viquez, L. D. (Noviembre de 2013). El Sistema y las reformas educativas en Honduras: El balance de dos decadas. (U. E. Distancia, Ed.) *Revista Calidad en la Educacion Superior*, 4(2), 144-169. Recuperado el Abril de 2017, de https://www.academia.edu/5329890/Las_Reformas_educativas_en_Honduras_El_balance_de_dos_d%C3%A9cadas
- Cleaves Reina , R. (30 de Marzo de 2017). *Honduras y su economía*. Obtenido de <http://www.hondurasysueconomia.com/?m=201704>
- Cohen,, L., Manion, L., & Morrison, K. (2007). *Research Methods in Education*. Routledge Taylor & Francis Group.
- Comisión Ad-Hoc , R., & Secretaria de Educacion . (2013). *Entendiendo la ley fundamental de Educación*.
- Covacevich, C. (2014). Cómo seleccionar un instrumento para evaluar aprendizajes estudiantiles. *Nota tecnica del BID*; 738.
- Cruz Cruz, L. (2015). Estado y perspectivas de la Educación en Honduras. *Primer Simposio Nacional de Educación: Estado actual y desafios de la Educación en Honduras*, (pág. 16). Tegucigalpa. Recuperado el 30 de Marzo de 2017
- Cruz, L. (2015). *Memoria Primer Simposio Educacion 2015*. Memoria de Simposio, Universidad Pedagogica Francisco Morazan , Instituto de Investigacion Educativa INIEES. Recuperado el 1 de Abril de 2017

- D. López, K. M., & O. Lever, C. (30 de Junio de 2016). Las evaluaciones estandarizadas del aprendizaje y la mejora de la calidad educativa. (M. Centro de Enseñanza Técnica y Superior CETYS Universidad, Ed.) *Temas Educacion*, 22(1), 132.
- Debera, L., & Nalbarte, L. (2006). Pruebas diagnósticas: una aplicación a la teoría de respuesta al ítem. *Serie documentos: Instituto Estadística, Universidad de la República de Uruguay*, 1-30.
- Downing, S., & Haladyna, T. (2006). *Handbook of test development*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Fast, M., & Vukmirovic, Z. (2016). Paradigmas y tendencias en medición y evaluación educativa. *Proceso de una evaluación nacional*.
- FEREMA. (2017). *Informe de progreso educativo*. Fundación para la Educación Ricardo Ernesto Maduro Andreu, Tegucigalpa, Honduras.
- Fernandez Ballesteros, R. (2000). Evaluación psicológica y test. *La evaluación psicológica en el año 2000*, 11-26.
- Fernandez Navas, M., Alcaraz Salarirche, N., & Sola Fernandez, M. (2017). Evaluación y Pruebas Estandarizadas: Una Reflexión sobre el sentido, utilidad y efectos de estas pruebas en el campo educativo. *Revista Iberoamericana de Evaluación Educativa*, 10, 51-67. Recuperado el 2019, de www.rinace.net/rie/
- Fernandez Navas, M., Alcaraz Salarirche, N., & Sola Fernandez, M. (2017). Evaluación y pruebas estandarizadas: una reflexión sobre el sentido, utilidad y efectos en el campo educativo. *Revista Iberoamericana de evaluación educativa*, 51-67.
- Ferrer, G., & Fiszbein, A. (2015). *¿Qué ha sucedido con los sistemas de evaluación de aprendizajes en América latina? Lecciones de la última década de experiencia*. Comisión para la educación de calidad para todos: Documentos de antecedentes, Banco Mundial.
- G.Lopez, B., & R. Monroy, J. A. (2009). Comprensión Lectora. *Revista Mexicana de Orientación Educativa REMO*, 37.

- Gaceta, L. (17 de Diciembre de 2014). Acuerdo N° 2188-SE-2014. 3. Tegucigalpa, Honduras.
- Gil, R. L. (2003, 2004, 2005, 2007, 2008). *Una Mirada al Derrotero de la Evaluacion Estandarizada en la Region Centroamericana*. OCIPE-IDEUCA.
- Goikoetxea, E. (2012). Las dificultades especificas del aprendizaje en al albor del siglo XXI. *Revista Electronica de Investigacion y Evaluacion Educativa RELIEVE*, 18(1), 13.
- Gomez Arbeo, B. (1990). *Evaluación criterial : una metodología útil para diagnosticar el nivel de aprendizaje de los alumnos*. Madrid: Narcea Ediciones.
- Gomez Lopez, B. E., & Monroy Romero, J. A. (2009). Comprension Lectora. *Revista Mexicana de Orientacion Educativa REMO*, 37.
- Gonzales , L., & Felicia, M. (2007). *Instrumentos de Evaluación Psicológica*. La Habana, Cuba: Editorial Ciencias Médicas.
- Greaney, V., & Kellaghan, T. (2016). *Evaluaciones nacionales del rendimiento académico* (Vol. Volumen I: Evaluacion de los niveles nacionales de rendimiento academico). (B. Mundial, Ed.) Washington, DC. doi:10.1596/978-1-4648-0743-5
- Gual, A. (2010). El mapa de competencias a adquirir. *Educacion Médica*, 13. Obtenido de http://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1575-181320100005000008&lng=es&tlng=es.
- Guerrero, E., Lucero , C., Segura, A., & Tovar Cuevas, J. (2013). Factores de riesgo asociados a bajo rendimiento academico en escolares de Bogotá. *Investigaciones Andina*, 15(26), 654-666.
- Hambleton, R., Swaminathan, J., & Rogers, H. (1991). *Fundamentos de la teoría de respuesta al ítem*. Newbury Park, California: Publicaciones sabias.
- Heredia Manrique, A. (2009). Dos formas diferenciadas de evaluación didactica: evaluacion normativa y criterial. 39-48.

- Hernández Sampieri, R., Fernandez Collado, C., & Baptista Lucio, M. (2014). *Metodología de la Investigación* (Sexta Edición ed.). Mexico.
- Hernández, R. (2009). Honduras 2008: La Evaluación de los Aprendizajes Vislumbra una luz de Esperanza en Educación. *Revista Iberoamericana de Evaluación Educativa*, 91.
- Hurtado Mondoñedo, L. (2018). Relacion entre los indices de dificultad y discriminacion. *Revista Digital de Investigacion en Docencia Universitaria*, 1-21.
- IEA. (2011). *Resultados TIMS 2011 Chile*. Asociación Internacional para la Evaluación del Logro Educativo, Santiago de Chile.
- INE, H. (2017). *Educacion 2016*. Recuperado el 1 de Abril de 2017, de http://www.ine.gob.hn/index.php?option=com_content&view=article&id=90
- INE, H. (Enero de 2017). *Educacion 2016*. Recuperado el 1 de Abril de 2017, de http://www.ine.gob.hn/index.php?option=com_content&view=article&id=90
- Instituto Nacional de evaluación Educativa. (2011). *PIRLS-TIMSS 2011*. Madrid.
- Jimenez Moreno, J. A. (2016). El papel de la Evaluacion a gran escala como politica de rendicion de cuentas en el Sistema Educativo Mexicano. *Revista Iberoamericana de Evaluacion Educativa*, 109-126.
- Kellaghan, T., Greaney, V., & Scott Murray, T. (2016). Utilización de los resultados de una evaluación nacional del rendimiento académico. En *Evaluaciones nacionales del rendimiento académico* (Vol. 5). Washington, DC: Banco Mundial. doi:10.1596/978-1-4648-0751-0.
- La Gaceta. (Abril de 2013). Decreto N° 33,106.
- Laboratorio Latinoamericano de la Evaluacion de la calidad (LLECE). (2014). *Tercer Estudio Regional Comparativo y Explicativo (TERCE)*.
- Lacave Rodero, C., Molina Diaz, A. I., Fernandez Guerrero, M., & Redondo Duque , M. A. (8 al 10 de Julio de 2015). Análisis de la fiabilidad y validez de un cuestionario docente. *Actas de las XXI Jornadas de la Enseñanza*, 8.

- Latorre , A., & Rincon, D. (2003). Bases Metodologicas de la investigacion educativa. Barcelona, España.
- Leiva Barajas, Y. (2011). Una reseña sobre la validez de constructo de pruebas referidas a criterio. *Perfiles Educativos UNAM*, vol. XXXIII(núm. 131, 2011).
- LLECE. (2009). Reglas para elaborar Items de formato de seleccion multiple. *Diseño de pruebas para la evaluacion educativa*, 40.
- Lopez, A. (2013). Alineacion entre las evaluaciones externas y los estandares academicos, el caso de las pruebas SABER, de Matematicas, Colombia. *RELIEVE*, 1-17.
- Lucio, G. (2009). *Una mirada al deterioro de la evaluacion estandarizada en la region centroamericana*. ocipe-ideuca.
- Mandeville, P. (Enero-Marzo de 2010). Tema 22: Muestreo multietápico. *Ciencia UANL Redalyc*, Vol. XIII(1), 102-105.
- Martinez Arias, M. (1995). Teoría de los tests psicológicos y educativos. *Dialnet*.
- Martinez Arias, M., & Hernandez LLoreda, M. (2006). *Psicometría*. Madrid: Alianza editorial.
- Martinez Rizo, F. (Octubre de 2011). evaluacion educativa y Pruebas estandarizadas. elementos para enriquecer el debate . *Revista de la Educacion Superior*, 30.
- Martinez Rizo, F., & Blanco, E. (2001). La evaluacion educativa en México: experiencias, avances y desafios . 1-32.
- Matas Terron, A. (2010). Introduccion al analisis de la Teoria Respuesta al Ítem. (Aidesoc.net, Ed.) *Aidesoc Formacion y Desarrollo*, 2-7.
- Mendoza, J., & Garza, J. (2009). La medición en el proceso de investigación científica: Evaluación de validez de contenido y confiabilidad. (I. d. Negocios, Ed.) *UANL*, 1-10.
- Meneses, J., Barrios, M., Bonillo, A., Cosculluela, A., Lozano, L., Turbany, J., & Valero, S. (2013). *Psicometria*. Barcelona: UOC Gran via de les Corts catalanes.

- Moncada, G., & Lopez, V. (2014). *Educación y neurociencias*. Linea e investigación y políticas educativas, Banco interamericano de Desarrollo (BID), Universidad Nacional autonoma de Honduras (UNAH), Tegucigalpa, Honduras.
- Mora T., F. (2013). *Solo se aprende lo que se ama*. Madrid, España.: Alianza Editoria D.L.
- Mora Vargas, A. (2004). la evaluacion educativa, conceptos, periodos y modelos. *Revista investigativa "Actualidades investigativas de educación"*, 29.
- Moreira Mora, T. E. (2008). Funcionamiento diferencial del ítem en prueba de matematica para educacion media. *Actualidades en psicologia*(22), 91-113.
- Moya, R., Pérez Gómez, Y., & Ruiz Cordovéz, R. (2016). La Teoría de Respuesta al Ítem para la evaluación del aprendizaje en matemática. *Edusol*, 16(55). Obtenido de <http://www.redalyc.org/articulo.oa?id=475753050020>
- Muñiz, J. (1996). *Teoria Clasica de los Test*. Madrid: Piramide.
- Muñiz, J. (2000). *Teoría Clásica de los Tests* (2da edicion ed.). España: Piramide.
- Muñiz, J. (2010). Teoria de los Test: Teoria Clasica y Teoria respuesta al ítem. *Papeles del psicologo*, 57-66.
- Muñiz, J., & Fonseca Pedrero, E. (2008). Construcción de instrumentos de medida para la evaluación universitaria. *DIALNET Revista de investigacion en Educacion*(5), 13-25.
- Muñiz, J., Cueto Garcia, E., Vicente Martinez, R., Moreno, R., & Fidalgo, A. (2005). Cuadernos de Estadística: Analisis de los ítems. Arco Libros - La Muralla, S.L. (1 de enero de 2005).
- Muñiz, J., Moreno Rodriguez, R., & Martinez, R. (2004). Directrices para la construccion de items de eleccion multiple. *Psicothema*, 16, 9. doi:490-497
- Muñoz, A., & Ocaña, M. (2017). Uso de estrategias metacognitivas para la comprensión linguistica. *Cuadernos de Lingüística Hispánica*, 223-244.
- Naciones Unidas. (2018). La agenda 2030 y los Objetivos de Desarrollo Sostenible: una oportunidad para America latina y el Caribe.

- Navarro, E. (2003). El rendimiento academico:concepto, investigacion y desarrollo. *REICE Revista Iberoamericana sobre calidad, eficacia y cambio en Educacion*, 1(2).
- OCDE. (2018). *PISA_D para el desarrollo*.
- OCDE, & PISA-D. (2018). Informe Resultados de Honduras PISA para el Desarrollo 2018.
- OCDE;. (2017). Marco de Evaluación y de Análisis de PISA para el Desarrollo : Lectura, matemáticas y ciencias.
- OECD. (2018). *PISA para el Desarrollo: resultados en foco*. Programa de Evaluación Internacional de Estudiantes.
- OEI. (2018). Neurodidactica en el aula: transformando la educación. *Revista Iberoamericana de Educación*, 78(1).
- OEI. (s.f.). *Biblioteca digital de la OEI*. Recuperado el 31 de Marzo de 2017, de <http://www.oei.es/historico/oeivirt/fp/cuad04a03p21.htm>
- ONE. (2009). Hacia una Cultura de Evaluacion. *CENSO*.
- OREALC, & UNESCO. (2015). *Aportes para la enseñanza de la escritura*. Tercer Estudio Regional Coparativo Explicativo TERCE, Santiago.
- Organización de la Naciones Unidas para la Educación, la Ciencia y la Cultura. (2015). *Informe de resultados TERCE, factores asociados*.
- Otzen, T., & Manterola, C. (2017). Sampling Techniques on a Population Study. *Scielo*, 35, 6.
- Pardo Adames, C. A., & Rocha Gaona, M. C. (2009). *Diseños de pruebas para la evaluacion educativa: Reglas para la elaborar items de formato de seleccion y de produccion*. Laboratorio Latinoamericano de Evaluacion de la Calidad de la Educación.
- Perez, C., & Carretero-Dios, H. (2005). Normas para el desarrollo y revision de estudios instrumentales. *International Journal of Clinical and Health Psychology*, 5(3), 521-551.

- PISA OCDE;. (2016). *PISA 2015 Resultados Clave*.
- PNUD. (2016). *Desarrollo humano para todas las personas: Informe Desarrollo Humano 2016*. Estados Unidos: PBM Graphics.
- PNUD. (2016). *Marco de Asistencia de las Naciones Unidas para el Desarrollo -MANUD 2017-2021*. Sistema de las Naciones Unidas. Recuperado el 27 de Marzo de 2017
- Popham, W. (1983). *Evaluación basada en criterios*. Madrid: Magisterio Español.
- Popham, W. (2002). *What Every Teacher Should Know about Educational Assessment*. (Allyn , & Bacon, Edits.) Boston.
- Popham, W. (2009). ¿Por qué las pruebas estandarizadas no miden la calidad educativa? *Programa de Promoción de la Reforma de America Latina y el Caribe PREAL*, 11.
- Posas, M. (2010). Estado de la Educación en Honduras . *UPNFM* , 55-56.
- Posas, M. (2012). Política Educativa y Reforma educativa en Honduras. *Estado de la Educación en Honduras*, 29-58. (U. P. Morazan, Ed.) Tegucigalpa, Honduras.
- Prado Delgado, V., & Medina Rivilla , A. (2014). Características de la enseñanza y evaluación de los aprendizajes en educación básica. *PAPELES Universidad Antonio Nariño, facultad de Ciencias de la Educación*, 6(11), 20-35.
- Quesada, R. (2008). *Como Planear la Enseñanza Estrategica* . LIMUSA.
- Ramírez Peradotto, P. (21 de Octubre de 2011). <http://autoestimaescolar.blogspot.com/>.
Obtenido de <http://autoestimaescolar.blogspot.com/2011/10/medicion-y-calificacion.html>
- Rapalo Castellano, R., Aguilar, C. R., & Mazariegos, K. (2005). Informe de progreso educativo Honduras. *Educación: el futuro es hoy.*, 54.
- Rápalo, R. (2010). Estado de la Educación en Honduras. *UPNFM*, 24-25.
- Ravela. (2007). *Las evaluaciones educativas que America Latina Necesita* .
- Ravela Pedro, P. A. (JULIO 2007). LAS EVALUACIONES EDUCATIVAS QUE AMERICA LATINA NECESITA. *OPREAL*.

- Ravela, P. (2006). Fichas didacticas para comprender las evaluaciones educativas. (P. Ravela, Ed.) *Programa de Promocion para la Reforma Educativa en America Latina y el caribe*.
- Républica de Honduras. (2012). Ley Fundamental de Educación.
- Roman, M., & Murillo, J. (2013). La Evaluacion integrada en el proceso de Aprendizaje. (C. Martínez Garrido, N. Hidalgo Farran, & V. Gonzales, Edits.) *Revista Iberoamericana de Evaluacion Educativa*, 6(1), 7-12. Recuperado el 15 de Febrero de 2017, de <http://www.rinace.net/riee/>
- Sanchez, C. H., & Reyes, C. (1986). Metodologia y diseño en la investigacion cientifica . Lima.
- SE. (2009). *Informe Nacional de Evaluación de los Aprendizajes 2008. Educación Básica: Primero y Segundo Ciclo*. Tegucigalpa.
- SE, & MIDEH. (2017). *Informe Nacional de Desempeño Académico*. Secretaria de Educación, Dirección General de Currículo y Evaluación , Tegucigalpa, Honduras.
- Secretaria de Educacion . (2003). *Curriculo Nacional Basico*. Tegucigalpa, Honduras.
- Secretaria de Educacion . (Noviembre de 2013). *Planes y Programas Bachilleratos en Ciencias y Humanidades* . Tegucigalpa , Honduras .
- Secretaria de Educacion . (2014). *Informe de Resultados del Rendimiento Académico de los egresados de Educación Media*. Tegucigalpa.
- Secretaria de Educacion , & Consejo Nacional de Educacion . (Noviembre de 2013). *Planes y Programas de Estudio de Educacion Media "Bachillerato en Ciencias y Humanidades"*. Tegucigalpa, Honduras.
- Secretaria de Educación. (2004). *Currículo Nacional Básico*. Honduras.
- Secretaría de Educación. (2013). *Planes y programas de estudio de educación media primer año, Bachillerato en Ciencias y Humanidades*. Tegucigalpa.
- Secretaria de educación. (2014). *Plan de estudios y programas curriculares del Bachillerato en Ciencias y Humanidades*.

- Secretaria de Educacion Honduras. (2007). *Planes y programas de Estudio de Educacion Media*. Tegucigalpa M. D. C.
- Secretaría de Educación, & MIDEH. (2017). *Informe Nacional de Desempeño Académico*. Secretaria de Educación, Tegucigalpa, Honduras.
- Secretaria de Educación;. (2015). *Informe Rendimiento Académico eficacia de los aprendizajes logrados por los estudiantes de educación media*. Secretaria de Educacion , Direccion General de Curriculo y Evaluación , Tegucigalpa, Honduras.
- Secretaria Educacion Honduras, & MIDEH. (2015). *Informe Nacional de Rendimiento Académico Español y Matemáticas 1° a 9° grado*.
- SHACES. (2013). *Manual de Acreditacion Institucional y Acreditacion de Carreras de Educacion Superior en Honduras*.
- Shiel, G., & Cartwright, F. (2016). Análisis de los datos de una evaluación nacional del rendimiento académico. En V. Greaney, & T. Kellaghan (Edits.), *Evaluaciones nacionales del rendimiento académico* (Vol. 4). Washington, DC: Grupo de Banco Mundial.
- Silva Trujillo, M. (23 de Enero-abril de 2014). El estudio de la comprensión lectora en Latinoamérica: necesidad de un enfoque en la comprension. (U. d. educación, Ed.) *Innovación Educativa*, vol. 14(64), 47-55.
- Solís, M. A. (2004). *Propuesta de Reforma de la Educación Media*. Tegucigalpa.
- Solís, M., & Moncada, G. (2010). Problemas de Equidad en el Sistema Educativo Hondureño. *Paradigma: Revista de Investigacion Educativa*(30), 205-229. Recuperado el 12 de Abril de 2017
- Solozarno Salas, J., & Montero Rojas , E. (2011). Construcción y validación de una prueba de comprensión de lectura mediante el Modelo de Rasch. *Revista Electrónica "Actualidades Investigativas en Educación"*, 11(2), 1-27.

- Tapia, V., & Luna, J. (1 de Septiembre de 2010). *Validacion de una prueba de habilidades de pensamiento para alumnos de secundaria y primer año de la Universidad* . Lima, Perú.
- Tedesco, J. C. (Septiembre de 2016). Diez notas sobre los sistemas de evaluación de los aprendizajes. (R. Opertti , É. Brylinski, & G. Magni, Edits.) *Cuestiones fundamentales y actuales del currículo, el aprendizaje y la evaluación*, 11.
- Treviño, E., Fraser, P., Meyer, A., Morawietz, L., Inostroza, P., Naranjo, E., & UNESCO/OREALC. (2016). *Inorme de Resultados TERCE*. Santiago.
- Tristan López, A., & Pedraza Corpus, N. (Octubre de 2017). La objetividad en las pruebas estandarizadas. *Revista Iberoamericana de Edvaluacion Educativa*.
- Tristán López, A., & Pedraza Corpus, N. Y. (25 de Febrero de 2017). La Objetividad en las Pruebas Estandarizadas. (I. d. (IEIA), Ed.) *Revista Iberoamericana de Evalusción educativa*, 1-21.
- UMCE. (Abril de 2017). Historia. Tegucigalpa, Honduras.
- UNAH. (15 de Noviembre de 2017). Nota de Prensa Universidad Nacional Autonoma de Honduras. *Resultado PAA 2016*. Tegucigalpa.
- UNAH, & Direccion de Sistema de Admision. (15 de Noviembre de 2016). Nota de Prensa. *Resultados de la PAA del 2016*. Tegucigalpa.
- UNESCO. (2014). *La lectura en la era móvil*. Estudio sobre la lectura móvil en los paises en desarrollo: una mirada desde México.
- UNESCO. (2014). La lectura en la era movil: un estudio sobre la lectura móvil en los paises en desarrollo. 1-9.
- UNESCO. (2018). *The impact of large scale learning assessment*. Paris.
- UNESCO UIS. (Septiembre de 2017). Mas de la mitad de los niños y adolescentes en el mundo no estan aprendiendo . *Ficha informativa N° 46 Instituto Estadistica de la UNESCO*, 25.

- UNESCO, & LLECE. (2016). *Informe de Resultados Tercer estudio comparativo explicativo*. Laboratorio Latinoamericano para la Evaluación de la Calidad Educativa (LLECE). doi:ISBN 978-92-3-300041-4
- UPNFM. (2011). Investigación Educativa en la UPNFM 2006-2010. *INIEES*, 234-235.
- USINIEH, & SEDUC. (2016). *Sistema Educativo Hondureño en Cifras: Periodo Academico 2015*. Anual, Unidad del Sistema Nacional de información Educativa de Honduras 2016, Francisco Morazan. Recuperado el 5 de Abril de 2017
- Vargas Castro, K., Jara Castro, M., Lozada Meza, M., & Dume Villacis, M. (2019). Influencia de la neurociencia en el aprendizaje de la lectoescritura. *Universidad Ciencia y Tecnología* (02-19), 33-38.
- Vergara Reyes, C. (25 de Enero de 2011). Concepciones de la evaluación de los Aprendizajes de Docentes destacados de Educación Básica A. *Revista Electrónica "Actualidades Investigativas en Educación"*, 11(1).
- Vukmirovic, Z. (2016). Conceptos básicos de la Teoría Respuesta al Ítem (IRT).
- Vukmirovic, Z., & Fast, M. (2017). Paradigmas y tendencias en medición y evaluación educativa. *Desarrollo de una evaluación nacional*.
- Webb, N., & Shavelson, R. (2005). Generalizability Theory. 599-612.
- Zeron, M. S. (2013). Las evaluaciones estandarizadas sus efectos entre países latinoamericanos. *Revista Latinoamericana de estudios educativos*.

ANEXOS

Anexo n° 1: Marco Operativo de la prueba español(lectura) 11° grado BCH, 2018



Marco operativo Prueba Español(Lectura) 11° grado BCH

Numero de ítems a diseñar

BLOQUE/ AREA CLAVE	N°	COMPONENTE/ SUB- AREA	N°	ESTANDAR/ HABILIDAD O INFORMACIÓN PRECISA QUE SE QUIERE EVALUAR	NIVEL COGNITIVO				NO. ÍTEMS A DISEÑAR			
					CONOCIMIENTO	COMPRENSIÓN	APLICACIÓN	TOTAL	A CONOCIMIENTO	B COMPRENSIÓN	C APLICACIÓN	TOTAL
Lectura	1	Tipos de texto	1.1.	Leen y utilizan con diversos propósitos textos narrativos, descriptivos, expositivos y persuasivos, tanto literarios como funcionales e informativos, de varias fuentes como libros, medios masivos de comunicación e internet, incluyendo contenido lingüístico y grafico	1	2	2	5	4	8	8	20

		1.2.	Demuestran comprensión de las ideas globales, principales, secundarias e inferenciales de un texto leído, incluyendo la interpretación de gráficos e iconos		0	4	1	5		0	16	4	20
		1.3.	Identifican el contexto, la intención comunicativa, el punto de vista y el estilo de los textos narrativos, descriptivos, expositivos y persuasivos		0	2	3	5		0	8	12	20
		1.4.	Interpretan información explícita e implícita de textos diversos para formular planteamientos en sentido crítico		0	3	2	5		0	12	8	20
		1.5.	Analizan diferentes puntos de vista y evalúan rasgos científicos, sociológicos y culturales de los textos		0	4	1	5		0	16	4	20
		1.6.	Comprenden, aprovechan y evalúan críticamente imágenes y textos, iconos verbales y de informática		2	4	1	7		8	16	4	28
	2	2.1.	Reconocen y entienden en textos leídos una variedad de palabras conocidas		1	1	1	3		4	4	4	12
		2.2.	Interpretan palabras desconocidas utilizando diversas estrategias		1	3	0	4		4	12	0	16
		2.3.	Interpretan en textos leídos el significado de términos técnicos, lenguaje no-estándar y palabras con múltiples significados		0	5	1	6		0	20	4	24

			2.4.	Reconocen e interpretan el lenguaje figurado, la intencionalidad del autor y otros aspectos semánticos de las palabras		1	1	3	5		4	4	12	20
						6	29	15	50		24	116	60	200

Fuente: Secretaria de Educación, Documento Estándares y Programaciones Español 11° . (2009)

Anexo nº 2: Guión metodológico: Taller de evaluación y redacción de ítems



UNAH
UNIVERSIDAD NACIONAL AUTÓNOMA DE HONDURAS

UNIVERSIDAD NACIONAL AUTÓNOMA DE HONDURAS
MAESTRÍA EN PSICOMETRÍA Y EVALUACIÓN EDUCATIVA



GOBIERNO DE LA REPÚBLICA DE HONDURAS

★ ★ ★ ★ ★

SECRETARÍA DE EDUCACIÓN

GUIÓN METODOLÓGICO				
Nombre del Taller	Taller de evaluación educativa y redacción de ítems	Duración de la actividad	3 días	Numero de Participantes
				100
Objetivos Generales	1. Validar las tablas de especificaciones de cada una de las pruebas a través de la revisión de especialistas del área del conocimiento de Español, Matemáticas, Ciencias Naturales y Sociales.			
	2. Contribuir en la formación de un banco de reactivos de opción múltiple para la construcción de pruebas estandarizadas del nivel básico y medio.			

Días	Objetivo	Contenido	Actividad	Recursos	Responsables	Tiempo		
						Hora de inicio	Hora de termino	Duración
	Explicar la importancia que tiene la evaluación como herramienta en el proceso de enseñanza-aprendizaje	La evaluación para el aprendizaje: una herramienta para la mejora educativa.	Exponer sobre la evaluación	Data show, computador Gafete Carpeta	Nanci Castro Dulce Guevara	08:30 am	9:00 am	1 hora

Miércoles 16/05/18	Conocer el Marco Normativo para la evaluación de los aprendizajes en Honduras Acuerdo 1796 –SE-2017.	Marco Normativo para la evaluación de los aprendizajes en Honduras Acuerdo 1796 –SE-2017.	Explicar el Acuerdo 1796 –SE-2017	Data show, computador Acuerdo 1796 –SE-2017.	José Luis Cabrera Director General de Currículo y Evaluación	09:00am	9:50am	50 minutos
	Espacio de preguntas y respuestas		Se le entregará un papelito a cada experto para que escriba las interrogantes que tenga en relación a la temática. Posteriormente, serán seleccionadas al azar para que la autoridad invitada dé respuesta.	Data show, computador Papel bond Lápiz, borrador	José Luis Cabrera moderador	9:50am	10:00am	10 minutos
	Receso					10:00 am	10:20am	20 minutos
	Describir los pasos para la construcción de una prueba que mida los conocimientos en las áreas curriculares de Español, Matemáticas, Ciencias Naturales y Ciencias Sociales	Diseño de una prueba de rendimiento nacional		Data show, computador Papel bond Lápiz, borrador	Nanci Castro Damary Sosa	10:20 am	11:00am	40 minutos

	A través del marco conceptual y operativo.	Marco conceptual y operativo para la construcción de una prueba.	Presentar los marcos conceptual y operativo para la construcción de una prueba.		Carlos Aguilar Isabel Hernández	11:00	12:00	1 hora
	Almuerzo					12:00m	01:00pm	1 hora
	Explicar la tabla de especificaciones de las áreas de español, Matemáticas, Ciencias Naturales y Ciencias Sociales.	Revisión de la Tabla de especificaciones Por área.	Revisar en equipos de 5 la Tabla de especificaciones por área, explicándoles que deben examinar los estándares que toman en cuenta al momento de dar las clases. Luego se selecciona un representante por grupo para explicar al consenso que llegaron.	Data show, computador Tabla de especificaciones Por área.	Cindy Arauz Darwin Dávila	01:00pm	02:20pm	1 hora, 20 minutos
	Validar la tabla de especificaciones de los estándares educativos a través de juicios de expertos en las áreas del conocimiento de español, Matemáticas,	Validación de la Tabla de especificaciones por área.	Después de discutir la tabla de especificaciones Con papel rotafolio se va a presentar una tabla de especificaciones final.	Data show, computador	Dulce Kesly	02:20pm	03:50pm	1 hora 10 minutos

	Ciencias Naturales y Ciencias Sociales.							
	Retroalimentación las actividades desarrollas durante el proceso para consolidar criterios.	Conclusiones, acuerdos y cierre de la jornada				03:30pm	4:00pm	30 minutos

Días	Objetivo	Contenido	Actividad	Recursos	Responsables	Tiempo		
						Hora de inicio	Hora de termino	Duración
Jueves 17/05/18	Establecer los criterios de redacción de ítems para la construcción de una prueba	Criterios para la redacción de ítems	Exponer los criterios de redacción de ítems para el diseño de las pruebas.	Data show, computador Manual de redacción de ítems	Cindy Arauz Kesly García	08:00 am	9:20 am	1 hora, 20 minutos
	Elaborar ítems según los criterios establecidos para la construcción de la prueba	Practica 1 y Discusión Redacción de ítems (Elaboración de 3 ítems)	Redactar y discusión ítems.	Data show, computador Plantillas de ítems Libros de texto Tabla de especificaciones	Integrantes Línea Investigación por área y docentes.	9:20 am	10:00am	40 minutos
	Receso					10:00 am	10:20 am	20 min.
	Explicar la metodología para la	Metodología para la redacción de ítems.	Guiar el procedimiento	Data show, computador Plantillas de ítems	Nanci Castro Damary Sosa	10:20 am	10:40 am	20 minutos

	redacción de ítems en las diferentes áreas.		para la redacción de ítems.					
	Crear 10 ítems de selección única para cada área curricular.	Construcción de 10 ítems, por área curricular.	Formular 10 ítems en pares por área y retroalimentar en grupo.	Data show, computador Plantillas de ítems Libros de texto Tabla de especificaciones	Darwin Isabel Carlos Dulce	10:40am	11:40 am	1 hora
	Revisar los ítems para la unificación criterios entre los expertos de cada área.	Discusión de ítems	Discutir que ítems se acepta, se rechaza o se mejora	Data show, computador Plantillas de ítems Libros de texto Tabla de especificaciones	Docentes por área curricular	11:40 am	12:00 am	40 min.
	Almuerzo					12:00	1:00	1 hora
	Construir 10 ítems de selección única para cada área curricular.	Redacción de 10 ítems por área curricular	Formular 10 ítems individual por área	Data show, computador Plantillas de ítems Libros de texto Tabla de especificaciones	Damary Sosa Nanci Castro Cindy Arauz Kesly García	01:00 pm	02:30 pm	1 hora 30
		Retroalimentación	retroalimentar en grupo			02:30 pm	2:50 pm	20 minutos
	Receso					02:50 pm	03:10 pm	20 minutos
	Construir 10 ítems de selección única por área curricular.	Redacción de 10 ítems individual y retroalimentación	Formular 10 ítems individualmente	Data show, computador Plantillas de ítems	Docentes por área curricular e integrantes	03:10 pm	03:50 pm	40 minutos

			por área y dar retroalimentación	Libros de texto Tabla de especificaciones	de la línea de investigación			
		Conclusiones, acuerdos y cierre de la jornada				03:50pm	4:00pm	
Días	Objetivo	Contenido	Actividad	Recursos	Responsables	Tiempo		
						Hora de inicio	Hora de termino	Duración
Viernes 18/05/18	Construir 10 ítems de selección única por área curricular.	Practica y revisión de 10 ítems individual y retroalimentación	Formular 10 ítems individualmente por área	Data show, computador Plantillas de ítems Libros de texto Tabla de especificaciones	Darwin Isabel Carlos dulce	08.00 am	09:30 am	1 hora 30 min.
		retroalimentación	dar retroalimentación			09:30 am	10:00 am	30min
	Receso					10:00 am	10:20 am	20 min.
	Construir 10 ítems de selección única por área curricular.	Redacción de 10 ítems en pares y retroalimentación	Formular 10 ítems individualmente por área y dar retroalimentación	Data show, computador Plantillas de ítems Libros de texto Tabla de especificaciones	Docentes por área curricular e integrantes de la línea de investigación	10:20 am	11:40 pm	1 hora 20 minutos
		Dar retroalimentación				11:40 am	12:00m	20 minutos
	Almuerzo					12:00 pm	01:00 pm	1 hora

	Revisar la actividad de creación de ítems para la retroalimentación de del taller.	Revisión y retroalimentación final en grupo	Monitorear el proceso de redacción de ítems	Data show, computador Plantillas de ítems Libros de texto Tabla de especificaciones	Docentes por área curricular e integrantes de la línea de investigación	01:00 pm	02:00 pm	1 hora y 10 minutos
	Agradecer a los expertos por área curricular y a los invitados especiales por su valiosa participación en el taller de redacción de ítems para la evaluación educativa	Palabras de Agradecimiento Entrega de Diplomas Cierre del Evento	Brindar palabras de Agradecimiento y Entrega de Diplomas	Data show, computador Plantillas de ítems Libros de texto Tabla de especificaciones		02:00 pm	03:00 pm	30 minutos

Anexo nº 3: Convocatoria Taller de Evaluación y redacción de ítems



Comayagüela M.D.C., 08 de Mayo de 2018.

OFICIO CIRCULAR N°066-DDEFM-2018

**DIRECTORA MUNICIPAL DISTRITO CENTRAL
DIRECTORES (AS) DISTRITALES No. 4, 5, 8, 9, 11 y 13
DIRECTORES DE CENTROS EDUCATIVOS**

Distinguidas (os) Funcionarias (os):

Reciban un cordial saludo, de la manera más atenta tengo a bien comunicarles que en fecha **16, 17 y 18 de mayo**, se ha programado el **Taller de Evaluación Educativa y Redacción de Ítems** dirigido a 80 docentes del Distrito Central de los niveles educativos de educación básica y media que cuentan con las especialidades de español, matemáticas, ciencias naturales y sociales.

Este proceso de formación se llevará a cabo en un horario de **8:00am a 4:00 p.m.**, en el lugar que ocupa el **Centro Regional de Formación Permanente**, antiguas instalaciones de la Normal Mixta Pedro Nufio. El día de **miércoles 16 de mayo** la jornada comenzara a las **7:30 a.m.**

Este proceso de formación docente se lleva a cabo en conjunto con la Universidad Nacional Autónoma de Honduras, a través del programa de Post Grado en Psicometría y Evaluación, con la colaboración del Centro Regional de Formación Permanente Centro Sur Oriente (CRFPCSO).

En virtud de ello les solicito convocar en tiempo y forma a los docentes de los centros educativos que se enlistan adjunto, considerando el siguiente perfil.

- Docentes que estén frente alumnos en los grados 6º, 9º del nivel de educación básica, preferiblemente con especialidad en Ciencias Sociales.
- Docentes frente a alumnos de 10º y 11º grado de las especialidades de español, matemáticas y ciencias naturales.
- Docentes proactivos, dinámicos, empáticos y que no estén próximos a jubilarse.
- Compromiso para asistir y permanecer durante la jornada completa del taller.

Agradeciendo su valiosa colaboración en lo anterior descrito, sin otro particular

Atentamente,



LICENCIADO HECTOR NAPOLEÓN BONILLA
DIRECTOR DEPARTAMENTAL DE EDUCACION
DE FRANCISCO MORAZAN
CC. Lic. Daniel Colindres/ Sub Director de Currículo y Evaluación

"Mejorando la convivencia escolar y la calidad educativa para contribuir a la convivencia social"

Col. Tiloarque, Comayagüela, M.D.C. Teléfonos (504) 2225-1259, 2225-1903

E-mail: ddefm.edu@gmail.com Facebook/[ddefm](https://www.facebook.com/ddefm) Twitter/[ddefm](https://twitter.com/ddefm) Blog: www.ddefm.blogspot.com

Anexo 4: Plantilla para la redacción de ítems



UNAH
UNIVERSIDAD NACIONAL
AUTÓNOMA DE HONDURAS

MAESTRÍA EN PSICOMETRÍA Y EVALUACIÓN EDUCATIVA



PLANTILLA DE REVISIÓN DE ÍTEMS

IDENTIFICACIÓN DEL ÍTEMS			
ASIGNATURA:			
BLOQUE:		COMPONENTE:	
ESTÁNDAR:			
NIVEL COGNITIVO:		(1) CONOCIMIENTO ☐ (2) COMPRENSIÓN ☐	
(3) APLICACIÓN ☐			
NIVEL EDUCATIVO:		MODALIDAD:	
BÁSICO ☐		BTP ☐	
MEDIO ☐		BCH ☐	
GRADO:		FECHA:	
NOMBRE DEL REDACTOR(DOCENTE):			
ÍTEM			
CLAVE:			
OBSERVACIONES			
Redacción clara (no ambigua) en todo el ítem.	Sí	No	
El reactivo corresponde al estándar descrito en la tabla de especificaciones	Sí	No	
Vocabulario adecuado en todo el ítem.	Sí	No	
Libre de problemas ortográficos.	Sí	No	
Cumple la norma de no presentar enunciados negativos.	Sí	No	
Cumple la norma de alternativas simétricas.	Sí	No	
Todas las opciones de respuesta son plausibles.	Sí	No	
Los textos y/o figuras adicionales son adecuados.	Sí	No	
Las alternativas están libres de pistas inadvertidas.	Sí	No	
Otros (especifique)			

Anexo 5: Mapa de la prueba operativa

Prueba Español, bloque: Lectura 11° grado BCH 2018

Área	Grado	Admón.	Forma	Posición	Función	ItemID	Tipo	Clave	Nicog	PMax	Componente	Estándar
ESP	11	2018	B	1	O	ESP11018_111BOB01	MD	a	B	1	Tipos de textos	Leen y utilizan con diversos propósitos textos narrativos, descriptivos, expositivos y persuasivos, tanto literarios como funcionales e informativos, de varias fuentes como libros, medios masivos de comunicación e internet, incluyendo contenido lingüístico y grafico
ESP	11	2018	B	2	O	ESP11018_111AOB02	MD	b	A	1	Tipos de textos	
ESP	11	2018	B	3	O	ESP11018_111COB03	MD	c	C	1	Tipos de textos	
ESP	11	2018	A	4	O	ESP11018_111COA03	MD	c	C	1	Tipos de textos	
ESP	11	2018	A	5	O	ESP11018_111BOA04	MD	a	B	1	Tipos de textos	
ESP	11	2018	A	6	A	ESP11018_112BAA05	MD	d	B	1	Tipos de textos	Demuestran comprensión de las ideas globales, principales, secundarias e inferenciales de un texto leído, incluyendo la interpretación de gráficos e iconos
ESP	11	2018	A	7	O	ESP11018_112COA06	MI	c	C	1	Tipos de textos	
ESP	11	2018	B	8	O	ESP11018_112BOB07	MD	d	B	1	Tipos de textos	
ESP	11	2018	A	9	O	ESP11018_112BOA07	MD	b	B	1	Tipos de textos	
ESP	11	2018	A	10	O	ESP11018_112BOA08	MD	a	B	1	Tipos de textos	
ESP	11	2018	B	11	A	ESP11018_113BAB09	MD	C	C	1	Tipos de textos	Identifican el contexto, la intención comunicativa, el punto de vista y el estilo de los textos narrativos, descriptivos, expositivos y persuasivos
ESP	11	2018	A	12	O	ESP11018_113COA10	MI	b	C	1	Tipos de textos	
ESP	11	2018	A	13	O	ESP11018_113BOA11	MD	c	B	1	Tipos de textos	
ESP	11	2018	B	14	O	ESP11018_113COB10	MI	c	C	1	Tipos de textos	
ESP	11	2018	A	15	O	ESP11018_113BOA12	MD	d	B	1	Tipos de textos	
ESP	11	2018	A	16	O	ESP11018_113BOA13	MD	d	B	1	Tipos de textos	
ESP	11	2018	A	17	O	ESP11018_114BOA14	MD	a	C	1	Tipos de textos	Interpretan información explícita e implícita de textos diversos para formular planteamientos en sentido critico
ESP	11	2018	A	18	A	ESP11018_114BAA15	MD	a	B	1	Tipos de textos	
ESP	11	2018	A	19	A	ESP11018_114BAA16	MD	d	B	1	Tipos de textos	
ESP	11	2018	A	20	A	ESP11018_114BAA17	MD	b	B	1	Tipos de textos	
			B	21	O	ESP11018_114BOB14	MD	a	C	1	Tipos de textos	
ESP	11	2018	A	22	O	ESP11018_115BOA18	MD	b	B	1	Tipos de textos	

ESP	11	2018	A	23	O	ESP11018_115BOA19	MD	a	B	1	Tipos de textos	Analizan diferentes puntos de vista y evalúan rasgos científicos, sociológicos y culturales de los textos
ESP	11	2018	A	24	O	ESP11018_121BOA20	MD	d	B	1	Vocabulario	Reconocen y entienden en textos leídos una variedad de palabras conocidas
ESP	11	2018	B	25	O	ESP11018_115BOB21	MD	b	B	1	Tipos de textos	Analizan diferentes puntos de vista y evalúan rasgos científicos, sociológicos y culturales de los textos
ESP	11	2018	B	26	o	ESP11018_115COB22	MD	b	C	1	Tipos de textos	Analizan diferentes puntos de vista y evalúan rasgos científicos, sociológicos y culturales de los textos
ESP	11	2018	A	27	O	ESP11018_115AOA23	MI	a	A	1	Vocabulario	Reconocen y entienden en textos leídos una variedad de palabras conocidas
ESP	11	2018	B	28	O	ESP11018_116BOB25	MI	c	B	1	Tipos de textos	Comprenden, aprovechan y evalúan críticamente imágenes y textos, iconos verbales y de informática
ESP	11	2018	B	29	O	ESP11018_116AOB26	MI	C	A	1	Tipos de textos	
ESP	11	2018	A	30	O	ESP11018_116AOA25	MI	a	A	1	Tipos de textos	
ESP	11	2018	A	31	O	ESP11018_116COA26	MI	a	C	1	Tipos de textos	
ESP	11	2018	A	32	O	ESP11018_116BOA27	MD	d	B	1	Tipos de textos	
			A	33	O	ESP11018_116BOA28	MD	c	B	1	Tipos de textos	
ESP	11	2018	A	34	O	ESP11018_116BOA29	MD	b	B	1	Tipos de textos	
ESP	11	2018	A	35	A	ESP11018_121AAA30	MD	c	A	1	Vocabulario	Reconocen y entienden en textos leídos una variedad de palabras conocidas
ESP	11	2018	A	36	A	ESP11018_122BAA31	MD	c	B	1	Vocabulario	Interpretan palabras desconocidas utilizando diversas estrategias
ESP	11	2018	A	37	O	ESP11018_124BOA32	MI	b	B	1	Vocabulario	Reconocen e interpretan el lenguaje figurado, la intencionalidad del autor y otros aspectos semánticos de las palabras
ESP	11	2018	A	38	O	ESP11018_124BOA33	MI	d	B	1	Vocabulario	
ESP	11	2018	B	39	A	ESP11018_122BAB34	MD	a	B	1	Vocabulario	Interpretan palabras desconocidas utilizando diversas estrategias
ESP	11	2018	B	40	O	ESP11018_122AAB35	MD	a	A	1	Vocabulario	

ESP	11	2018	B	41	O	ESP11018_122BAB36	MD	b	B	1	Vocabulario	
ESP	11	2018	B	42	O	ESP11018_123CAB37	MD	c	C	1	Vocabulario	Interpretan en textos leídos el significado de términos técnicos, lenguaje no-estándar y palabras con múltiples significados
ESP	11	2018	A	43	O	ESP11018_123COA38	MD	a	C	1	Vocabulario	Interpretan en textos leídos el significado de términos técnicos, lenguaje no-estándar y palabras con múltiples significados
ESP	11	2018	B	44	A	ESP11018_123COB38	MD	c	C	1	Vocabulario	
ESP	11	2018	B	45	O	ESP11018_123BOB39	MD	d	B	1	Vocabulario	Interpretan en textos leídos el significado de términos técnicos, lenguaje no-estándar y palabras con múltiples significados
ESP	11	2018	A	46	O	ESP11018_124COA40	MI	b	C	1	Vocabulario	Interpretan en textos leídos el significado de términos técnicos, lenguaje no-estándar y palabras con múltiples significados
ESP	11	2018	B	47	A	ESP11018_124BOB41	MI	a	B	1	Vocabulario	Reconocen e interpretan el lenguaje figurado, la intencionalidad del autor y otros aspectos semánticos de las palabras
ESP	11	2018	B	48	O	ESP11018_124BOB42	MI	b	B	1	Vocabulario	
			A	49	O	ESP11018_123BAA42	MI	c	B	1	Vocabulario	
ESP	11	2018	B	50	O	ESP11018_124COB43	MI	b	C	1	Vocabulario	