

UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA

Instituto de Investigación y Desarrollo Educativo



“Evaluación criterial del área metodológica de la carrera de Psicología de la UABC”

*Tesis que para obtener el grado de
MAESTRO EN CIENCIAS EDUCATIVAS*

Presenta:

Juan Carlos Pérez Morán

Ensenada, B.C. Septiembre de 2010.

UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA

Instituto de Investigación y Desarrollo Educativo



“Evaluación criterial del área metodológica de la carrera de psicología de la UABC”

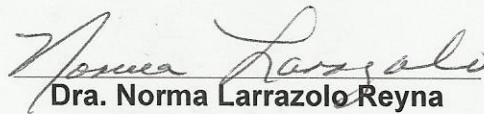
Tesis que para obtener el grado de

MAESTRO EN CIENCIAS EDUCATIVAS

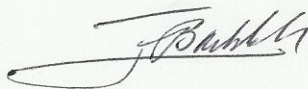
Presenta:

Juan Carlos Pérez Morán

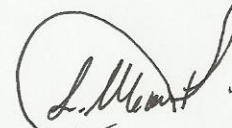
Aprobado por:


Dra. Norma Larrazolo Reyna

Directora de la tesis



Dr. Eduardo Backhoff Escudero
Sinodal



Dr. Luis Ángel Contreras Niño
Sinodal

Ensenada, B.C. Septiembre de 2010

Ensenada, B.C. a 21 de septiembre de 2010

ASUNTO: Voto aprobatorio sobre trabajo
de tesis de grado de Maestría.

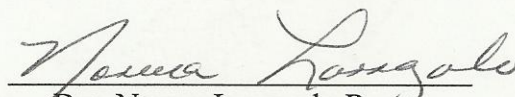
Dra. Ma. Guadalupe Tinajero Villavicencio
Coordinadora de la Maestría en Ciencias Educativas
Presente.

Después de haber efectuado una revisión minuciosa sobre el trabajo de tesis presentado por el ciudadano Juan Carlos Pérez Morán para poder presentar la defensa de su examen y obtener el grado de Maestría en Ciencias Educativas me permito comunicarle que he dado mi VOTO APROBATORIO, sobre su trabajo intitulado:

“Evaluación criterial del área metodológica de la carrera de psicología de la UABC”.

Esperando reciba el presente de conformidad, quedo de Usted.

Atentamente


Dra. Norma Larrazolo Reyna

Ensenada, B.C. a 21 de septiembre de 2010

ASUNTO: Voto aprobatorio sobre trabajo
de tesis de grado de Maestría.

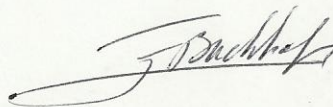
Dra. Ma. Guadalupe Tinajero Villavicencio
Coordinadora de la Maestría en Ciencias Educativas
Presente.

Después de haber efectuado una revisión minuciosa sobre el trabajo de tesis presentado por el ciudadano Juan Carlos Pérez Morán para poder presentar la defensa de su examen y obtener el grado de Maestría en Ciencias Educativas me permito comunicarle que he dado mi VOTO APROBATORIO, sobre su trabajo intitulado:

“Evaluación criterial del área metodológica de la carrera de psicología de la UABC”.

Esperando reciba el presente de conformidad, quedo de Usted.

Atentamente



Dr. Eduardo Backhoff Escudero

Ensenada, B.C. a 21 de septiembre de 2010

ASUNTO: Voto aprobatorio sobre trabajo
de tesis de grado de Maestría.

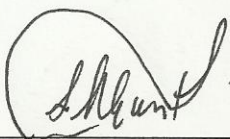
Dra. Ma. Guadalupe Tinajero Villavicencio
Coordinadora de la Maestría en Ciencias Educativas
Presente.

Después de haber efectuado una revisión minuciosa sobre el trabajo de tesis presentado por el ciudadano Juan Carlos Pérez Morán para poder presentar la defensa de su examen y obtener el grado de Maestría en Ciencias Educativas me permito comunicarle que he dado mi VOTO APROBATORIO, sobre su trabajo intitulado:

“Evaluación criterial del área metodológica de la carrera de psicología de la UABC”.

Esperando reciba el presente de conformidad, quedo de Usted.

Atentamente



Dr. Luis Ángel Contreras Niño

UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA

Instituto de Investigación y Desarrollo Educativo



“Evaluación criterial del área metodológica de la carrera de Psicología de la UABC”

Tesis que para obtener el grado de

MAESTRO EN CIENCIAS EDUCATIVAS

Presenta:

Juan Carlos Pérez Morán

Ensenada, B.C. Septiembre de 2010.

Agradecimientos

Agradezco a Dios por la existencia y a mis padres por apoyarme en cada momento de mi vida y por inculcarme que con trabajo arduo se alcanzan los sueños.

A mi esposa por su paciencia, comprensión y ayuda, pero principalmente por mostrarme que el amor es siempre un motor para vivir y salir adelante.

A mis hijos por ser la inspiración y la esperanza viva en mi existencia.

A mis hermanos Daniel, Guillermo y Omar que siempre están conmigo acompañándome y compartiendo sus experiencias.

A mi directora de tesis la Dra. Norma Larrazolo, a quien admiro y respeto mucho, agradezco todo su apoyo incondicional en cada una de las etapas de mi formación a lo largo de la maestría.

A cada uno de mis asesores y profesores de la maestría por su invaluable apoyo, especialmente al Dr. Eduardo Backhoff y al Dr. Luis Ángel Contreras que compartieron conmigo su amplio conocimiento.

A Guadalupe Delgadillo y a Nora Fuentes por su incondicional apoyo y por ser luz y guía en mi desarrollo profesional.

A la Dra. Sara Cortés y a la Dra. Guadalupe Guzmán, por haberme apoyado en momentos claves de mi proyecto de tesis.

A todos los que estuvieron conmigo siempre, muchas gracias.

Juan Carlos

ÍNDICE

<i>Agradecimientos</i>	ii
<i>Índice de tablas</i>	v
<i>Índice de figuras</i>	vi
<i>Relación de anexos</i>	vii
<i>Siglas y acrónimos</i>	ix
RESUMEN	1
I INTRODUCCIÓN	2
1.1 Definición del problema.....	3
1.2 Antecedentes de la investigación.....	8
1.3 Objetivos del estudio.....	11
1.3.1 Objetivos generales.....	11
1.3.2 Objetivos específicos.....	11
1.4 Justificación del estudio.....	12
1.4.1 Beneficios.....	13
1.4.2 Alcances.....	14
II MARCO TEÓRICO	15
2.1 Antecedentes de la evaluación del aprendizaje.....	15
2.1.1 Evaluación estandarizada del aprendizaje a nivel internacional.....	17
2.1.2 La evaluación del aprendizaje en escuelas superiores de México.....	22
2.2 Aspectos teóricos de la evaluación del aprendizaje.....	29
2.2.1 Definición y clasificación de las evaluaciones.....	30
2.2.2 Características básicas de las pruebas educativas del aprendizaje.....	38
2.2.3 Modelos teórico-metodológicos para el desarrollo de estrategias evaluativas del aprendizaje.....	51

III	MÉTODO DE EVALUACIÓN.....	62
4.1	Modelo para evaluar del área metodológica de la carrera de Psicología.....	62
4.2	Fase I. Diseño y validación del examen de trayecto.....	70
	Etapa 1.1 Definición del dominio de resultados pretendidos por el currículum.....	71
	Etapa 1.2 Análisis del currículum.....	76
	Etapa 1.3 Desarrollo del plan de evaluación.....	80
	Etapa 1.4 Producción, validación y pilotaje de ítems.....	85
4.3	Fase II. Reestructuración y ajuste del examen.....	92
	Etapa 2.1 Revisión de ítems.....	93
	Etapa 2.2 Estructuración del modelo de examen.....	95
IV	RESULTADOS DE LA EVALUACIÓN.....	98
5.1	Fase III. Análisis de los resultados de la evaluación.....	98
	Etapa 3.1 Análisis primario.....	98
	Etapa 3.2 Análisis de los resultados.....	102
	Características de la ejecución de los examinados en las líneas de formación curricular.....	109
	La socialización de los resultados.....	117
V	CONCLUSIONES.....	119
6.1	Aportaciones de la tesis.....	123
6.2	Limitaciones del estudio.....	125
6.3	Futuras investigaciones.....	127
VI	REFERENCIAS.....	130
VII	ANEXOS	

ÍNDICE DE TABLAS

Tabla I	Avances observados en materia de evaluación estandarizada del aprendizaje en la educación superior en México hasta 1999.....	25
Tabla II	Exámenes diagnósticos desarrollados por el CENEVAL para la licenciatura.....	26
Tabla III	Exámenes que se aplican actualmente en la FIM de la UABC.....	27
Tabla IV	Características generales de los tests normativos y criterial.....	41
Tabla V	Características de los tests de respuesta construida y de respuesta	42
Tabla VI	Características generales de los exámenes de pequeña escala y de gran escala (Contreras, 2000).....	43
Tabla VII	Descripción de los principales criterios que definen la calidad de un test de gran escala (Contreras, 2000; Popham, 1990).....	46
Tabla VIII	Componentes de definición de los EE (Jornet y Backhoff, 2008).....	48
Tabla IX	Ejemplos de etiquetas usuales en EE.....	48
Tabla X	Tipología de estándares de interpretación (Linn, 1994).....	49
Tabla XI	Modelo de Anthony Nitko para crear exámenes nacionales de referencia criterial y referencia normativa alineados con el currículum.....	52
Tabla XII	Modelo de INEE para crear Exámenes de Calidad de Logro Educativo nacionales alineados con el currículum...	55
Tabla XIII	Modelo de CENEVAL para el desarrollo de exámenes de ingreso, egreso y de trayecto del nivel medio superior y superior.....	57
Tabla XIV	Modelo de Contreras (2009) para desarrollar pruebas de gran escala de referencia criterial alineadas con el currículum.....	59
Tabla XV	Comparación de cuatro diferentes modelos metodológicos de evaluaciones nacionales e internacionales.....	60
Tabla XVI	Modelo para el desarrollo de la evaluación institucional de trayecto del área metodológica.....	64
Tabla XVII	Áreas de control de calidad, estándares y criterios para el desarrollo de la evaluación del área metodológica.....	69

Tabla XVIII	Redes curriculares del área metodológica y cantidad de contenidos que las integran.....	79
Tabla XIX.	Versión final de la tabla de especificaciones generales del ETAM.....	82
Tabla XX.	Resultados del análisis del piloteo de ítems del ETAM.....	90
Tabla XXI.	Taxonomía de fallas más comunes y decisiones adoptadas para el mejoramiento de los ítems (Contreras, 2003).....	93
Tabla XXII.	Fallas presentadas por los ítems de la prueba.....	94
Tabla XXIII.	Resultado del análisis de ítems del ETAM.....	99
Tabla XXIV.	Estándares de calidad técnica e Indicadores psicométricos del ETAM.....	101
Tabla XXV.	Porcentaje de alumnos por semestre y grupo escolar: 6to, 7mo y 8vo semestre de la carrera de Psicología de la FMP.....	103
Tabla XXVI.	Estadísticos descriptivos básicos y resultados del ETAM por semestre y grupo.....	104
Tabla XXVII.	Descripción genérica de las competencias académicas que logran los estudiantes en cada nivel de logro educativo (Jornet y Backhoff, 2008).....	108
Tabla XXVIII.	Porcentaje de estudiantes en los cuatro niveles de logro educativo del área metodológica de la carrera de Psicología de la FMP.....	109
Tabla XXIX.	Porcentaje de aciertos por eje curricular.....	111
Tabla XXX	Porcentaje de aciertos por red curricular.....	112
Tabla XXXI	Resultados puntuales de la ejecución de los estudiantes en el ETAM.....	114

ÍNDICE DE FIGURAS

Figura 1	Reticulación sobre la toma de decisiones concernientes al proceso de delimitación del contenido a evaluar.....	74
Figura 2	Retícula del contenido del área metodológica de la carrera de Psicología.....	78
Figura 3	Comparación de las medias en el área metodológica entre los grupos de 6to, 7mo y 8vo semestre.....	106
Figura 4	Distribución de las puntuaciones obtenidas en el ETAM.....	107
Figura 5	Porcentaje de estudiantes por semestre en los cuatro niveles de logro educativo del área metodológica de la carrera de Psicología de la FMP.....	110

RELACIÓN DE ANEXOS

Anexo 1	Ejemplo de justificaciones de la importancia de los contenidos para la elaboración de ítems.
Anexo 2	Ejemplo de una especificación puntual de ítem
Anexo 3	Ajustes de ítems
Anexo 4	Ejemplo de portada del ETAM
Anexo 5	Formato de salida del análisis de ítems del ITEMAN
Anexo 6	Resultados puntuales por ítem en la evaluación 2010-1 del ETAM

SIGLAS Y ACRÓNIMOS

AERA	Asociación Americana de Investigación Educativa
APA	Asociación Americana de Psicología
ACT	American College Testing
ANUIES	Asociación Nacional de Universidades e Instituciones de Enseñanza Superior
CCE	Comité Coordinador del Examen
CDE	Comité Diseñador del Examen
CER	Comité Elaborador de Reactivos
CENEVAL	Centro Nacional de Evaluación para la Educación Superior, A.C
CICESE	Centro de Investigación Científica y de Educación Superior de Ensenada
CIEES	Comités Interinstitucionales para la Evaluación de la Educación Superior
COMAEM	Consejo para la Acreditación de la Educación Superior en México, A.C.
CONPES	Coordinación Nacional para la Planeación de la Educación Superior
CRESST	National Center for Research on Evaluation, Standards and Student Testing
DGESPE	Dirección General de Educación Superior para Profesionales en Educación
DE	Dominio Educativo
EE	Estándares de Interpretación
EGEL	Exámenes de Egreso de Licenciatura
ESEA	Educación Elemental y Secundaria
ETAM	Examen de Trayecto del Área Metodológica
EGC-LEPRE	Examen General de Conocimiento de la Licenciatura en Educación Preescolar
EGC-LEPRI	Examen General de Conocimiento de las Licenciatura en Educación Primaria
Excale	Exámenes de la Calidad y el Logro Educativos
EXHCOBA	Examen de Habilidades y Conocimientos Básicos
EXANI I	Examen Nacional de Ingreso a la Educación Media Superior
EXANII II	Examen Nacional de Ingreso a la Educación Superior
EXANI III	Examen Nacional de Ingreso al Posgrado
FMP	Facultad de Medicina y Psicología
FIM	Facultad de Ingeniería de Mexicali
IES	Instituciones de Enseñanza Superior
IEA	Asociación Internacional de Evaluación
IIDE	Instituto de Investigación y Desarrollo Educativo
NCEM	Consejo Nacional para la Medición en Educación
NAEP	National Assessment of Educational Progress

NCLB	No Child Left Behind
OCDE	Cooperación y el Desarrollo Económico
PAA	Prueba de Aptitud Académica
PC	Puntuación de Corte
PISA	Programme for International Student Assessment
PIRLS	Progress in International Reading Literacy Study
ProGEN	Programa de Fortalecimiento de la Gestión Estatal
ProFEN	Programas de Fortalecimiento de la Escuela Normal
SEP	Secretaría de Educación Pública
SESI	Subsecretaría de Educación Superior e Investigación Científica
SAT	Collage Board, y el Scholastic Aptitud Test
SACMEQ	South African Consortium For the Monitoring of Educational Quality
SERCE	Segundo Estudio Regional Comparativo y Explicativo
SIMCE	Sistema de Medición para la Calidad Educativa
TIMS	Third International Mathematic and Science Study
TOEFL	Test of English as a Foreign Language
LLECE	Laboratorio Latinoamericano de la Evaluación de Calidad de la Educación
GRE	Graduate Record Examination
UNAM	Universidad Nacional Autónoma de México
UA	Unidades Académicas
UAM	Universidad Autónoma de México
UABC	Universidad Autónoma de Baja California
UEE	Unidad de Evaluación Educativa
UNESCO	Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura

RESUMEN

El presente trabajo tuvo dos objetivos principales: a) evaluar el nivel de aprendizaje en el área metodológica de los estudiantes al concluir satisfactoriamente los créditos de las asignaturas pertenecientes al área metodológica de la Carrera de Psicología de la Facultad de Medicina y Psicología (FMP), de la Universidad Autónoma de Baja California (UABC) y b) desarrollar y validar un examen institucional de trayecto de referencia criterial alineado al currículum.

Para cubrir estos dos objetivos principales, fue necesario realizar previamente una serie de revisiones sobre fundamentos normativos y teóricos que ayudaran a la comprensión del desarrollo de evaluaciones educativas en el nivel superior de educación. Después, se buscó y adoptó un modelo teórico-metodológico de evaluación, a través de una revisión de diferentes autores especializados en esta área. Principalmente se realizó una adaptación del modelo de Nitko (1994) para desarrollar evaluaciones nacionales de referencia criterial y referencia normativa alineada al currículum para la certificación y selección de estudiantes. Una vez realizadas las revisiones sobre fundamentos normativos y teóricos, se procedió al desarrollo de la evaluación de trayecto. El modelo evaluativo se estructuró en tres Fases principales. La primera Fase consistió en el diseño, validación y piloteo del examen de trayecto. En la segunda fase se estructuró y ajustó el examen para su posterior aplicación a la población de estudio. En la última Fase se realizó el análisis de los resultados de la evaluación y se socializaron los resultados.

Durante el análisis de los resultados se describió la ejecución de los examinados en el currículum. Primero se calculó el porcentaje de aciertos que presentaron los examinados por semestre, por nivel de logro, por eje y por red curricular. Después se puntualizó en los resultados por cada uno de los contenidos evaluados contrastando los ejes y las redes curriculares. Por último se socializaron los resultados presentándolos en un seminario ante los académicos de la FMP y en la página electrónica oficial de la institución (<http://medicina.tij.uabc.mx/>).

I. INTRODUCCIÓN

Dentro del quehacer de las Instituciones de cualquier nivel educativo se destaca la tarea del proceso de evaluación, debido a que es indispensable, importante e inevitable. La evaluación del proceso de enseñanza y aprendizaje debe considerarse como una actividad necesaria, en tanto que aporta a los directivos de las instituciones educativas, a los profesores, a los alumnos y a la sociedad en general un mecanismo de realimentación que permite la autorregulación y el conocimiento de los factores que promueven dicho proceso (Díaz y Hernández, 2002).

Para cualquier Organización educativa la evaluación es una actividad histórica, sistematizada y de gran importancia que valora el hecho educativo de las instituciones (Carrion, 1991). Para Popham (1993) la evaluación educativa es un proceso sistematizado que consiste en evaluar formalmente la calidad educativa. Por su parte Livas (1988, en Antillón, 2003) señala que la evaluación es un proceso en donde se obtienen datos en forma sistematizada y objetiva, en donde la interpretación de la información resultante apoya directamente la creación de políticas para impulsar la calidad educativa. Para Galli (1991) el proceso de evaluación requiere de instrumentos de medición con alta calidad técnica que ayuden a la identificación de los puntajes de aprovechamiento, aportando evidencias empíricas del aprendizaje de los estudiantes.

El presente trabajo se centró en la evaluación del aprendizaje del área metodológica de la Carrera de Psicología de la Facultad de Medicina y Psicología (FMP), de la Universidad Autónoma de Baja California (UABC). Como base para realizar dicha evaluación, se desarrolló y validó el Examen de Trayecto del Área Metodológica (ETAM), el cual se concibe como un examen institucional de referencia criterial alineado al currículum y de bajo impacto.

1.1 Definición del problema

La evaluación educativa vista desde el ámbito de circunscripción, es un mecanismo que permite verificar la pertinencia y la calidad de los programas educativos. Así, la actividad de evaluación se circunscribe tanto en el ámbito de la relación profesor-alumno dentro del aula, como en los diferentes niveles de los planes de estudio, de la estructura administrativa, y de los espacios físicos de las instituciones educativas.

En particular, en la investigación evaluativa, se espera que el evaluador intervenga de algún modo en la realidad evaluada y como resultado de su acción se propicie algún tipo de cambio. Cook y Richart (1997) argumentan que lo anterior es una característica particular de la investigación evaluativa que la distingue de otros tipos de investigación. De hecho, Weiss(1987) ya había expresado algo parecido al plantear que lo que distingue a la investigación evaluativa no es el método o la materia de la investigación, si no, la intención y la finalidad con que se lleva a cabo. Desde esta perspectiva, el objetivo de la investigación evaluativa, es medir los efectos de un programa en comparación con las metas que se propuso alcanzar, a fin de contribuir a la toma de decisiones para mejorar la operación futura del programa en cuestión. En la misma línea Ruthman (1983) plantea que la investigación evaluativa consiste en la aplicación de procedimientos científicos para acumular evidencias validas y fiables sobre la manera y grado en que un conjunto de actividades especificas produce resultados o efectos concretos.

Delimitando el ámbito de la evaluación a la educación superior, encontramos que en México, los Comités Interinstitucionales para la Evaluación de la Educación Superior (CIEES), asentaron en la segunda edición del Marco de Referencia para la Evaluación de las Ciencias de la Salud varios criterios básicos sobre el proceso de la investigación en evaluación (CIEES, 2000). Un criterio a resaltar en dicho marco de referencia, es la necesidad de la participación de todos los implicados en el proceso de evaluación. Lo anterior da pie a la conformación de cuerpos colegiados

para el desarrollo de estrategias evaluativas de calidad. Cabe señalar, que el trabajo de cuerpos colegiados en el diseño, operación y análisis de resultados de la evaluación, es un factor fundamental para obtener información de calidad sobre la operación de cualquier plan de estudios (SEP, 2007). También dicho trabajo fomenta la auténtica toma de conciencia respecto del papel que desempeña cada quien dentro de una institución.

Por otra parte, en el mismo contexto de la evaluación de la educación superior, se atienden cuatro diferentes ámbitos: 1) el nivel individual, que es el proceso de evaluación más fino y constituye la base para la innovación académica; 2) la evaluación de los programas educativos y departamentos, que es esencial para fortalecer la calidad académica y la eficiencia de las funciones sustantivas; 3) la evaluación institucional, que es la que permite impulsar el desarrollo de cada casa de estudios, y 4) la evaluación del sistema nacional de educación superior y de los subsistemas que lo componen. (Tunnermann, 2003) Los cuatro ámbitos de evaluación son elementos fundamentales para la mejora continua del proceso educativo y para la acreditación de la calidad de la educación superior. Sin embargo, es muy importante delimitar la estrategia evaluativa con el fin de tener un mejor control de sus procesos y productos. En lo particular el ámbito de la evaluación de programas educativos y departamentos en el nivel superior, juega un papel importante en el fortalecimiento de la calidad académica.

Por otra parte, evaluar los programas educativos, permite a cualquier institución educativa verificar si el alumno está adquiriendo las competencias y objetivos preestablecidos en el plan de estudios en cuestión. También ayuda a verificar si las estrategias utilizadas por el profesor promueven la adquisición de los objetivos de aprendizaje. Cabe señalar que las Instituciones de Enseñanza Superior (IES) de México, no tienen un currículum común y universal, por lo que las estrategias evaluativas alineadas al currículum deben particularizarse para cada una de ellas (Gago y Mercado, 1995).

La UABC, como miembro de la Asociación Nacional de Universidades e Instituciones de Enseñanza Superior (ANUIES) y como una institución que busca la calidad de todos sus programas educativos en forma continua, se somete a diversas evaluaciones realizadas por instituciones acreditadoras. Dos importantes instituciones que evalúan la calidad educativa en México y que realizan gran parte de la acreditación de la UABC, son el Consejo para la Acreditación de la Educación Superior en México, A.C. (COMAEM) y el Comité Interinstitucional para la Evaluación de la Educación Superior (CIEES). La participación de estas instituciones en la evaluación que realiza la UABC se ve reflejada en las políticas educativas que se establecen en los documentos oficiales de la institución.

El Estatuto Escolar de la UABC publicado en la gaceta universitaria el 5 de agosto del 2006, en el título tercero trata la evaluación del aprendizaje. El capítulo primero del título tercero de dicho estatuto, trata el objeto de la evaluación y las escalas de calificación. En particular el artículo 63 define el objeto de la evaluación. A continuación se presenta dicha normatividad:

Artículo 63: La evaluación de los procesos de aprendizaje tiene por objeto:

- I. Que las autoridades universitarias, los académicos y alumnos dispongan de la información adecuada para evaluar los resultados del proceso educativo y propiciar su mejora continua;
- II. Que los alumnos conozcan el grado de aprovechamiento académico que han alcanzado y, en su caso, obtengan la promoción y estímulo correspondiente, y
- III. Evidenciar las competencias adquiridas durante el proceso de aprendizaje.

Asimismo el capítulo tercero en los artículos 79 y 80, trata el tema de las evaluaciones institucionales:

Artículo 79.- La Universidad aplicará periódicamente evaluaciones de carácter institucional que revelen el grado de aprendizaje de los alumnos inscritos en un programa educativo, con el propósito de disponer de la información adecuada para valorar los resultados del proceso educativo y propiciar su mejora continua.

Artículo 80.- Son evaluaciones de carácter institucional:

- I. Los exámenes departamentales;
- II. Los exámenes de trayecto;
- III. Los exámenes de egreso;
- IV. Los demás que se determinen para cumplir con los propósitos en el que precede.

Cabe señalar que las evaluaciones a las que se refieren las fracciones anteriores son obligatorias para el alumno. Por otra parte, es necesario para el objeto de estudio de la presente investigación, puntualizar la normatividad institucional referente a los exámenes de trayecto:

Artículo 80.- Los exámenes de trayecto son una variante de la evaluación departamental, que tienen como propósito específico evaluar las competencias académicas adquiridas por los alumnos al terminar una o más etapas de formación del plan de estudios en el que se encuentren inscritos, o en los periodos escolares específicos que determine la unidad académica.

Artículo 81.- Los exámenes de trayecto serán aplicados en forma simultánea a todos los alumnos inscritos en la etapa de formación o periodo escolar a evaluar. La unidad académica establecerá el valor en calificación o créditos que se asignarán a los alumnos evaluados por exámenes de trayecto, así como los procedimientos de asignación, conversión o transferencia de créditos correspondientes.

Respecto a lo mencionado en el artículo 81 sobre el establecimiento de créditos para los alumnos evaluados por el examen de trayecto, los cuerpos colegiados decidieron en forma consensuada establecer una lógica diferente. Dicha decisión se fundamentó en el propósito con que se desarrolló el examen, el cual, es de diagnóstico y de bajo impacto. Es decir, los resultados de la evaluación de trayecto del área metodológica sirvieron para conocer en qué medida estaban siendo logrados los conocimientos esperados por los estudiantes al concluir satisfactoriamente los créditos de las asignaturas pertenecientes al área metodológica de la Carrera de Psicología de la UABC. Además la evaluación no tuvo consecuencias directas para los alumnos. Su finalidad principal fue obtener evidencias empíricas sobre la operación del currículum.

1.2 Antecedentes de la investigación

El campo de la evaluación, en específico el referente al desarrollo de pruebas psicológicas y educativas, ha crecido con rapidez en el mundo occidental desde la década de los años 20. Actualmente, se elaboran y distribuyen en el mercado una gran infinidad de test (Aiken, 1996). Después de la Segunda Guerra Mundial, las pruebas estandarizadas, en especial las que evalúan el desempeño académico, se extendieron por todo el mundo. Durante el siglo XX, no se detuvo el avance en ésta área. También durante el siglo pasado, se desarrollaba paralelamente la estadística y el análisis factorial.

En lo estrictamente psicométrico, los conceptos como la confiabilidad y validez, necesarios para la calidad de las pruebas se mantuvieron estables a lo largo del siglo XX. Sin embargo, el panorama cambio debido a que estos conceptos se aplicaban ingenuamente desde la teoría clásica, la cual progresivamente se enriqueció por planteamientos más sofisticados como la teoría de la generalizabilidad (*generalizability*) y la de respuesta al ítem (*item response theory*) (Nunally y Berstein, 1994).

En las últimas décadas han surgido innovaciones en el campo de la psicometría. Una de las innovaciones más importante es la redefinición del concepto de validez. Dicho concepto de validez, ha pasado desde un concepto que se distinguía por su criterio concurrente o predictivo y de constructo, a una concepción unitaria, con varias fuentes de evidencia (Martínez, 2001).

En Estados Unidos (E.U.) las pruebas que comprenden versiones diferentes y que son aplicadas en múltiples ocasiones, han tenido una gran tradición. Entre las más destacadas se encuentran la *American College Testing* (ACT) o mejor conocida como *Collage Board*, y el *Scholastic Aptitud Test* (SAT). Algunas Instituciones de Educación Superior (IES) privadas y públicas en México, utilizan

aún estas pruebas en la versión puertorriqueña (Backhoff, Larrazolo y Rosas, et al, 2000).

En Latinoamérica y África la tendencia en el desarrollo de las pruebas educativas es la misma. Además de México, desde 1981 Chile destaca por la introducción del *Sistema de Medición para la Calidad Educativa* (SIMCE). En Argentina, Brasil, Colombia, Honduras, Bolivia y toda la América Central construyeron pruebas a gran escala. *La Oficina Regional* de la UNESCO ha presionado para que se logre este fin. Por otra parte, en África se han desarrollado sistemas de evaluación con pruebas de gran escala: el *South African Consortium For the Monitoring of Educational Quality* (SACMEQ) (Martínez, 2001). Dicho sistema evalúa poblaciones de 6to grado únicamente. Esto es porque la mayoría de los países participantes tienen tasas de repetición relativamente altas y hay una gran variación en la edad de los examinados (Ross y Makuwa, 2007).

En la actualidad las instituciones de educación superior en México son conscientes de la gran necesidad de contar con exámenes de buena calidad técnica que muestren evidencias del desempeño y el logro académico de sus alumnos. El Centro Nacional de Evaluación para la Educación Superior, A.C. (CENEVAL), actualmente reúnen una gran diversidad de exámenes con distintos enfoques y objetivos, como por ejemplo, la serie de Exámenes de Egreso de Licenciatura (EGEL) o los Exámenes Diagnósticos de Licenciatura (CENEVAL, s.f.a). El desarrollo de nuevos exámenes tiene el fin de responder a la creciente demanda por parte de las IES sobre aspectos de evaluación, principalmente aspectos referidos a la creación de instrumentos de medición de alta calidad técnica que aporten información pertinente para la toma de decisiones.

En el caso particular de la Facultad de Medicina y Psicología (FMP) de la UABC, Campus Tijuana, en el año 2007 inicio la gestión y construcción de un sistema interno de evaluación. Durante el proceso de acreditación de la FMP como Programa de Calidad, y en respuesta a las observaciones hechas por el COMAEM y el CIEES, conformó el Departamento de Evaluación. Los encargados del

departamento después de un diagnóstico de la problemática institucional en el tema de evaluación, tomaron como uno de sus principales objetivos a corto plazo, la construcción de exámenes departamentales y de trayecto para las etapas básica, disciplinaria y terminal con el fin de mejorar y ampliar la tarea evaluativa.

Por lo tanto, los encargados de la evaluación de la FMP decidieron darse a la tarea de construir y desarrollar instrumentos de medición con la mejor calidad técnica posible. Esto fue necesario para identificar objetivamente los conocimientos y habilidades adquiridos por los alumnos, y obtener información sobre la operación del currículum escolar y del plan de estudios. Así la FMP buscó cumplir con las observaciones hechas por las instituciones acreditadoras.

La UABC como uno de sus objetivos expuestos en su Plan de Desarrollo Institucional 2007-2010 (UABC, 2007) prevé la importancia del desarrollo de exámenes de trayecto con el fin de mejorar la calidad educativa. En este sentido, fue de suma importancia que, ante la necesidad creciente de evidenciar el desempeño y los logros académicos durante las diferentes etapas de la carrera de psicología, se construyeran instrumentos de medición con características técnicas de alto nivel. En particular, el desarrollo de instrumentos de medición de logro educativo que reúnan las suficientes evidencias de validez, confiabilidad, normalización y estandarización para poder realizar una evaluación de calidad.

Para el presente trabajo, se tomaron de referencia dos estudios importantes en el ámbito del desarrollo de estrategias evaluativas. Uno de ellos, es el trabajo realizado por Backhoff (1992) que muestra el desarrollo del Examen de Habilidades y Conocimientos Básicos (EXHCOBA) para los aspirantes a la Universidad Autónoma de Baja California. El segundo estudio de referencia es el presentado por Contreras (2009) sobre el desarrollo y validación de un modelo para elaborar exámenes criteriosales de gran escala alineados con el currículum. Dichos estudios aportan las bases para el desarrollo de una estrategia de evaluación, como es el caso de la presente tesis.

1.3 Objetivos del estudio

Tomando en cuenta lo mencionando, se plantearon los siguientes objetivos:

1.3.1 Objetivos generales:

- Evaluar el nivel del aprendizaje en el área metodológica de los alumnos que cursaron la etapa disciplinaria de la carrera de Psicología de la Facultad de Medicina y Psicología de la UABC.
- Desarrollar y validar un examen de trayecto institucional de referencia criterial alineado al currículum para el área metodológica.

1.3.2 Objetivos específicos:

- Documentar los antecedentes de la evaluación educativa en México y del desarrollo y validación de exámenes de logro educativo de referencia criterial.
- Desarrollar un examen de trayecto institucional de referencia criterial alineado al currículum para el área metodológica.
- Obtener evidencias de validez del examen de trayecto.
- Conocer los niveles de logro educativo en el área metodológica que alcanzan los estudiantes de la carrera de psicología al concluir la etapa disciplinaria.
- Contribuir en la generación de las condiciones necesarias para establecer un proceso permanente de evaluación de la calidad del aprendizaje.

1.4 Justificación del estudio

Si se pretende mejorar la calidad de la evaluación en la educación superior en nuestro país, una de las vías para alcanzar dicho objetivo, es desarrollar pruebas educativas con una alta calidad técnica. Desde esta perspectiva, es necesario que cualquier estrategia evaluativa se someta a procesos de desarrollo, análisis y administración adecuados para generar una excelente calidad de los resultados (Backhoff, 2001). En concordancia con lo anterior, el propósito del presente estudio fue la implementación de una estrategia evaluativa institucional apoyada en el desarrollo de un examen que cubriera los estándares básicos de calidad técnica internacionales.

La aplicación del examen fue de gran importancia debido a que los resultados sirvieron para realimentar a los directivos, a los profesores, a los estudiantes y a la sociedad en general sobre las habilidades y los conocimientos adquiridos por los estudiantes en el área metodológica. Por otra parte, la experiencia adquirida por los cuerpos colegiados en el desarrollo del examen, facilitó el desarrollo de estrategias evaluativas en otras áreas sustantivas del plan de estudio de la institución. Puntualizando, un instrumento de medición que arroje información precisa del aprendizaje con respecto al currículum escolar, tiene un valor estratégico para la institución educativa que lo aplica. En este caso para la Facultad de Medicina y Psicología (FMP).

También cabe señalar, que anterior a este estudio, no hay evaluaciones en la FMP que aporten evidencias empíricas de la operación del currículum en el área metodológica. Tampoco hay instrumentos de evaluación institucionales que presenten una buena calidad técnica a la luz de estándares internacionales para el desarrollo de pruebas psicológicas y educativas. Los únicos instrumentos utilizados para evaluar el área metodología en la FMP, son los Exámenes Generales de Egreso de la Licenciatura (EGEL) desarrollados por académicos del CENEVAL. Sin

embargo, dichos exámenes son aplicados con el fin de acreditar a los estudiantes al terminar la carrera y no con el fin de dar información sobre la formación de los futuros psicólogos en la fase intermedia de su licenciatura, sobre los conocimientos y habilidades intelectuales en el área metodológica que se considera básica para su formación profesional.

1.4.1 Beneficios:

- **Prácticos:** El presente estudio, ofrece el diseño y validación de un instrumento que proporcionó evidencias empíricas sobre la operación del currículum, específicamente en el área metodológica de la carrera de Psicología de la UABC.
- **Institucionales:** Da cuenta a través del análisis de los resultados la operación del currículum del área metodológica. En específico se ilustra el procedimiento para dar cuenta de la ejecución de los examinados en los ejes, líneas de formación y contenidos específicos en que se clasificó según su relevancia por cuerpos colegiados el contenido del área metodológica de la carrera de psicología. También se ofrece información para la toma de decisiones de manera directa a alumnos, profesores y directivos.
- **Sociales:** De manera indirecta aporta información a los familiares, al estado y a la sociedad en general sobre el nivel de aprendizaje en una área tan relevante para la formación del psicólogo como es el área metodológica.
- **Teóricos:** Presenta la adaptación de un modelo teórico-metodológico para el desarrollo de un examen de logro educativo pertinente a un contexto de aplicación en particular. Sin embargo, dicho modelo puede generalizarse a otras áreas disciplinarias y a otros planes de estudio de la educación superior con las previas revisiones y adaptaciones necesarias para el caso.

1.4.2 Alcances:

- Institucional y regional, ya que la evaluación de la formación metodológica de los estudiantes de la Carrera de Psicología se encuentra contextualizada sobre un único currículum operante en la Facultad de Medicina y Psicología de la UABC.
- Interdisciplinario, puesto que contribuyó en la generación de las condiciones necesarias para, eventualmente, evaluar un mayor rango de diferentes áreas de formación relevantes para el perfil del estudiante de Psicología de la FMP.
- Permanente, siempre y cuando el desarrollo y aplicación del examen se establezca en un proceso continuo de evaluación de la calidad del aprendizaje en las diferentes áreas curriculares de relevancia para la formación del psicólogo.
- Limitado, porque como anteriormente se dijo, la evaluación tuvo como objeto determinar el nivel de aprendizaje en un área del conocimiento contextualizada en un currículum único operante en la Facultad de Medicina y Psicología de la UABC.

II MARCO TEÓRICO

Este capítulo engloba dos apartados importantes: primero, los antecedentes de las evaluaciones estandarizadas del aprendizaje y segundo, los aspectos teóricos de la evaluación del aprendizaje. En el primer apartado, se menciona en lo particular, el desarrollo de las evaluaciones estandarizadas del aprendizaje a nivel internacional y la evaluación del aprendizaje en las escuelas superiores de México. En el segundo apartado, se mencionan por su parte, la definición y clasificación de las evaluaciones, las características básicas de las pruebas educativas del aprendizaje y algunos modelos teóricos-metodológicos nacionales e internacionales para el desarrollo de estrategias evaluativas del aprendizaje.

2.1 Antecedentes de la evaluación del aprendizaje

El desarrollo histórico de la evaluación educativa del aprendizaje ha sido basto, pues se trata de una disciplina que ha sufrido profundas transformaciones conceptuales a lo largo de la historia y, sobre todo, a lo largo del siglo XX. Algunos autores como Cabrera (1986) y Salvador (1992) mencionan en sus trabajos tres grandes épocas. El punto de donde partieron para establecer las épocas históricas, es la figura de Tyler en el segundo cuarto del Siglo XX. Esta época, es descrita como “el nacimiento de la verdadera evaluación”. La época anterior a la de Tyler se le denomina “de precedentes” o “antecedentes” y a la posterior “de desarrollo”.

Otros autores como Guba y Lincoln (1989), establecen cuatro generaciones del desarrollo histórico de la evaluación. La primera generación es llamada de “medición” y llega hasta el primer tercio del siglo XX, la segunda es la de la “descripción”, la tercera la del “juicio” o “valoración”, y la actual, apoyada en el

enfoque constructivista y en las necesidades de los actores sociales implicados en la evaluación.

Durante la generación de los *precedentes* de la evaluación, se aplicaron estrategias evaluativas sin ningún referente teórico para valorar, diferenciar y seleccionar a estudiantes. Entre las aplicaciones evaluativas más reconocidas en esta generación, se encuentran los procedimientos que se empleaban en la China imperial, hace más de tres mil años, para seleccionar a los altos funcionarios. En Grecia, el tratado más importante de evaluación es el Tetrabiblos, que se atribuye a Ptolomeo. Después, durante la edad media, se introducen los exámenes formales en las universidades. Para el Renacimiento, se utilizaron procedimientos selectivos en los exámenes para “ingenios de las ciencias” (McReynold, 1975, como aparece en Escudero, 2003).

A finales de la época de los *precedentes* de la evaluación, a inicios del siglo XIX, se establecen los sistemas nacionales de educación. Tal realidad exigió nuevas formas de evaluar y de hacer frente a las necesidades de la sociedad creciente. En las naciones industrializadas, a través de la acreditación de un examen de estado, se entregaron diplomas de graduación. En los Estados Unidos, en 1845, comienza a utilizarse las primeras técnicas evaluativas de tipo test que intentan evaluar en forma rutinaria las destrezas de lecto-escritura. Sin embargo, estos exámenes aún carecían de procedimientos estandarizados y objetivos (Escudero, 2003). Por último, a mediados de la etapa de los precedentes, Rice en 1897, realizó la primera investigación evaluativa en educación a través de un análisis comparativo sobre la ortografía entre escuelas. Dicho estudio, se enfocó en las puntuaciones obtenidas en los tests.

2.1.1 Evaluación estandarizada del aprendizaje a nivel internacional

El campo de la evaluación, en específico, el desarrollo de pruebas psicológicas y educativas, creció con rapidez en el mundo occidental durante la década de los años 20 (Aiken, 1996). Después de la segunda guerra mundial, los tests estandarizados, en especial aquellos del desempeño académico, se extendieron por todo el mundo. Durante el siglo XX el avance en esta área no se detuvo.

Aunado al desarrollo de los test, durante el siglo XX, se desarrolló la estadística avanzada, fortaleciendo de esta manera la consolidación de calidad técnica de los test. Algunos de los avances estadísticos más sobresalientes durante esta época, son las primeras formulaciones de Spearman y los análisis factoriales (Melia, 1990). En lo estrictamente psicométrico, los conceptos como la confiabilidad y la validez, se mantuvieron estables. Hay que señalar, que los trabajos pioneros de Spearman de principios del siglo XX, dan las raíces a la Teoría Clásica de los Test. Dicha teoría, progresivamente se enriqueció por planteamientos más sofisticados como la teoría de la generalizabilidad (*generalizability*) y la de respuesta al ítem (*Item response theory*) (Nunally y Berstein, 1994).

Para comienzos del siglo XX, los tests psicológicos tuvieron mayor impacto, siendo probablemente la obra de Thorndike en 1904, la de mayor influencia. En Francia, destacan los trabajos de Alfred Binet y de Stanford sobre tests de capacidades cognitivas. Después, durante la primera guerra mundial, Otis y Yerkes en 1917 dirigen un equipo que construye tests colectivos de inteligencia general (Alfa para lecto-escritores y Beta para analfabetos) e inventarios de personalidad (Melia, 1990). Terminada la Guerra, las pruebas psicológicas sirven a fines sociales aplicándose para 1923 la primera prueba estandarizada de rendimiento, el Stanford Achievement Test (Hamilton y Koretz, 2002, como aparecen en Martínez, 2003). Después de la guerra hasta los años sesenta del siglo pasado hubo un gran desarrollo de pruebas educativas en Estados Unidos. Sin embargo, los resultados

solo se utilizaron para decisiones en el ámbito individual y no para evaluar sistemas educativos (Martínez, 2003).

A mediados del siglo pasado, entre 1932 y 1940, se generó una revolución en el ámbito de la evaluación educativa. Smith y Tyler en 1942 publicaron el famoso *Eight Year Study of Secondary Education* para la *Progressive Education Association*. Debido al impacto de este estudio, Tyler es considerado como el padre de la evaluación educativa. Tal estudio, plantea la necesidad de una evaluación científica que sirva para perfeccionar la calidad de la educación. También Tyler en el mismo estudio, expuso su concepción sobre el currículum, de tal manera que integró su método sistemático de evaluación educativa, y lo definió como el proceso surgido para determinar en qué medida han sido alcanzados los objetivos previamente establecidos en el currículum. Para 1970, en España, con la ley General de Educación, se extendieron los planteamientos de Tyler (Escudero, 2003). Por estos y otros motivos, esta época se consideró la del nacimiento de la verdadera evaluación.

Para 1958, ya durante la etapa del *desarrollo*, en Estados Unidos, se promulgó una nueva ley fundamenta el desarrollo de programas educativos, los medios para evaluarlos, y la rendición de cuentas. Se crea así, la *National Assessment of Educational Progress* (NAEP). La labor principal de dicha institución, era evaluar los programas compensatorios impulsados por el Título I de la Ley de Educación Elemental y Secundaria (ESEA) (Hamilton y Koretz, 2002, como aparecen en Martínez 2003). Durante las dos últimas décadas del siglo XX, el uso de pruebas de rendimiento estandarizadas en gran escala, se extendió en muchos países del mundo. Específicamente en los Estados Unidos, se adoptaron progresivamente nuevas características de la evaluación. Por ejemplo, las pruebas se utilizaron para formular juicios evaluativos en estudiantes, escuelas, sistemas educativos, estados, regiones y en el país entero. Además, se vuelve notorio el uso de las pruebas para realizar comparaciones internacionales (Martínez, 2003).

En forma paralela, al desarrollo de las pruebas educativas en el campo de la psicometría, en 1985, la Asociación Americana de Psicología (APA) en colaboración con la Asociación Americana de Investigación Educativa (AERA) y el Consejo Nacional para la Medición en Educación (NCME) (1999) publicaron el manual: “Estándares para las pruebas educativas y psicológicas”. Dicho documento, subraya la importancia en la procuración del bienestar de los individuos que contestan las pruebas. También señala la necesidad de evitar el mal uso de los instrumentos de evaluación. En particular, el establecimiento de los estándares, fortaleció el desarrollo de evaluaciones estandarizadas de alta calidad, principalmente las de gran escala a nivel nacional e internacional (Backhoff, 2006).

El desarrollo de evaluaciones estandarizadas durante las dos últimas décadas del siglo XX, se extendió a nivel regional y mundial en una amplia gama de países con diversas culturas y formas de gobierno. Una prueba de lo dicho, es la creciente participación de los países en las evaluaciones internacionales como la Programme for International Student Assessment (PISA), la Third International Mathematics and Science Study (TIMSS) y el Progress in International Reading Literacy Study (PIRLS). Otra prueba, es la participación de países en evaluaciones regionales tales como el Segundo Estudio Regional Comparativo y Explicativo (SERCE) (en América Latina), el Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ) (en África) y además, el desarrollo de diversos sistemas y subsistemas nacionales de evaluación.

Algunas de las evaluaciones mencionadas, parten de organismos internacionales como la UNESCO y la Organización para la Cooperación y el Desarrollo Económico (OCDE). Sin embargo, la organización más antigua en este rubro, es la Asociación Internacional de Evaluación (IEA), que desde los años cincuenta inició aplicaciones de exámenes en países europeos. Actualmente la IEA, se conoce como prueba TIMSS, la cual se utiliza en México desde 1990 (Díaz, 2006).

Para el 2002, el Presidente George Bush, gestionó la adopción de la última versión de la Ley de Educación Elemental y Secundaria, comúnmente conocida como *No Child Left Behind* (NCLB) (Martínez, 2003). La noción de esta ley, es que todos los estados americanos deben establecer un conjunto de estándares curriculares y un sistema de pruebas estatales alineadas a dichos estándares (*Test-Based Accountability Systems*). Dichas evaluaciones, debían ser aplicadas a todos los alumnos de 3ro a 8vo grado del nivel básico.

Consecuentemente, los sistemas nacionales de evaluación estandarizada en América Latina, se desarrollaron con fuerza. Ravela *et al.* (2008) en su trabajo titulado “Las Evaluaciones Educativas que América Latina Necesita” hacen un recuento de las evaluaciones de la región durante las últimas dos décadas. En resumen:

- Dieciséis países, durante los últimos 12 años, estaban participando en el Segundo Estudio Regional en 3ro y 6to de primaria que realiza la OREAL/UNESCO: Argentina, Brasil, Chile, Colombia, Costa Rica, Cuba, Ecuador, El Salvador, Guatemala, México, Nicaragua, Panamá, Paraguay, Perú, República Dominicana y Uruguay.
- Seis de estos países participaron en PISA 2006 y otros tres estaban por incorporarse al ciclo PISA 2009.
- Otros países de la región, participaron en los estudios de Matemáticas y Ciencias (TIMSS), Lectura (PIRLS) y Educación Cívica que lleva adelante la *International Association for the Evaluation of Educational Achievement* (IEA).
- Durante los últimos 12 años, ha estado activa en la región, la red de los sistemas de evaluación organizada en torno al Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (LLECE) de OREAL/UNESCO, en la que participaron casi todos los países Latinoamericanos.

- Durante ese tiempo, se constituyó el Grupo Iberoamericano de PISA, integrado por Argentina, Brasil, Chile, Colombia, España, México, Portugal y Uruguay.

En general, las evaluaciones estandarizadas de gran escala en la región latinoamericana son más frecuentes y mejores. Sin embargo, las principales debilidades se observan en la estrategia de difusión, en el uso de los resultados y en la calidad técnica de las evaluaciones (Ravela *et al.* 2008).

En Estados Unidos hay una gran tradición de pruebas con versiones diferentes que se aplican en múltiples ocasiones. Por ejemplo, la *American College Testing* (ACT) o mejor conocida como *Collage Board*, el *Scholastic Aptitud Test* (SAT). El propósito de dichas pruebas es la admisión a las universidades. Cabe señalar que el *SAT* es administrado por primera vez en el año de 1926 utilizando pruebas con ítems de opción múltiple (Angoff, 1971). Otro ejemplo, es el Graduate Record Examination (GRE) para ingresar al posgrado y el Test of English as a Foreign Language (TOEFL) para acreditar el conocimiento del inglés como segundo idioma. (Backhoff y Tirado, 1992).

Hasta mediados del siglo pasado, la construcción de pruebas era básicamente una labor intuitiva y sin un sistema válido. Después de esta etapa, la elaboración de pruebas entró en una etapa científica durante la cual han ocurrido una infinidad de cambios tecnológicos que impactan enormemente la vida escolar (Madsen, como aparece en Contreras, (2000). Los avances logrados durante los últimos 25 años, en lo relativo a modelos psicométricos y en los métodos y procedimientos evaluativos, han sido muy importantes (Contreras, 2000).

2.1.2 La evaluación del aprendizaje en escuelas superiores de México

La evaluación del aprendizaje de los alumnos es de gran importancia para las instituciones educativas, sin embargo, la evaluación se puede realizar de diversas maneras. En México, esta labor, ha correspondido a los maestros, generalmente en forma individual, dejando a un lado el trabajo colegiado y también la aplicación a grandes cantidades de estudiantes. La evaluación de la educación superior no ha sido la excepción. Hasta los años sesenta, comenzaron a aplicarse pruebas estandarizadas a gran escala para el ingreso a la Universidad Nacional Autónoma de México (UNAM).

En México, el primer instrumento de evaluación técnicamente válido, es administrado en el área de la salud en 1963, el *Examen Profesional Teórico de la Facultad de Medicina* por la Universidad Autónoma de México (UAM) (Gago, 1998). De dicho examen, a mediados de los sesenta, se desarrolló un banco de reactivos de opción múltiple para el examen de titulación de la carrera de médico cirujano. Sin embargo, la aportación más importante, fue la construcción de exámenes de ingreso a nivel bachillerato y licenciatura. Debido, a que con estos exámenes se construyó la base que daría paso a la creación del Centro Nacional de la Evaluación para la Educación Superior, A. C. (CENEVAL) en 1994 (Martínez, 2001).

Otras universidades, como la Universidad Autónoma de Coahuila y la Universidad Autónoma de Aguascalientes, han empleado exámenes de admisión desarrollados en el extranjero. Por ejemplo, la *Prueba de Aptitudes Académicas* (PAA), desarrollada en Puerto Rico (Backhoff y Tirado, 1992). Sin embargo, este es solo uno de los varios ejemplos que hay en México. Otras Instituciones de carácter privado, como el *Instituto Tecnológico de Monterrey* y la *Universidad de las Américas*, comenzaron a utilizar servicios de la *Oficina de la College Board* con la utilización de la *Prueba de Aptitud Académica* (PAA) (Martínez, 2001).

Por otra parte la Universidad Autónoma de Baja California (UABC) dispone actualmente de un Examen de Habilidades y Conocimientos Básicos (EXHCOBA), el cual se administra con propósito de selección para el ingreso a la universidad. El desarrollo del examen inicio en 1992 en su versión de lápiz y papel. Desde entonces se ha aplicado oficialmente en la UABC y en otras IES mexicanas como: la Universidad de Guanajuato, la Universidad de Sonora, la Universidad Autónoma de Zacatecas, la Universidad Autónoma de Querétaro, la Universidad Autónoma de Nayarit, el Centro de Investigación Científica y de Educación Superior de Ensenada (CICESE) y como Examen diagnóstico en la Escuela Nacional de Estudios Profesionales, Iztacala (UNAM) (Backhoff, 2001).

Para 1993 y 1994, el EXHCOBA en su versión computarizada, se aplicó a diferentes universidades tanto nacionales como extranjeras con propósitos de admisión e investigación. De esta manera, se aplicó a instituciones de educación superior nacionales como las universidades de Sonora, Guanajuato, Coahuila y Aguascalientes. También, se ha aplicado en Universidades extranjeras con propósitos únicamente de investigación. Por ejemplo, en la Universidad Estatal de San Diego (Estados Unidos) y en la Universidad Simon Frasier (Canadá) (Backhoff, 2001).

No fue hasta 1991, cuando la Secretaría de Educación Pública (SEP), inició un proceso de evaluación para las universidades a nivel nacional. Para tal tarea, se crearon los Comités Interinstitucionales de Evaluación de la Educación Superior (CIEES). A través de la creación de dichos comités, se estructuró un modelo de evaluación que permitiera a las instituciones de educación del país, efectuar un análisis interno y externo de sus funciones con el fin de elevar la calidad educativa. Otro de los fines de la creación de dichos comités fue la acreditación de las instituciones educativas, respondiendo con ello a las necesidades crecientes de evaluación en las IES del país.

A partir de los Comités Interinstitucionales para la Evaluación de la Educación Superior (CIEES), se fortaleció el campo de la evaluación diagnóstica en las Instituciones de Educación Superior (IES) del país. Estos han tenido un papel importante en los trabajos de constitución y colaboración CENEVAL (Rosario, Marúm, Vargas, Arroyo, y González, 2006). En 1993, el Secretariado Conjunto de la Coordinación Nacional para la Planeación de la Educación Superior (CONPES) y el Consejo Nacional de la Asociación Nacional de Universidades e Instituciones de Educación Superior (ANUIES), propusieron crear el CENEVAL con el fin de que se responsabilizara de la evaluación de los resultados.

En 1994, se creó entonces el CENEVAL como un organismo no gubernamental y autofinanciable, que llevaría a cabo las acciones necesarias para aplicar los diferentes exámenes requeridos para la evaluación de la educación media superior y superior. Dichos exámenes, son el Examen Nacional de Ingreso a la Educación Media Superior (EXANI I); el Examen Nacional de Ingreso a la Educación Superior (EXANI II); el Examen Nacional de Ingreso al Posgrado (EXANI III) y el Examen General de Egreso de la Licenciatura (EGEL). Para 1999, 137 instituciones aplicaron alguno de los diferentes EGEL (ANUIES, 2000).

Díaz (2006), en su trabajo sobre el análisis de las pruebas masivas y de sus diferencias técnicas, realiza un resumen general en materia de evaluaciones estandarizadas en México y de sus avances observados hasta 1999. Dicho resumen, aunque no es exhaustivo, da un panorama general sobre el tema (véase Tabla I).

Tabla I. Avances observados en materia de evaluación estandarizada del aprendizaje en la educación superior en México hasta 1999.

Ámbito individual	Tipo de evaluación	Certificación
Ingreso	<ul style="list-style-type: none"> -Exámenes de ingreso diseñados por las propias IES -Exámenes diseñados por el CENEVAL (EXANI I, II y III) -Exámenes de admisión desarrollados en el extranjero como es la <i>Prueba de Aptitudes Académicas</i> (PAA) -Examen Nacional de Aspirantes a Residencias Médicas -Examen de Ingreso a la Educación Media Superior (COMIPEMS) -Examen de Ingreso al Nivel Superior en el Estado de Aguascalientes -Examen de Habilidades y conocimientos Básicos (EXHCOBA) 	
Trayecto	<ul style="list-style-type: none"> -IES (a cargo del personal académico) -IES (Exámenes Departamentales) 	
Egreso	<ul style="list-style-type: none"> -IES (a cargo del personal académico) -Examen de Egreso (ITESM) -CENEVAL (EGC/EGEL) 	<ul style="list-style-type: none"> -IES (otorgamiento del título o grado) -CENEVAL (certificación social: EGC/EGEL)

Con la reestructuración orgánica de la SEP, planteada en el Plan Nacional de Desarrollo 2001-2006, se gestaron cambios importantes. Se incorporaron el Programa de Fortalecimiento de la Gestión Estatal (ProGEN) y los Programas de Fortalecimiento de la Escuela Normal (ProFEN). Dado lo anterior, en el 2003 la Dirección General de Educación Superior para Profesionales en Educación (DGESPE), estableció convenios de colaboración con el CENEVAL para el desarrollo de los Exámenes Generales de Conocimiento de las Licenciaturas en Educación Preescolar y Primaria (EGC-LEPRE y EGC-LEPRI). Ambas pruebas basadas en competencias, han permitido a través de sus resultados contar con una estimación del nivel de logro alcanzado por los estudiantes de 8vo semestre en términos de las competencias específicas (Leyva, 2010).

Actualmente, hay en el mercado una cantidad básica de exámenes de buena calidad técnica requerida por las IES Mexicanas para hacer frente al compromiso de la evaluación de calidad. Sin embargo no hay suficientes pruebas de trayecto con la calidad técnica básica para poder diagnosticar el aprendizaje adquirido en un punto intermedio dentro de un plan de estudios de las licenciaturas. Por tal motivo la ANUIES recomendó al CENEVAL agregar a sus actuales series de exámenes una que se oriente a la evaluación de conocimientos, habilidades y competencias de carácter disciplinar y genérico o transversal y que incida en el diagnóstico dentro de un determinado corte en la licenciatura, lo cual es uno de sus objetivos para el 2010 (CENEVAL, s.f.b). Algunos de estos exámenes ya desarrollados a la fecha, se presentan en la Tabla II.

Tabla II. *Exámenes diagnósticos desarrollados por el CENEVAL para la licenciatura.*

Tipo de examen	Exámenes	Propósito	Momento de aplicación
Examen intermedio de la licenciatura	-EXIL de ciencias básicas de las ingenierías	-Identificar el nivel de dominio del alumno sobre conocimientos y habilidades intelectuales que son fundamentales en la formación del futuro ingeniero	-Fase intermedia de la licenciatura
Exámenes por campo de conocimiento	-Estadísticas Básicas (Extra-es), Gestión empresarial, Administración, Ciencia política.	-Determinar el nivel de dominio de los contenidos básicos de un campo de conocimiento afín a varias carreras profesionales	-Fase intermedia y al final de la formación profesional
Examen de competencias genéricas	-Examen de Competencia Comunicativa y Pensamiento Crítico Nivel Licenciatura (ECC y PEC antes Excal)	-Identificar el nivel de dominio del alumno sobre habilidades fundamentales para el desarrollo académico y el buen desempeño	-En cualquier momento de La formación profesional y en más de una ocasión, posibilitando las comparaciones en el tiempo.

En el ámbito institucional, La UABC ha desarrollado diferentes mejoras en la planeación y evaluación del aprendizaje. En general diferentes Unidades

Académicas (UA) de la UABC aplican año con año diferentes exámenes para evaluar el logro académico de sus estudiantes. Un ejemplo representativo es la Facultad de Medicina y Psicología (FMP) de Tijuana, la cual aplica desde el año 2006 exámenes departamentales y de trayecto a todos sus estudiantes de los dos programas educativos que ofrece, el de medicina y el de psicología.

Otro ejemplo digno de mencionar es el de la Facultad de Ingeniería Campus Mexicali (FIM). En particular en marzo de 2004 respondió a la convocatoria para diseñar instrumentos de evaluación colegiada del aprendizaje hecha por la Subsecretaría de Educación Superior e Investigación Científica (SESIC) de la SEP y por la misma UABC. En total, fueron apoyados cinco proyectos financiados para construir exámenes colegiados. Los proyectos seleccionados recibieron capacitación y asesoría para la construcción de los exámenes por parte del Instituto de Investigación y Desarrollo Educativo (IIDE) de la UABC. La construcción de los instrumentos, dio inicio con el caso de los modelos de examen para la materia de Matemáticas I (Cálculo Diferencial) en la Facultad de Ingeniería de Mexicali (FIM) (Encina, Rivera, y Contreras, 2005). En la Tabla III se muestran los exámenes que aplica la FIM actualmente.

Tabla III. *Exámenes que se aplican actualmente en la FIM de la UABC.*

Ámbitos individual	Tipo de evaluación
Examen de Ubicación	-Examen de Ubicación de Matemáticas (EXUMAT)
Ingreso	-Examen de Habilidades y conocimientos Básicos (EXHCOBA)
Exámenes colegiados	-Examen de Matemáticas I y II, calculo diferencial, calculo Integral y de circuitos I y II
Examen Intermedio de la Licenciatura	-Examen Intermedio de Licenciatura para Ciencias Básicas de Ingeniería (EXIL-CBI). De Cálculo diferencial, integral, multivariable, ecuaciones diferenciales, estática y dinámica.
Egreso	-EGEL de ingeniería civil, ciencias computacionales, eléctrico, electrónico, mecánico e ingeniería industrial.

Hasta aquí, se puede afirmar que en el desarrollo de las pruebas educativas hay avances considerables, aún cuando la evaluación sistematizada se inició en 1942 con el trabajo de Tyler. En general, los avances de las evaluaciones estandarizadas del aprendizaje, se enfocan en el avance teórico y tecnológico del desarrollo de dichas evaluaciones. También se han hecho grandes avances en la aplicación de estrategias evaluativas en los niveles básicos de educación tanto a nivel nacional como internacional. Sin embargo, en México, aún falta llevar estos avances en forma sistematizada a la educación superior. Actualmente, en gran parte de las universidades del país, se cuenta con exámenes para el ingreso y egreso de sus planes de estudio, pero cuentan con muy pocos exámenes intermedios, departamentales y de trayecto de buena calidad técnica.

2.2. Aspectos teóricos de la evaluación del aprendizaje.

Este apartado trata aspectos teóricos sobre evaluación del aprendizaje. El propósito principal, es presentar un marco de referencia de forma organizada para facilitar la comprensión de los métodos evaluativos, sus características, sus usos y las posibilidades de dichas evaluaciones. En general, la teoría de las evaluaciones educativas, se ha construido a la par de la teoría psicométrica. La teoría y los desarrollos psicométricos, son un cuerpo sustantivo de conocimiento científico claramente estructurado y acumulativo (Melia, 1990). A diferencia de otras disciplinas psicológicas y sociales, la psicometría se estructura en forma ordenada desde sus principales líneas y crece fundada en cada uno de los nuevos desarrollos y aplicaciones propios de la disciplina.

Como sucede en otros campos de la ciencia, las evaluaciones tienen su propio vocabulario. Por lo que es necesario acercarse a sus diferentes definiciones y clasificaciones. Una primera clasificación a analizar, trata la dicotomía *estandarizada* contra la *no estandarizada*. Una evaluación estandarizada tiene un método riguroso para su aplicación y calificación, está diseñada por expertos en el campo y se aplica a un grupo representativo de la población.

Dependiendo los objetivos de la evaluación estandarizada, se pueden establecer distintos tipos de criterios o normas de calificación a partir de los puntajes brutos resultantes de la prueba aplicada al grupo de estandarización. Dichas normas, sirven como base para interpretar las calificaciones de personas a quienes se les aplica subsecuentemente. Por otra parte, las evaluaciones no estandarizadas, son desarrolladas de forma intuitiva e informal por los maestros, y en general, son más comunes que las evaluaciones estandarizadas (Aiken, 1996).

2.2.1 Definición y clasificación de las evaluaciones

Las definiciones de la evaluación se establecen según la perspectiva u objeto de evaluación. De esta manera hay definiciones de la evaluación basadas en el método, en el objeto de evaluación, en el objetivo a alcanzar, etcétera. Desde la perspectiva teórico-metodológica, la evaluación estandarizada podría definirse como una actividad metódica que consiste en una colección de conceptos y prácticas que demuestran su utilidad y, que se aplican de modo que pueden contribuir al perfeccionamiento de las actividades prácticas (Scriven, 1994). También, desde esta perspectiva, la investigación evaluativa se define como un conjunto de procedimientos para la recolección y análisis de datos. Así, la evaluación, aumentan la posibilidad de probar, más que de afirmar, el valor de una acción social (Suchman, 1967 en Rosario *et al.* 2006).

Por otra parte, desde la perspectiva del logro de objetivos, la principal función de la evaluación sería determinar el valor o el grado de éxito en la consecución de un objetivo determinado (Donadedian, 1966). De esta manera, la conceptualización de la evaluación, se basaría en el análisis de las discrepancias entre lo que es y lo que debe ser, es decir, la disparidad entre una situación deseada o esperada y otra existente o real (Kaufman y English, 1979). En general la definición de la evaluación desde esta perspectiva se describiría como el análisis e interpretación valorativa de los resultados y logros del programa y sus correspondientes procesos (Chacón, Anguera y López, 2000).

Hay perspectivas de la evaluación que ven a ésta como un proceso con un fin específico. Stufflebeam (1987, en Escudero, 2003) define a la evaluación como el proceso de identificar, obtener y proporcionar información útil y descriptiva acerca del valor y el mérito de las metas, la planificación, la realización y el impacto de un objeto determinado, con el fin de servir de guía para la toma de decisiones, solucionar los problemas de responsabilidad y promover la comprensión de los

fenómenos implicados. También dentro de esta perspectiva de la evaluación, hay concepciones más complejas en donde se da importancia a la interacción entre los actores de dicho proceso. Así, la evaluación interactiva como ejemplo, se puede visualizar como un proceso de investigación participativa que analiza la organización, el funcionamiento y desarrollo de un programa en relación con sus objetivos, las expectativas de sus participantes y los resultados obtenidos.

Por otra parte, según Alvira (1985) evaluar resulta simplemente de realizar un juicio sobre la actividad, recurso o resultado. También, en una forma simplista, evaluar es fijar el valor de cosa; para hacerlo se requiere efectuar un procedimiento el cual compara aquello a evaluar de un criterio o patrón determinado (Franco, 1988).

Rosario y sus colaboradores (2006), clasifican a la evaluación en tres tipos. El primer tipo de evaluación, se da según el momento en que se evalúa, pudiendo ser la evaluación aplicada antes de la toma de decisiones, durante la toma de decisiones o posteriori. El segundo tipo, se da según la función que cumple la evaluación, ya sea sumativa o formativa. La evaluación sumativa hace referencia al cumplimiento de los objetivos y a los resultados de los programas educativos e investiga los efectos y los compara con las necesidades de los usuarios o beneficiarios. Por otra parte, la formativa da un seguimiento durante el proceso de ejecución de un programa o proyecto y proporciona información acerca de cómo se desarrolla dicho proceso. El tercer y último tipo de la clasificación, se refiere a la procedencia de los evaluadores, pudiéndose mencionar cuatro tipos:

- Evaluación externa. Es aquella que ocurre a evaluadores que no pertenecen, ni están vinculados a la institución ejecutora del programa o proyecto que se evalúa.
- Evaluación interna. Es la que realiza con la participación de personas que pertenecen a la institución promotora o gestora del programa o proyecto a evaluar pero que no son directamente responsables de la ejecución.

- Evaluación mixta. Este tipo de evaluación es una combinación de las anteriores; se realiza por equipo de trabajo que incluye evaluadores externos y evaluadores internos.
- Evaluación en donde las actividades evaluativas están a cargo de aquellos implicados directamente en la ejecución o en la realización del programa o proyecto, y que valoran y enjuician sus propias actividades para determinar si cumplen las metas propuestas.

Desde una perspectiva más amplia, se entienden como atributos distintivos de la evaluación: su finalidad de apoyar la toma de decisiones, el alcance holístico, la dimensión axiológica y el carácter participativo de los involucrados. Desde aquí la evaluación implica generar y analizar información relevante con el fin de que la elección de acciones y su instrumentación, estimulen los procesos de mejoramiento y cambio del programa evaluado. Dicho proceso, supone que la información obtenida, se proporciona en forma oportuna a los responsables de los diferentes programas (Llarena, 1994).

Uno de los objetivos de la evaluación, es que sus resultados se puedan observar de forma directa e inmediata. De esta forma, se da la posibilidad de obtener una visión amplia del fenómeno evaluado, dejando en evidencia un conjunto de aspectos importantes y esenciales del trabajo de las instituciones. En específico, la evaluación estandarizada es de gran ayuda para las instituciones educativas. En general la evaluación de tipo estandarizada aplicada al sector educativo muestra información sobre: a) en qué medida los alumnos están aprendiendo lo que se espera de ellos al finalizar ciertos grados o niveles, y b) el efecto que tiene en los logros educativos las inversiones en programas educativos, los cambios en la estructura del sistema, los cambios curriculares, los programas de formación, la adquisición de material educativo, etc. (Ravela *et al.* 2008).

Una buena estrategia o un buen sistema de evaluación de aprendizajes, puede aportar información valiosa para las instituciones educativas y ser un

instrumento clave de mejora. De tal manera que los resultados de la evaluación, enriquezcan la comprensión de la situación educativa, de la operación del currículum y de la toma de decisiones. Tal información, es un recurso básico para diversos actores sociales involucrados directa e indirectamente en el proceso de enseñanza-aprendizaje de una sociedad (Ravela *et al.* 2008). Por ejemplo:

- Las autoridades y los encargados de formular políticas educativas pueden comprender mejor la problemática de enseñanza aprendizaje y así desarrollar políticas pertinentes para apoyar el trabajo de las escuelas.
- Los directivos y docentes pueden, a partir de una mirada externa sobre los logros educativos en el conjunto del sistema, comprender mejor lo que se está logrando y que no están logrando sus propios estudiantes, como están aprendiendo y que dificultades tienen.
- Los supervisores y encargados de la formación de docentes pueden apoyarse en la información sobre los logros y dificultades educativas en el conjunto del sistema para estudiar en profundidad las debilidades en el enfoque o en la práctica de la enseñanza que probablemente generan algunas de las insuficiencias constatadas en los aprendizajes.
- Los padres y madres de estudiantes, adecuadamente informados, pueden comprender mejor que se espera que sus hijos aprendan, que es lo que están logrando y que pueden hacer para colaborar con la escuela y con el aprendizaje de sus hijos.
- La ciudadanía en general, estará mejor informada sobre lo que acontece al interior del sistema educativo y, por lo tanto, estará más atenta a los temas y problemas de la educación.

Actualmente, la evaluación focalizada en educación superior, se considera como la fase del proceso de planeación educativa. Dicha fase, consiste en la

recolección, organización y análisis de la información que será utilizada para emitir juicios de valor respecto al sistema educativo. Los juicios de valor sobre el hecho educativo, se basan en la acción de cotejar periódicamente los objetivos y metas del mismo a fin de realimentar las actividades de construcción e instrumentación del sistema (Llarena, 1994). Desde esta perspectiva, la evaluación tiene como finalidad ofrecer información objetiva y oportuna para la adecuada toma de decisiones y para la formación objetiva de un juicio de valor el cual estará relacionado con una serie de parámetros (estándares de calidad) deseables. En consecuencia, la evaluación debe permitir como principal función la adopción de decisiones informadas y respaldadas adecuadamente. Al respecto, siguiendo la misma línea, Llarena (1994) propone tres funciones principales de la evaluación educativa:

1. *Función diagnóstica.* Consiste en la posibilidad de identificar problemas y necesidades, así como verificar el estado en el que se encuentra una actividad o servicio con respecto a una norma ideal de sus perspectivas.
2. *Función de realimentación.* Se relaciona con los mecanismos creados para localizar los elementos esenciales del sistema de educación superior que afectan el desarrollo de las funciones sustantivas del mismo. Constituye un proceso permanente de verificación que permite reajustar en forma continua y sistemática una actividad o programa, con el fin de asegurar la consecución oportuna de las metas y objetivos propuestos e identificar situaciones inesperadas y susceptibles a prevenirse.
3. *Función de verificación.* Consiste en un punto culminante del proceso de evaluación, el cual permite constatar el grado de realización de los objetos al término de un programa o proyecto.

Por otra parte, Ravela (2000) hace énfasis sobre la clasificación de la evaluación según las distintas finalidades o necesidades del sistema educativo o de la sociedad. En general propone una clasificación de los propósitos de las evaluaciones en seis grandes tipos:

1. acreditación y certificación;
2. selección;
3. toma de decisiones “blandas” o decisiones de mejora;
4. toma de decisiones “duras”;
5. establecimiento de incentivos y sanciones y,
6. rendición de cuentas.

Todos y cada uno de los tipos de evaluación aquí mencionados, son cruciales para la mejora del sistema educativo. Ahora bien, la mejora de la calidad de los sistemas educativos no solamente se logrará al realizar buenas evaluaciones estandarizadas, sino que, además, es necesario que las evaluaciones de los docentes en las aulas, las evaluaciones a los mismos docentes, las evaluaciones de selección para los directivos y supervisores, etc., se desarrollen en las mejores condiciones posibles y con la mejor calidad.

Ahora bien, para fines del presente estudio, es necesario detenerse a profundizar sobre las evaluaciones para la toma de decisiones “blandas” o decisiones de mejora. Ravela (2006, p. 24) señala que “todas las evaluaciones pueden –y deberían– ser utilizadas con el fin de comprender mejor la realidad con la que se está trabajando y contribuir a mejorarla”. Para tal tipo de evaluación se utiliza el término de “evaluación formativa”. Dicho tipo de evaluación tiene el propósito principal de servir como base para la toma de decisiones. Una vez obtenidos los resultados de la evaluación, se generan acciones para la mejora del foco evaluativo. Ravela (2006), presenta varios ejemplos posibles para aclarar lo que sería la aplicación de la evaluación formativa en la educación:

- Las evaluaciones de los alumnos realizadas por su profesor con el fin de comprender el proceso de aprendizaje y apoyar a cada alumno individualmente en función de sus dificultades.
- Las evaluaciones de los docentes cuya finalidad es brindarles orientación profesional para desarrollar mejor su labor de enseñanza y de relación con los alumnos.
- Las evaluaciones de centros educativos cuya finalidad es desencadenar procesos de diálogo entre los actores de las instituciones para buscar caminos de mejora en relación a los problemas existentes.
- Las evaluaciones del currículo o de proyectos específicos cuya finalidad es detectar necesidades de cambio y mejora.
- Las evaluaciones nacionales e internacionales de logros educativos cuyo propósito es contribuir a identificar las principales debilidades de los sistemas educativos y orientar la reflexión y formulación de las políticas educativas.

Como se puede ver, en todos los ejemplos mencionados no hay ni una consecuencia formal específica para los actores sociales en cuestión. Es por esto que a estos tipos de evaluación se les denomina también de “bajo impacto” o de “decisiones blandas”. El fin de la evaluación de bajo impacto es apoyar en la comprensión del fenómeno educativo y así generar una cultura evaluativa en donde se realicen acciones y decisiones para la mejora de la calidad educativa.

Por otra parte, la evaluación conforma un proceso intencional que responde a necesidades naturales de las personas, de los grupos humanos y de las instituciones (Rosario, et al 2004). Sin embargo, hay una subcultura de resistencia a la evaluación formal en casi todos los sectores sociales, por ejemplo en el sector educativo —principalmente por las autoridades educativas— hay una aversión a los procesos evaluativos. También en general, las personas, los grupos y las

instituciones, suelen tener aversión a los intentos formales y sistematizados de evaluación, formando mecanismos de bloqueo, similares a los que establecen los gerentes de empresas privadas ante las auditorías externas (Ravela, 2000).

La evaluación en el ámbito educativo, como cualquier proceso institucional, presenta diversas limitaciones y debilidades. Principalmente cuando la evaluación como proceso se efectúa constantemente y es afectado por diversos factores presentes en el contexto. Dichas debilidades no son únicas ni exclusivas de las evaluaciones estandarizadas, sino que también están presentes en los otros tipos de evaluación.

Para finalizar sobre lo concerniente a la definición y clasificación de la evaluaciones, se presenta a continuación algunos los problemas principales que enfrenta la formación de la cultura de evaluación. Al respecto Álvarez (1997), plantea los siguientes ejemplos:

- Falta de conceptos claros sobre evaluación.
- Desconocimiento de sus diferentes funciones a finalidades con las que se puede utilizar.
- Confusión del proceso de evaluación con el proceso de investigación.
- Desvinculación del proceso de evaluación del proceso general de planeación, de la toma de decisiones y de otros procesos sociales e institucionales.
- Selección inadecuada de métodos y técnicas de evaluación o de análisis.
- Desconocimiento de las características principales que debe tener la información de evaluación: validez, confiabilidad, utilidad, oportunidad.
- Falta de precisión de los aspectos a evaluar en cada campo y de los tipos y enfoques que resultan más apropiados en cada caso.
- Utilización inadecuada de los resultados de la evaluación.
- Ausencia de alguna de las condiciones que hacen factible un proceso de evaluación.

- Condiciones del entorno social que imponen la subcultura de resistencia a la evaluación formal, subcultura del fraude o evasión de responsabilidades.

2.2.2 Características básicas de las pruebas educativas del aprendizaje

Las pruebas para evaluar el aprendizaje, tienen en forma general varias características, aplicaciones y formas de clasificación. Aiken (1996) en su texto *Test Psicológicos y Evaluación*, muestra la gran variedad de test y de formas de aplicación. También proporciona una clasificación de las pruebas que se divide en: pruebas de ubicación/diagnostico, pruebas de competencia del lenguaje, pruebas afectivas y las pruebas de aprovechamiento. En especial, en relación al objeto de estudio de la presente tesis, se puede mencionar que algunas de las aportaciones de las pruebas de aprovechamiento: a) hacen un muestreo del nivel de aprendizaje del estudiante a lo largo de varias áreas de habilidad general; b) estiman lo que un estudiante sabe y hace en áreas específicas como resultado de la escolaridad; c) permiten la comparación del progreso de los estudiantes en una escuela o proyecto; y d) los resultados pueden usarse para tomar decisiones de egreso de los estudiantes (de un programa), medir el crecimiento de los individuos, proporcionar información sobre el mejoramiento del programa y dar indicaciones sobre su efectividad.

Por su parte Ravela (2006), en su propuesta de clasificación de la evaluación educativa, profundiza en uno de los elementos que definen a una evaluación de aprovechamiento académico, la formulación de juicios de valor acerca de los individuos, instituciones o realidades evaluados. En general explica y ejemplifica tres enfoques principales para formular estas valoraciones:

1. Enfoque normativo. compara las posiciones relativas entre individuos y grupos.

2. Enfoque de progreso o crecimiento. Compara el avance o retroceso de un individuo o grupo respecto a su propio desempeño en un momento anterior.
3. Enfoque criterial. comparar el desempeño demostrado por cada individuo con una definición clara y exhaustiva del dominio evaluado y del nivel de desempeño deseable.

Cada uno de estos enfoques, son utilizados normalmente en forma complementaria en las aulas. Sin embargo, se encuentra un mayor desarrollo tanto teórico como técnico en las evaluaciones estandarizadas a gran escala. En lo particular, el enfoque mayormente utilizado a pequeña y a gran escala es el criterial, dado que permite dar una mejor respuesta a las principales preguntas que los diversos actores involucrados en el quehacer educativo esperan que los sistemas nacionales de evaluación respondan (Ravela, 2006).

En contraste con la clasificación mencionada, Popham (1990) señala solo dos enfoques para la evaluación del desempeño académico, los de referencia normativa y los de referencia criterial. Aunado a esto, también señala que los tests educativos se aplican en tres rubros sustantivos: 1) en la evaluación de individuos para decidir aspectos instruccionales o vocacionales, 2) en la evaluación de programas, y 3) para evaluar la instrucción. Desde esta clasificación, en el enfoque normativo, se ordenan a los casos (individuos, instituciones o subsistemas) evaluados con el fin de compararlos entre sí, es decir, se interpretan los resultados de la unidad de análisis con el grupo de comparación. Las pruebas que más se utilizan dentro de este enfoque son las relacionadas con la selección. En estas pruebas no importa tanto qué es lo que un individuo “sabe” o domina, sino si “sabe” más o menos que los otros. El foco no está ubicado en describir los conocimientos y competencias de los individuos, sino en especificar su posición en el conjunto. En resumidas cuentas, el fin es seleccionar a los mejores candidatos para el ingreso a una institución educativa.

Por otra parte, de acuerdo a Ravela (2006, p. 44) en el enfoque criterial, se le da mayor importancia a *“la comparación del desempeño del estudiante con una definición clara y precisa de lo que se espera que conozca y sea capaz de hacer en un determinado dominio”* (por ejemplo, comprensión de los métodos cuantitativos). Para presentar los resultados de las evaluaciones criterioles se definen distintos niveles de logro del dominio en cuestión (avanzados, alrededor de la media, por debajo de lo esperado, etc.) y se busca determinar en qué nivel se encuentra cada individuo.

Así, las pruebas alineadas al currículum son de referencia criterial, porque se elaboran con una metodología adaptada para evaluar con precisión los contenidos curriculares (conocimientos o competencias) que se definen en los planes y programas de estudio oficiales. La alineación de las pruebas al currículum implica determinar los objetivos establecidos por el currículum, asociar las acciones de evaluación con los contenidos esenciales, definir el dominio curricular sobre el cual se desarrollará la prueba y precisar los resultados de aprendizaje (Contreras, 2009).

Como se ha visto en los párrafos anteriores, hay varios criterios para la clasificación de pruebas que se emplean para evaluar el aprendizaje. Para Contreras (2000), en el ámbito de la evaluación educativa es posible identificar, al menos, tres sistemas de clasificación de dichas pruebas: 1) pruebas referidas a una norma y/o referidas a un criterio, 2) pruebas diferenciadas por el tipo de respuesta construida o de respuesta seleccionada, y 3) las pruebas que se distinguen por su escala de aplicación, que son las de pequeña y las de gran escala.

En la Tabla IV se contrastan los exámenes de referencia normativa y de referencia criterial con base en sus características generales y sus propósitos. Básicamente, la distinción entre ambos tipos de prueba, depende de la manera en que se interpreta la ejecución de un examinado ante el test (Popham, 1990; Ravela, 2006).

Tabla IV. *Características generales de los tests normativos y criterios.*

Tipo de test	Juicio de valor	Propósitos educativos
Normativo	-El foco de atención es ordenar a los individuos, instituciones o subsistemas evaluados con el fin de compararlos entre sí	-Selección de estudiantes -Asignación de recursos a gran escala
Criterial	-Consiste en privilegiar la comparación del desempeño de un individuo con una definición clara y precisa de lo que se espera que conozca y sea capaz de hacer en un determinado dominio	-Evaluación de programas -Certificación de competencias académicas -Diagnóstico y diseño instruccional

La referencia a un criterio y la referencia a una norma, no son esquemas que se excluyen necesariamente, sino que pueden ser complementarios. Ambos pueden ser empleados dentro de un mismo test para lograr un conocimiento claro de un examinado, siempre y cuando se sigan procedimientos específicos de calidad para su elaboración, particularmente la satisfacción del criterio de validez y el conjunto de sus evidencias (Martínez, 2001).

Ahora bien, el segundo sistema de clasificación establecido por Contreras (2000) describe dos tipos de tests, los de respuesta construida y los de respuesta seleccionada. Esta clasificación, se refiere al tipo de formato de los ítems que conforman la prueba. Aunque en un examen pueden ser incluidos diferentes tipos de preguntas, en el contexto de la evaluación, será de gran importancia establecer qué tipo de ítems se utilizará según las limitaciones que se presenten en el contexto de evaluación. En la Tabla V se describen los principales tipos de tests de respuesta construida y de respuesta seleccionada, además de una breve descripción de estos.

Tabla V. Características de los *tests de respuesta construida y de respuesta seleccionada*.

Tipo de test	Tipo de ítem	Descripción
Respuesta construida	-Ensayo	-Pide al estudiante integrar libremente y por escrito, lo que sabe acerca de un tema
	-Respuesta breve	-El estudiante responde con una palabra, símbolo o frase corta
	-Ejecución	-Enfatiza lo que el estudiante puede hacer, no lo que sabe, en un contexto auténtico
Respuesta seleccionada	-Opción múltiple	-Pide al estudiante elegir la respuesta correcta o la mejor opción entre las que se le ofrecen
	-Respuesta alterna	-Pide al estudiante juzgar la verdad o falsedad de proposiciones
	-Asociación	-Pide al estudiante relacionar conceptos u ordenar fases o eventos

Las pruebas de respuesta múltiple y las de respuesta construida, presentan ciertas particularidades. Las limitaciones de los formatos ante el contexto de aplicación, ha sido un punto decisivo para que los desarrolladores de pruebas en general opten por emplear casi exclusivamente pruebas que contengan reactivos de respuesta seleccionada, particularmente los de opción múltiple. Uno de los posibles motivos para tal decisión, es que los exámenes estructurados con reactivos de opción múltiple son relativamente más fáciles para su administración y calificación a gran escala. Sin embargo, Contreras (2000) argumenta que las pruebas de respuesta construida resultan más apropiadas para el contexto de la instrucción, en donde se necesita diagnosticar las competencias puntuales logradas durante el proceso de enseñanza-aprendizaje.

Otra de las clasificaciones que menciona Contreras (2000) en varios de sus trabajos, es la de pequeña escala y de gran escala. Esta distinción resulta relevante

debido al gran impacto social que pueden tener los exámenes. En la Tabla VI se describen ambos tipos de instrumento.

Tabla VI. Características generales de los *exámenes de pequeña y gran escala* (Contreras, 2000).

Tipo de examen	Contexto de uso	Propósitos de la evaluación	Consecuencias de su aplicación	Requisitos técnicos
Pequeña escala	-Salón de clases	-Ubicación, monitoreo del aprendizaje	-Usualmente mínimas o moderadas	-Mínimos, usualmente básicos
Gran escala	-Más de un plantel, usualmente en un estado, región o país	-Seleccionar, rendir cuentas de la gestión, certificar el logro educativo	-Usualmente muy poderosas, en particular para alumnos y profesores	-Altos, con procedimientos relativamente sofisticados

Como puede observarse, la diferencia básica entre ambos tipos de examen radica en los propósitos de la evaluación y en el contexto de su uso. Popham (1990) en particular señala, que los exámenes de gran escala son esenciales cuando se trata de decidir cómo distribuir recursos financieros a gran escala, comparar estudiantes de diferentes instituciones y áreas geográficas, seleccionar o certificar a una amplia cantidad de individuos, etc.

Otra de las características esenciales de las pruebas educativas o psicológicas, son los estándares de calidad técnica y psicométrica. Al respecto la normatividad más importante a nivel internacional, se establece en el manual “Estándares para las pruebas educativas y psicológicas” elaborado en 1985 por la Asociación Americana de Psicología (APA), en colaboración con la Asociación Americana de Investigación Educativa (AERA) y el Consejo Nacional para la Medición en Educación (NCME) en Estados Unidos.

En la misma línea, el National Center for Research on Evaluation, Standards and Student Testing (CRESST, 1994), en Estados Unidos, desarrolló criterios para revisar la calidad técnica de las evaluaciones que apelan a la complejidad cognitiva,

la calidad del contenido, significatividad, propiedad del lenguaje, transferencia y generabilidad, justicia, confiabilidad y consecuencias pretendidas. Cabe señalar, que el énfasis reciente en la calidad educativa, en la noción de asumir la responsabilidad y el incremento del impacto social de la evaluación del aprendizaje, obligan cada vez más a quienes elaboran las pruebas y, en general a los educadores, a estar interesados e informados en los elementos técnicos que definen la calidad de una prueba educativa (Contreras, 2000).

Para 1999, la versión de 1985 del manual “Estándares para las pruebas educativas y psicológicas” fue revisada y reestructurada. La nueva versión tiene más material de referencia en profundidad en cada uno de sus apartados, tiene también una mayor cantidad de normas especializadas y un glosario e índices ampliados. En esta última versión se ve reflejada la Ley Federal de Educación actual en los Estados Unidos, la medición de las tendencias que afectan a la validez, lo concerniente a las pruebas para personas con discapacidad o diferente ámbito lingüístico, los nuevos tipos de pruebas y los usos de las pruebas ya existentes.

Las normas establecidas en el manual, fueron escritas para profesionales y expertos en el área del desarrollo de pruebas. Estas normas se ocupan de cuestiones profesionales y técnicas del desarrollo de las pruebas y su uso en la educación, la psicología y el empleo. Dicho manual, es una referencia muy importante para los elaboradores de pruebas profesionales, patrocinadores, editores, usuarios, políticos, empresarios, y estudiantes de educación y psicología. A continuación se presentan algunos puntos esenciales que trata el manual:

Parte I. Construcción de la prueba, evaluación y documentación.

- a) Validez.
- b) La fiabilidad y errores de medición.
- c) Desarrollo del examen y revisión.
- d) Escalas, normas, y la comparabilidad de la puntuación.

- e) Administración de la prueba, puntuaciones, y presentación de informes.
- f) Documentación de apoyo para las pruebas.

Parte II. Equidad en la prueba.

- g) La equidad y el uso de las pruebas.
- h) Los derechos y responsabilidades de los participantes de la prueba.
- i) Las pruebas individuales de diversos entornos lingüísticos.
- j) Pruebas de individuos con discapacidades.

Parte III. Las aplicaciones de pruebas.

- k) Las responsabilidades de los usuarios de la prueba.
- l) Pruebas psicológicas y evaluación.
- m) Pruebas y evaluación de la educación.
- n) Pruebas en el empleo y la acreditación.
- o) Pruebas en evaluación de programas y políticas públicas.

Para Ravela (2006), son dos los problemas principales que deben cuidarse respecto a los estándares de calidad técnica y que son comunes para cualquier evaluación educativa. Uno de ellos es la validez y el otro la confiabilidad. El concepto de validez refiere al grado en que los juicios de valor que se formulan en la evaluación están adecuadamente sustentados en evidencia empírica y están efectivamente relacionados con el “referente” definido para la evaluación. El concepto de confiabilidad refiere a la precisión de las medidas y de la evidencia empírica empleada en la evaluación.

En la Tabla VII, se describen algunos de los criterios, especialmente importantes, que deben satisfacer los exámenes de gran escala, ya sean normativos o criteriosales. El desarrollo de este tipo de instrumentos usualmente queda reservado a las autoridades educativas, centros de investigación educativa y organismos especializados, que tienen las condiciones y el interés en hacerlo o están obligados a ello.

Tabla VII. Descripción de los principales criterios que definen la calidad de un test de gran escala (Contreras, 2000; Popham, 1990).

Criterio de calidad	Descripción	Principales tipos
Estandarización	-Consiste en emplear procedimientos uniformes para administrar y calificar el test, así como para interpretar las calificaciones de manera que resulten comparables los resultados de los diferentes examinados	-Administración (locales, resguardo, instrucciones) -Calificación (automática, corrección por adivinar, estándares y punto de corte) -Interpretación de datos
Validez	-Grado en que un test mide lo que dice. Técnicamente, se refiere a la obtención de evidencias que soportan las inferencias basadas en los puntajes obtenidos en el test (Es el indicador más importante de la calidad de un test)	-Contenido -Criterio -Constructo
Confiabilidad	-Se refiere a la consistencia o reproducibilidad de los puntajes del test. - Puede ser de ocasión a ocasión, de prueba a prueba, de ítem a ítem, de juez a juez, etc.	-Índice de estabilidad (test-retest) -Correlación de puntajes en formas paralelas o alternas -Consistencia interna (homogeneidad de los ítems)
Descripción de la conducta	-Formulación explícita de los conocimientos, habilidades, aptitudes y actitudes del examinado que se pretende medir con el test	-Descripción breve (objetivo conductual usualmente) -Descripción detallada (especificaciones de ítems)
Extensión de la prueba	-Especificación del número de ítems por conducta medida en el examen. Depende de la importancia de la decisión involucrada	-Tipo de estimación de la ejecución del examinado -Tipo de impacto (grado de afectación a los sujetos)
Alcance de la prueba	-Se refiere a la amplitud del atributo que mide el test.	-Restringido (más precisa la descripción y menos ítems) -Amplio (menos precisa la descripción y más ítems)
Datos comparativos	-Se refiere a la cantidad y calidad de los datos normativos que permiten interpretar apropiadamente la ejecución de los examinados	-Marco referencial clave de los tests normativos -Marco referencial básico de los tests criterios
Ausencia de sesgos	-Se refiere a detectar y eliminar un funcionamiento diferencial de los ítems, ante grupos diferentes de examinados, que no depende del grado de conocimiento o habilidad que se mide	-Ofensa (el ítem retrata a un grupo estereotípicamente) -Penalización (un grupo falla el ítem aunque posee la misma habilidad que otro)

El tema de los estándares de interpretación, es otro de los temas que deben tomarse en cuenta en general para pruebas de rendimiento educativo, y en específico para el desarrollo de pruebas de referencia criterial. Al respecto, Jornet y González (2009) en su trabajo titulado “*Evaluación criterial: determinación de Estándares de Interpretación (EE) para pruebas de rendimiento educativo*” hacen una revisión del estado de la cuestión acerca de los métodos para determinar estándares en pruebas de referencia criterial. En dicha revisión se analizan el concepto de estándar, sus componentes y los problemas implicados en las tareas de juicio en estos métodos. También, revisan los procedimientos para diseñar descriptores de estándares y los métodos para identificar puntuaciones de corte.

En particular, para que se puedan interpretar de forma adecuada los puntajes de las pruebas, el desarrollo de la prueba debe estar bien integrado. Principalmente se debe disponer de un adecuado análisis y especificación del *Dominio Educativo* (DE) a evaluar. Una vez obtenido el DE, se debe tener claro el sistema de interpretación de puntuaciones, y específicamente los métodos. Ahora bien, antes de hablar de los diferentes métodos para determinar los estándares de interpretación, es necesario diferenciar el concepto de *estándares* del concepto de *puntuaciones de corte*.

El término estándar, se utiliza únicamente en referencia al sistema de criterios de interpretación, la definición teórica de los niveles de desempeño, logro o competencia, mientras que el término *Puntuación de Corte* (PC) indica el puntaje en la prueba que ayuda a diferenciar entre dos niveles de desempeño (Jornet y González, 2009). En la Tabla VIII se muestra una diversidad de los componentes de la definición de los EE y en la Tabla IX se muestran ejemplos de etiquetas usuales en EE.

Tabla VIII. Componentes de definición de los EE (Jornet y Backhoff, 2008).

Componente	Descripción
Categorías o etiquetas relativas a niveles de desempeño	-Generalmente politómicos (con 3 a 6 categorías), se describen por etiquetas alusivas al nivel de dominio o simplemente con números.
Descripción de los niveles de desempeño	-Relativas al tipo de aprendizaje característico de los sujetos clasificables en cada nivel.
Puntuaciones de corte	-Las puntuaciones que en la prueba sirven para diferenciar entre cada uno de los niveles de desempeño.
Ítems característicos	-Ítems que son capaces de realizar los sujetos de un determinado nivel de desempeño.

Tabla IX. Ejemplos de etiquetas usuales en EE.

Origen	Etiquetas
National Assessment of Educational Progress	-Básico, Competente, Avanzado.
Terranova (2da ed.) (CTB/McGraw Hill)	-En camino, Progresando, Cerca de la competencia, Competente, Avanzado.
Pruebas de rendimiento del Estado de Ohio	-Limitado, Básico, Competente, Acelerado, Avanzado.
Estado de California, Test de California	-Muy por debajo del nivel básico, Debajo del nivel básico, básico, Competente, Avanzado.
Estado de Texas, estándares de valoración de Texas de conocimientos y destrezas	-No llega al nivel usual, llega al nivel usual, rendimiento destacado.
INCE (Instituto Nacional de Calidad y Evaluación): Diagnóstico del sistema educativo español, 1998	-Sin etiquetas, identificación mediante la cualificación de la escala numérica.
Proyecto PISA	-Niveles numéricos
SERCE (LLECE)	-Cuatro niveles numéricos.
EXCALE (INEE, México)	-Por debajo del nivel básico, Básico, Medio, Avanzado.

(Adaptación de Jornet y González, 2009 desde Cizeck, Bunch y Koons, 2004, p. 34).

Según Linn (1994), hay cuatro tipos de estándares de interpretación. La lógica de dicha tipología depende del objetivo a que se dirijan los estándares, el foco de evaluación en que se utilicen y la dirección o plan de los desarrolladores (jueces) en el proceso. De tal manera, las actividades para la determinación de los estándares requiere enfocarse en la aclaración precisa de la tipología de EE que necesita la institución que desarrolla la evaluación (Jornet y González, 2009). En la Tabla X se muestran los cuatro tipos de EE establecida por Linn (1994).

Tabla X. *Tipología de estándares de interpretación (Linn, 1994).*

Tipo de estándar	Descripción
Exhortación	-Representan metas deseables de logro a las que debe tender la mejora de un sistema educativo o de los estudiantes.
Ejemplificación o muestra de rendimiento	-Representan las habilidades o competencias características de diversos niveles de ejecución.
Rendición de cuentas para educadores	-Representan metas curriculares precisas, orientando la evaluación hacia el contraste entre el currículum diseñado, el implementado y los logros educativos.
Certificación del logro del estudiante	-Identifican un nivel mínimo de competencia del logro de estudiantes.

La definición de calidad de aprendizaje es el componente más cualitativo del proceso de determinación de EE. Principalmente, porque los elementos del EE como las categorías de contenido, las descripciones de lo que son competentes los individuos de cada nivel de desempeño y la selección de ítems característicos para cada uno de ellos, se desarrollan mediante procesos de juicio. Con relación a lo anterior, se han desarrollado tres enfoques principales para la determinación de los estándares de evaluación, los cuales se clasifican según Jornet y González, (2009) en: a) dependientes exclusivamente del análisis del DE, b) dependientes fundamentalmente del comportamiento empírico en la prueba, y c) dependientes del DE y del funcionamiento de la prueba (mixtos).

Los estándares evaluativos establecidos como dependientes exclusivamente del análisis del dominio evaluativo, funcionarían como referentes desde donde se desarrollan los ítems, de tal manera que dichos estándares, se pueden diseñar al inicio del desarrollo de la prueba. Por otra parte, los que dependen sólo del comportamiento empírico de la prueba, se realizan comúnmente con los datos de la última aplicación de la prueba. El procedimiento se basa en analizar los ítems característicos de cada nivel de habilidad total en la prueba y establecer cortes en la escala, de forma que se diseña el descriptor de cada nivel a partir de los ítems que componen cada nivel. Por último, una variación de los enfoques anteriores es aquélla fundamentada en el dominio de evaluación, pero con procesos mixtos de juicio-empírico, de tal forma que su diseño se basa en el análisis del DE, pero se informa adicionalmente del comportamiento empírico de los ítems.

2.2.3 Modelos teóricos-metodológicos para el desarrollo de estrategias evaluativas del aprendizaje

En este último apartado del marco teórico, se revisan cuatro modelos metodológicos evaluativos de gran relevancia a nivel internacional y nacional. Primero, a nivel internacional, se describe el modelo de Nitko (1994) para crear exámenes nacionales de referencia criterial y referencia normativa alineados con el currículum. Después a nivel nacional, se describen tres modelos importantes: 1) el modelo del INEE para el desarrollo de los Exámenes de la Calidad y el Logro Educativos (Excale); 2) el modelo del CENEVAL para el desarrollo de exámenes de ingreso, egreso y de trayecto del nivel medio superior y superior, como la serie de los Exámenes de Egreso a la licenciatura (EGEL), los Exámenes Nacionales de Ingreso a la Educación Media Superior, superior y posgrado (EXANI I, II y III) y los Exámenes Intermedios de la licenciatura (EXCIL) y 3) el modelo desarrollado por Contreras (2009) para desarrollar exámenes de referencia criterial alineados al currículum. Cada uno de estos modelos fue desarrollado con diferentes fines, sin embargo, algunas de sus características son semejantes.

El modelo de Nitko (1994) tiene el fin de crear exámenes nacionales de referencia criterial y referencia normativa alineados con el currículum. De esta manera, el currículum es el foco evaluativo. Por lo tanto el modelo evaluativo se opera con base a los objetivos de aprendizaje establecidos en el currículum para decidir qué se debe evaluar, y cómo desarrollar el examen. Esencialmente el modelo busca a lo largo de su desarrollo obtener evidencias que den seguridad de que la ejecución del examinado ante el test, es una evidencia que nos permite inferir válidamente como sería su ejecución. Según Contreras (2000) en el modelo de Nitko el juicio humano es clave para lograr tales evidencias, por lo que se encuentra presente en cada una de sus etapas. En general, dicho modelo, propone un proceso de desarrollo que consiste en nueve etapas. En la Tabla XI, se muestra el Modelo de Nitko para el desarrollo de estrategias evaluativas del aprendizaje.

Tabla XI. Modelo de Anthony Nitko para crear exámenes nacionales de referencia criterial y referencia normativa alineados con el currículum.

Fases	Etapas y procedimientos generales
1. Definir el dominio de resultados que pretende el currículum	-Análisis del contenido curricular
2. Analizar el currículum	-Elaboración de un mapa curricular del dominio de contenido a evaluar -Establecimiento de concordancia
3. Desarrollar un plan de evaluación	-Muestreo de resultados de aprendizaje a evaluar -Diseño de prototipos de especificaciones de ítems -Establecimiento de comités de examen
4. Desarrollarla especificación de ítems	-Diseño de especificaciones de ítems
5. Producir y validar ítems	-Crear ítems según las especificaciones -Revisar la congruencia ítem-especificación -Ensayo empírico y revisión de ítems
6. Ensamblar los exámenes	-Estructuración de una muestra de ítems representativa del dominio curricular
7. Establecer estándares	-Definición de estándares de ejecución, justos, mínimos y que representen ejecuciones comparables en el tiempo
8. Análisis primario	-Especificación de estándares de calidad para el contenido del ítem, para la calidad técnica del ítem y para la calidad integral de los puntajes del examen
9. Análisis secundario	-Diseño y aplicación de procedimientos para calificar el examen y analizar los resultados
9. Análisis secundario	-Análisis de resultados para elaborar los reportes de resultados por escuela y municipio, por eje temático del plan de estudios y por destinatario

En el modelo evaluativo del INEE, el desarrollo de exámenes nacionales alineados con el currículum presenta ciertas particularidades que lo diferencian de otros modelos. En particular, el INEE desarrolló una nueva generación de pruebas nacionales para evaluar las habilidades y conocimientos de los estudiantes de educación básica y media superior, que se conocen con el nombre de *Exámenes de la Calidad y el Logro Educativos* (Excale). La evaluación realizada por la institución tiene como fin último contribuir al conocimiento sobre la calidad de la educación básica y media superior en México, es decir, conocer el rendimiento académico de los estudiantes a niveles estatal y nacional, así como los factores de contexto que lo explican, emitir juicios de valor para apoyar la toma de decisiones documentada y hacer una rendición de cuentas a la sociedad mexicana sobre el estado que guarda la educación nacional. Una forma en que se atiende este objetivo es a través de la evaluación de los aprendizajes logrados por los estudiantes en determinadas asignaturas y grados escolares (INEE, 2006).

La aplicación de los Excale confirma los resultados de evaluaciones anteriores, y proporciona información nueva, por su diseño alineado al currículum y porque incluye niveles de competencia definidos en forma rigurosa. Con los resultados obtenidos en dichas pruebas, el INEE puede obtener en qué medida los alumnos consiguen alcanzar los objetivos establecidos en los planes y programas de estudio al final de la primaria y de la secundaria.

En general, las pruebas producto de la operación del modelo evaluativo del INEE, contienen las siguientes características básicas: 1) son de tipo criterial, 2) están alineados al currículum nacional y 3) tienen un diseño matricial. Tanto lo referente a la evaluación criterial como la alineación al currículum, se ha tratado en párrafos anteriores, sin embargo, es necesario puntualizar lo concerniente al diseño matricial de pruebas. Las pruebas matriciales son diseñadas para evaluar gran cantidad de contenidos sin someter a los estudiantes a largas jornadas de resolución de pruebas. Para ello, se requiere que para el dominio curricular completo que se desea evaluar se construyan un grupo de reactivos que cubran

dicho dominio. Después ese grupo de ítems son divididos en subconjuntos y distribuidos entre los estudiantes, de tal manera que cada alumno contesta sólo parte del examen. De esta manera se limita la cantidad de ítems que responde cada estudiante, reduciendo el tiempo de la aplicación, a la vez que se cubre la totalidad de contenidos curriculares.

Por otro lado, siendo los Excale pruebas de gran escala, sus reactivos son básicamente de *respuesta seleccionada*, con un diseño de opción múltiple. Sin embargo, para evaluar algunas habilidades y conocimientos también se utilizan reactivos de respuesta abierta o construida. En general, dicho modelo, propone un proceso de metodológico de evaluación que consiste en nueve etapas. En la Tabla XII, se muestra el Modelo utilizado por el INEE para el desarrollo de los Excale.

Tabla XII. Modelo del INEE para crear Exámenes de Calidad de Logro Educativo nacionales alineados con el currículum.

Fases	Etapas y procedimientos generales
1. Planeación general	<ul style="list-style-type: none"> -Diseño del plan de evaluación -Aprobación por la junta directiva -Diseño y elaboración de cuestionarios de contexto -Aprobación por el consejo técnico -Diseño y desarrollo del sistema informático -Aprobación de especialistas del INEE
2. Estructuración de los Excale	<ul style="list-style-type: none"> -Diseño de las pruebas -Aprobación por especialistas del INEE -Especificaciones de reactivos -Aprobación de revisores del INEE
3. Construcción de reactivos de Excale	<ul style="list-style-type: none"> -Elaboración de reactivos Especialistas del INEE revisan y aprueban -Validación de reactivos -Aprobación de especialistas del INEE Piloteo de reactivos y cuestionarios de contexto -Aprobación de revisores del INEE
4. Conformación de los Excale	<ul style="list-style-type: none"> -Selección de reactivos e integración de bloques y formas -Aprobación de especialistas del INEE -Edición, armado e impresión -Aprobación de INEE
5. Aplicación de Excale	<ul style="list-style-type: none"> -Muestreo poblacional -Aprobación de especialistas del INEE -Administración de las pruebas y captura de los resultados -Aprobación de especialistas del INEE
6. Análisis e interpretación de resultados de los Excale	<ul style="list-style-type: none"> -Análisis iniciales de resultados -Aprobación de especialistas del INEE -Establecimiento de niveles de logro -Aprobación de especialistas del INEE -Elaboración del informe técnico de resultados de aprendizaje -Aprobación de especialistas del INEE
7. Validación de los Excale	<ul style="list-style-type: none"> -Estudios de validez de procesos y resultados de las pruebas (se deciden reactivos para descartar, revisar y para publicar) -Reinicia la fase II

En general las evaluaciones del CENEVAL son clasificadas como externas, de alto impacto y sensibles a los procesos de instrucción, ya que las personas que los sustentan afrontan consecuencias directas derivadas de los resultados que obtienen. La metodología que utiliza para el desarrollo de exámenes de ingreso, egreso de la educación media superior y superior, se basa en un modelo secuencial en el que los productos de una fase se convierten en los insumos de la siguiente (CENEVAL, 2008).

El modelo considera diversas acciones de verificación técnica en cada una de las fases tratando de apegarse a estándares internacionales. Para que las actividades de cada etapa se den por concluidas, los productos elaborados deben cumplir con los estándares de calidad que exige el propio modelo. Lo anterior se asegura mediante la revisión técnica continua. Por otra parte, el CENEVAL se declara como imparcial en sus actividades, argumentando que la estructura y el contenido de sus exámenes son determinados con el apoyo de cuerpos colegiados externos, provenientes de diversas instituciones y organismos; cada prueba cuenta con un consejo técnico y con comités académicos de apoyo, que busca proporcionar validez y legitimidad del desarrollo de la evaluación.

Los contenidos a evaluar se definen con el fin de que se ajusten a los cambios que sobrevienen en las áreas disciplinares o profesionales. Principalmente los reactivos se ajustan a competencias profesionales establecidas por organizaciones gremiales y educativas. El CENEVAL para garantizar la validez y homogeneidad del proceso de evaluación en la certificación de los distintos profesionales, procura afinar instrumentos de evaluación únicos, en cuyo diseño y elaboración participan de manera sistemática las principales organizaciones gremiales (colegios, asociaciones, federaciones, etcétera) de cada una de las profesiones. De tal manera que los exámenes generales de egreso para la licenciatura tienen un referente único para su construcción, desde donde se definen estándares para comparar el nivel de logro de los egresados contra un referente

común y, en consecuencia, una interpretación homogénea de los resultados. En general, dicho modelo, propone un proceso metodológico que consiste en nueve etapas. En la Tabla XIII, se muestra el Modelo utilizado por el CENEVAL para el desarrollo de exámenes.

Tabla XIII. *Modelo de CENEVAL para el desarrollo de exámenes de ingreso, egreso y de trayecto del nivel medio superior y superior.*

Fases	Etapas y procedimientos generales
1. Diseño	-Proyecto de evaluación -Diseño de la evaluación -Delimitación del contenido
2. Construcción	-Especificaciones de reactivos -Elaboración de reactivos
3. Verificación	-Revisión técnica de reactivos -Validación externa de reactivos -Revisión de estilos -Piloteo de reactivos -Calibración de reactivos
4. Ensamblaje	-Integración de versiones -Revisión editorial -Impresión del instrumento (opcional)
5. Aplicación	-Plan de aplicación -Registro de sustentantes -Capacitación -Almacenamiento y distribución -Aplicación del instrumento impreso (opcional) -Aplicación en línea (opcional) -Recepción y clasificación
6. Calificación	-Calificación de sustentantes -Definición de niveles de desempeño y puntos de corte (opcional)
7. Reportes	-Reportes a sustentantes -Reportes institucionales
8. Mantenimiento	-Consolidación de la información -Actualización del contenido del examen -Actualización de los parámetros de los reactivos
9. Complementos	-Cuestionario de contexto -Guía de examen para el sustentante -Manual técnico del instrumento -Plan de operación interna -Manual de aplicación

Otro de los modelos que se describen en este apartado es el creado por Contreras (2009) para desarrollar pruebas de gran escala de referencia criterial alineadas con el currículum. El modelo fue elaborado en la Unidad de Evaluación Educativa (UEE), de la UABC, como parte de los trabajos que realiza la unidad para satisfacer las necesidades evaluativas del Sistema Educativo Estatal de Baja California. En dicho modelo se requieren de la operación de un conjunto de procedimientos curriculares, psicométricos y estadísticos con el fin de lograr productos necesarios para los subsecuentes procedimientos. En particular, el logro de los estándares de calidad técnica se encuentra entrelazado en los procedimientos y etapas del mismo modelo.

El modelo se encuentra estructurado en cuatro ejes o planos principales; 1) el metodológico, 2) el temporal, 3) el de productos y 4) el plano de control de estándares. En total, el plano metodológico se encuentra integrado por 55 procedimientos específicos organizados en seis fases o etapas. El plano temporal, está condicionado en un tiempo aproximado de 12 a 24 meses según el tipo y las características del examen que se construya. El plano de los productos, se constituye de los resultados de la operación de cada uno de los procedimientos particulares. El desarrollo de dichos procedimientos es llevado a cabo por diferentes comités, de los cuales uno es el responsable técnico de la prueba (Comité Coordinador del Examen, CCE). En el plano técnico, se estructura de los principales estándares, criterios y medidas definidos nacional e internacionalmente para orientar el desarrollo de pruebas a gran escala. Para el desarrollo de las pruebas criteriosales alineadas al currículum, cada uno de los planos del modelo confluye presentando interacciones entre los elementos de las cuatro dimensiones que resultan significativas.

En general, el modelo evaluativo creado por Contreras (2009), presenta un proceso metodológico y psicométrico tetradimensional, sin embargo, con fines presentación en la Tabla XIV, se muestra solo el plano metodológico.

Tabla XIV. *Modelo de Contreras (2009) para desarrollar pruebas de gran escala de referencia criterial alineadas con el currículum.*

Fases	Etapas y procedimientos generales
1. Definición de resultados que pretende el currículum	<ul style="list-style-type: none"> -Selección y capacitación del Comité Coordinador del Examen (CCE) -Selección y capacitación del Comité Diseñador del Examen -Determinación del universo de contenido
2. Análisis del currículum	<ul style="list-style-type: none"> -Análisis del currículum -Determinación del universo de medida -Justificación del universo de medida
3. Desarrollo del plan de evaluación	<ul style="list-style-type: none"> -Diseño de la estructura del examen -Selección y capacitación del Comité Elaborador de Especificaciones de Ítems del Examen (CEEIE) -Elaboración de especificaciones de ítems -Selección y capacitación del Comité Elaborador de Reactivos del Examen (CERE) -Elaboración de ítems según especificaciones
4. Producción y validación de ítems	<ul style="list-style-type: none"> -Selección y capacitación del Comité Evaluador de Ítems del Examen (CEIE) -Evaluación de la congruencia ítem-especificación-currículum -Evaluación del sesgo y calidad técnica de ítems y modelos de examen -Piloteo y análisis de ítems y modelos -Estructuración final del examen -Capacitación de aplicadores del examen -Determinación de la muestra de examinados -Aplicación del examen a gran escala
5. Análisis primario de resultados	<ul style="list-style-type: none"> -Lectura de hojas de respuesta -Estructuración de la base de datos -Selección y capacitación del Comité para el Análisis, Reporte e Interpretación de Resultados del Examen -Calificación del examen -Análisis final de ítems y modelos
6. Análisis secundario de resultados	<ul style="list-style-type: none"> -Análisis desagregados de los resultados por área curricular -Análisis desagregado de resultados por variable de contexto -Análisis desagregado de resultados por destinatario -Elaboración del reporte por destinatario -Difusión de la información

Para resumir la presentación de los modelos metodológicos de evaluación descritos durante este apartado, en la Tabla XV se muestran algunas de sus características sustantivas de cada uno de ellos.

Tabla XV. Comparación de cuatro diferentes modelos metodológicos de evaluaciones nacionales e internacionales.

Modelo	Enfoque evaluativo	Criterio de referencia	Propósito educativo	Escala	Tipo de reactivos
Modelo de Nitko (1994)	Normativo y criterial	Alineación con el currículum	De alto impacto: -Rendición de cuentas -Toma de decisiones “duras”	Gran escala a nivel nacional	Variedad extensa de tipos de reactivo según se requiera
Modelo del INEE para el desarrollo de los Excale	Criterial	Alineación con el currículum	De bajo impacto: -Rendición de cuentas -Toma de decisiones “blandas”	Gran escala a nivel nacional	Básicamente de <i>respuesta seleccionada</i> , con un diseño de opción múltiple y de respuesta abierta o construida.
Modelo del CENEVAL para el desarrollo de la serie de EGEL, EXANI y EXCIL	Normativos y criteriales	Competencias profesionales establecidas por organizaciones gremiales y educativas	De alto impacto: -Selección -Acreditación y certificación -Toma de decisiones “duras” -Rendición de cuentas	Gran escala a nivel institucional y nacional	Básicamente de <i>respuesta seleccionada</i> .
Modelo desarrollado por Contreras (2009)	Criterial	Alineación con el currículum	De bajo impacto: -Rendición de cuentas -Toma de decisiones “blandas”	Gran escala a nivel nacional	Variedad extensa de tipos de reactivo según se requiera

Cada uno de los modelos metodológicos mencionados, presentan ciertas particularidades que los hacen aptos para ciertos fines y contextos evaluativos. En general, los modelos en su totalidad o una buena cantidad de sus procedimientos,

son generalizables para el desarrollo de diversos tipos de pruebas educativas. Sin embargo, algunos aspectos y características de dichos modelos, solo pueden ser aplicados en el contexto para el que se desarrollaron. Por lo tanto, trasladar a diferentes contextos evaluativos cualquier metodología desarrollada en un contexto particular sin realizar las adaptaciones pertinentes es un grave error.

Toda institución que pretenda desarrollar y aplicar una estrategia evaluativa, debe comprometerse con los principios de transparencia y rendición de cuentas. En particular, es necesario que los resultados de las evaluaciones se difundan de manera que se tengan en cuenta las características de cada evaluación, sus alcances y límites, para evitar inducir a error a quienes los consulten (Martínez, F., 2006). Esta posición es necesaria, pues tiene que ver con el compromiso de una difusión de resultados responsable, que debe ser a la vez transparente y congruente con las características de la evaluación en cuestión.

III. MÉTODO DE EVALUACIÓN

El objetivo de este capítulo es presentar el modelo teórico-metodológico para evaluar el aprendizaje logrado del área metodológica de los estudiantes de la carrera de psicología de la Facultad de Medicina y Psicología (FMP) de la Universidad Autónoma de Baja California (UABC). Es necesario aclarar de antemano que el modelo de evaluación contiene tres fases de las cuales solo las dos primeras se presentan en este capítulo. La Fase III será abordada en el siguiente capítulo cuando se hable del análisis de los aprendizajes logrados en el área metodológica.

El presente capítulo se encuentra estructurado en tres secciones: en la primera sección, se muestra el modelo para evaluar el aprendizaje logrado en el área metodológica. En la segunda sección se presenta la Fase I del modelo para el diseño y el piloteo del Examen de Trayecto del Área Metodológica (ETAM). Por último en la tercera sección se describe la Fase II del modelo la cual atiende la reestructuración y el ajuste del ETAM.

4.1 Modelo para evaluar el área metodológica de la carrera de Psicología

El modelo teórico-metodológico de evaluación utilizado en el presente estudio, es una adaptación de modelos evaluativos y psicométricos reconocidos internacionalmente. El modelo del estudio partió del modelo de Nitko (1994) para desarrollar evaluaciones nacionales de referencia criterial y referencia normativa alineada al currículum para la certificación y selección de estudiantes. También se partió de la metodología para la construcción de test, tanto criteriosales como normativos, propuesta por James Popham (1990), del manual para el desarrollo de

pruebas, propuesto por Steven Downing y Thomas Haladyna (2006) y del modelo adaptado por Contreras (2009) para desarrollar pruebas criterioles alineadas al currículum.

Los modelos metodológicos mencionados en el párrafo anterior, conciben al currículum como el fundamento desde donde se decide:

- Qué contenidos son esenciales de evaluar.
- El procedimiento necesario para evaluar.
- El desarrollo de la prueba o instrumento que se utilizará para dicha evaluación.
- La aplicación del examen.
- El análisis de los resultados.

En la siguiente página se muestra el modelo metodológico adaptado que se usó en el presente estudio.

Tabla XVI Modelo para el desarrollo de la evaluación institucional de trayecto del área metodológica.

Fase I	Etapas	Procedimientos
Diseño, validación y piloteo del examen	1.1 <i>Definición del dominio de resultados que pretende el currículo</i>	<ul style="list-style-type: none"> -Constituir el Comité Coordinador del Examen (CCE). -Establecimiento de los estándares de calidad de la evaluación. -Seleccionar y coordinar al Comité Diseñador del Examen (CDE). -Realizar un primer análisis del currículo. -Determinar el universo de contenido a evaluar.
	1.2 <i>Análisis del currículo</i>	<ul style="list-style-type: none"> -Analizar a profundidad el currículo. -Elaborar la retícula del contenido a evaluar. -Determinar el universo de medida. -Justificar el universo de medida.
	1.3 <i>Desarrollo del plan de evaluación</i>	<ul style="list-style-type: none"> -Diseñar la estructura del examen y elaborar las especificaciones generales de ítems. -Elaborar las especificaciones puntuales de ítems (etapa general reducida a procedimiento).
	1.4 <i>Producción, validación y pilotaje de ítems</i>	<ul style="list-style-type: none"> -Seleccionar y coordinar al Comité Elaborador de Ítems (CEI). -Elaborar ítems según especificaciones. -Evaluar la congruencia ítem-especificación-curriculum y posibles sesgos en los ítems (procedimientos efectuados informalmente). -Evaluar calidad técnica de ítems -Estructurar la primera versión del examen. -Pilotear a pequeña escala los ítems. -Análisis de la calidad técnica de ítems y del modelo de examen
Fase II	Etapas	Procedimientos
Ajustes de ítems del examen	2.1 <i>Revisión de ítems (procedimiento incorporado como etapa general)</i>	<ul style="list-style-type: none"> -Especificar estándares de calidad para el contenido del ítem, para la calidad técnica de ítems y para la calidad integral de los puntajes del examen. -Analizar las fallas de ítems. -Ajustar los ítems según el tipo de falla.
	2.2 <i>Estructuración del modelo de examen</i>	<ul style="list-style-type: none"> -Estructurar el banco de ítems muestra representativos del dominio curricular. -Estructurar la versión final del examen (etapa general reducida a procedimiento). -Determinar la muestra de examinados. -Capacitar a los aplicadores. -Aplicar el examen institucional.

Fase III	Etapas	Procedimientos
Análisis de los resultados de la evaluación	3.1 <i>Análisis primario</i>	<ul style="list-style-type: none"> -Leer las hojas de respuesta y capturar los resultados en una base de datos para su análisis. -Calificar los resultados del examen. -Analizar la calidad técnica de los ítems y del modelo de examen.
	3.2 <i>Análisis de resultados</i>	<ul style="list-style-type: none"> -Analizar la ejecución de los examinados por ejes curriculares y por redes temáticas estructuradas. -Determinar los estándares de interpretación de los resultados. -Elaborar reportes de resultados del examen apropiados a las necesidades de información de los destinatarios. -Socialización de resultados.

El Modelo para el desarrollo de la evaluación institucional del área metodológica de la carrera de Psicología es un modelo teórico-metodológico alineado al currículum de referencia criterial y de bajo impacto (Nitko, 1994; Popham 1990 y Ravela, 2006). Desde otra perspectiva metodológica el modelo evaluativo se puede clasificar como un estudio descriptivo desde un enfoque cuantitativo utilizando un modelo simple de evaluación el cual sólo tiene en cuenta la dimensión de los resultados académicos de los alumnos (Escudero, 1997). Durante el estudio no se controló, seleccionó ni ordenó a los individuos ni a los grupos. Tampoco se obligó a los estudiantes para que realizaran una serie de pruebas de rendimiento para comprobar su preparación y entonces acreditarlos. La finalidad de la evaluación consistió en obtener a través de una evaluación de calidad evidencias empíricas sobre la operación del currículum, y así conocer el nivel de logro educativo en el área y la etapa del plan de estudios en cuestión.

En otro punto, como se dijo anteriormente el modelo teórico-metodológico que aquí se presenta es una adaptación de otros modelos. Las adaptaciones consistieron principalmente en la reducción de etapas del modelo propuesto por Nitko (1994). En general se redujeron dos de las nueve etapas originales integrándose como procedimientos en el modelo evaluación, se ascendió un procedimiento específico a etapa general y se eliminaron procedimientos de la tercera fase. En definitiva quedaron tres fases y ocho etapas generales. La primera fase constó de cuatro etapas, y la segunda fase al igual que la tercera constó de dos etapas. Las adaptaciones se justifican debido a la necesidad de ajustar la metodología a las limitaciones y alcances del presente estudio.

Cada una de las adaptaciones contiene ciertas particularidades. En específico las etapas generales reducidas a procedimientos son el **desarrollo de las especificaciones de ítems** y el **ensamblaje de exámenes**. La primera etapa reducida se incorporó como procedimiento a la etapa general **Desarrollo del plan de evaluación**, y la segunda etapa reducida se incorporó de igual forma como procedimiento a la etapa **Estructuración del modelo de examen**. Los motivos de las adaptaciones descritas hasta este momento tienen que ver con la necesidad de reducir el tiempo requerido para dicha tarea y conservar la secuencia lógica y la consistencia de los procedimientos involucrados.

Otra de las adaptaciones al modelo consistió en ascender el procedimiento **revisión de ítems** a etapa general ocupando el 5to lugar dentro de la secuencia de las etapas del modelo adaptado. El motivo para darle énfasis a este procedimiento se justifica en la relevancia y el tiempo otorgado a las tareas de revisión y ajuste de ítems dentro del presente proyecto de evaluación.

En la misma línea de rendición de cuentas sobre las adaptaciones del modelo metodológico es necesario aclarar que la etapa general **Establecer estándares de ejecución** fue reducida a procedimiento. La reducción de dicha etapa, tiene su fundamento debido a que no se contaba en su momento de la

cantidad adecuada de jueces. Por tal motivo, se realizó dicha acción en la Fase III al momento del análisis de los resultados de la evaluación. Cabe señalar, que la actividad fue realizada por una sola persona integrante del Comité Coordinador del Examen que realizó la función de juez. Dicha actividad, se basó en el enfoque de establecimiento de estándares de interpretación (EE) dependiente del comportamiento empírico en la prueba. El procedimiento se basó en analizar los ítems característicos de cada nivel de habilidad total en la prueba y establecer cortes en la escala, de forma que se diseñó el descriptor de cada nivel a partir de los ítems que componen cada nivel. En específico, el método partió de un cuadernillo de ítems ordenados por su dificultad empírica. La tarea del juez fue identificar los ítems que actúan como punto de inflexión entre dos niveles de desempeño definidos según la división de los niveles de dificultad. Este tipo de enfoque, permite ajustes más realistas de los EE al tener como referencia la dificultad empírica de los ítems (Jornet y González, 2009).

En otro punto, para el desarrollo del examen se establecieron diversos criterios de calidad técnica y estándares psicométricos. De antemano, es necesario señalar que el examen se concibió como una evaluación de mediana escala debido a que solo se aplicó en una institución, en un solo plantel escolar y con una cantidad reducida de grupos (Aiken, 1996). Sin embargo, desde el inicio del desarrollo del examen se adoptaron en su mayoría los estándares de calidad técnica para el desarrollo de pruebas nacionales de gran escala.

Uno de los principales estándares de calidad que se siguieron en el desarrollo del examen fue la validez de contenido. Dicho criterio es el más importante para definir la calidad de evaluaciones de referencia criterial alineadas al currículum (Contreras, 2003, Nitko, 1994). En el presente estudio, se interconstruyó dentro del modelo metodológico de evaluación el proceso de validación tal y como lo propone Nitko (1994). De esta manera el objetivo del modelo metodológico fue

asegurar que el resultado de la ejecución de los estudiantes ante el examen, permitiera inferir en forma válida el dominio del currículum.

Cabe señalar que uno de los procedimientos que proporciona evidencias empíricas de la validación del examen, es el resultado del proceso de validación de la prueba mediante jueces. Sin embargo dicha actividad fue realizada en forma limitada por dos de los integrantes del Comité Coordinador del Examen. Las limitaciones tienen que ver principalmente con la cantidad reducida de jueces y con el uso de formatos de tipo cualitativo para dichas evaluaciones. Las actividades realizadas por los dos jueces con el fin de garantizar la relación ítem-representación-curriculum consistieron principalmente en la revisión y corrección general y cualitativa de la calidad del contenido y de la calidad técnica de cada ítem del test.

En general se tomó la organización de áreas de control de calidad construida por Contreras (2000) para el aseguramiento de la calidad del desarrollo de exámenes criterioles. Dichas áreas de control son las mismas propuestas por Nitko (1994) en su modelo. En particular son tres las áreas de control: a) calidad del contenido de los ítems del test, b) calidad técnica de cada ítem y c) calidad de las calificaciones de la prueba. En la Tabla XVII se muestran las áreas de control de calidad y los estándares establecidos para el desarrollo de la presente evaluación.

Tabla XVII. *Áreas de control de calidad, estándares y criterios para el desarrollo de la evaluación del área metodológica.*

Área de control	Estándar establecido
Calidad del contenido de los ítems del test	<ul style="list-style-type: none"> -Corrección del contenido -Corrección de la respuesta correcta -Relevancia e importancia de la tarea a ejecutar -Congruencia del ítem del test con el contenido -Correspondencia del ítem del test con su especificación
Calidad técnica de cada ítem del test	<ul style="list-style-type: none"> -Escritura de ítems sin defecto -Vocabulario apropiado -Evitación de estereotipos étnicos y de género (ofensa) -Evitación de sesgo -Dificultad ($P > 0.05$ y <0.95) -Discriminación bajos-altos ($D \Rightarrow 0.2$) -Discriminación ítem-total ($R_{bis} \Rightarrow 0.20$)
Calidad de las calificaciones del test	<ul style="list-style-type: none"> -Alta confiabilidad ($\alpha \Rightarrow 0.85$)

Los estándares adoptados para el desarrollo de la evaluación del área metodológica son solo algunos de los más importantes. Otro criterio que puede adoptarse sería el concerniente a la correspondencia del ítem con la categoría de pensamiento. También, una vez aplicado el examen a lo largo del tiempo, se puede obtener indicadores más robustos sobre el índice de confiabilidad y la validez convergente, por mencionar algunos. Sin embargo los criterios aquí mencionados quedan fuera del alcance del estudio. Primero, porque no se contó con especialistas en el análisis cognitivo de los ítems requeridos para conseguir evidencias sobre la correspondencia ítem-categoría de pensamiento; y segundo, para los otros indicadores mencionados se requiere estudios longitudinales (Contreras, 2000).

4.2 Fase I. Diseño, validación y piloteo del examen

Definido entonces el modelo para el desarrollo de la evaluación institucional y los estándares de calidad que se deben perseguir durante el desarrollo del mismo, se constituyó el Comité Coordinador del Examen (CCE). Las funciones del CCE fueron las siguientes:

- Diseño y desarrollo general de la prueba.
- Control de calidad de actividades y productos del proceso.
- Constitución y capacitación de los subcomités requeridos en las diferentes etapas del desarrollo del instrumento.
- Elaboración de materiales y facilitación de herramientas para los trabajos especializados de cada subcomité.
- Pilotaje del instrumento
- Análisis de los datos resultantes.

El CCE fue integrado por profesores de tiempo completo de la FMP e investigadores del Instituto de Investigación y Desarrollo Educativo (IIDE) de la UABC. El perfil de los miembros del CCE se fundamenta en la experiencia en el desarrollo de pruebas de referencia criterial y referencia normativa de gran escala y en el conocimiento teórico y práctico del desarrollo de exámenes a nivel superior. En específico uno de los integrantes del CCE es encargado de la evaluación de la FMP.

La Fase I del modelo metodológico de evaluación tuvo como finalidad contar con un instrumento piloteado y calibrado. En específico, como resultado de las

acciones realizadas por los diferentes comités en la primera fase, se obtuvieron evidencias de la calidad técnica de los ítems y del modelo de la prueba en su conjunto. Una vez revisados y contrastados los índices de la calidad técnica de los ítems y del modelo del examen con estándares de calidad previamente establecidos, se procedió a la Fase II en donde se tomaron decisiones sobre ajustes y correcciones a los ítems. La Fase I del desarrollo de la prueba constó de cuatro etapas sustantivas:

1.1 Definición del dominio de resultados pretendidos por el currículo.

1.2 Análisis en profundidad del currículo.

1.3 Desarrollo del plan de evaluación.

1.4 Producción, validación y pilotaje de los ítems.

Etapa 1.1 Definición del dominio de resultados pretendidos por el currículum

Una de las primeras tareas del CCE fue constituir y coordinar al Comité Diseñador del Examen (CDE). Los integrantes del CDE durante su participación en el desarrollo del examen se responsabilizaron de las siguientes funciones: a) realizar el primer análisis curricular del área de competencia metodológica de la carrera de Psicología, b) determinar el universo de contenido a evaluar, c) analizar a profundidad el currículum, d) elaborar la retícula del contenido a evaluar, e) determinar y justificar el universo de medida, y por último, f) diseñar la estructura del examen y elaborar las especificaciones generales de los ítems realizando un muestreo de los objetivos del aprendizaje a evaluar.

El CDE fue integrado por profesores de tiempo completo y de asignatura de la FMP quienes trabajaron en forma asincrónica y que a su vez fueron coordinados

estratégicamente por el CCE. El CDE se caracterizó por tener miembros especializados en el área de competencia metodológica y psicólogos, así como de profesores que imparten las materias de métodos en psicología y que conocen la operación del currículum. También, el CDE se conformó de especialistas en técnicas de reticulación necesarias para el análisis a profundidad del currículum y en el diseño de especificaciones puntuales de ítems.

Las actividades que realizaron los integrantes del CDE consistieron en analizar el área del currículum en cuestión, detectar y estructurar el contenido importante a evaluar, elaborar las especificaciones de ítems y finalmente juzgar la congruencia ítem-especificación. Durante todas las actividades que realizaron los integrantes del CDE fueron apoyados directa y estratégicamente por miembros del CCE que procuraron respetar la independencia sobre las decisiones y observaciones que tomaba el CDE. En específico, los integrantes del CCE apoyaron en cuestiones teóricas y técnicas del desarrollo del examen, y en la comunicación y conciliación entre los comités. El fin de dichas acciones, es conseguir que cada procedimiento de la evaluación se realice en apego a los estándares de calidad previamente establecidos.

En general, se cuidó la independencia de los diferentes comités y grupos de especialistas en cuanto a juicios, observaciones y toma de decisiones en el desarrollo de sus actividades. Nitko (1994) en su metodología para el desarrollo de exámenes nacionales propone conformar comités de especialistas independientes que realicen los diferentes procedimientos y que a través de la discusión consoliden la concordancia de los productos e insumos propios de las diversas etapas del desarrollo del examen. El fin de cuidar la independencia entre los comités es propiciar la objetividad y consolidar consensos para recabar evidencias de validez (Contreras, 2009).

Los materiales y técnicas utilizados por los diferentes comités durante el desarrollo del examen fueron aportados por el CCE. Algunos materiales y manuales

se obtuvieron de instituciones nacionales que se dedican al desarrollo de pruebas educativas tales como el Instituto Nacional para la Evaluación de la Educación (INEE) y del Centro Nacional de la Evaluación para la Educación Superior, A. C. (CENEVAL) por mencionar algunas. En lo particular se utilizaron los manuales de dichas instituciones para el análisis curricular, la elaboración de especificaciones de ítems y para la elaboración de los ítems de la prueba. Cabe señalar que los manuales son de libre acceso y se pueden encontrar en las páginas electrónicas oficiales de las instituciones mencionadas. A continuación se describen los materiales utilizados como apoyo para el desarrollo del ETAM:

- Software especializado (Microsoft Office Visio 2003) para el desarrollo de la retícula del contenido curricular a analizar.
- Bibliografía sobre de las teorías curriculares y las teorías educativas del proceso humano de información.
- Manuales para el análisis curricular, la elaboración de especificaciones de ítems y para la elaboración de ítems.
- Materiales curriculares (plan de estudios, programas de las unidades de aprendizaje del eje metodológico, bibliografía básica y planes de trabajo de los cursos propios del área metodológica).
- Formatos, guías y prototipos para elaborar las justificaciones y especificaciones de los ítems (**Anexos 1 y 2**).

Una de las primeras decisiones importantes tomadas por integrantes del CDE —tras un primer análisis del currículum institucional— fue la primera delimitación y muestreo del dominio curricular sobre el cual se desarrolló el ETAM. Las técnicas utilizadas para esta primera delimitación fueron el análisis de contenido y una adaptación de la técnica de reticulación. El análisis de contenido se utilizó para seleccionar las metas académicas sustantivas y pertinentes a evaluar, las cuales se encuentran en los documentos oficiales curriculares. Una vez identificados los contenidos en cuestión se llevó a cabo la técnica de reticulación adaptada para apoyar la toma de decisiones concernientes al proceso de delimitación. La Figura 1 constituye el conjunto de decisiones aquí descrito.

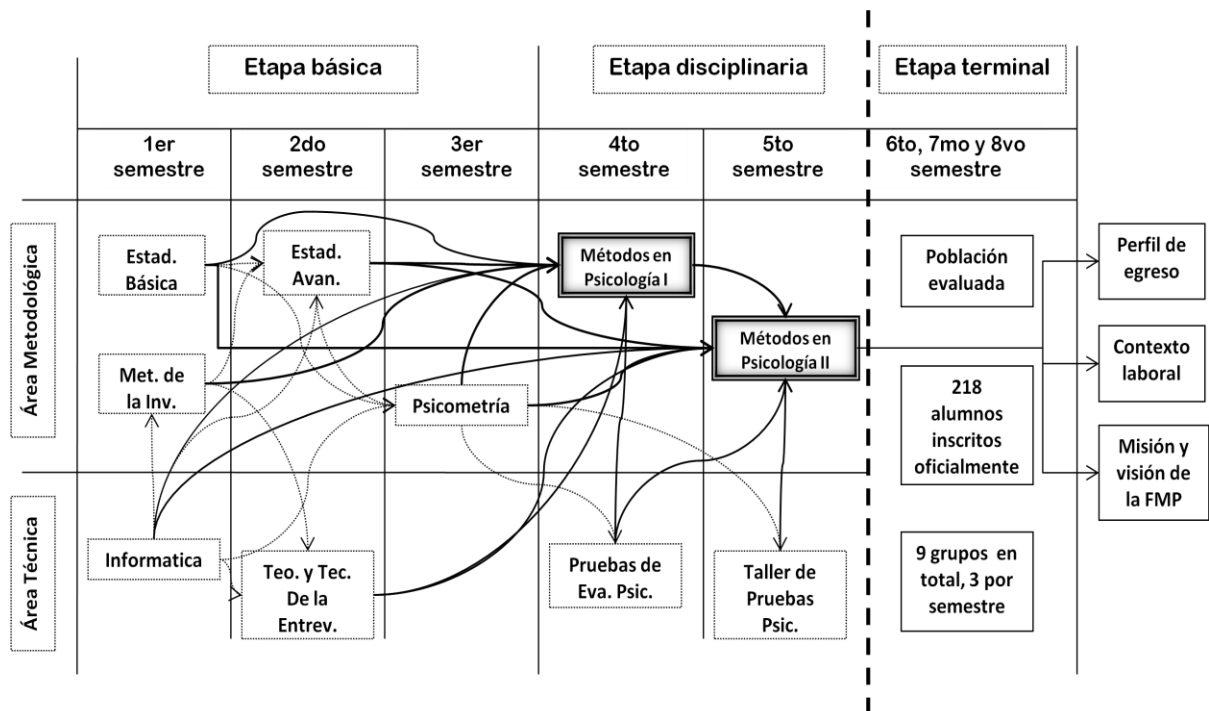


Figura 1. Reticulación sobre la toma de decisiones concernientes al proceso de delimitación del contenido a evaluar.

En resumen, el primer paso del proceso de delimitación del contenido curricular a evaluar, resultó en la selección de dos de las unidades de aprendizaje de las seis que integran oficialmente el área metodológica dentro del plan de estudios. Las dos unidades de aprendizaje seleccionadas por los expertos del CDE fueron *Métodos en psicología I* y *Métodos en psicología II*. Dichas unidades de aprendizaje terminan de impartirse al finalizar la etapa disciplinaria en el 4to y 5to semestre de la carrera de psicología. Delimitar el universo de contenido a evaluar en solo dos de las seis unidades de aprendizaje, se justificó en el hecho de no contar con expertos en el manejo de los contenidos del área metodológica y de especialistas en el manejo de contenidos según el tipo de aprendizaje y el nivel taxonómico.

Es necesario aclarar, que durante el proceso de delimitación del contenido curricular, los integrantes del CCE y del CDE decidieron que los estudiantes que cursaban los últimos tres semestres del plan curricular fueran la población del estudio. Se eligió entonces a todos los estudiantes inscritos oficialmente en los 6tos, 7mos y 8vos semestres debido a que ya habían cursado la etapa disciplinaria y dentro de ella el área metodológica.

Cabe señalar que para la delimitación y selección del universo de contenido a evaluar se discutió ampliamente sobre la validez del examen. Los expertos de cada uno de los comités coincidieron que las dos unidades seleccionadas son una muestra representativa del currículum por lo que la delimitación no atenta contra la validez y la calidad del examen. Un criterio a favor de tal decisión es que las unidades de aprendizaje seleccionadas integran y sintetizan los conocimientos adquiridos a lo largo de la etapa básica y disciplinaria en el área metodológica.

Contreras (2000) comenta que la calidad de un examen criterial es juzgada y contrastada continuamente con los objetivos educativos pretendidos por el currículum. Lo anterior tiene sentido cuando recordamos que la validez de contenido

es un indicador de alta relevancia para el desarrollo de evaluaciones criterioles. El concepto de validez, como se comentó en capítulos anteriores, se refiere al grado en que los juicios de valor que se formulan en la evaluación están adecuadamente sustentados en evidencia empírica y están efectivamente relacionados con el “referente” definido para la evaluación.

Dicho lo anterior se puede decir que durante el proceso de delimitación y selección de las unidades de aprendizaje *Métodos en psicología I* y *Métodos en psicología II*, se siguieron los estándares de calidad establecidos para el desarrollo del examen. En especial se cuidó el criterio de la validez de contenido, el cual según Ravela (2006) se logra cuando la prueba cubre adecuadamente la diversidad de conocimientos y competencias definidos en el referente.

Etapa 1.2 Análisis del currículum

Durante el análisis a profundidad del currículum se definió el objeto de medida sobre el que se desarrolló el examen. El comité diseñador fue el encargado de efectuar un análisis curricular puntual con el fin de estructurar en una retícula los resultados de aprendizaje importantes para ser evaluados.

En lo particular con el fin de seguir manteniendo los criterios de validez del examen, durante la presente etapa se revisaron y sintetizaron diferentes documentos de planeación curricular (principalmente los programas de las unidades de aprendizaje en cuestión), materiales instruccionales en los que se apoyan los profesores que imparten clases en el área metodológica y bibliografía especializada en las Ciencias Sociales y del Comportamiento. La actividad tiene como fin identificar y sintetizar los presupuestos y estrategias educativas relevantes para obtener el objeto de medida del examen a través de un análisis de contenido exhaustivo y de la estrategia de reticulación.

Cabe argumentar que de acuerdo con Alvarado y Robredo (1984), la técnica de la reticulación ayuda a resolver el problema de la determinación y secuenciación en los planes y programas de estudio, así como en el análisis de contenido para la organización del proceso de la enseñanza-aprendizaje. Tal técnica fue de gran ayuda para que los integrantes del CDE identificaran y organizaran el objeto de medida.

Los resultados del análisis a profundidad del currículum se condensaron en un modelo reticular (véase Figura 2). El análisis consistió en representar gráficamente los contenidos de las unidades de aprendizaje de *Métodos en psicología I* y *Métodos en psicología II*. La retícula se estructuró en dos ejes principales, el eje horizontal que describe a los grados educativos (4to y 5to semestre) y, el eje vertical que define los niveles curriculares (macrorretícula, mesorretícula, microrretícula y picorretícula). Como se puede observar los ejes temáticos representan la estratificación oficial de los contenidos. Tal estratificación se da en **Unidades** (I, II, III,... etc.), **Temas** (1.1, 1.2,... etc.), **Subtemas** (1.1.1, 1.1.2,... etc.) y **Pico temas** (1.1.1.1, 1.1.1.2,... etc.).

Por otra parte, durante el proceso del análisis de contenido curricular, los integrantes del comité diseñador estructuraron redes de contenidos a través de un proceso inductivo. Fueron diez redes temáticas emergentes formadas por contenidos que se interrelacionan formando macro-estructuras en los diferentes niveles tanto a lo ancho como a lo largo del currículum. Cada una de las redes temáticas aparece marcada en la estrategia reticular con un color distintivo. En la Figura 2 se muestra la retícula resultado del procedimiento de análisis del currículum a profundidad.

Métodos de investigación en psicología I

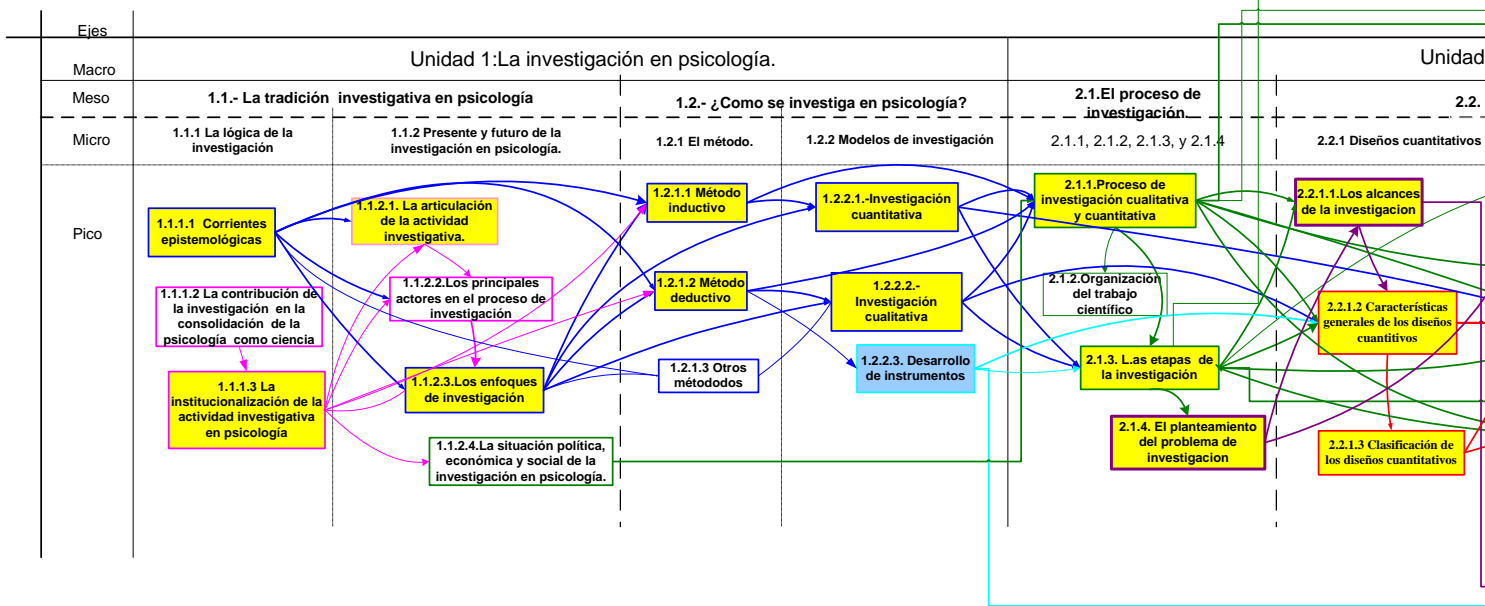
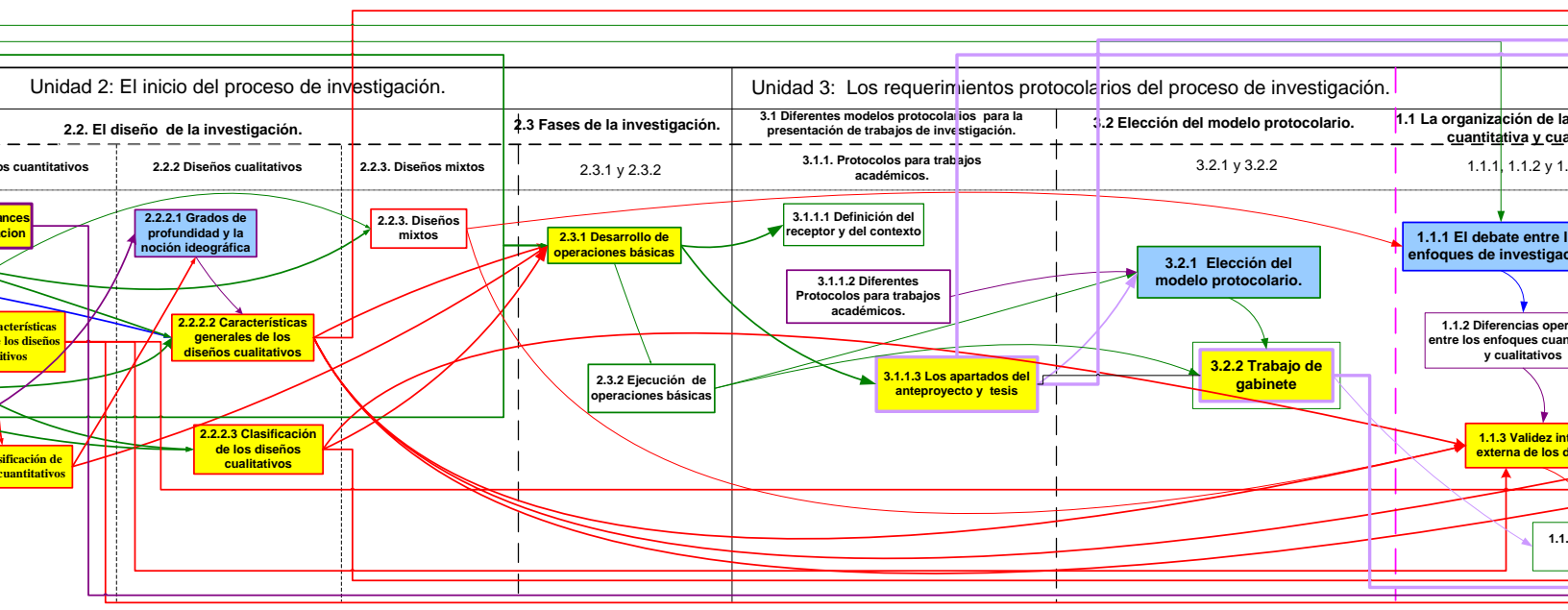
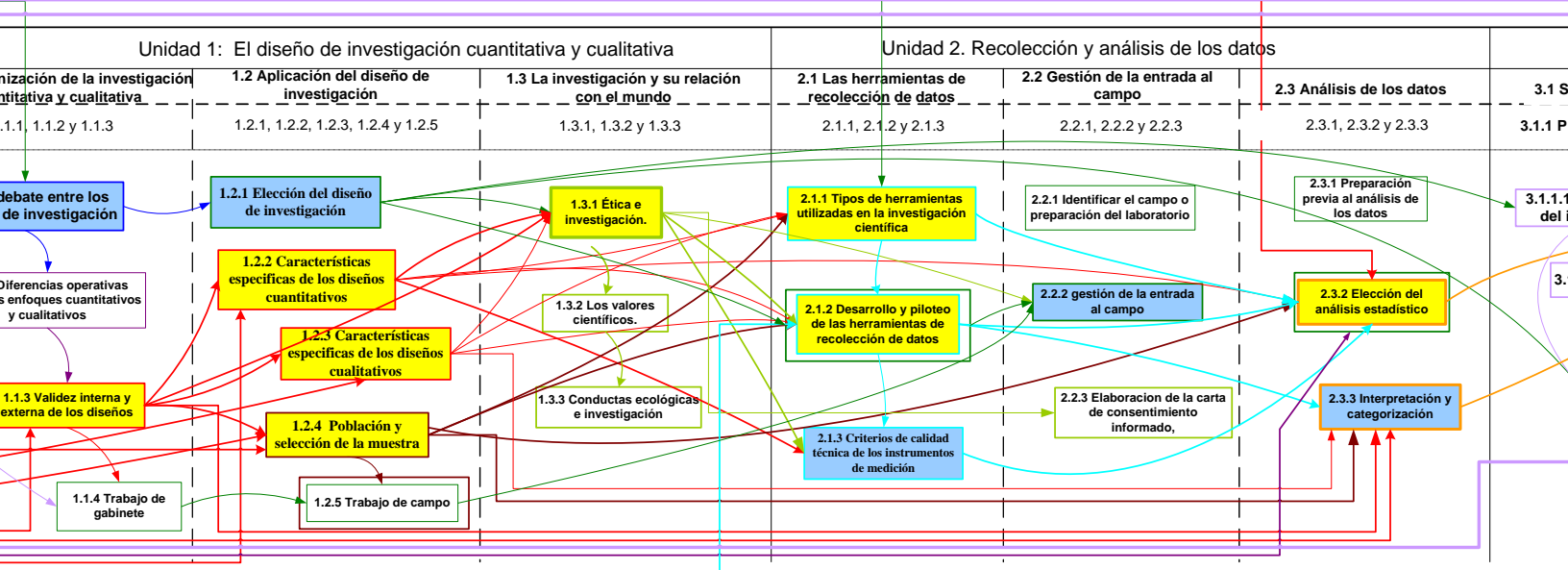
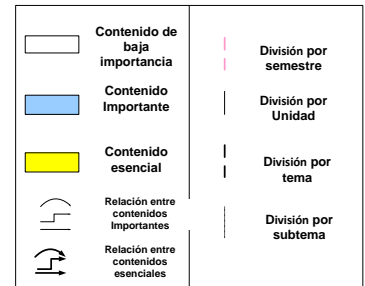
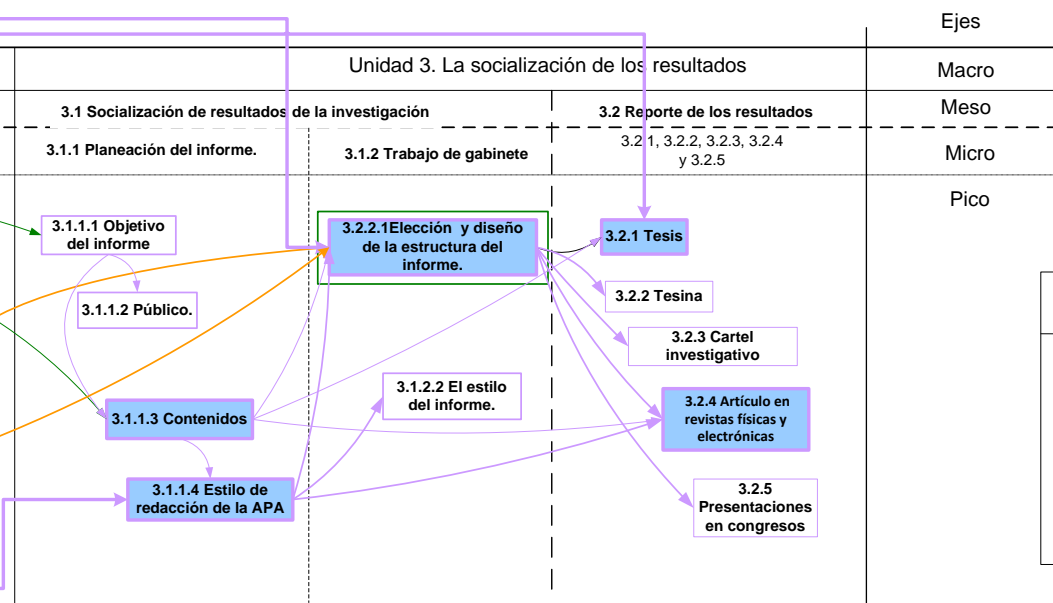


Figura 2. Retícula del contenido del Área Metodológica de la Carrera de Psicología.





en psicología II



Son varios los análisis que podemos hacer a partir de las redes de contenido estructuradas. Como se puede observar en la estrategia de reticulación los contenidos que integran cada una de las redes temáticas del área metodológica se interrelacionan a lo largo y ancho del currículum. Otro aspecto que se puede observar en cuanto a la estructura de las redes temáticas, es la respuesta a una lógica instruccional relativamente consistente. Cabe señalar que la lógica con que fueron planificadas las unidades de aprendizaje de Métodos en Psicología I y II se fundamenta en el diseño instruccional basado en el desarrollo de proyectos de investigación. Por lo que el currículum presenta contenidos referentes a actividades y competencias que se interrelacionan a lo largo del mismo en busca de objetivos educativos integradores.

La detección de redes temáticas, fue una actividad sustantiva para detectar contenidos no detectables a simple vista y que son relevantes para su evaluación debido a su importancia en el contexto curricular. En la Tabla XVIII se muestran las redes curriculares resultantes del análisis curricular:

Tabla XVIII *Redes curriculares del área metodológica y cantidad de contenidos que las integran.*

No.	Redes curriculares	Cantidad de contenidos
1	Técnicas de muestreo	2
2	Historia y ciencia en psicología	4
3	Lógica de la investigación	8
4	Proceso de investigación	17
5	Planteamiento del problema	5
6	Diseño de investigación	8
7	Análisis de los datos	2
8	Herramientas y técnicas de recolección de datos	4
9	Redacción académica	13
10	Ética de la investigación	4
Total:		67

Otro de los procedimientos relevantes en esta etapa fue la definición y selección de los contenidos esenciales e importantes que conformaron el universo de medida. Los integrantes del CDE decidieron de manera consensual la relevancia de los contenidos identificando el tipo de relaciones con otros contenidos y la cantidad de estas relaciones. También la asignación de relevancia de los contenidos se basó en criterios sobre la naturaleza pedagógica y disciplinaria. Los criterios y reglas que indican la relevancia de los contenidos se especifica en la estrategia de reticulación (véase Figura 2).

Etapa 1.3 Desarrollo del plan de evaluación

En la presente etapa, se elaboró el plan de evaluación para construir el examen. El plan de evaluación consistió de tres actividades principales. En la primera actividad se redujo el universo de contenido para la obtención de una muestra de medida representativa del currículum. En específico en esta actividad, se elaboraron las justificaciones y especificaciones generales de los ítems de la prueba (véase Anexo 1 y Tabla XIX). Para la segunda actividad se diseñaron las justificaciones y especificaciones puntuales de los ítems. Por último como tercera actividad se capacitó al Comité Elaborador de Reactivos (CER).

Durante esta etapa los integrantes del comité diseñador continuaron con el proceso de delimitación del objeto de medida. Aún después de dos acciones anteriores para delimitar el currículum había una cantidad de contenidos mayor de la que era posible evaluar en una sola ocasión. Fue necesario entonces obtener una muestra representativa del dominio curricular estructurado. Para realizar el muestreo intencional, los integrantes del CDE consideraron los contenidos relevantes del currículum que fueron definidos en la estrategia de reticulación antes efectuada.

Finalmente descritas las partes del currículum a evaluar, se determinó la distribución que tendría cada sección del examen. Tal distribución de contenidos se estructuró según su organización en el currículum. Cada una de las decisiones y de los resultados del proceso de delimitación del contenido curricular se integró en una tabla de especificaciones generales del examen. En la Tabla XIX se muestran las especificaciones generales de los ítems del ETAM.

Tabla XIX Versión final de la tabla de especificaciones generales del ETAM.

Eje curricular (asignatura)	Sub-eje curricular (unidad temática)	Contenidos a evaluar	Red curricular a la que pertenece el contenido	No. de especificaciones puntuales	No. de ítems	Posición del ítem en el examen	
Métodos en Psicología I	Unidad 1: La investigación en Psicología.	1.1.1.1 Corrientes epistemológicas	Lógica de la investigación	2	2	19 y 4	
		1.1.1.3 La institucionalización de la actividad investigativa en Psicología	Historia y ciencia en Psicología	1	1	20	
		1.1.2.1. La articulación de la actividad investigativa.	Lógica de la investigación	2	2	2 y 3	
		1.1.2.3. Los enfoques de investigación		1	1	5	
		1.2.1.1 Método inductivo		1	1	33	
		1.2.1.2 Método deductivo		1	1	6	
		1.2.2.1. Investigación cuantitativa		2	5	22, 11, 12, 13 y 14	
		1.2.2.2. Investigación cualitativa		1	1	9	
	8	—	11	14	—		
	Unidad 2: El inicio del proceso de investigación	2.1.1. Procesos de investigación	Proceso de investigación	1	2	16 y 21	
		2.1.3. L.as etapas de la investigación		2	2	28 y 7	
		2.1.4. El planteamiento del problema de investigación	Planteamiento del problema	5	5	29, 8, 30, 24 y 27	
		2.2.1.1. Los alcances de la investigación		2	2	31 y 34	
		2.2.1.2 Características generales de los diseños cuantitativos	Diseño de investigación	3	4	56, 15, 26, 36,	
		2.2.1.3 Clasificación de los diseños cuantitativos		1	1	37	
		2.2.2.2 Características generales de los diseños cualitativos		1	1	38	
		2.2.2.3 Clasificación de los diseños cualitativos.		1	1	10	
		2.3.1 Desarrollo de operaciones básicas	Proceso de investigación	1	1	68	
	9	—	17	19	—		
	Unidad 3: Los requerimientos protocolarios del proceso de investigación	3.1.1.3 Los apartados del anteproyecto y tesis	Redacción académica	1	1	55	
		3.2.2 Trabajo de gabinete (fuentes de información y redacción de referencias)		2	2	67 y 41	
		2	—	3	3	—	
	Métodos en Psicología II	Unidad 1: El diseño de investigación cuantitativa y cualitativa	1.1.3 Validez interna y externa de los diseños	Diseños de investigación	2	4	25, 32, 60, 42,
			1.2.2 Características específicas de los diseños cuantitativos		2	4	43, 40, 35, 46,
			1.2.3 Características específicas de los diseños cualitativos		1	2	49 y 50
			1.2.4 Población y selección de la muestra	Técnicas de muestreo	5	7	61, 53, 54, 1, 58, 57 y 59
			1.3.1 Ética e investigación.	Ética de la investigación	1	2	47 y 48
5		—	11	19	—		
Unidad 2. Recolección y análisis de los datos		2.1.1 Tipos de herramientas utilizadas en la investigación científica	Herramientas de recolección de datos	1	4	62, 63, 51 y 52	
		2.1.2 Desarrollo y piloteo de las herramientas de recolección de datos		1	2	23 y 64	
		2.3.2 Elección del análisis estadístico	Análisis de los datos	5	7	65, 39, 44, 45 66, 17 y 18	
3		—	7	13	—		
Totales		5	27	—	49	68	—

En la misma Tabla XIX se muestran los ejes curriculares oficiales, la red temática a la que pertenece el contenido y el total de especificaciones e ítems de la prueba. En cuanto a los ejes curriculares podemos ver que en la primera, segunda y cuarta columna se describen los ejes y sub-ejes del currículum del área metodológica que fueron seleccionados para ser evaluados. Nótese que para la asignatura de *Métodos en Psicología II* solo se evaluaron la unidad I y II. La Unidad III no se evaluó debido a que ninguno de sus contenidos fue considerado por los integrantes del CDE como esencial. La cantidad total de contenidos en los ejes curriculares sin tomar en cuenta su nivel de relevancia se muestra en la tercera columna.

La distribución de la cantidad de especificaciones por eje o red temática no es equitativa. En lo que respecta a los ejes temáticos, para *La investigación en psicología* se elaboraron once especificaciones; para *El inicio del proceso de investigación* diecisiete; para *Los requerimientos protocolarios del proceso de investigación* tres; para *El diseño de investigación cuantitativa y cualitativa* once; y para *Recolección y análisis de los datos* siete.

De la misma forma cada una de las redes temáticas contiene una cantidad no equivalente de especificaciones; la red temática *Lógica de la investigación* contiene diez especificaciones; *Historia y ciencia en psicología* una especificación; *Proceso de investigación* cuatro especificaciones; *Planteamiento del problema* siete especificaciones; *Diseños de investigación* once especificaciones; *Redacción académica* tres especificaciones; *Técnicas de muestreo* cinco especificaciones; *Ética de la investigación* una especificación; *Herramienta de recolección de datos* dos especificaciones y *Análisis de datos* cinco especificaciones.

En las tres últimas columnas que muestra la Tabla XIX se presenta el total de especificaciones puntuales de ítems, la cantidad de reactivos elaborados para cada

especificación y la posición asignada al ítem en la prueba. Nótese como dato interesante, que en los dos últimos ejes por unidad temática se elaboraron aproximadamente dos ítems por especificación en contraste con los otros ejes temáticos en donde se elaboraron aproximadamente un ítem por cada una de las especificaciones.

Cabe señalar que debido a las condiciones modestas con que se desarrolló el presente estudio, los integrantes del CDE decidieron únicamente evaluar los contenidos relacionados con el logro educativo. Tal decisión se realizó aún cuando en el programa se marcaban contenidos de otro tipo de evaluación como los de procedimiento y de opinión. También es necesario aclarar bajo las mismas condiciones antes descritas que se elaboraron únicamente ítems de opción múltiple.

Como resultado de las decisiones de los comités de expertos la extensión de la prueba fue corta. Tanto los integrantes del CDE como los del CCE decidieron considerar la evaluación como de bajo impacto y estimar la ejecución de los estudiantes solo en forma general. Como apoyo para la toma de decisiones sobre la cantidad de ítems que constituirían el ETAM, los integrantes del CCE realizaron un piloteo con 68 ítems muestra de 72 que se elaboraron. El tiempo de duración del examen fluctuó entre 1 y 1.8 horas. Lo cual no superó el tiempo de 2 horas que se encuentra establecido en la legislación interna de la FMP para la aplicación de exámenes departamentales. Además durante el tiempo que duró el piloteo los estudiantes no mostraron cansancio o falta de atención. En total se decidió elaborar 49 especificaciones para producir 68 ítems. Las condiciones específicas para la realización del piloteo se muestran en el siguiente apartado.

Una vez creado el plan de evaluación los integrantes del CDE elaboraron las especificaciones puntuales para la construcción de cada uno de los reactivos de la prueba. Dentro de cada especificación puntual de cada ítem se procuró describir los contenidos a evaluar y sus características teóricas y técnicas contextualizadas en el

currículum. El principal propósito de las especificaciones puntuales fue informar en forma específica sobre el contenido a evaluar a los elaboradores de ítems. Otro de los propósitos de las especificaciones puntuales fue proporcionar a la comunidad académica y estudiantil un marco de referencia para la interpretación de la ejecución de los estudiantes. En el **Anexo 2** se muestra un ejemplo de una especificación de ítem completa. Como aclaración, el prototipo de las especificaciones fue de nivel medio de complejidad con el fin de poseer una herramienta más adecuada y menos limitada para la elaboración de las especificaciones puntuales de los ítems (Nitko, 1994 y Popham, 1990 en Contreras, 2000).

Al final de la presente etapa se integró el Comité Elaborador de Reactivos (CER). Tal comité fue constituido por cinco profesores de tiempo completo y de asignatura pertenecientes al área metodológica de la carrera de psicología. Dos de ellos eran expertos en la elaboración de reactivos. Los integrantes del CER que no eran expertos en la elaboración de ítems, fueron capacitados por integrantes del CCE. La capacitación se realizó en forma asincrónica y en línea a través de los correos electrónicos.

Etapa 1.4 Producción, validación y pilotaje de ítems

Después de integrar el Comité Elaborador de Reactivos (CER) se procedió a la producción, validación y piloteo de ítems que integraron el examen. Basándose en las especificaciones puntuales de los ítems los integrantes del CER elaboraron 68 ítems. Los integrantes del CCE acompañaron a los integrantes del CER en la elaboración de cada uno de los ítems con el fin de proceder según las normas de calidad establecidas por organismos internacionales como la AERA y el NCME, y así consolidar la validez de la prueba. Para probar empíricamente los ítems se realizó como ya se dijo anteriormente un pilotaje con 68 ítems. Dicho pilotaje se

realizó con una muestra de estudiantes que se encontraba finalizando la etapa disciplinaria de la carrera y con ello los créditos pertenecientes al área metodológica.

La elaboración de ítems tuvo características particulares. Es necesario especificar que una de las tareas primordiales que realizaron los expertos del CDE fue la de garantizar la relación *ítem-representa-curriculum*. Los integrantes del CCE constantemente apoyaron a los integrantes del CDE para garantizar la calidad de dicha relación y así la validez del examen. Para la elaboración de los ítems los integrantes del CER se dividieron las especificaciones según su área de especialidad. Algunos de los integrantes realizaron más ítems muestra de los que se establecía en la especificación. El fin de dicha acción fue poseer una cantidad mayor de ítems de diferente estilo lingüístico pero con el mismo sentido del contenido a evaluar. Sin embargo, como se vio en párrafos anteriores solo 68 de los 72 ítems muestra que se elaboraron fueron utilizados para formar parte del examen. Tal y como se estableció previamente por los comités de expertos.

En lo que respecta al pilotaje de los ítems fue necesario realizar distintas actividades previas a su implementación. Dichas actividades consistieron en la estructuración de la primera versión del ETAM y la selección de la muestra de estudiantes. No fue necesaria la capacitación de los aplicadores del examen puesto que fue realizada por integrantes del CCE dentro de los cuales se encuentra un responsable del área de evaluación de la FMP. En cuanto a la estructuración del examen, el Comité Coordinador del ETAM estructuró los ítems siguiendo la planificación establecida previamente por comités de expertos en las especificaciones generales del examen.

Después de haber realizado el piloteo de los ítems y del modelo de examen, se capturaron los resultados y se efectuó el análisis de los mismos. El Comité Coordinador del Examen (CCE) fue el encargado de tal procedimiento. En

particular, se realizó la lectura de las hojas de respuestas utilizando un lector óptico. Dicho dispositivo tecnológico facilitó la captura de la información contenida en las hojas de respuesta. Una vez procesada la información se archivó en una base de datos electrónica proporcionada por el mismo dispositivo.

Contenida la información en la base de datos se procedió a su edición para adaptarla según los requerimientos del software especializado para el análisis de ítems. El software usado para el análisis psicométrico de los resultados del piloteo es el Item and Test Analysis Program (ITEMAN) (tm) Versión 3.50 (1993) el cual es un producto de la Assessment Systems Corporation. Dicho software se desarrolló desde la perspectiva de la Teoría Clásica de los Test (Crocker y Algina, 1986).

La finalidad del análisis psicométrico de los ítems del ETAM es calibrarlos y estimarlos a la luz de los estándares de calidad técnica adoptados en el presente estudio. Los indicadores psicométricos que puntualmente se analizaron son el *índice de dificultad*, el *índice de discriminación* (altos-bajos), el *coeficiente de correlación puntual-biserial* (R_{pbis}) y el *coeficiente de consistencia interna* (α de Cronbach). El procedimiento para la obtención de los indicadores psicométricos mencionados consistió de cuatro ecuaciones principales. La primera que se requirió para la obtención del *índice de dificultad* del reactivo fue la ecuación (1):

$$p_i = A_i / N_i$$

En esta ecuación (1) p_i es el *índice de dificultad* del reactivo, A_i es la cantidad de aciertos en el reactivo y N_i es la cantidad de aciertos más la cantidad de errores en el reactivo. La ecuación (2) que se utilizó para obtener el *índice de discriminación* (altos-bajos) fue:

$$D_i = \frac{GA_i - GB_i}{N_{\text{grupo mayor}}}$$

En esta ecuación (2) D_i es el *índice de discriminación* del reactivo i , GA_i es la cantidad de aciertos del reactivo i del 27% de examinados que obtuvieron las puntuaciones más altas en el examen, GB_i la cantidad de aciertos del reactivo i del 27% de examinados que obtuvieron las puntuaciones más bajas en el examen, y N es la cantidad de personas en el grupo más cuantioso (GA_i o GB_i). La ecuación (3) que se utilizó para obtener el *coeficiente de correlación puntual-biserial* fue:

$$r_{pbis} = \frac{\bar{x}_1 - \bar{x}_0}{S_x} * \sqrt{\frac{n_1 n_0}{n(n-1)}}$$

En esta ecuación (3) \bar{x}_1 es la media de las puntuaciones totales de aquellos que respondieron correctamente el ítem, \bar{x}_0 es la media de las puntuaciones totales de aquellos que respondieron incorrectamente el ítem, S_x es la desviación estándar de las puntuaciones totales, n_1 es la cantidad de casos que respondieron correctamente al ítem, n_0 es la cantidad de casos que respondieron incorrectamente al ítem y n es igual a $n_1 + n_0$. Por último, la ecuación (4) que se utilizó para obtener el *coeficiente de consistencia interna* (α de Cronbach) del instrumento fue:

$$\alpha = \left(\frac{n}{n-1} \right) \left(\frac{\sigma_i^2 - \sum \sigma_i^2}{\sigma_i^2} \right)$$

En esta ecuación (4) α es el coeficiente de consistencia interna, n es la cantidad de ítems de la prueba, σ_i^2 es la varianza de las puntuaciones de la prueba y $\sum \sigma_i^2$ es la sumatoria de las varianzas de los reactivos. En la Tabla XX se muestra

el resultado del análisis de ítems del ETAM tras su piloteo a la muestra de estudiantes. En este primer resultado se identifican los indicadores psicométricos: total de aciertos por ítem, *índice de dificultad*, *índice de discriminación* (bajos-altos), el *coeficiente de correlación puntual-biserial* (R_{pbis}) y el *coeficiente de consistencia interna* (α de Cronbach).

Tabla XX. Resultados del análisis del piloteo de ítems del ETAM.

N=98 K=68							
ítem	P	D	Rbis	ítem	P	D	Rbis
1	0.55	0.36	0.36	36	0.47	0.12	0.20
2	0.71	0.16	0.17	37	0.33	0.21	0.23
3	0.75	0.05	-0.05	38	0.33	0.35	0.37
4	0.86	0.33	0.55	39	0.44	0.22	0.33
5	0.88	0.08	0.18	40	0.49	0.61	0.61
6	0.70	0.34	0.38	41	0.74	0.30	0.42
7	0.47	0.53	0.55	42	0.38	0.37	0.46
8	0.45	0.29	0.37	43	0.64	0.51	0.49
9	0.55	0.43	0.50	44	0.25	0.43	0.48
10	0.64	0.47	0.47	45	0.64	0.43	0.51
11	0.70	0.35	0.46	46	0.38	0.04	0.02
12	0.08	0.00	-0.05	47	0.50	0.21	0.22
13	0.18	0.37	0.47	48	0.33	0.35	0.37
14	0.46	0.24	0.31	49	0.22	0.23	0.32
15	0.66	0.35	0.26	50	0.40	0.28	0.14
16	0.62	0.31	0.42	51	0.64	0.51	0.65
17	0.36	0.62	0.64	52	0.58	0.09	0.28
18	0.50	0.47	0.53	53	0.70	0.34	0.35
19	0.39	0.17	0.21	54	0.89	0.25	0.44
20	0.76	0.24	0.44	55	0.29	0.26	0.37
21	0.47	0.36	0.37	56	0.77	0.51	0.56
22	0.78	0.43	0.58	57	0.72	0.59	0.86
23	0.57	0.36	0.32	58	0.72	0.34	0.35
24	0.89	0.19	0.49	59	0.60	0.19	0.28
25	0.68	0.59	0.57	60	0.62	0.43	0.33
26	0.60	0.52	0.53	61	0.83	0.58	0.92
27	0.80	0.16	0.25	62	0.98	0.08	0.37
28	0.51	0.29	0.23	63	0.81	0.26	0.30
29	0.36	0.33	0.31	64	0.58	0.32	0.46
30	0.76	0.27	0.47	65	0.43	0.18	0.30
31	0.16	-0.01	0.09	66	0.60	0.60	0.61
32	0.56	0.58	0.67	67	0.47	0.09	0.15
33	0.42	0.44	0.42	68	0.57	0.26	0.31
34	0.11	0.15	0.45				
35	0.77	0.18	0.41				
				Promedio	0.56	0.32	0.39
				$\alpha = 0.70$			

D: Índice de discriminación (N altos=64, N bajos=62)

α : coeficiente de consistencia interna (α de Cronbach)

P: Índice de dificultad

Rbis: Coeficiente de discriminación

No satisfacen algunos de los estándares de calidad

Al final de la primera columna de la Tabla XX, se muestra el promedio del índice de dificultad de los 68 ítems que integran la prueba piloto. Dicho promedio fue de 0.56. Al respecto es necesario agregar que los ítems de la prueba presentaron diversos niveles de dificultad, desde un mínimo de 0.98 (muy fácil) hasta un máximo de 0.08 (muy difícil). En específico solo el ítem 62 resultó ser muy fácil de responder.

En la misma línea, al final de la tercera columna de la Tabla XX, se muestra el promedio del índice de discriminación de la diferencia entre la dificultad de los ítems del subgrupo de examinados que obtuvo las calificaciones más altas y del subgrupo de examinados que obtuvo las calificaciones más bajas. Dicho promedio fue de 0.32. Sin embargo dieciséis de los ítems que integraron el ETAM presentaron un índice de discriminación bajos-altos positivo, pero por debajo del estándar establecido ($D \Rightarrow 0.2$) y el ítem 31 obtuvo una discriminación bajos-altos negativa. En total el 25% de los ítems presentaron una discriminación por debajo de los estándares establecidos.

Con respecto al *coeficiente de correlación puntual-biserial* (R_{pbis}), en la última columna de la Tabla XX se muestra que el promedio de dicho coeficiente es de 0.39. Sin embargo, aunque el coeficiente es adecuado según el estándar establecido, en seis ítems el coeficiente fue positivo pero bajo y en dos de ellos (3 y 12) fue negativo.

Para terminar con el análisis de los indicadores psicométricos de la prueba obtenidos en el piloteo, se puede decir que el *coeficiente de confiabilidad* α de Cronbach fue de 0.70. Dicho coeficiente se encuentra por debajo del estándar establecido ($\alpha \Rightarrow 0.85$).

4.3 Fase II. Ajustes de ítems del examen

Una vez realizada la prueba empírica a pequeña escala y el análisis de la calidad técnica de los ítems y del modelo de examen, los integrantes del CCE y del CDE realizaron ajustes de los ítems y reestructuraron la versión final del examen. El objetivo de la Fase II fue contar con ítems y un modelo de examen que presentaran mejores características técnicas.

En lo particular durante la Fase II se realizaron los siguientes procedimientos: a) especificación de estándares de calidad para el contenido del ítem, para la calidad técnica del ítem y para la calidad integral de los puntajes del examen, b) análisis de las fallas de los ítems, c) ajuste de los ítems según el tipo de falla, d) estructuración del banco de ítems representativos del dominio curricular, e) ensamblaje de la versión final de examen, f) determinación de la muestra de examinados, g) capacitación de los aplicadores y h) aplicación del examen institucional.

Para el logro de los objetivos de la Fase II del modelo de evaluación se operaron dos etapas generales que contienen cada uno de los procedimientos antes mencionados. Así la Fase II del estudio evaluativo constó de las siguientes etapas:

2.1 Revisión de los ítems.

2.2 Estructuración del modelo de examen.

Etapa 2.1 Revisión de ítems

Durante esta etapa, los integrantes del Comité Coordinador del Examen junto con algunos Miembros del Comité Diseñador realizaron un análisis minucioso de los indicadores psicométricos de los ítems y del modelo de examen. Apoyándose en el resultado del análisis se identificaron algunos de estos indicadores que no cumplían con los estándares de calidad establecidos. El objetivo de la revisión es identificar el tipo de falla que presentan los ítems para realizar los ajustes necesarios.

Para la revisión, se tomó como base la *taxonomía de fallas más comunes y decisiones adoptadas para el mejoramiento de los ítems propuesta* por Contreras (2000). Como se observa en la Tabla XXI la taxonomía mencionada considera una revisión minuciosa de ítems a través del análisis de sus *índices de dificultad y discriminación* (bajos-altos), de los *coeficientes de correlación puntual-biserial* (*R_{pbis}*), de los distractores y, de la redacción y edición de cada uno de los ítems.

Tabla XXI. *Taxonomía de fallas más comunes y decisiones adoptadas para el mejoramiento de los ítems (Contreras, 2003).*

Tipo de falla	Ejemplos	Decisiones adoptadas para el mejoramiento de los ítems
Complejidad cognitiva	-Muy difícil	-Cuidadoso análisis del contenido del ítem y, en su caso, redacción más categórica, hacer más o menos atractivas las opciones o sustituirlas.
	-Muy fácil	
Discriminación errónea	-Negativa	-Cambiar la opción que atrae a los que saben, cambiar distractores poco elegidos.
	-Baja	-Cambiar la opción que parece está creando el problema.
Edición	-Escritura confusa	-Corregir errores mecanográficos, hacer dibujos más claros.
Redacción	-Conceptos complejos	-Simplificar las conceptualizaciones, cambiar la opción confusa.
Mixta		-Hacer más categóricas la base y la respuesta correcta, o sustituir distractores.

El resultado de la revisión minuciosa de los ítems fue la identificación de diversos tipos de fallas. La falla de mayor aparición fue la *discriminación* (bajos-altos) errónea con 28 ítems que representan al 26.5% de la prueba. La falla con menor ocurrencia fue la correspondiente a ítems con un alto nivel de dificultad por arriba del estándar establecido y por criterio de los integrantes del CCE. En contraste nótese que el 50% de los ítems presentaron una buena calidad técnica según los indicadores psicométricos. En la Tabla XXII se muestran los tipos de fallas que presentaron los ítems de la prueba.

Tabla XXII. Fallas presentadas por los ítems de la prueba.

ítems de la prueba que presentan fallas	No. ítems	%	Tipo de falla
2, 3, 5, 12, 15, 19, 24, 27, 31, 34, 35, 36, 46, 52, 59, 62, 65 y 67.	18	26.5	Discriminación general (bajos-altos) errónea
2, 3, 5, 12, 31, 46, 50 y 67.	8	11.8	Discriminación general (<i>Rpbis</i>) errónea
6, 8, 13, 15, 20, 23, 29, 33, 49, 53, 54, 56, 63 y 64.	14	20.6	Discriminación errónea solo en distractores
12, 13 y 31.	3	2.9	Ítems difíciles
3, 5, 54, 56, y 62	5	7.4	Ítems fáciles
15, 26, 35, 36, 39, 46, 49 y 67	8	11.8	Redacción confusa de la base o de algún distractor
ítems con buena calidad	Total	%	
1, 4, 7, 9, 10, 11, 14, 16, 17, 18, 21, 22, 25, 28, 30, 32, 37, 38, 39, 40, 41, 42, 43, 44, 45, 47, 48, 51, 55, 57, 58, 60, 66 y 68	34	50.0	

Una vez identificadas las fallas de los ítems se realizaron los ajustes pertinentes. Las principales decisiones tomadas fueron la conservación o modificación de cada uno de los ítems con fallas. En cuanto a los ajustes en el **Anexo 3** se muestran puntualmente las decisiones tomadas para cada uno de los ítems que presentaron fallas.

También es necesario agregar, en lo que respecta a los ajustes y modificaciones de los ítems, que los integrantes del CCE realizaron mejoras de edición y formato a la mayoría de los ítems del ETAM. Particularmente se realizaron correcciones de contenido o errores conceptuales y de vocabulario inapropiado para el nivel de los examinados. También se revisaron errores sobre estereotipos étnicos o de género, de tal forma que no penalizaran injustamente u ofendiera a ningún grupo en particular.

Etapas 2.2 Estructuración del modelo de examen

Después de la revisión minuciosa de los ítems y de sus ajustes en base a las fallas identificadas, el Comité Coordinador del ETAM estructuró los ítems corregidos en la última versión del examen antes de su aplicación oficial. No hubo modificaciones en las especificaciones generales de los ítems, por lo que quedó exactamente como se estableció en la **Etapas 1.3 Desarrollo del plan de evaluación**. Cabe señalar que cada uno de los ítems formó parte de un banco de ítems que fue resguardado por los encargados de evaluación de la FMP.

Con esta última acción, se dio por concluido el desarrollo del examen. Los integrantes del Comité Coordinador consideraron oportuno aplicar el ETAM en forma oficial dado que a su juicio reunía suficientes evidencias empíricas sobre su calidad técnica y por lo tanto puede ser utilizado en el monitoreo permanente de los estudiantes. Con lo anterior se da cumplimiento a uno de los principales objetivos del estudio. Sin embargo, los especialistas en psicometría recalcaron la necesidad de mantener en un proceso de calidad continua el examen. Principalmente recomendaron que antes de la siguiente aplicación del ETAM se estructurara un comité de expertos independiente que evaluara la alineación ítem-currículo a través de un procedimiento estandarizado y objetivo. Dicho procedimiento es necesario para consolidar la validez del examen.

En el **Anexo 4** se muestra la portada del ETAM con el fin de ilustrar su estructuración. Los demás ítems del examen no se muestran con el fin de mantener la confidencialidad e integridad del instrumento y con ello la posibilidad de nuevas aplicaciones. La estructura general del examen presenta las siguientes características:

- Un encabezado con el nombre de la institución y del examen
- Una breve presentación sobre el dominio del examen.
- Una sección de instrucciones generales para responder.
- En la misma página da inicio el examen con la presentación del ítem 1.

Los últimos procedimientos de la **Etap 2.2 Estructuración del modelo de examen**, son la definición de la población a evaluar y la aplicación del examen. La primera acción, fue prevista durante el desarrollo de la **Etap 1.1 Definición del dominio de resultados pretendidos por el currículum** por los integrantes del CCE y del CDE. La segunda acción, tuvo una planificación corta debido a que no hubo necesidad de capacitar a los aplicadores.

La población a evaluar fue conformada por 218 alumnos inscritos oficialmente en los últimos tres semestres (6to, 7mo y 8vo) de la carrera de Psicología de la FMP durante el ciclo escolar de enero-septiembre del 2010. Cada uno de los semestres se divide a su vez en tres grupos con un aproximado de 27 alumnos por grupo. La evaluación fue oficialmente decretada como obligatoria y de bajo impacto por los directivos de la FMP. Uno de los criterios importantes fue que todos y cada uno de los participantes del estudio evaluativo tuvieran aprobados todos sus créditos del área metodológica.

La capacitación de los aplicadores fue mínima debido a que los profesores de la carrera de Psicología fueron quienes aplicaron el examen en horas de clase.

Todos los aplicadores sin excepción alguna eran psicólogos experimentados. Por lo tanto, la capacitación solo consistió en proporcionar información sobre la logística general de la aplicación de los exámenes y de la recepción y entrega de los mismos. Cabe señalar que parte de los integrantes del Comité Coordinador estuvo durante todo el tiempo de la aplicación del examen, procurando que se llevara a cabo según las normas establecidas para resguardar los exámenes y la confidencialidad de la información.

En especial la logística para la aplicación de exámenes a lápiz y papel es una tarea ampliamente conocida y dominada por el personal de la institución. Tal dominio es el logro de una experiencia de más de 4 años de aplicación de exámenes departamentales. De la misma forma todos los recursos tecnológicos (copiadoras, impresoras, computadoras, lector óptico, etc.) como los materiales (hojas blancas, hojas de respuestas, lápices, etc.) utilizados en la aplicación fueron facilitados por la institución. Lo anterior, agilizó en gran manera la aplicación del ETAM.

IV. RESULTADOS DE LA EVALUACIÓN

5.1 Fase III. Análisis de los resultados de la evaluación

Durante este capítulo se presenta la Fase III del modelo de evaluación del área metodológica. El capítulo se estructura en dos apartados esenciales que describen las etapas generales contenidas en esta fase. En el primer apartado, se presenta el análisis primario que consistió en la revisión de ítems y del modelo de examen. El propósito de esta etapa, es observar las características técnicas presentadas por el ETAM una vez que fue aplicado oficialmente en la población de estudio. Así, identificados dichos criterios de calidad, se pueden tomar las decisiones respectivas para el uso e interpretación de los resultados. Para el segundo apartado, se muestra el análisis de los resultados de la evaluación en forma puntual, es decir, se analiza la ejecución de los examinados por semestre escolar y por áreas curriculares. Por último, se describe la elaboración del reporte de resultados y la socialización de los mismos.

Etapa 3.1 Análisis primario

Después de haber realizado la aplicación del ETAM a los estudiantes, se capturaron los resultados y se efectuó el análisis de ítems. El Comité Coordinador del Examen (CCE) fue el encargado de tal procedimiento. En particular, se realizaron los mismos procedimientos de captura y análisis de resultados descritos en la etapa del piloteo. En la Tabla XXIII, se muestra el resultado del análisis de ítems del ETAM tras su aplicación a la población de estudio. También en el **Anexo 5**, se muestran los resultados del análisis de ítems en el formato de salida del ITEMAN. En este primer resultado, se identifican los indicadores psicométricos: total de aciertos por ítem, *índice de dificultad*, *índice de discriminación* bajos-altos, y el *coeficiente de correlación puntual-biserial* (R_{pbis}). Más adelante se presenta en la Tabla XXIV la


descripción del coeficiente de consistencia interna de la prueba así como una síntesis de los indicadores de calidad técnica aquí mencionados.

Tabla XXIII. Resultado del análisis de ítems del ETAM.

N= 218 K=68					36	55	0.25	-0.01	-0.01
ítem	TA	p	D	Rpbis	37	67	0.31	0.44	0.5
1	87	0.40	0.17	0.22	38	47	0.22	0.2	0.32
2	116	0.53	0.3	0.33	39	134	0.61	0.43	0.42
3	105	0.48	0.29	0.31	40	44	0.20	0.28	0.37
4	114	0.52	0.14	0.21	41	187	0.86	0.06	0.15
5	214	0.98	0.03	0.32	42	76	0.35	0.03	0.07
6	154	0.71	0.15	0.18	43	99	0.45	0.24	0.24
7	103	0.47	0.33	0.31	44	67	0.31	0.1	0.06
8	160	0.73	0.29	0.34	45	13	0.06	0.1	0.37
9	81	0.37	0.5	0.54	46	47	0.22	-0.04	-0.04
10	166	0.76	0.15	0.21	47	99	0.45	0.09	0.11
11	154	0.71	0.4	0.48	48	83	0.38	0.17	0.23
12	69	0.32	0.3	0.38	49	52	0.24	0.37	0.47
13	90	0.41	0.43	0.47	50	138	0.63	0.21	0.19
14	126	0.58	0.4	0.44	51	123	0.56	0.24	0.2
15	22	0.10	0.26	0.55	52	170	0.78	0.15	0.18
16	183	0.84	0.28	0.44	53	137	0.63	0.29	0.37
17	78	0.36	0.3	0.29	54	168	0.77	0.15	0.23
18	76	0.35	0.09	0.19	55	125	0.57	0.24	0.31
19	188	0.86	0.23	0.39	56	37	0.17	0.23	0.42
20	184	0.84	0.17	0.24	57	143	0.66	0.45	0.51
21	92	0.42	0.52	0.53	58	117	0.54	0.3	0.3
22	159	0.73	0.47	0.52	59	114	0.52	0.3	0.31
23	185	0.85	0.3	0.47	60	58	0.27	0.31	0.33
24	197	0.90	0.08	0.12	61	173	0.79	0.17	0.28
25	72	0.33	0.34	0.43	62	136	0.62	0.34	0.38
26	109	0.50	0.27	0.31	63	138	0.63	0.18	0.19
27	83	0.38	0.08	0.1	64	96	0.44	0.4	0.42
28	141	0.65	0.1	0.11	65	87	0.40	0.49	0.54
29	42	0.19	0.09	0.14	66	104	0.48	-0.03	0.01
30	146	0.67	0.47	0.52	67	131	0.60	0.11	0.16
31	207	0.95	0.11	0.42	68	124	0.57	0.05	0.12
32	111	0.51	0.41	0.41	Promedios: 0.51 0.24 0.31				
33	35	0.16	0.04	0.09					
34	18	0.08	0.19	0.54					

TA: Total de aciertos por ítem

D: Índice de discriminación (n altos=64, n bajos=62)

 No satisfacen los estándares de calidad

p: Índice de dificultad

Rpbis: Coeficiente de correlación puntual-biserial

Al final de la primera columna de la Tabla XXIII, se muestra el promedio del índice de dificultad de los 68 ítems que integran la prueba. Dicho promedio fue de 0.51. Al respecto es necesario agregar que los ítems de la prueba presentaron diversos niveles de dificultad, desde un mínimo de 0.98 (muy fácil) hasta un máximo de 0.06 (muy difícil). Es necesario aclarar que el ETAM es un examen de referencia criterial alineado al currículum y que por lo tanto puede presentar ítems con diferentes niveles de dificultad cada vez que sean una muestra representativa del currículum. En específico solo dos ítems (5 y 31) resultaron ser muy fáciles de responder.

El promedio del índice de discriminación bajos-altos fue de 0.24. Sin embargo veintiséis de los ítems que integraron el ETAM presentaron un índice de discriminación bajos-altos positivo, pero por debajo del estándar establecido ($D \geq 0.2$) y otros tres (36, 46 y 66) obtuvieron una discriminación negativa. En total el 43% de los ítems presentaron una discriminación bajos-altos por debajo de los estándares establecidos. En cuanto al índice de discriminación solo el 4% de los ítems presentaron un índice de discriminación negativo.

El promedio del coeficiente de correlación puntual-biserial de la *Rpbis* es de 0.31. Dicho coeficiente es adecuado según el estándar de calidad técnico establecido ($Rpbis \geq 0.20$). Sin embargo en diecisiete ítems fue positivo pero bajo y en dos de ellos (36 y 46) fue negativo. Dado lo anterior, puede decirse que el coeficiente *Rpbis* presentó fallas en 28% de los ítems del examen al no cumplir cabalmente con los estándares de calidad establecidos. En específico el 3% de los ítems presentaron un coeficiente *Rpbis* negativo, lo cual es una falla grave que deberá de corregirse antes de realizar nuevas aplicaciones del examen.

La Tabla XXIV, además de mostrar una síntesis de los indicadores de calidad técnica del examen hasta ahora descritos, muestra también el coeficiente de confiabilidad de la prueba. Los estadígrafos utilizados para obtener dicho coeficiente

fueron el α de Cronbach y el KR-21 de Kuder Richardson. El α de confiabilidad del examen fue de 0.715, lo cual está por debajo del estándar establecido ($\alpha \Rightarrow 0.85$). Es necesario aclarar que los problemas de consistencia interna de la prueba están asociados a los problemas que presentan los índices de discriminación de los ítems. Así que una vez corregidos dichos problemas en un proceso de mejora continua es muy probable que incremente sin necesidad de agregar más ítems a la prueba.

Tabla XXIV. Estándares de calidad técnica e indicadores psicométricos del ETAM.

Calidad técnica de la prueba		K= 68	
Concepto	Estándares de calidad	Indicadores psicométricos	Observaciones
Índice de dificultad	$P > 0.05$ y < 0.95	\bar{X} de $p = 0.51$	-El promedio del índice de dificultad es adecuado para un examen de referencia criterial. Solo dos ítems (5 y 31) resultaron muy difíciles.
Índice de discriminación	$D \Rightarrow 0.2$	\bar{X} de $D = 0.24$	-El promedio del índice de discriminación bajos-altos es el adecuado, sin embargo en 26 ítems fue positiva pero baja y en 3 de ellos fue negativa.
Coefficiente de correlación biserial-puntual de la prueba	$R_{pbis} \Rightarrow 0.20$	\bar{X} de $R_{pbis} = 0.31$	-El promedio de la R_{pbis} es adecuado, sin embargo en 17 ítems fue positiva pero baja y en 2 fue negativa.
Coefficiente de correlación biserial de los distractores	Distractores: $- 0 < \sqrt{\quad}$	9 ítems – $0 < \sqrt{\quad}$	-La correlación correcta-total fue negativa en 2 ítems y en 17 de ellos fue positiva pero menor que la de otro distractor.
Coefficiente de confiabilidad α de Cronbach	$\alpha \Rightarrow 0.85$	$\alpha = 0.715$	-El coeficiente de confiabilidad α de Cronbach del examen se encuentra por debajo del criterio, sin embargo tampoco es desdeñable.

Sobre el análisis primario de los resultados, se puede decir que el examen exhibe una dificultad media para los examinados debido a que el 30% de los ítems presentó un índice de dificultad media entre 0.40 y 0.60. Por otra parte, 23 ítems mostraron un alto nivel de dificultad entre 0.06 y 0.38, en contraste con 24 ítems (35%) con un bajo nivel de dificultad entre 0.61 y 0.98.

Además, el 65% de los ítems del ETAM muestran una buena calidad técnica. Solo dos de 68 ítems que integran la prueba resultaron con una dificultad fuera del estándar, tres ítems con una discriminación altos-bajos negativa, y dos más con un coeficiente de correlación ítem-total negativa. Aproximadamente un 35% de los ítems del ETAM pueden mejorarse y corregirse a la luz de las cualidades técnicas observadas. Sin embargo tales fines se encuentran fuera de los alcances del presente estudio. Al respecto el coordinador el CCE recomendó a los encargados del departamento de evaluación de la FMP mantener la prueba en un proceso de calidad y mejora continua.

Etapas 3.2 Análisis de los resultados

Después de haber obtenido los indicadores de calidad técnica de la prueba, los integrantes del CCE efectuaron un análisis de los resultados del examen. Los procedimientos específicos que integraron esta etapa general, fueron el análisis de la ejecución de los examinados por semestre escolar y por áreas curriculares, la elaboración del reporte de resultados y la socialización de los mismos.

Para lograr los propósitos descritos, se realizaron varios procedimientos. En el análisis de los datos, se requirió del paquete estadístico SPSS (Statistical Package for the Social Sciences, versión 10.5 en español). En el SPSS se codificaron cada una de las variables del presente estudio. Primero la base de datos en formato (*.xls) que se estructuró durante la etapa del análisis primario, se editó para su traslado a formato (*.sav). Se revisaron las especificaciones de los parámetros preestablecidos por el programa SPSS previos a la ejecución de las variables. Una vez establecidas las variables del estudio, se procedió al análisis estadístico de los datos.

Para lograr los propósitos del estudio evaluativo, se realizó —como ya se comentó al final del capítulo del método de evaluación— un censo de la población.

En total, se aplicó el examen a 218 estudiantes inscritos oficialmente en los últimos tres semestres (6to, 7mo y 8vo) de la carrera de psicología de la FMP durante el ciclo escolar de enero-septiembre del 2010. Cada semestre se divide en tres grupos con un aproximado de 27 alumnos por grupo. La evaluación fue oficialmente decretada como obligatoria y de bajo impacto por los directivos de la FMP con el fin de contrastar el currículum diseñado, el implementado y los logros educativos. Así, el diseño del censo de población tuvo como propósito conocer los niveles de logro educativo de los estudiantes de la carrera de psicología en el área metodológica.

Como se muestra en la Tabla XXV, en definitiva, el censo quedó conformado por 218 estudiantes de 6to, 7mo y 8vo semestre de la carrera de psicología.

Tabla XXV. *Porcentaje de alumnos por semestre y grupo escolar: 6to, 7mo y 8vo semestre de la carrera de Psicología de la FMP.*

Semestre	N	%	Grupo	N	%
6to	75	34.4	261	28	12.8
			262	24	11
			263	23	10.5
7mo	72	33	271	23	10.5
			272	21	9.6
			273	28	12.8
8vo	71	32.6	281	25	11.5
			282	20	9.2
			283	26	11.9
Totales	218	100%	9	218	100%

En la Tabla XXVI se muestran con claridad los resultados del aprendizaje de los alumnos, que se relacionan con los estratos por semestre utilizados. El ordenamiento de los resultados de aprendizaje de mayor a menor puntuación por semestre fue el siguiente: 6to semestre, 8vo semestre y 7mo semestre. A nivel de

grupos, el ordenamiento de los resultados de aprendizaje de mayor a menor puntuación fue: 263, 261, 262, 282, 283, 272, 273, 281 y 271.

Tabla XXVI. Estadísticos descriptivos básicos y resultados del ETAM por semestre y grupo.

Semestre	%de aciertos	Error de la media	Grupo	% de aciertos	Desviación estándar	Rango de aciertos
6to	57.05	1.17	261	57	9.2	35.3-72
			262	52.8	10.4	35.3-70.6
			263	61.5	9.5	42.7-80.9
7mo	47.01	.80	271	45.8	6.6	30.9-57.4
			272	48.9	7.0	41.1-66.2
			273	46.7	6.9	30.9-57.4
8vo	49.02	1.13	281	46.0	8.7	30.9-61.8
			282	50.7	7.1	38.2-64.7
			283	50.6	11.5	30.9-73.5
Media =51.2		Moda= 51.5 Mediana= 50	Desviación tip.= 9.96		Varianza= 99.2	
Asimetría= .337		Curtosis= -.352	Min.= 30.9, Max.= 80.9		Rango= 50	

En la misma Tabla XXVI se muestran las puntuaciones promedio de los dos niveles de estratos escolares. En el estrato correspondiente al semestre escolar, el 6to semestre se encuentra 10 puntos por encima del 7mo semestre y 8 puntos por encima del 8vo y el 7mo semestre se encuentra a solo 2 puntos por debajo del 8vo semestre. Por otra parte, a nivel de los grupos, el puntaje más alto corresponde al 263 con una puntuación de 61.5. La puntuación más baja entre los grupos corresponde al 271 con 45.8 puntos. También nótese, que el porcentaje promedio de aciertos en general oscila en un rango de 30.9 a 80.9 dando una diferencia entre uno y otro puntaje de 50% de aciertos. Téngase en cuenta la brecha entre el nivel de logro educativo del grupo 261 con respecto a los grupos del 7mo y 8vo semestre,

la cual aproximadamente para cada uno de ellos es de casi una desviación estándar de la distribución poblacional.

Tomando en cuenta la población evaluada, los grupos de estudiantes de sexto semestre alcanzaron puntuaciones más altas que los grupos de séptimo y octavo. Lo anterior, parece ir en contra de la lógica del proceso de escolarización, sin embargo, hay que tomar en cuenta que en 4to y 5to semestre se toman las últimas asignaturas (Métodos en Psicología I y II) correspondientes al área metodológica. Dichas materias integran los conocimientos en el área metodológica a lo largo de la etapa básica y disciplinaria de la carrera de Psicología. Lo anterior fue uno de los argumentos para que en el presente estudio se desarrollara el ETAM en base a estas asignaturas. También hay que señalar que a lo largo de los tres últimos semestres de la carrera (6to, 7mo y 8vo) ya no se imparte ninguna asignatura que tenga que ver directamente con el área metodológica. Aun, con todo lo anterior, no hay evidencias suficientes para suponer que exista un factor de recencia o de falta de sensibilidad de la prueba sobre la escolaridad. Por otra parte, los grupos de 7mo y 8vo semestre presentan puntuaciones más o menos equivalentes en el área metodológica.

La Figura 3 muestra la media obtenida de las puntuaciones en el ETAM por cada uno de los grupos evaluados. En total son nueve los grupos, tres por cada uno de los semestres mencionados. En particular los grupos de 6to semestre en su conjunto presentaron un puntaje promedio visiblemente más alto que los grupos de 7mo y 8vo semestre.

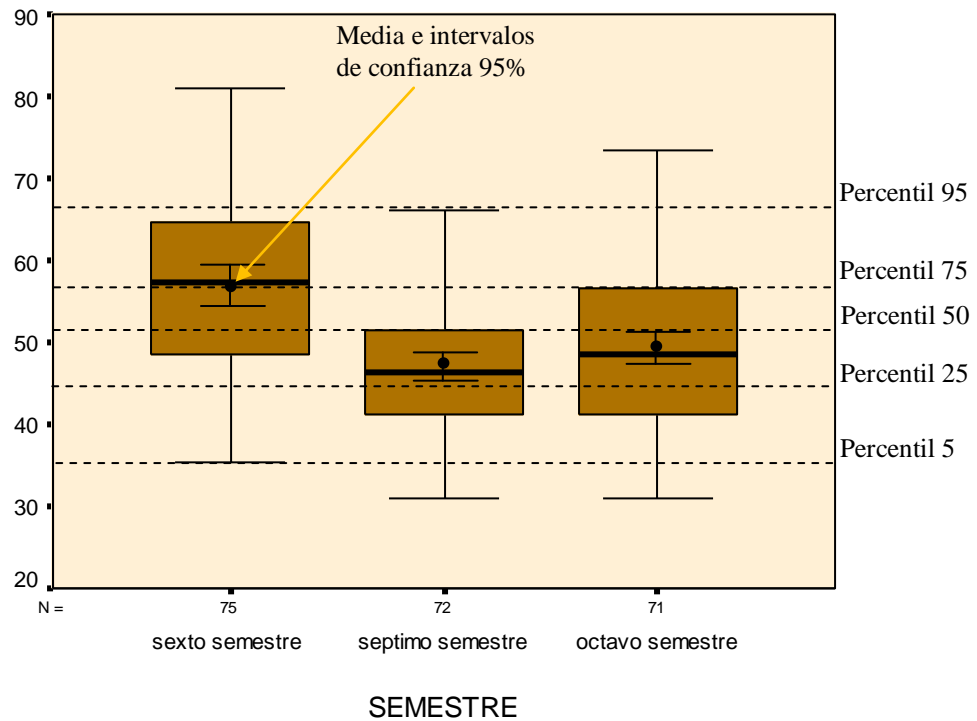


Figura 3. Comparación de los resultados obtenidos en el área metodológica de 6to, 7mo y 8vo semestre.

Con los datos obtenidos en el análisis de los resultados de la prueba, se puede decir que los estudiantes de 6to, 7mo y 8vo semestres de la carrera de Psicología tuvieron un dominio medio del área metodológica con una media de aciertos de 34.8 que corresponde a un porcentaje promedio de aciertos del 51.2%. En el 6to semestre la puntuación media de los alumnos queda significativamente por arriba de la media global de 51.2 puntos. Por otra parte, las puntuaciones medias de los alumnos del 7mo y 8vo semestre quedan por debajo de la media global, sin embargo las diferencias no son significativas ni muy grandes: 2 y 4 puntos por debajo. En cambio el 6to semestre presenta una diferencia mas abultada en referencia a la media global: 6 puntos por arriba. Cabe señalar, que no se puede concluir simplemente, a partir de los resultados del ETAM, que un estrato a nivel de los grupos o de los semestres escolares es mejor que otro en términos de la calidad educativa, puesto que no hay datos para afirmar dichas hipótesis. Lo que

sí se puede decir es que los aprendizajes de los estudiantes son mejores en unos grupos y semestres que en otros, independientemente de la razón a la que se le atribuya este hecho.

Otro dato importante que podemos observar en la Tabla XXVI, es el referente a las características de la curva de las distribuciones de los resultados del examen. El puntaje de asimetría es de 0.337 y el de la curtosis de -0.352. Lo anterior indica que los valores se agrupan ligeramente a la izquierda de la curva (por debajo de la media) y la curva de la distribución de las frecuencias es ligeramente planicúrtica. En la Figura 6 podemos observar la curva de la distribución de las puntuaciones obtenidas en el ETAM.

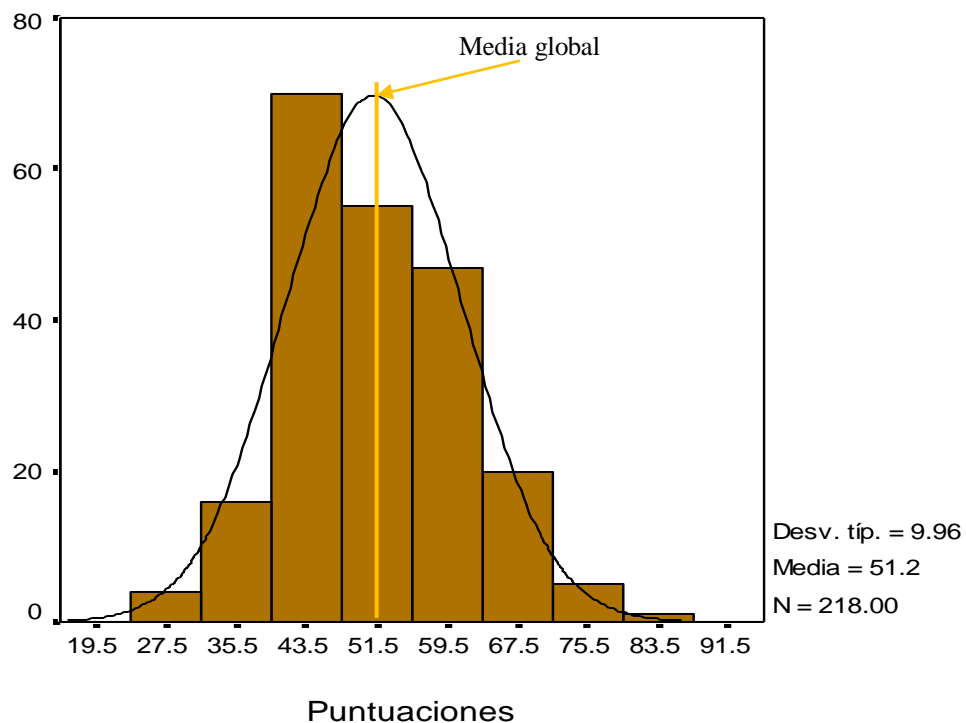


Figura 4. Distribución de puntuaciones obtenidas en el ETAM.

Otro punto importante en el análisis de los resultados, es el establecimiento de puntos de corte (PC) y de estándares de interpretación (EE). Dichas actividades

son necesarias para definir los niveles de logro educativo en un área curricular. En el presente estudio, para definir los PC para los resultados del ETAM, se partió de la dificultad media de los ítems para después ajustar los cortes próximos a $\pm .15$ de dificultad. Para la determinación de EE se adoptaron los niveles de logro de los EXCALE del Instituto Nacional para la Evaluación de la Educación (INEE) de México. Como se vio durante el fundamento teórico del estudio, las etiquetas de rendimiento académico del Excale son cuatro: Por debajo del nivel básico, Básico, Medio y Avanzado (Jornet y Backhoff, 2008). En la Tabla XXVII se describen los niveles de logro académico y las competencias académicas para cada nivel.

Tabla XXVII. Descripción genérica de las competencias académicas que logran los estudiantes en cada nivel de logro educativo (Jornet y Backhoff, 2008).

Niveles de logro	Descripción
Por debajo del nivel básico	-Indica carencias importantes en los conocimientos, habilidades y destrezas escolares que expresan una limitación para continuar progresando satisfactoriamente en el área metodológica de la carrera de Psicología
Básico	-Indica un dominio imprescindible (suficiente, mínimo, esencial, fundamental o elemental) de conocimientos, habilidades y destrezas escolares necesarios para poder seguir progresando satisfactoriamente en el área metodológica de la carrera de Psicología
Medio	-Indica un dominio sustancial (adecuado, apropiado, correcto o considerable) de conocimientos, habilidades y destrezas escolares que pone de manifiesto un buen aprovechamiento de lo previsto en el currículum en el área metodológica de la carrera de Psicología.
Avanzado	-Indica un dominio muy avanzado (intenso, inmejorable, óptimo o superior) de conocimientos, habilidades y destrezas escolares que refleja el aprovechamiento máximo de lo previsto en el currículum en el área metodológica de la carrera de Psicología.

Características de la ejecución de los examinados en las líneas de formación curricular

Uno de los principales objetivos del estudio fue dar cuenta del aprendizaje logrado por los estudiantes en el área metodológica de la carrera de Psicología. Es necesario recordar que la presente estrategia evaluativa se definió como una evaluación institucional de referencia criterial alineada al currículum. Por lo tanto, los resultados más relevantes del aprendizaje de los estudiantes se presentan en el siguiente orden: niveles de logro educativo a nivel institucional y por semestre, y diferencias por sexo. Como se puede apreciar en la Tabla XXVIII, a nivel institucional, casi tres de cada cien estudiantes (2.7%) se encuentran por debajo del nivel básico; poco menos de la mitad (42.7%) se ubica en el nivel básico; de forma equivalente otro 42.7% se ubica en el nivel medio; y sólo el 11.9% en el nivel avanzado.

Tabla XXVIII. *Porcentaje de estudiantes en los cuatro niveles de logro del área metodológica de la carrera de Psicología de la FMP.*

PC/media de $P=.51 \pm 0.15$	Intervalo de % de aciertos del PC	N	%
Por debajo del nivel básico	1-34	6	2.7
Básico	35-49	93	42.7
Medio	50-75	93	42.7
Avanzado	76-100	26	11.9
Totales		N=218	100%

Como se puede apreciar en la Figura 5, las diferencias que distinguen a los semestres entre sí, se puede apreciar considerando que por debajo del nivel básico se encuentran, en orden descendente, el 2.8% de los estudiantes de séptimo semestre y el 5.6% de los estudiantes de octavo semestre. Nótese que ninguno de

los estudiantes del sexto semestre obtuvo una puntuación por debajo del nivel básico. Además, este semestre, es el que presenta el mayor porcentaje de estudiantes en el nivel avanzado con un 26.6% en comparación con séptimo y octavo semestre que presentaron un porcentaje de estudiantes en el nivel avanzado de 1.4% y 7% respectivamente.

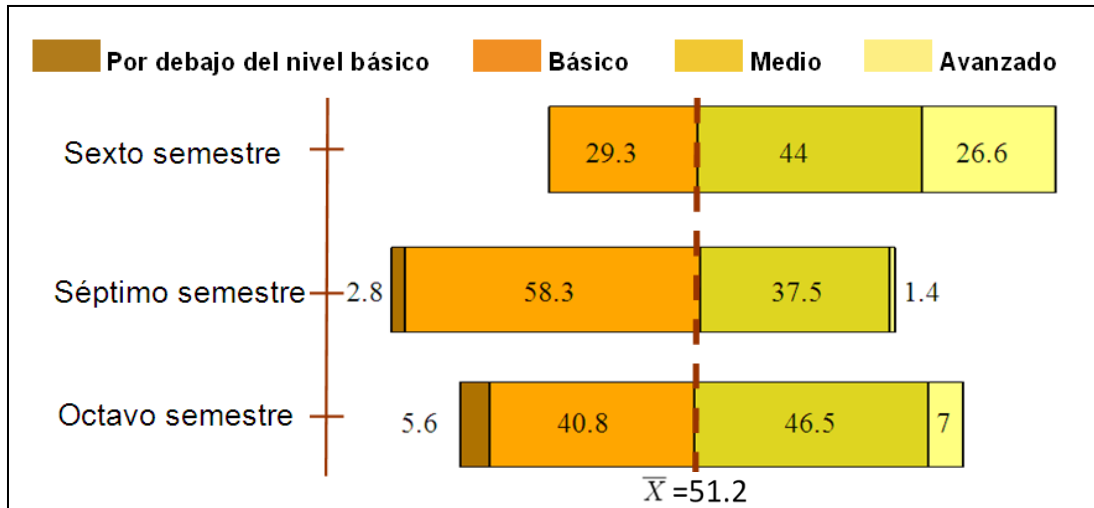


Figura 5. Porcentaje de estudiantes por semestre en los cuatro niveles de logro del área metodológica de la carrera de Psicología de la FMP.

En análisis de los resultados por sexo presenta ciertas particularidades que deben ser analizadas cuidadosamente. De inicio, la población de mujeres y hombres que fueron evaluados no es equivalente. Fueron 171 mujeres en contraste con 47 hombres que presentaron la prueba. Cabe señalar que la carrera de psicología tiene una gran demanda por parte de estudiantes de sexo femenino, es decir, tradicionalmente hay más mujeres que hombres inscritos en esta carrera. La población de mujeres obtuvo un puntaje promedio de 54.4% de aciertos, mientras que los 47 hombres obtuvieron un puntaje promedio de 53.7% de aciertos. Al respecto no se realizó algún estudio comparativo entre hombres y mujeres debido a la gran diferencia de individuos entre uno y otro grupo. Dicho lo anterior, es

necesario señalar que se debe tener mucho cuidado en realizar interpretaciones de los resultados por sexo.

En términos de contenidos curriculares, las Tablas XXIX y XXX muestran los resultados en las diferentes líneas de formación. Los porcentajes que aparecen en las tablas están ponderados según la cantidad de ítems que evalúan el dominio de cada eje y red curricular.

Tabla XXIX. Porcentaje de aciertos por eje curricular.

Ejes curriculares oficiales del área metodológica		% de aciertos
	IU.1 La investigación en psicología.....(k= 15)	58.6
Métodos en Psicología I	IU.2 El proceso de investigación.....(k= 19)	48.3
	IU.3 Requerimientos protocolarios del proceso de investigación.....(k= 3)	67.7
Métodos en Psicología II	IIU.1 El proceso de investigación cuantitativa y cualitativa.....(k= 19)	45.2
	IIU.2 Recolección y análisis de datos.....(k= 13)	49.6
K=68	Media poblacional	51.1

En la tabla XXIX se puede observar que en el dominio presentado por los estudiantes en el eje curricular **IU.3 Requerimientos protocolarios del proceso de investigación** el porcentaje ponderado de aciertos fue de 67.7%. En forma descendente la ejecución en el eje curricular **IU.1 La investigación en psicología** fue de 58.6%. En los restantes ejes curriculares la ejecución de los estudiantes presentó un porcentaje ponderado de aciertos entre 45.2% y 49.6%.

Por otra parte el dominio de los estudiantes en las redes curriculares presentó una mayor variación en la cantidad de aciertos obtenidos en cada una de ellas. El porcentaje ponderado de aciertos de las redes curriculares presentó un

puntaje mínimo de 33.9% y un puntaje máximo de 84%. Las redes curriculares **Diseños de investigación, Análisis de los datos y La ética de la investigación** presentaron la menor proporción de aciertos con 33.9%, 36.6% y 41.7% respectivamente (véase la Tabla XXX).

Porcentaje de aciertos por red curricular.

Redes curriculares del área metodológica	% de aciertos
1. Técnicas de muestreo.....(k= 7)	61.5
2. Historia y ciencia en psicología.....(k= 1)	84
3. Lógica de la investigación.....(k=13)	56.6
4. Proceso de investigación.....(k= 5)	59
5. Planteamiento del problema.....(k= 7)	59.9
6. Diseño de investigación.....(k=17)	33.9
7. Análisis de los datos.....(k= 7)	36.6
8. Herramientas de recolección de datos.....(k= 6)	64.8
9. Redacción académica.....(k= 3)	67.7
10. Ética de la investigación.....(k= 2)	41.7
K=68	Media poblacional 51.1

En la Tabla XXX se muestran también las redes curriculares que presentaron un porcentaje promedio entre 56.6% y 59.9%. Dichas redes son: **Planteamiento del problema, La lógica de la investigación y Proceso de investigación**. Las redes curriculares que obtuvieron el mayor porcentaje de aciertos fueron **Técnicas de muestreo, Recolección de datos y Redacción académica**. Sin embargo la red curricular con mayor cantidad de aciertos fue **Historia de la psicología** con un porcentaje ponderado de aciertos de 84%. Es importante señalar que esta red curricular solo contiene un reactivo, tal y como se decidió entre los diferentes comités encargados del desarrollo del examen.

De forma más puntual, en la Tabla XXXI se muestra el dominio presentado por los estudiantes en cada uno de los contenidos considerados esenciales por los integrantes del Comité Diseñador del Examen.

Tabla XXXI. Resultados puntuales de la ejecución de los estudiantes en el ETAM.

Asignatura	Eje curricular	Contenidos a evaluar	Red curricular	Posición del ítem	% de aciertos por ítem
Métodos en Psicología I (54%)	Unidad 1: La investigación en Psicología (59%)	1.1.1.1. Corrientes epistemológicas (69%)	Lógica de la investigación (57%)	19	86
				4	52
		1.1.1.3. La institucionalización de la actividad investigativa en Psicología (84%)	Historia y ciencia en Psicología (84%)	20	84
		1.1.2.1. La articulación de la actividad investigativa (50%)	Lógica de la investigación	2	53
				3	48
		1.1.2.3. Los enfoques de investigación (98%)		5	98
		1.2.1.1. Método inductivo (16%)		33	16
		1.2.1.2. Método deductivo (71%)		6	71
				22	73
		1.2.2.1. Investigación cuantitativa (55%)		11	71
		12		32	
		13	41		
		14	58		
		9	37		
		8	—	14	57
	Unidad 2: El inicio del proceso de investigación (48%)	2.1.1. Procesos de investigación (63%)	Proceso de investigación (59%)	16	84
				21	42
		2.1.3. Las etapas de la investigación (56%)		28	65
			7	47	
		2.1.4. El planteamiento del problema de investigación (57%)	Planteamiento del problema (56%)	29	19
				8	73
				30	67
				24	90
				27	38
		2.2.1.1. Los alcances de la investigación (51%)	Diseño de investigación (34%)	31	95
				34	8
		2.2.1.2. Características generales de los diseños cuantitativos (25%)		56	17
		15		10	
	2.2.1.3. Clasificación de los diseños cuantitativos (31%)	26		50	
		36		25	
	2.2.2.2. Características de los diseños cualitativos (22%)	37	31		
	2.2.2.3. Clasificación de los diseños cualitativos (76%)	38	22		
	2.3.1. Desarrollo de operaciones básicas (57%)	10	76		
	68	57			
	9	—	19	49	
Unidad 3: los requisitos protocolarios del proceso de investigación (68%)	3.1.1.3. Los apartados del anteproyecto y tesis (57%)	Redacción académica (68%)	55	57	
	3.2.2. Trabajo de gabinete (fuentes de información y redacción de referencias) (73%)		67	60	
			41	86	
	2	—	3	68	
Métodos en Psicología II (47%)	Unidad 1: El diseño de investigación cuantitativa y cualitativa (45%)	1.1.3. Validez interna y externa de los diseños (36%)	Diseños de investigación	25	33
				32	51
				60	27
				42	35
				43	45
		1.2.2. Características específicas de los diseños cuantitativos (27%)	40	20	
			35	25	
		1.2.3. Características específicas de los diseños cualitativos (42%)	46	22	
			49	24	
			50	63	
	1.2.4. Población y selección de la muestra (61%)	Técnicas de muestreo (62%)	61	79	
			53	63	
			54	77	
			1	40	
			58	54	
			57	66	
	59	52			
	1.3.1. Ética e investigación (41%)	Ética de la investigación (42%)	47	45	
		48	38		
		5	—	19	45
Unidad 2. Recolección y análisis de los datos (50%)	2.1.1. Tipos de herramientas utilizadas en la investigación científica (65%)	Herramientas de recolección de datos (65%)	62	62	
			63	63	
	2.1.2. Desarrollo y piloteo de las herramientas de recolección de datos (64%)		51	56	
			52	78	
		23	85		
		64	44		
	2.3.2. Elección del análisis estadístico (37%)	Análisis de los datos (37%)	65	40	
			39	61	
			44	31	
			45	6	
66			48		
17			36		
	18	35			
	3	—	13	50	
Totales	6	27	—	68	51

 Líneas curriculares e ítems que representan valores extremos de puntuación.

En términos de habilidades y conocimientos, los resultados obtenidos por los estudiantes de sexto, séptimo y octavo semestre de la carrera de Psicología indican, que en términos generales:

- La Comprensión de los Diseños de Investigación es deficiente. Sólo tres de cada diez alumnos tienen una buena probabilidad de resolver reactivos que implican desarrollar una comprensión a profundidad de la operación de los diseños de investigación, principalmente los diseños cuantitativos. Muchos de los estudiantes pueden identificar características y procedimientos generales de los diseños de investigación, sin embargo muy pocos pueden identificar las cuestiones más puntuales y específicas como la simbología de los diseños experimentales y las diferencias entre uno y otro tipo de diseño.
- En cuanto al análisis de los datos, la comprensión y aplicación de las pruebas estadísticas es apenas el mínimo esencial de conocimientos, habilidades y destrezas necesarios para poder seguir progresando satisfactoriamente en el área metodológica de la carrera de Psicología. Solo tres de diez alumnos pude identificar el tipo de prueba estadística, comprender como se aplica, conocer el nivel de medición requerido para efectuar la prueba y aplicar la prueba estadística para la solución de una problemática específica.
- Con respecto a la comprensión del proceso de investigación tanto cuantitativo como cualitativo indica un dominio elemental de conocimientos, habilidades y destrezas. En particular la mayoría de estudiantes tiene dificultades para desarrollar en forma ordenada y organizada los procedimientos necesarios para aplicar el método de investigación, principalmente lo relacionado a las técnicas de muestreo, el desarrollo de herramientas de recolección de datos y el análisis psicométrico de instrumentos.

- La interpretación, comprensión y aplicación de la ética de la investigación presenta un dominio mínimo. Solo cuatro de cada diez estudiantes comprende el juicio ético para determinar si la investigación debe ser llevada a cabo o no. También, la capacidad para comprender las implicaciones de no proceder con ética y de no utilizar el consentimiento informado es deficiente. Solo cuatro de diez estudiantes es capaz de juzgar de forma más o menos adecuada el juicio ético para realizar la investigación científica.
- Por último, los contenidos que obtuvieron los puntajes extremos más altos en la prueba fueron el relacionado al conocimiento de hechos históricos que aportaron a la institucionalización de la actividad investigativa en la psicología y, el relacionado a la comprensión de las nociones generales de los enfoques de investigación. El porcentaje promedio de aciertos obtenido en dichos contenidos fue de 84% y 98% respectivamente. Nótese que los dos contenidos aquí mencionados solo contienen un ítem para su evaluación.

En forma puntual, en el Anexo 6 se muestran los resultados puntuales por ítem. Cada uno de los ítems que se presentan son un ejemplo muestra de los ítems que se presentaron en la evaluación del área metodológica. En dicho anexo también se pueden observar varios recuadros con los siguientes datos:

- Codificación de los contenidos evaluados por asignatura y eje curricular
- Contenidos de formación red curricular.
- Ítem muestra con su respectiva respuesta correcta (*).
- Posición del ítem en el contexto de la prueba.
- Porcentaje de estudiantes que contestaron correctamente el ítem.
- Niveles de logro al que corresponde el ítem.

La socialización de los resultados

En este apartado se presentan las dos últimas actividades y procedimientos del modelo de evaluación. Dichas actividades fueron la elaboración de los reportes de los resultados del aprendizaje logrado por los estudiantes y la socialización de los resultados. En particular, los reportes de los resultados deben ser apropiados a las necesidades de información de los destinatarios. Tales acciones fueron realizadas por los integrantes del Comité Coordinador del ETAM. Como ya se comentó al inicio de la tesis la evaluación del proceso de enseñanza y aprendizaje debe considerarse como una actividad necesaria, en tanto que aporta a los directivos de las instituciones educativas, a los profesores, a los alumnos y a la sociedad en general un mecanismo de realimentación que permite la autorregulación y el conocimiento de los factores que promueven dicho proceso. Dado lo anterior, se puede decir que estas actividades son de gran relevancia en el contexto del estudio.

Para decidir el tipo de información que debería presentar el informe de los resultados se tomó en cuenta las recomendaciones establecidas por Contreras (2009) y Ravela (2006). El informe de los resultados del ETAM fue elaborado con el fin de informar a los directivos de la institución, a los planeadores del currículum y a los encargados de evaluación de la institución que en su totalidad son los mismos profesores de tiempo completo, y a los estudiantes. Es necesario señalar que la FMP es una institución local con un margen de representatividad curricular más o menos reducido en contraste con instituciones que tienen un currículum estatal o nacional y que operan con una cantidad elevada de estudiantes. Por lo anterior no es necesario elaborar reportes a funcionarios públicos como inspectores, cada vez que las funciones propias de estos son realizadas por los mismos directivos de la institución.

En resumen se elaboró un único informe de los resultados del ETAM el cual consistió en los siguientes puntos:

- Promedio global y porcentaje de aciertos por línea de formación para cada uno de los estudiantes evaluados.
- Promedio general y niveles de logro de los estudiantes, por semestre y por grupo.
- Porcentaje general de aciertos por eje y red curricular.
- Promedio de ejecución de estudiantes en grupos de preguntas que evalúan contenidos y nodos claves.
- Indicadores de la calidad técnica de los ítems y del modelo de examen.

Por último, la actividad concerniente a la socialización de los resultados consistió en una serie de pasos. Primero se entregó a los directivos un informe impreso de los resultados del ETAM y se les dio una breve descripción de los mismos. Como segundo paso se entregó a cada uno de los grupos un documento electrónico que contenía el informe de los resultados y de igual manera se les brindó una breve descripción de dichos resultados. El tercer paso consistió en dar a los encargados de evaluación de la FMP el informe de resultados del ETAM y toda la información recopilada y utilizada para operar el modelo de evaluación. De esta manera se ayudó al desarrollo y consolidación de otras estrategias evaluativas. En el siguiente paso se presentó en un seminario ante los académicos de la FMP, la estrategia evaluativa desarrollada para evaluar el Área Metodológica de la Carrera de Psicología y los resultados de la misma. Como último paso, el informe de los resultados se subió a la página de la facultad (<http://medicina.tij.uabc.mx/>) siendo así el primer esfuerzo histórico de la institución a través de medios electrónicos que tiene que ver con la socialización y la rendición de cuentas de los resultados del aprendizaje de sus estudiantes.

V. CONCLUSIONES

Este capítulo presenta las conclusiones del estudio para cada uno de los objetivos principales planteados y para cada una de las etapas del modelo metodológico de evaluación. También presenta las aportaciones significativas de la tesis para la Facultad de Medicina y Psicología, las limitaciones del estudio, las sugerencias para futuras evaluaciones y el desarrollo posterior del ETAM.

Sin duda, cerrar un trabajo realizado durante dos años no es fácil, más aún cuando se trató de la aplicación y adaptación de un modelo metodológico de evaluación para el desarrollo de pruebas nacionales a gran escala. Dicho modelo requirió del desarrollo de procesos complejos los cuales se rigieron por estándares de calidad internacionales. En gran medida la dificultad principal del estudio consistió en las condiciones humildes en las que se aplicó la evaluación y en la adaptación del modelo a un contexto institucional dentro de un plan de estudios en particular.

Para el monitoreo del aprendizaje dentro de la FMP, se requiere del desarrollo de estrategias evaluativas de buena calidad que realimenten la operación del currículum. A pesar de los grandes esfuerzos que realizan los encargados de evaluación de la Facultad de Medicina y Psicología no cuentan con dichas evaluaciones. Por lo tanto, el desarrollo de una evaluación bajo modelos metodológicos consolidados responde a las necesidades institucionales de contar con evaluaciones de buena calidad. De esta manera, con la construcción y aplicación de exámenes desarrollados bajo la lógica anterior, se obtienen evidencias empíricas fiables que realimenten a los involucrados del proceso de enseñanza-aprendizaje sobre la operación del currículum. De ahí el interés en desarrollar un estudio que contribuya a resolver esta problemática.

Este trabajo tuvo dos objetivos principales: a) evaluar el nivel del aprendizaje en el área metodológica de los estudiantes que cursan las etapas disciplinaria y terminal de la carrera de psicología de la Facultad de Medicina y Psicología de la UABC y b) desarrollar y validar un examen de trayecto institucional de referencia criterial alineado al currículum para poder llevar a cabo la evaluación propuesta. En forma más detallada se establecieron cinco objetivos específicos. Durante los siguientes párrafos se da cuenta de los resultados logrados para cada uno de ellos.

De acuerdo con el primer objetivo específico establecido y con los resultados obtenidos de las revisiones sobre los antecedentes la evaluación educativa en México y el desarrollo de exámenes estandarizados, se identificaron las transformaciones conceptuales a lo largo de la historia y, sobre todo, a lo largo del siglo XX. Principalmente la descripción que establecieron Guba y Lincoln (1989) sobre las cuatro generaciones del desarrollo histórico de la evaluación, permitió clarificar los antecedentes de los diferentes enfoques evaluativos en la educación. También, con la revisión se pudo esclarecer el proceso de consolidación de la evaluación estandarizada del aprendizaje a nivel internacional y nacional. En lo particular, se logró identificar diferentes políticas sobre evaluación educativa en el nivel superior, diversos modelos metodológicos de evaluación, y algunos de los organismos más importantes que rigen los estándares de calidad para el desarrollo de pruebas educativas de gran escala.

En particular, sobre el desarrollo de evaluaciones educativas en el nivel superior, se identificaron varios proyectos importantes. Sobre evaluaciones educativas que destacan a nivel internacional se pueden mencionar la *American College Testing* (ACT) o mejor conocida como *Collage Board*, y el *Scholastic Aptitud Test* (SAT), las dos desarrolladas en Estados Unidos. A nivel nacional destacan el Examen de Egreso de Licenciatura (EGEL), desarrollado y aplicado por el Centro Nacional de Evaluación para la Educación Superior, A.C. (CENEVAL). En el plano institucional, se pueden mencionar proyectos trascendentales como El *Examen de habilidades y Conocimientos Básicos* (EXHCOBA), el cual se administra con

propósito de selección para el ingreso a la universidad. Otro proyecto importante es el Examen de Ubicación de Matemáticas (EXUMAT, 2.0) el cual se desarrolló en el Instituto de Investigación y Desarrollo Educativo (IIDE) y fue diseñado por el Dr. Eduardo Backhoff, por el Dr. Lázaro Dibut Toledo y por el L.C.C. José Luis Ramírez. Dicho examen tiene como objetivo identificar los conocimientos y habilidades matemáticas de los estudiantes de los niveles medio-superior y superior.

En el ámbito institucional, se puede mencionar la iniciativa general plasmada en el Plan de Desarrollo Institucional 2003-2006, en donde se promovió el desarrollo de exámenes colegiados. La Coordinación de Formación Básica fue la encargada de la promoción de los exámenes y es actualmente la instancia encargada del diseño de los sistemas de evaluación de la Universidad Autónoma de Baja California. Durante esta revisión puntual, se identificó la normatividad de la FMP sobre las evaluaciones institucionales. Los principales logros de la revisión fueron, la identificación de los requisitos relativos a la evaluación de los procesos de aprendizaje y, la clasificación y requisitos de las evaluaciones de carácter institucional.

Por último, para el logro del primer objetivo específico, en lo que respecta a la revisión de diferentes modelos metodológicos de evaluación, se logró la consolidación de la postura metodológica del presente estudio. Así, con una comprensión más amplia de los modelos de evaluación se pudo tomar la decisión de realizar una adaptación de modelos como el desarrollado por Nitko (1994) para realizar evaluaciones nacionales de referencia criterial y referencia normativa alineada al currículum para la certificación y selección de estudiantes, y el adaptado por Contreras (2009) para el desarrollo de pruebas de referencia criterial alineadas al currículum.

Por otra parte, con relación al segundo objetivo específico, se puede decir que se logró en buenos términos el desarrollo del ETAM. Algunos de los productos resultantes del proceso de desarrollo de la prueba, son la delimitación del contenido a evaluar (Figura 1) y la estrategia de reticulación en donde se condensan cada una

de las decisiones tomadas por los comités desarrolladores de la prueba en cuanto a la selección de contenidos esenciales para su evaluación (ver Figura 2). También la tabla de especificaciones generales y las especificaciones puntuales de los ítems (véase Tabla XIX y Anexo 2). Por último, para el logro del segundo objetivo específico, en la Tabla XX se muestran los resultados del pilotaje de los ítems del ETAM. Dado lo anterior se puede decir que en lo general los indicadores psicométricos del ETAM presentan buena calidad técnica a la luz de los estándares establecidos. Solo el coeficiente de confiabilidad α de Cronbach de 0.70, resultado por debajo del estándar establecido ($\alpha \Rightarrow 0.85$).

Para el logro del tercer objetivo específico, se presentaron varias limitaciones, por lo tanto se puede decir que no se logró cabalmente el objetivo pretendido. Aunque se realizaron diferentes actividades por expertos para la consolidación de la validez del examen, dichas actividades no se realizaron de la forma más óptima. Por ejemplo, no se realizó un procedimiento estandarizado de jueceo por expertos para obtener evidencias empíricas de la validez del ETAM. En cambio, se realizó una revisión cualitativa de la congruencia ítem-representación-curriculum y de la correspondencia del ítem del test con su especificación. Además se realizaron ajustes a los ítems, que consistieron en la corrección de fallas que atentaban a la calidad del contenido y a la calidad técnica de los ítems (Véase Tabla XXII y Anexo 3).

Con respecto al cuarto objetivo específico del estudio, en cuanto a la Identificación desde un modelo simple de evaluación del nivel de aprendizaje en el área metodológica de los estudiantes de la carrera de psicología, es importante mencionar que se logró el objetivo de forma satisfactoria. El ETAM se aplicó a 218 alumnos inscritos oficialmente en los últimos tres semestres (6to, 7mo y 8vo) de la carrera de psicología de la FMP durante el ciclo escolar de enero-septiembre del 2010. Uno de los criterios importantes fue que todos y cada uno de los participantes del estudio tuvieran aprobados todos sus créditos del área metodológica. En general se puede decir que se obtuvieron resultados puntuales de las líneas

curriculares evaluadas. En específico se detectaron los contenidos en donde los alumnos obtuvieron bajos y altos puntajes dentro del área metodológica (véase Tablas de la XXVI a la XXXI). Tales resultados fueron de suma importancia para la realimentación de la operación del currículum de la FMP.

En el quinto y último objetivo específico, se puede mencionar que uno de los mayores logros del presente estudio, fue el de contribuir a la generación de las condiciones necesarias para establecer un proceso permanente de evaluación de la calidad del aprendizaje. Principalmente, el presente proyecto de evaluación contribuyó a la generación de las bases para la aplicación continua del ETAM y en la cada vez mayor consolidación de cuerpos colegiados capacitados para desarrollar estrategias evaluativas de buena calidad. De esta manera se logra generar una mayor cultura de la evaluación educativa y un mecanismo que permita verificar la pertinencia y la calidad de los programas educativos en cuestión.

6.1 Aportaciones de la tesis

A manera de reflexión, se puede decir que la evaluación del aprendizaje para cualquier institución educativa es indispensable, importante e inevitable. La evaluación del proceso de enseñanza y aprendizaje debe considerarse como una actividad necesaria, en tanto que aporta a los directivos de las instituciones educativas, a los profesores, a los alumnos y a la sociedad en general un mecanismo de realimentación sobre la operación del currículum que permite la autorregulación de dicho proceso (Díaz Barriga y Hernández, 2002).

La metodología desarrollada en esta investigación concibe al currículum como el fundamento desde donde se decide qué contenidos son esenciales de evaluar, el procedimiento necesario para evaluar, el desarrollo de las pruebas o instrumentos que se utilizarán para dicha evaluación, la aplicación de los instrumentos de medición, el análisis de los resultados y la socialización de los

mismos. Dicho lo anterior, a continuación se describen las aportaciones del presente estudio evaluativo:

- Análisis, selección y consulta de la literatura relacionada con el desarrollo de evaluaciones educativas en el nivel superior de educación. En este rubro se revisaron algunas experiencias sobre el aprendizaje en el contexto internacional, nacional, e institucional.
- Adaptación de un modelo metodológico de evaluación. En particular se desarrollo una metodología que se puede generalizar a otras áreas disciplinarias dentro de la misma FMP, a otras Unidades Académicas (UA) de la UABC, a los programas de Fortalecimiento de los Procesos y Mecanismos Institucionales de Seguimiento y Evaluación que gestiona la Coordinación de Formación Básica que es la instancia responsable de promover el diseño de un sistema de evaluación en la UABC y a otras posibles Instituciones de Educación Superior del país.
- Diseño y validación de un instrumento para evaluar el área metodológica de la carrera de psicología. El ETAM es un producto novedoso resultado del trabajo colegiado por académicos de la FMP y del IIDE. Dicho examen, es uno de los primeros instrumentos que se aplican en la FMP con el fin de evaluar un área de competencia específica en una etapa intermedia del plan de estudios de la Carrera de Psicología. Además es el único de los exámenes que se aplican en la UA en cuestión que alcanza altos estándares de calidad técnica.
- Evaluación de alta relevancia por estar dirigida a un área del conocimiento esencial dentro del contexto curricular de la carrera de psicología. En específico se ilustró el procedimiento para dar cuenta de la ejecución de los examinados en los ejes, líneas de formación y contenidos específicos en que se clasificó según su relevancia por cuerpos colegiados el contenido del área

metodológica. También se proporcionó información para la toma de decisiones de manera directa a alumnos, profesores y directivos.

- Establecimiento de bases para la aplicación permanente del ETAM. El objetivo es el monitoreo de la calidad del aprendizaje a lo largo del tiempo en el área metodológica de la carrera de psicología de la FMP.
- De manera directa se aportó información a los familiares y a la sociedad en general sobre el nivel de aprendizaje en un área tan relevante para la formación del psicólogo como es el área metodológica.
- Creación de una sinergia en la FMP al involucrar a autoridades y académicos en la evaluación colegiada, lo cual contribuyó en la generación de una cultura evaluativa.

6.2 Limitaciones del estudio

En contraste, se puede decir que el estudio presentó varias limitaciones. La primera de ellas consistió en las dificultades propias de la adaptación y operación de la metodología. Una segunda limitación, consistió en no contar con los suficientes expertos para realizar en forma óptima las diferentes actividades propias del desarrollo de la estrategia evaluativa. Otra de las limitaciones, se basó en la dificultad para lograr cabalmente algunos estándares internacionales para el desarrollo de pruebas educativas, como el proceso estandarizado de validación por jueces.

Como ya se comentó, la principal limitación del estudio recayó en que los diferentes modelos teóricos-metodológicos para la elaboración de exámenes alineados con el currículum, son inicialmente elaborados con el fin de desarrollar instrumentos a gran escala a nivel nacional, y para la evaluación del logro educativo de la educación básica. Por obvias razones, instrumentar el desarrollo de un examen alineado al currículum en el nivel superior, en una institución, en un solo

plantel escolar y con una cantidad reducida de grupos, requirió de un enorme esfuerzo en el trabajo de adaptación de dichos modelos.

Una de las principales dificultades de la adaptación del modelo, se focalizó en el ajuste de ciertos procedimientos estandarizados en donde se requiere idóneamente una considerable cantidad de expertos que apoyen el logro de los objetivos y de los estándares internacionales para el desarrollo de pruebas criteriales a gran escala. Debido a las condiciones modestas en que se realizó el estudio, en la operación del modelo adaptado, la mayoría de dichos procedimientos fueron adscritos al Comité Coordinador del Examen. Como se comentó en el apartado del método de evaluación, dicho comité constaba de una cantidad reducida de expertos. Sin embargo el desarrollo y la aplicación del ETAM, exigió un alto grado de participación y especialización técnica por parte de los cuerpos colegiados para alcanzar los objetivos del estudio. Dicha participación no se efectuó en los mejores términos debido a que la mayoría de los integrantes de los comités eran profesores de tiempo completo con otras tantas actividades a realizar y que carecían de conocimientos especializados necesarios para ciertas actividades del desarrollo de la prueba.

Cabe señalar, que en estos tipos de estudios, es común que existan restricciones asociadas principalmente a los recursos humanos involucrados en los comités técnicos que elaboran los exámenes. La principal razón es que tales limitaciones obedecen a la dificultad para conseguir especialistas en el contenido que tengan disponibilidad de tiempo para involucrarse en actividades sobre el análisis curricular y en la elaboración de especificaciones de ítems (Contreras, 2003). Además, la mayoría de los modelos psicométricos empleados en el desarrollo de pruebas educativas de gran escala, exige una cantidad indispensable de comités para el desarrollo de las mismas, lo cual no siempre se consigue.

Como consecuencia de las limitaciones ya mencionadas, el proceso de validación por jueces, no se realizó en forma estandarizada. Las limitaciones tienen que ver principalmente con la cantidad reducida de jueces y con el uso de formatos

no estandarizados para dichas evaluaciones. La actividad fue realizada en forma austera por tres de los integrantes del Comité Coordinador del Examen. Las actividades realizadas por los jueces con el fin de garantizar la relación ítem-representa-currículum, consistieron principalmente en una revisión general y cualitativa de la congruencia entre los diferentes productos del plan evaluativo desarrollado (ítems, modelos, especificaciones, muestra de contenidos, retícula, currículum), la revisión y corrección general de la calidad del contenido y de la calidad técnica de cada ítem del test. Sin embargo no se realizó ninguna medición objetiva que recabara evidencias empíricas de dicha actividad.

De la misma manera, la cantidad reducida de los integrantes de los comités fue un factor fundamental para la delimitación de los contenidos y de los diferentes aprendizajes a evaluar que establece el enfoque por competencias. En particular, el enfoque educativo por competencias —al cual se adscribe la FMP— propone fomentar en los estudiantes aprendizajes de conocimiento, de reflexión crítica, meta-cognitivos, procedimentales y actitudinales. De esta manera, es necesario que una estrategia evaluativa integral considere cada uno de los objetivos de aprendizaje establecidos en el currículum oficial. Sin embargo, en el presente estudio, no se pudo contar con expertos en el manejo de contenidos según el tipo de aprendizaje y el nivel taxonómico, por lo que solo se evaluaron los niveles de conocimiento, comprensión y aplicación del área metodológica y quedó pendiente la evaluación de aprendizajes de tipo procedimental y de opinión.

6.3 Futuras investigaciones

Finalmente, se proponen recomendaciones para futuras investigaciones que permitan el seguimiento y desarrollo de estrategias evaluativas de buena calidad. Principalmente para su aplicación en el nivel superior, en donde se requieren un sin fin de mejoras en el rubro de la evaluación educativa:

- Pilotear y aplicar dos o más versiones del ETAM con el fin de maximizar y aprovechar los productos logrados por los cuerpos colegiados encargados del desarrollo de la estrategia evaluativa.
- Mantener el ETAM en un proceso de mejora continua tras su aplicación. Para efectuar la consiguiente revisión y mejoramiento de los ítems y modelos.
- Realizar análisis de los indicadores psicométricos del examen agregando el modelamiento Rasch y el análisis gráfico de ítems.
- Recoger evidencias de validación por juicio de expertos realizando un proceso estandarizado y riguroso.
- Realizar una estrategia evaluativa integral incluyendo cuestionarios de contexto y tests para evaluar la opinión de los alumnos y el aprendizaje de procedimientos del área metodológica, tal y como lo establece el plan de estudios por competencias de la institución.
- Incorporar otros indicadores psicométricos del examen, a la luz de los estándares internacionales para el desarrollo de test educativos.
- Exponer el análisis y revisión de los resultados a los destinatarios clave, de tal forma que el seguimiento de la evaluación y la actualización del ETAM por parte de los académicos del FMP mantenga la calidad de la estrategia evaluativa.
- Realizar una réplica de este estudio utilizando para ello dos o más versiones del examen.

Además, es recomendable considerar en futuras investigaciones:

- Socializar los resultados a través de foros, talleres, seminarios y en general en eventos académicos de tipo presencial y virtual.

- Consolidar cuerpos académicos capacitados para el desarrollo de estrategias evaluativas integrales.
- Extender la metodología desarrollada en el presente trabajo, a otras áreas académicas dentro de la FMP y a otras unidades académicas de la UABC con características similares, que están desarrollando la evaluación colegiada del aprendizaje.

REFERENCIAS BIBLIOGRAFICAS

- Aiken, L.R. (1996). *Tests psicológicos y evaluación*. México: Prentice Hall Hispanoamericana.
- Alvarado, F. y M. Robredo (1984). "La reticulación como una aplicación de la teoría de sistemas, en la elaboración de planes y programas de estudio". *Revista de la educación superior*, 52, 71-82.
- Álvarez, I. (1997). *Modelo para una evaluación integral de las políticas sobre gestión de calidad en la educación Superior*. Gestión y Estrategia. UAM. Nos. 11 y 12.
- Alvira, F. (1985). La investigación evaluativa: una perspectiva experimentalista. *Revista Española de Investigaciones Sociológicas*, núm. 29.
- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (1999). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.
- Antillón, M. (2003). *Igualación de Tres Versiones del Examen de Habilidades y Conocimientos Básicos (EXHCOBA)*. Tesis de maestría publicada en línea. Universidad Autónoma de Baja California, Ensenada, México. Consultada el 8 de noviembre de 2008 en: <http://148.231.200.29/blogs/mce/files/2010/09/Luz-Elena-Antillon-Macias.pdf>
- ANUIES, (2000). *La educación superior en el siglo XXI. Líneas estratégicas de desarrollo. Una propuesta de la ANUIES*. México: Autor.
- Backhoff, E. (2006). Exámenes de la Calidad y el Logro Educativos (Excale): nueva generación de pruebas nacionales. Instituto Nacional para la Evaluación de la Educación. Cuaderno de Investigación No 7. México: INEE.
- Backhoff, E. (2001). *Desarrollo, validación e implementación de un sistema para la selección de estudiantes a la Universidad Autónoma de Baja California*. Aguascalientes: Universidad Autónoma de Aguascalientes.
- Backhoff, E. y Tirado, F. (1992). *Desarrollo del Examen de Habilidades y Conocimientos Básicos (EXHCOBA)*. *Revista de la Educación Superior*, (83), 95-117.
- Backhoff, E., Larrazolo, N. y Rosas, M. (2000). Nivel de dificultad y poder de discriminación del Examen de Habilidades y Conocimientos Básicos

- (EXHCOBA). *Revista Electrónica de Investigación Educativa*, 2 (1). Consultado el día 24 de marzo de 2010 en el world wide web: <http://redie.ens.uabc.mx/vol2no1/contenido-backhoff.html>
- Backhoff, E., A. Sánchez, M. Peón, L. Monroy y M.L. Tanamachi (2006). "Diseño y desarrollo de los exámenes de la calidad y el logro educativos". *Revista Mexicana de Investigación Educativa*, 11(29), 617-638.
- Cabrera, F. (1986). *Proyecto docente sobre técnicas de medición y evaluación educativas*. Barcelona: Universidad de Barcelona.
- Chacon, S., Anguera, M. T. y Lopez, J. (2000). Diseños de evaluación de programas: Bases metodológicas. *Psicothema*, ano/vol.12, Oviedo: Universidad de Oviedo. Pp.127-131.:<http://redalyc.uaemex.mx/pdf/727/72797033.pdf>
- Carrión (1991). Evaluación educacional: Panorama actual. En Secretaría de Educación Pública. Evaluación educacional (Antología): Vol. 1. *Fundamentos Teóricos de la Evaluación*, 3ra., reimpresión. México: Secretaría de Educación Pública.
- CENEVAL (s.f.a). *Acerca del CENEVAL y los exámenes generales para el egreso de la licenciatura*. México: autor.
- CENEVAL (s.f.b). Exámenes aplicados mayo-agosto 2008. *El papel de la evaluación, Boletín CENEVAL*. Recuperado el 20 de marzo de 2010, de: http://archivos.ceneval.edu.mx/archivos_portal/2624/NEnumero19.pdf.pdf
- CENEVAL (2008). Metodología para la elaboración de instrumentos de medición. *El papel de la evaluación, Boletín CENEVAL*. Recuperado el 20 de marzo de 2010, de: http://archivos.ceneval.edu.mx/archivos_portal/2627/NEnumero17.pdf.pdf
- CIEES, (2000). Marco de referencia para la evaluación CIEES, *Comité de Ciencias de la Salud* (2ª. ed.).México: CIEES.
- Contreras, L. (2000) *Desarrollo y pilotaje de un examen de español para la educación primaria en Baja California*. Tesis de maestría publicada en línea. Universidad Autónoma de Baja California, Ensenada, México. Consultada el 24 octubre de 2008 en: <http://148.231.200.29/blogs/mce/files/2010/09/Luis-Angel-Contreras-Nino.pdf>
- Contreras, L.A. (2009). Manual para el desarrollo y elaboración exámenes criterios de gran escala alineados con el currículum. Ensenada: UEE.
- Cook, T. D. y Reichardt, C. (1997). *Métodos cualitativos y cuantitativos en investigación evaluativa*. Madrid: Ediciones Morata.

-
- COPAES (2003). *Antecedentes, situación actual y perspectivas de la evaluación y acreditación de la educación superior en México*. Documento para IESALC-UNESCO. México: Autor
- CRESST (1994). *Assesmtment Profile-State Summary. Evaluation Comment*. National Center for Research on Evaluation, Standars and Student Testing.
- Crocker, L. y Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. Holt, Rinehart and Winston.
- Díaz, A. (2006). Las pruebas masivas. Análisis de sus diferencias técnicas. COMIE. *Revista Mexicana de Investigación Educativa*, 11 (29), 583-615.
- Díaz Barriga, F. y Hernández, G. (2002). *Estrategias docentes para un aprendizaje significativo. Una interpretación constructivista* (2ª. ed.). México: McGraw Hill.
- Downing, S. M. y Haladyna, T. M. (Eds.) (2006). *Handbook of test development*. Mahwah, NJ: Erlbaum.
- Encinas, J. A., Rivera, R. E. y Contreras, L. A. (2005). *Evaluación colegiada del aprendizaje en la Universidad Autónoma de Baja California: Construcción de un examen criteria de gran escala para evaluar el dominio de conceptos y procedimientos del Calculo Diferencial*. Ponencia en el VIII Congreso Nacional de Investigación Educativa. Hermosillo, Sonora. Octubre-Noviembre 2005.
- Escudero, T. (1997). Enfoques modélicos y estratégicos de evaluación de centros educativos. *Revista Electrónica de Investigación y Evaluación Educativa*, 3(1). Documento en línea, recuperado en: http://ww.uv.es/RELIEVEv3n1_1.htm.
- Escudero, T. (2003). Desde los tests hasta la investigación evaluativa actual. Un siglo, el XX, de intenso desarrollo de la evaluación en educación. *RELIEVE*, 9 (1), 11-43. Documento en línea, recuperado en: http://www.uv.es/RELIEVE/v9n1/RELIEVEv9n1_1.htm
- Franco, R. (1988). *Algunas reflexiones sobre la evaluación del desarrollo*. Santiago de Chile: ILPES.
- Gago, A. y Mercado, R. (1995). La evaluación en la educación superior mexicana. *Revista de Educación Superior*, 14 (96).
- Gago, A. (1998). Acceso a la Universidad: El caso mexicano. La Universidad. *Boletín Informativo de la Secretaría de Políticas Universitarias SPU*, 14. Consultado el 17 de noviembre de 2009 en: <http://www.ses.me.gov.ar/documentos/publicaciones/lu14.pdf>.

-
- Gago, A. (2000). El CENEVAL y la evaluación externa de la educación en México. *Revista Electrónica de Investigación Educativa*, 2 (2). Consultado el 3 de octubre de 2009 en: <http://redie.ens.uabc.mx/vol2no2/contenido-gago.html>.
- Galli, A. (1991). Evaluación educacional: panorama actual. En Secretaría de Educación Pública. *Evaluación educacional (Antología): Vol. 1. Fundamentos Teóricos de la Evaluación*, 3ra., reimpresión, pp. 27-29. México: Secretaría de Educación Pública (SEP).
- Guba, E.G. y Lincoln, Y.S. 1989. *Fourth Generation Evaluation*. Newbury Park: Sage.
- Steven M. Downing and Thomas M. Haladyna (Eds.) (2006). *Handbook of test development*. Mahwah, NJ: Erlbaum.
- Haladyna, T. M. (2004). *Developing and validating multiple-choice test items* (3rd). Mahwah, NJ: Lawrence Erlbaum Associates.
- INEE (2006). *El aprendizaje del español, las matemáticas y la expresión escrita en la educación básica en México: Sexto de primaria y tercero de secundaria*. Resumen ejecutivo. México: Autor
- Jornet, J. M. y Backhoff, E. (2008). Modelo para la determinación de niveles de logro y puntos de corte de los exámenes de la calidad y el logro educativos (Excale). *Colección Cuadernos de Investigación*, 30. México D.F.: INEE.
- Jornet, J. M. y Gonzalez, J. (2009). *Evaluación criterial: determinación de estándares de interpretación (EE) para pruebas de rendimiento educativo*. España: Servicios de Publicaciones de la Universidad de Valencia
- Kaufman, R. y English, F. (1979). *Needs Assessment. Concept and Application*. Englewood Cliffs: Educational Technology Pub.
- Leyva, Y. E. (2010). La evaluación como recurso estratégico para la mejora de la práctica docente ante los retos de una educación basada en competencias. *Revista Iberoamericana de Evaluación Educativa*, 3 (1) .Consultada el 19 de septiembre de 2010 en: http://www.rinace.net/riee/numeros/vol3-num1_e/art18.pdf
- Linn, R. L. (1994). *The likely impact of performance standards as a function of uses: From rhetoric to sanctions*. En *Proceedings of the Joint Conference on Standard Setting for Large- Scale Assessments* (pp. 267-276). Washington, DC.
- Livas, G. (1988). *Análisis e interpretación de los resultados de la evaluación educativa*. (2ª. ed.). México: Trillas.

-
- Llarena, R. (1994). La evaluación de la educación superior en México. *Revista de Educación Superior*, (89), 36-56.
- Martínez, F. (2001). Evaluación educativa y pruebas estandarizadas. Elementos para enriquecer el debate". *Revista de Educación Superior*, 30 (120), 71-85.
- Martínez, F. (2003). Pruebas y rendición de cuentas. *Cuaderno No. 12*. México: INEE. Consultado el 28 de febrero de 2009 en: <http://www.inee.edu.mx/index.php/publicaciones/resultados-de-investigacion/cuadernos-de-investigacion/3708>.
- Martínez, F. (2006). Pruebas de selección y pruebas para evaluar escuelas. Nuevas consideraciones sobre su uso y la difusión de resultados. *Cuadernos de Investigación*, N° 21. México, INEE.
- Martínez, M. A. (2008). *Informe de actividades Enero-diciembre de 2008/Facultad de Ingeniería Campus Mexicali de la Universidad Autónoma de Baja California*. Mexicali: Autor
- McReynold, P. (1975). *Advances in Psychological Assessment, vol. III*. San Francisco: Jossey-Bass.
- Melia, J. L. (1990). *La construcción de la Psicometría como ciencia teórica y aplicada*. Valencia: Cristobal Serrano.
- Nitko, A. (1994, julio). A model for developing curriculum-driven criterion-referenced and norm-referenced national examinations for certification and selection of students. *Ponencia presentada en la conferencia internacional sobre evaluación y medición educativas, de la Asociación para el Estudio de la Evaluación Educativa en Sudáfrica (ASSESA)*.
- Nunnally, J. C. y Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). Nueva York: McGraw-hill.
- Popham, W. (1990). *Modern educational measurement: practical guidelines for educational leaders*. (2nd Edition). Needham Heights, MA: Allyn y Bacon (402p)
- Popham, W, (1993). *Educational evaluation*. 3ra. Ed. Nueva York: Allyn and Bacon.
- Ravela, P. (2000): *Para comprender las evaluaciones educativas. Fichas didácticas*, Santiago de Chile: PREAL.
- Ravela, P. (2006). *"Fichas didácticas para comprender la evaluaciones educativas"*, PREAL. Chile: Ed. San Marino.

-
- Ravela, P., Arregui, P. Valverde, G., Wolfe, R., Ferrer, G., Martínez, F., Aylwin, M. y Wolff, L. (2008). Las Evaluaciones Educativas que América Latina Necesita. *Documento de trabajo N. 40*. Santiago de Chile: PREAL.
- Ross, K. N., y Makuwa, D. (2009). *What is SACMEQ?* Paris: UNESCO-IIEP.
- Rosario, V. M., Marúm, E., Vargas, R., Arroyo, J. y González, V. (2006). Acreditación y certificación de la educación superior: *Experiencias, realidades y retos para las IES*. México: Autor
- Rutman, L. y Mowbray, G. (1983). *Understanding program evaluation*. Beverly Hills, Ca.: Sage
- Salvador, L. (1992). *Proyecto docente. Universidad de Cantabria*. Santander: Autor
- Scriven, M (1967). *The Methodology of Evaluation, en Perspectives on Curriculum Evaluation*. Chicago: Rand McNally.
- Scriven, M. (1994). Evaluation as a discipline. *Studies in Educational Evaluation*, 20, 1, 147-166.
- SEP, (2007). Programa Sectorial de Educación 2007-2012. México, D. F.: *Diario Oficial del la Federación*. Consultado el 22 de agosto de 2010 en: <http://www.sep.gob.mx/work/apps/site/programa/plan.zip>
- Stufflebeam, D. y Shinkfield, A. J. (1987). *Evaluación sistemática: guía teórica y práctica*. Madrid: Paidós/MEC.
- Tyler, R. W. (1950, 1967). *Basic principles of curriculum and instruction*. Chicago: University of Chicago Press.
- Tunnermann, C. (2003). La universidad ante los retos del siglo XXI. Mérida, Yucatán, México: Universidad Autónoma de Yucatán.
- UABC, (2006). Estatuto escolar de la Universidad Autónoma de Baja California. Mexicali: Autor
- UABC, (2007). *Plan de Desarrollo Institucional 2007-2010*. Universidad Autónoma de Baja California. Mexicali: Autor.
- Weiss, C. (1987). *Investigación Evaluativa, métodos para determinar la eficiencia de los programas de acción*. México: Editorial Trillas.

Anexo 1

Ejemplos de justificaciones de la importancia de los contenidos para la elaboración de ítems.

Tabla de justificaciones de la importancia de los contenidos			
Área metodológica			
Contenido a evaluar en el examen	Relevancia asignada	Razones que justifican la decisión: <ul style="list-style-type: none"> • Tipos de relaciones con los contenidos curriculares • Objetivo del contenido • Tipo de relevancia del contenido curricular y números de ítems • Puntos específicos de atención 	Observaciones
Métodos de investigación en psicología I	/		
Unidad 1. La investigación en psicología.	/		
Tema: 1.1. - La tradición investigativa en psicología	/		
Subtema: 1.1.1 La lógica de la investigación	/		
Picotema: 1.1.1.1 Corrientes epistemológicas	Esencial	Es un contenido fuente que proporciona cinco servicios. Es esencial para los aprendizajes posteriores por su naturaleza pedagógica. Por ser un contenido esencial se elaboraran dos especificaciones, de la primera especificación se generara un ítem atendiendo en específico el concepto de epistemología. De la segunda especificación se generaran dos ítems que deberán atender la identificación y discriminación entre diferentes corrientes epistemológicas a través sus características generales.	Seleccionado por censo
Picotema: 1.1.1.3 La institucionalización de la actividad investigativa en psicología	Esencial	Es un contenido base que no recibe ningún servicio y proporciona cinco. Por ser un contenido esencial se elabora 2 especificaciones para dos ítems. Comprender el proceso de institucionalización de la actividad investigativa en psicología. Las especificaciones atenderán a identificar las diferentes etapas del proceso de institucionalización del trabajo investigativo.	Seleccionado por censo

Anexo 2

Ejemplo de una especificación puntual de ítem.



Universidad Autónoma de Baja California
Facultad de Medicina y Psicología



Especificaciones puntuales de los ítems del Examen de Trayecto del Área Metodológica (ETAM) de la Carrera de Psicología.

Elaborado por: Juan Carlos Pérez Moran	Fecha: 11 de noviembre de 2009
Revisada por: Norma Larrazolo Reyna	Fecha: Enero de 2009
No. de especificación: 1	No. identificación del reactivo: MPI4DMCC111.1, MPI4DMCC111.2
I. datos de identificación del reactivo	
Asignatura: Métodos en Psicología I	Semestre y etapa: 4to, Disciplinaria
área específica de conocimiento: Métodos cuantitativos y cualitativos	Nombre de la unidad: La investigación en psicología
Competencia de la Unidad: Explicar el significado, las expectativas y la formalización de la investigación científica en forma crítica. Extrapolar el significado, las expectativas y la formalización de la investigación a la realidad científica y profesional de la disciplina psicológica.	
Contenido específico a evaluar: 1.1.1. Corrientes epistemológicas	
Relevancia asignada al contenido: Esencial (X) importante () baja importancia ()	
II. Descripción general del contenido	
Interpretación y delimitación del sentido del contenido: Para lograr un entendimiento amplio de las diferentes visiones o formas de ver la realidad desde un enfoque científico, es necesario que el alumno desarrolle una lógica científica sobre la actividad investigativa. Por tal motivo es necesario que en primera instancia sepa definir el concepto de Epistemología. También es importante que reconozca diferentes corrientes epistemológicas relevantes para su objeto de estudio, entre ellas el idealismo-materialismo, empirismo-racionalismo; la noción epistemología interpretacionismo: hermenéutica, fenomenología, naturalismo, constructivismo, psicoanálisis; y la noción epistemología explicacionista: positivismo, postpositivismo, neopositivismo, pragmatismo. Por último también es necesario que conozca características generales de las diferentes corrientes epistemológicas.	

Tipo y nivel de conocimiento o habilidad involucrada: Comprender y explicar ideas y conceptos

III. Plantillas de los reactivos

Plantilla 1	Base del reactivo: Rama de la filosofía cuyo objeto de estudio es el conocimiento científico. Se le conoce también como Teoría del conocimiento.
--------------------	--

Vocabulario e información textual, gráfica o tabular a emplear: Ninguna

Opción correcta: d) Epistemología

Distractores.

Opción incorrecta 1: a) Fenomenología

Opción incorrecta 2: b) Ontología

Opción incorrecta 3: c) Metodología

Plantilla 2	Base del reactivo: Esta corriente epistemológica sostiene que sólo lo material existe independientemente de nuestra conciencia, y que ésta es un fenómeno derivado de procesos objetivos que afectan a la materia. Uno de sus grandes representantes del siglo XIX es Karl Marx.
--------------------	--

Vocabulario e información textual, gráfica o tabular a emplear:

Opción correcta: b) Materialismo

Distractores.

Opción incorrecta 1: a) Positivismo

Opción incorrecta 2: c) Racionalismo

Opción incorrecta 3: d) Empirismo

Bibliografía consultada:

1. Hernández, R. (2006) *Metodología de la investigación*. (4ª ed.) México, D.F.: Edit. Mc Graw Hill.
2. Rojas, R. (1994) *Guía para realizar investigaciones sociales*. (14ª ed.) México D.F.: Edit. P y V,
3. Shaughnessy, J., Zechmeister, E. & Zechmeister, J. (2007) *Métodos de investigación en psicología*. (2ª ed.) México D.F.: Edit. Mc Graw Hill,

Comentarios de revisión:

IV. Reactivos elaborados para la especificación MPI4DMCC111.1, MPI4DMCC111.2

4.1. Reactivos elaborados para la plantilla 1(MPI4DMCC111.1)

Ítem muestra 1:

Se ocupa de problemas tales como las circunstancias históricas, psicológicas y sociológicas que llevan a la obtención del conocimiento científico, y los criterios por los cuales se lo justifica o invalida.

- a) Fenomenología b) Ontología c) Metodología d) Epistemología

Ítem muestra 2:

Rama de la filosofía cuyo objeto de estudio es el conocimiento científico. Se le conoce también como Teoría del conocimiento.

- a) Fenomenología b) Ontología c) Metodología d) Epistemología

4.2. Reactivos elaborados para la plantilla 2 (MPI4DMCC111.2)

Ítem muestra 3:

Esta corriente epistemológica sostiene que la realidad extramental no es cognoscible tal como es en sí misma, y que el objeto del conocimiento está preformado o *construido* por la actividad cognoscitiva. Se le considera en contraposición al materialismo.

- a) Fenomenología b) Constructivismo c) Racionalismo d) Idealismo

Ítem muestra 4:

Enfatiza el papel de la experiencia, ligada a la percepción sensorial, en la formación de los conceptos y justifica la posibilidad de la validez de las teorías a través del experimento, en relación con la experiencia posible. Se considera en contraposición al llamado racionalismo.

- a) Fenomenología b) Materialismo c) Positivismo d) Empirismo

Ítem muestra 5:

Es una corriente filosófica que acentúa el papel de la razón en la adquisición del conocimiento y que apareció en Francia en el siglo XVII, formulada por René Descartes.

- a) Interpretacionismo b) Idealismo c) Racionalismo d) Cognoscitivismo

Ítem muestra 6:

Es una corriente o escuela filosófica que afirma que el único conocimiento auténtico es el conocimiento científico, y que tal conocimiento solamente puede surgir de la afirmación positiva de las teorías a través del método científico. Apareció en Francia a inicios del siglo XIX de la mano del pensador francés Augusto Comte.

- a) Positivismo b) Materialismo c) Fenomenología d) Empirismo

Ítem muestra 7:

Es una corriente epistemológica que analiza y estudia los fenómenos lanzados a la conciencia, es decir, las esencias de las cosas. Edmund Husserl fue considerado el padre de esta disciplina.

- a) Interpretacionismo b) Idealismo c) Fenomenología d) Constructivismo

Anexo 3

Ajustes de ítems

No.	Calidad técnica general del ítem				Calidad técnica por distractor					Tipo de falla	Recomendaciones sobre el ítem
	P	D	Bis.		Prop. Total	Bajos	Altos	Bis.	Clave		
1	0.55	0.36	0.36	a	0.02	0.00	0.00	-0.21		Ninguna	
				b	0.55	0.33	0.69	0.36	*		
				c	0.36	0.50	0.31	-0.11			
				d	0.06	0.17	0.00	-0.73			
2	0.71	0.16	0.17	a	0.11	0.12	0.15	0.04		Discriminación errónea	Cambiar la opción que atrae a los que saben.
				b	0.71	0.58	0.74	0.07	*		
				c	0.03	0.04	0.00	-0.22			
				d	0.16	0.27	0.11	-0.22			
3	0.75	0.05	-0.05	a	0.13	0.15	0.07	-0.12		Complejidad cognitiva (fácil) y discriminación (bajos-altos) baja y discriminación (rbis). Negativa.	Hacer más o menos atractivas las opciones o sustituirlas. Cambiar la opción que atrae a los que saben.
				b	0.12	0.08	0.11	0.21			
				c	0.75	0.77	0.81	-0.05	*		
				d	0.00	0.00	0.00	0.00			
4	0.86	0.33	0.55	a	0.03	0.11	0.00	-0.75		Ninguna	
				b	0.08	0.11	0.00	-0.22			
				c	0.86	0.67	1.00	0.55	*		
				d	0.03	0.11	0.00	-0.62			
5	0.88	0.08	0.18	a	0.09	0.12	0.04	-0.14		Complejidad cognitiva (fácil), discriminación (bajos-altos) y discriminación (rbis) baja.	Hacer más o menos atractivas las opciones o sustituirlas. Cambiar la opción que atrae a los que saben, cambiar distractores poco elegidos.
				b	0.88	0.85	0.93	0.18	*		
				c	0.01	0.04	0.00	-0.84			
				d	0.01	0.00	0.04	0.45	?		
6	0.70	0.34	0.38	a	0.05	0.10	0.00	-0.47		Discriminación (rbis) errónea en un distractor	Cambiar distractores poco elegidos.
				b	0.00	0.00	0.00				
				c	0.26	0.35	0.12	-0.27			
				d	0.70	0.55	0.88	0.38	*		
7	0.47	0.53	0.55	a	0.02	0.08	0.00	-1.00		Ninguna	
				b	0.47	0.17	0.69	0.55	*		
				c	0.47	0.75	0.31	-0.41			
				d	0.04	0.00	0.00	0.04			
8	0.45	0.29	0.37	a	0.36	0.54	0.30	-0.30		Discriminación (rbis) errónea en un distractor	Cuidadoso análisis del contenido del ítem y, en su caso, redacción más categórica, y cambiar la opción que parece está creando el problema.
				b	0.18	0.19	0.11	-0.17			
				c	0.01	0.00	0.04	0.30			
				d	0.45	0.27	0.56	0.37	*		

9	0.55	0.43	0.50	a	0.11	0.15	0.04	-0.34		Ninguna	
				b	0.13	0.23	0.07	-0.27			
				c	0.21	0.27	0.11	-0.28			
				d	0.55	0.35	0.78	0.50	*		
10	0.64	0.47	0.47	a	0.11	0.15	0.04	-0.32		Ninguna	
				b	0.64	0.42	0.89	0.47	*		
				c	0.03	0.04	0.00	-0.13			
				d	0.22	0.38	0.07	-0.36			
11	0.70	0.35	0.46	a	0.05	0.08	0.04	-0.30		Ninguna	
				b	0.21	0.35	0.07	-0.40			
				c	0.70	0.50	0.85	0.46	*		
				d	0.04	0.08	0.04	-0.16			
12	0.08	0.00	-0.05	a	0.08	0.08	0.07	-0.05	*	Complejidad cognitiva (difícil), discriminación (bajos-altos) baja y discriminación (rbis).negativa.	Hacer más o menos atractivas las opciones o sustituirlas, Cambiar la opción que atrae a los que saben.
				b	0.21	0.08	0.37	0.23	?		
				c	0.25	0.23	0.26	0.20			
				d	0.46	0.62	0.30	-0.31			
13	0.18	0.37	0.47	a	0.45	0.54	0.26	-0.28		Complejidad cognitiva (difícil), discriminación errónea en un distractor, y redacción confusa	Cuidadoso análisis del contenido del ítem y, en su caso, redacción más categórica, y cambiar la opción que parece está creando el problema.
				b	0.29	0.31	0.22	-0.06			
				c	0.18	0.08	0.44	0.47	*		
				d	0.08	0.08	0.07	0.06			
14	0.46	0.24	0.31	a	0.46	0.42	0.67	0.31	*	Ninguna	
				b	0.08	0.08	0.00	-0.25			
				c	0.29	0.27	0.22	-0.02			
				d	0.17	0.23	0.11	-0.30			
15	0.66	0.35	0.26	a	0.06	0.17	0.00	-0.54		Discriminación errónea del distractor y, edición y redacción confusa de la base	Cambiar el distractor que atrae a los altos y mejorar la redacción de la base
				b	0.19	0.00	0.15	0.24			
				c	0.66	0.50	0.85	0.26	*		
				d	0.09	0.33	0.00	-0.60			
16	0.62	0.31	0.42	a	0.25	0.35	0.11	-0.32		Ninguna	
				b	0.09	0.15	0.07	-0.34			
				c	0.04	0.00	0.00	-0.04			
				d	0.62	0.50	0.81	0.42	*		

17	0.36	0.62	0.64	a	0.11	0.11	0.08	-0.23		Ninguna	
				b	0.19	0.33	0.15	-0.25			
				c	0.33	0.56	0.15	-0.35			
				d	0.36	0.00	0.62	0.64	*		
18	0.50	0.47	0.53	a	0.31	0.56	0.23	-0.37		Ninguna	
				b	0.14	0.11	0.08	-0.17			
				c	0.50	0.22	0.69	0.53	*		
				d	0.06	0.11	0.00	-0.39			
19	0.39	0.17	0.21	a	0.39	0.31	0.48	0.21	*	Discriminación (bajos-altos) baja y redacción confusa.	Cuidadoso análisis del contenido del ítem y, en su caso, redacción más categórica. Cambiar la opción que atrae a los que saben.
				b	0.09	0.15	0.04	-0.10			
				c	0.21	0.23	0.19	-0.14			
				d	0.30	0.31	0.30	-0.06			
20	0.76	0.24	0.44	a	0.08	0.19	0.04	-0.57		Discriminación errónea en un distractor	Cuidadoso análisis del contenido del ítem y, en su caso, redacción más categórica. Cambiar la opción que atrae a los que saben.
				b	0.08	0.12	0.04	-0.38			
				c	0.76	0.65	0.89	0.44	*		
				d	0.08	0.04	0.04	0.02			
21	0.47	0.36	0.37	a	0.08	0.04	0.04	-0.01		Ninguna	
				b	0.24	0.23	0.15	-0.07			
				c	0.47	0.35	0.70	0.37	*		
				d	0.21	0.38	0.11	-0.44			
22	0.78	0.43	0.58	a	0.08	0.12	0.04	-0.06		Ninguna	
				b	0.09	0.23	0.04	-0.52			
				c	0.78	0.50	0.93	0.58	*		
				d	0.05	0.15	0.00	-0.73			
23	0.57	0.36	0.32	a	0.23	0.42	0.15	-0.24		Discriminación errónea en distractor d).	Agregar un distractor más plausible
				b	0.19	0.25	0.15	-0.20			
				c	0.57	0.33	0.69	0.32	*		
				d	0.00	0.00	0.00				
24	0.89	0.19	0.49	a	0.00	0.00	0.00	0.00		Discriminación (bajos-altos) baja.	Cuidadoso análisis del contenido del ítem y, en su caso, redacción más categórica, y cambiar la opción que parece está creando el problema.
				b	0.89	0.77	0.96	0.49	*		
				c	0.03	0.08	0.00	-0.68			
				d	0.08	0.15	0.04	-0.32			
25	0.68	0.59	0.57	a	0.02	0.08	0.00	-0.58		Ninguna	
				b	0.68	0.33	0.92	0.57	*		
				c	0.19	0.33	0.08	-0.31			
				d	0.11	0.25	0.00	-0.50			

26	0.60	0.52	0.53	a	0.06	0.25	0.00	-1.00		Edición y redacción confusa en distractores	Cambiar el distractor que atrae a los altos y mejorar la redacción de la base
				b	0.32	0.42	0.23	-0.11			
				c	0.02	0.08	0.00	-0.58			
				d	0.60	0.25	0.77	0.53	*		
27	0.80	0.16	0.25	a	0.01	0.04	0.00	-0.39		Discriminación (bajos-altos) baja.	Cuidadoso análisis del contenido del ítem y, en su caso, redacción más categórica, hacer más o menos atractivas las opciones o sustituirlas.
				b	0.16	0.15	0.07	-0.16			
				c	0.80	0.77	0.93	0.25	*		
				d	0.03	0.04	0.00	-0.26			
28	0.51	0.29	0.23	a	0.05	0.10	0.04	-0.21		Ninguna	
				b	0.35	0.42	0.38	0.02			
				c	0.51	0.29	0.58	0.23	*		
				d	0.09	0.19	0.00	-0.48			
29	0.36	0.33	0.31	a	0.12	0.23	0.07	-0.21		Discriminación errónea en un distractor	Cambiar la opción que parece está creando el problema.
				b	0.28	0.38	0.19	-0.29			
				c	0.36	0.19	0.52	0.31	*		
				d	0.24	0.19	0.22	0.05			
30	0.76	0.27	0.47	a	0.17	0.27	0.11	-0.41		Ninguna	
				b	0.01	0.04	0.00	-0.46			
				c	0.76	0.58	0.85	0.47	*		
				d	0.05	0.12	0.04	-0.25			
31	0.16	-0.01	0.09	a	0.14	0.19	0.15	-0.25		Discriminación (bajos-altos) negativa, discriminación baja (rbis) y dificultad cognitiva alta	Cambiar la opción que atrae a los que saben, cambiar distractores poco elegidos.
				b	0.41	0.38	0.33	0.11	?		
				c	0.16	0.15	0.15	0.09	*		
				d	0.29	0.27	0.37	-0.02			
32	0.56	0.58	0.67	a	0.08	0.11	0.08	-0.25		Ninguna	
				b	0.03	0.00	0.00	-0.07			
				c	0.33	0.78	0.23	-0.61			
				d	0.56	0.11	0.69	0.67	*		
33	0.42	0.44	0.42	a	0.21	0.27	0.15	-0.09		Discriminación errónea en un distractor.	Cambiar la opción que atrae a los que saben.
				b	0.32	0.50	0.15	-0.41			
				c	0.42	0.19	0.63	0.42	*		
				d	0.05	0.04	0.07	0.09			
34	0.11	0.15	0.45	a	0.12	0.15	0.15	0.00		Discriminación (bajos-altos) errónea.	Cuidadoso análisis del contenido del ítem y, en su caso cambiar la opción que parece está creando el problema.
				b	0.11	0.08	0.11	-0.01			
				c	0.67	0.73	0.56	-0.22			
				d	0.11	0.04	0.19	0.45	*		

35	0.77	0.18	0.41	a	0.11	0.17	0.00	-0.55		Discriminación (bajos-altos) errónea, edición y redacción confusa de los distractores.	Cuidadoso análisis del contenido del ítem y sustituir la opción que tiene el problema.
				b	0.06	0.17	0.08	-0.43			
				c	0.77	0.67	0.85	0.41	*		
				d	0.06	0.00	0.08	0.23			
36	0.47	0.12	0.20	a	0.21	0.08	0.31	0.19		Discriminación (bajos-altos) baja y redacción confusa.	Cuidadoso análisis del contenido del ítem y, en su caso, redacción más categórica y, hacer más o menos atractivas las opciones o sustituirlas y cambiar el distractor que atrae a los altos.
				b	0.06	0.00	0.00	0.21			
				c	0.47	0.42	0.54	0.20	*		
				d	0.26	0.50	0.15	-0.50			
37	0.33	0.21	0.23	a	0.33	0.33	0.54	0.23	*	Ninguna	
				b	0.14	0.11	0.08	-0.05			
				c	0.33	0.22	0.23	-0.03			
				d	0.19	0.33	0.15	-0.22			
38	0.33	0.35	0.37	a	0.06	0.00	0.00	-0.33		Ninguna	
				b	0.08	0.11	0.08	0.00			
				c	0.33	0.11	0.46	0.37	*		
				d	0.53	0.78	0.46	-0.33			
39	0.44	0.22	0.33	a	0.44	0.35	0.58	0.33	*	Ninguna	
				b	0.38	0.39	0.27	-0.22			
				c	0.08	0.13	0.08	-0.24			
				d	0.08	0.13	0.08	-0.04			
40	0.49	0.61	0.61	a	0.49	0.08	0.69	0.61	*	Ninguna	
				b	0.26	0.25	0.15	-0.14			
				c	0.04	0.17	0.00	-1.00			
				d	0.21	0.50	0.15	-0.34			
41	0.74	0.30	0.42	a	0.06	0.03	0.08	0.16		Ninguna	
				b	0.08	0.16	0.04	-0.40			
				c	0.74	0.58	0.88	0.42	*		
				d	0.12	0.23	0.00	-0.47			
42	0.38	0.37	0.46	a	0.26	0.42	0.23	-0.34		Ninguna	
				b	0.30	0.33	0.23	-0.13			
				c	0.38	0.17	0.54	0.46	*		
				d	0.06	0.08	0.00	-0.17			
43	0.64	0.51	0.49	a	0.64	0.42	0.92	0.49	*	Ninguna	Procurar en el ITEMAN definir el ítem con 3 distractores solamente.
				b	0.21	0.33	0.00	-0.48			
				c	0.15	0.25	0.08	-0.20			
				d							

44	0.25	0.43	0.48	a	0.31	0.33	0.15	-0.33		Ninguna	
				b	0.25	0.11	0.54	0.48	*		
				c	0.08	0.11	0.00	-0.08			
				d	0.36	0.44	0.31	-0.06			
45	0.64	0.43	0.51	a	0.17	0.25	0.08	-0.47		Ninguna	
				b	0.13	0.17	0.08	-0.13			
				c	0.06	0.17	0.00	-0.34			
				d	0.64	0.42	0.85	0.51	*		
46	0.38	0.04	0.02	a	0.04	0.00	0.08	0.28	?	Discriminación (bajos-altos) y (rbis) errónea y edición y redacción confusa.	Cuidadoso análisis del contenido del ítem y corrección del contenido de la base, sustituir opciones principalmente la opción c). Cambiar la opción que atrae a los que saben.
				b	0.38	0.50	0.54	0.02	*		
				c	0.09	0.08	0.08	0.06			
				d	0.49	0.42	0.31	-0.11			
47	0.50	0.21	0.22	a	0.08	0.00	0.08	0.12		Ninguna	
				b	0.50	0.33	0.54	0.22	*		
				c	0.36	0.56	0.31	-0.21			
				d	0.06	0.11	0.08	-0.23			
48	0.33	0.35	0.37	a	0.06	0.00	0.00	-0.03		Ninguna	
				b	0.08	0.11	0.08	0.00			
				c	0.33	0.11	0.46	0.37	*		
				d	0.53	0.78	0.46	-0.33			
49	0.22	0.23	0.32	a	0.08	0.22	0.08	-0.48		Discriminación errónea de un distractor y edición y redacción confusa.	Cuidadoso análisis del contenido del ítem y corrección del contenido de la base, sustituir opciones principalmente la opción b). Cambiar la opción que atrae a los que saben.
				b	0.64	0.56	0.69	0.11			
				c	0.06	0.22	0.00	-0.54			
				d	0.22	0.00	0.23	0.32	*		
50	0.40	0.28	0.14	a	0.06	0.17	0.00	-0.51		Discriminación (rbis) errónea.	Cambiar la opción que tiene el problema
				b	0.4	0.33	0.62	0.14	*		
				c	0.49	0.42	0.38	0.07			
				d	0.02	0.00	0.00	-0.05			
51	0.64	0.51	0.65	a	0.64	0.33	0.85	0.65	*	Ninguna	
				b	0.17	0.42	0.15	-0.44			
				c	0.13	0.08	0.00	-0.12			
				d	0.06	0.17	0.00	-0.86			
52	0.58	0.09	0.28	a	0.25	0.44	0.31	-0.27		Discriminación (bajos-altos) errónea.	Cambiar la opción que parece está creando el problema.
				b	0.11	0.00	0.15	0.05			
				c	0.58	0.44	0.54	0.28	*		
				d	0.06	0.11	0.00	-0.31			

53	0.70	0.34	0.35	a	0.02	0.08	0.00	-0.58		Discriminación errónea en distractores	Cuidadoso análisis del contenido del ítem y hacer más o menos atractivas las opciones o sustituirlas.
				b	0.06	0.17	0.00	-0.76			
				c	0.21	0.17	0.08	0.00			
				d	0.70	0.58	0.92	0.35	*		
54	0.89	0.25	0.44	a	0.00	0.00	0.00			Complejidad cognitiva (fácil). Discriminación errónea en distractores a) y b).	Cuidadoso análisis del contenido del ítem. Hacer más o menos atractivas las opciones o sustituirlas. Hacer más o menos atractivas las opciones o sustituirlas
				b	0.00	0.00	0.00				
				c	0.11	0.25	0.00	-0.44			
				d	0.89	0.75	1.00	0.44	*		
55	0.29	0.26	0.37	a	0.29	0.16	0.42	0.37	*	Ninguna	
				b	0.40	0.48	0.35	-0.16			
				c	0.19	0.23	0.12	-0.20			
				d	0.12	0.10	0.12	-0.01			
56	0.77	0.51	0.56	a	0.17	0.33	0.08	-0.35		Complejidad cognitiva (fácil) y discriminación errónea de un distractor.	Hacer al distractor b) más atractivo, es decir elaborar un distractor más plausible.
				b	0.00	0.00	0.00				
				c	0.06	0.25	0.00	-0.65			
				d	0.77	0.42	0.92	0.56	*		
57	0.72	0.59	0.86	a	0.11	0.22	0.08	-0.44		Ninguna	
				b	0.72	0.33	0.92	0.86	*		
				c	0.03	0.11	0.00	-0.68			
				d	0.14	0.33	0.00	-0.73			
58	0.72	0.34	0.35	a	0.06	0.17	0.00	-0.30		Ninguna	
				b	0.72	0.58	0.92	0.35	*		
				c	0.06	0.08	0.00	-0.25			
				d	0.15	0.17	0.08	-0.21			
59	0.60	0.19	0.28	a	0.60	0.50	0.69	0.28	*	Discriminación (bajos-altos) baja.	Cambiar la opción que atrae a los que saben.
				b	0.23	0.25	0.23	-0.03			
				c	0.02	0.08	0.00	-0.37			
				d	0.15	0.17	0.08	-0.35			
60	0.62	0.43	0.33	a	0.19	0.25	0.08	-0.05		Ninguna	
				b	0.13	0.17	0.08	-0.38			
				c	0.06	0.17	0.00	-0.27			
				d	0.62	0.42	0.85	0.33	*		
61	0.83	0.58	0.92	a	0.02	0.08	0.00	-0.37		Ninguna	
				b	0.09	0.33	0.00	-0.99			
				c	0.06	0.17	0.00	-0.47			
				d	0.83	0.42	1.00	0.92	*		

62	0.98	0.08	0.37	a	0.00	0.00	0.00			Complejidad cognitiva (fácil) y discriminación (bajos-altos) baja.	Cuidadoso análisis del contenido del ítem y hacer más categórica la base. Hacer más o menos atractivas las opciones o sustituirlas.
				b	0.00	0.00	0.00				
				c	0.98	0.92	1.00	0.37	*		
				d	0.02	0.08	0.00	-0.37			
63	0.81	0.26	0.30	a	0.14	0.22	0.08	-0.05		Discriminación errónea en distractor d).	Cambiar la opción que parece está creando el problema.
				b	0.81	0.67	0.92	0.30	*		
				c	0.06	0.11	0.00	-0.62			
				d	0.00	0.00	0.00				
64	0.58	0.32	0.46	a	0.17	0.33	0.00	-0.49		Discriminación errónea en distractor d).	Cambiar la opción que parece está creando el problema.
				b	0.58	0.44	0.77	0.46	*		
				c	0.25	0.22	0.23	-0.19			
				d	0.00	0.00	0.00				
65	0.43	0.18	0.30	a	0.15	0.19	0.08	-0.31		Discriminación (bajos-altos) errónea.	Cambiar la opción que parece está creando el problema.
				b	0.43	0.35	0.54	0.30	*		
				c	0.29	0.39	0.19	-0.21			
				d	0.10	0.06	0.19	0.18			
66	0.60	0.60	0.61	a	0.04	0.08	0.00	-0.44		Ninguna	
				b	0.60	0.25	0.85	0.61	*		
				c	0.21	0.42	0.08	-0.39			
				d	0.15	0.25	0.08	-0.35			
67	0.47	0.09	0.15	a	0.47	0.45	0.54	0.15	*	Discriminación (bajos-altos) y (rbis) baja y redacción confusa de la base	Simplificar la redacción de la base y redacción más categórica.
				b	0.41	0.39	0.42	0.00			
				c	0.06	0.10	0.04	-0.25			
				d	0.06	0.03	0.00	-0.22			
68	0.57	0.26	0.31	a	0.05	0.10	0.00	-0.45		Ninguna	
				b	0.03	0.03	0.00	-0.35			
				c	0.35	0.32	0.19	-0.14			
				d	0.57	0.55	0.81	0.31	*		

Anexo 4 Ejemplo de portada del ETAM



Universidad Autónoma de Baja California Facultad de Medicina y Psicología



Examen de Trayecto del Área Metodológica (ETAM) de la carrera de psicología.

Cuadernillo de reactivos (Modelo 1/ Otro)

Instrucciones Generales: En esta prueba te presentamos una serie de reactivos acordes a los temas vistos durante el área *Metodológica*. La prueba comprende contenidos de materias como *estadísticas, psicometría, metodología de la investigación y métodos en psicología I y II*. El tiempo aproximado de contestación de esta prueba es de 45 a 60 minutos. **Lee cada uno de los reactivos y contesta correctamente según corresponda.**

No escribas en este cuadernillo. Anota tus contestaciones en la hoja de respuestas correspondiente. Te recomendamos utilizar lápiz del número 2, y tener sacapuntas y borrador a la mano.

Al contestar en la hoja de respuestas, asegúrate de que el número del enunciado concuerde con el número de la respuesta. **Asegúrate de rellenar bien sus marcas.** Si deseas cambiar alguna respuesta, borra por completo lo que quieras modificar.

1. Subgrupo de población que se divide en **segmentos**, seleccionando una muestra para cada uno de ellos.
 - a) Muestreo simple
 - b) Muestreo estratificado
 - c) Muestreo por racimos
 - d) Muestreo dirigido
2. Principales **objetivos** de investigación científica en psicología.
 - a) Identificar las necesidades de la investigación formal y su relevancia social
 - b) Generar conocimiento de tipo teórico y solucionar problemas en contextos específicos
 - c) Analizar de los datos obtenidos de la realidad y socializar los resultados
 - d) Comprobar hipótesis y consolidar la investigación
3. Conjunto de **procesos sistemáticos y empíricos** que se aplican al estudio de un fenómeno real.
 - a) Investigación cualitativa
 - b) Investigación acción
 - c) Investigación científica
 - d) Investigación experimental

Item and Test Analysis Program -- ITEMAN (tm) Version 3.50

Item analysis for data from file C:\PROGRA~1\PROGRA~1\ITEMAN\TRAY.DAT
 Date: 03/16/10 Time: 4:28 pm

***** ANALYSIS SUMMARY INFORMATION *****

Scale Definition Codes: DICHOT = Dichotomous MPOINT = Multipoint/Survey

Scale: 0

 Type of Scale DICHOT
 N of Items 68
 N of Examinees 218

***** CONFIGURATION INFORMATION *****

Type of Correlations: Biserial
 Correction for Spuriousness: NO
 Ability Grouping: YES
 Subgroup Analysis: NO
 Express Endorsements As: PROPORTIONS
 Score Group Interval Width: 1

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics					
		Prop. Correct	Disc. Index	Biser.	Alt.	Prop. Total	Endorsing		Biser.	Key
						Low	High			
1	0-1	.40	.17	.22	A	.05	.06	.05	-.11	
					B	.40	.38	.55	.22	*
					C	.51	.50	.40	-.11	
					D	.03	.05	.00	-.30	
					Other	.00	.00	.00	-.45	
2	0-2	.53	.30	.33	A	.18	.25	.13	-.20	
					B	.53	.36	.66	.33	*
					C	.07	.06	.06	-.00	
					D	.21	.33	.15	-.26	
					Other	.00	.00	.00	-.09	
3	0-3	.49	.29	.31	A	.09	.14	.03	-.22	
					B	.10	.13	.06	-.12	
					C	.49	.39	.68	.31	*
					D	.32	.34	.21	-.20	
					Other	.00	.00	.00	.52	
4	0-4	.52	.14	.21	A	.41	.45	.37	-.15	
					B	.02	.05	.02	-.26	
					C	.04	.06	.03	-.12	

					Anexo 5					
					D	.52	.44	.58	.21	*
					Other	.00	.00	.00		
5	0-5	.98	.03	.32	A	.00	.00	.00	-.14	
					B	.98	.97	1.00	.32	*
					C	.00	.02	.00	-.55	
					D	.01	.02	.00	-.21	
					Other	.00	.00	.00		
6	0-6	.71	.15	.18	A	.02	.02	.06	.26	
					B	.00	.00	.00		
					C	.27	.38	.18	-.23	
					D	.71	.61	.76	.18	*
					Other	.00	.00	.00		
7	0-7	.47	.33	.31	A	.06	.06	.03	-.09	
					B	.47	.34	.68	.31	*
					C	.41	.53	.27	-.23	
					D	.06	.06	.02	-.20	
					Other	.00	.00	.00	.01	

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics					
		Prop. Correct	Disc. Index	Biser.	Alt.	Prop. Total	Endorsing Low	Endorsing High	Biser.	Key
8	0-8	.73	.29	.34	A	.13	.16	.05	-.14	
					B	.12	.25	.06	-.37	
					C	.01	.02	.02	-.14	
					D	.73	.58	.87	.34	*
					Other	.00	.00	.00		
9	0-9	.37	.50	.54	A	.10	.23	.02	-.49	
					B	.01	.03	.02	-.30	
					C	.50	.56	.32	-.25	
					D	.37	.14	.65	.54	*
					Other	.01	.00	.00	-.40	
10	0-10	.76	.15	.21	A	.12	.19	.15	-.09	
					B	.76	.64	.79	.21	*
					C	.03	.06	.03	-.19	
					D	.09	.11	.03	-.22	
					Other	.00	.00	.00		
11	0-11	.71	.40	.48	A	.08	.17	.03	-.40	
					B	.18	.31	.10	-.36	
					C	.71	.47	.87	.48	*
					D	.04	.05	.00	-.19	
					Other	.00	.00	.00		
12	0-12	.32	.30	.38	A	.08	.08	.03	-.11	
					B	.10	.13	.03	-.19	
					C	.50	.58	.42	-.21	
					D	.32	.22	.52	.38	*
					Other	.00	.00	.00		
13	0-13	.41	.43	.47	A	.27	.36	.11	-.31	
					B	.23	.22	.21	-.06	
					C	.41	.25	.68	.47	*
					D	.08	.17	.00	-.41	
					Other	.00	.00	.00		
14	0-14	.58	.40	.44	A	.58	.42	.82	.44	*
					B	.06	.08	.03	-.19	
					C	.23	.30	.08	-.29	
					D	.13	.20	.06	-.29	

Anexo 5

Other .00 .00 .00

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics					
		Prop. Correct	Disc. Index	Biser.	Alt.	Prop. Total	Endorsing		Biser.	Key
						Low	High			
15	0-15	.10	.26	.55	A	.10	.14	.10	-.09	
					B	.37	.38	.29	-.06	
					C	.10	.00	.26	.55	*
					D	.43	.47	.34	-.15	
					Other	.01	.00	.00	.04	
16	0-16	.84	.28	.44	A	.84	.69	.97	.44	*
					B	.05	.11	.00	-.43	
					C	.09	.17	.00	-.39	
					D	.02	.03	.03	-.06	
					Other	.00	.00	.00		
17	0-17	.36	.30	.29	A	.12	.16	.10	-.09	
					B	.25	.25	.19	-.04	
					C	.27	.31	.15	-.23	
					D	.36	.27	.56	.29	*
					Other	.00	.00	.00	-.34	
18	0-18	.35	.09	.19	A	.36	.41	.35	-.04	
					B	.25	.30	.24	-.17	
					C	.35	.28	.37	.19	*
					D	.03	.00	.02	-.04	
					Other	.01	.00	.00	.09	
19	0-19	.86	.23	.39	A	.86	.72	.95	.39	*
					B	.10	.20	.02	-.40	
					C	.03	.05	.02	-.19	
					D	.01	.03	.02	-.14	
					Other	.00	.00	.00		
20	0-20	.84	.17	.24	A	.06	.08	.02	-.20	
					B	.02	.06	.00	-.42	
					C	.84	.75	.92	.24	*
					D	.07	.09	.06	-.04	
					Other	.00	.00	.00	-.55	
21	0-21	.42	.52	.53	A	.03	.03	.00	-.18	
					B	.15	.22	.05	-.35	
					C	.42	.19	.71	.53	*
					D	.40	.56	.24	-.30	
					Other	.00	.00	.00		

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics					
		Prop. Correct	Disc. Index	Biser.	Alt.	Prop. Total	Endorsing		Biser.	Key
						Low	High			
22	0-22	.73	.47	.52	A	.04	.05	.02	-.18	
					B	.19	.42	.05	-.50	
					C	.73	.47	.94	.52	*
					D	.04	.06	.00	-.25	
					Other	.00	.00	.00		
23	0-23	.85	.30	.47	A	.08	.14	.02	-.29	
					B	.03	.08	.00	-.39	

					Anexo 5					
					C	.85	.69	.98	.47	*
					D	.04	.09	.00	-.46	
					Other	.00	.00	.00		
24	0-24	.90	.08	.12	A	.07	.08	.03	-.16	
					B	.90	.88	.95	.12	*
					C	.02	.03	.02	.10	
					D	.00	.02	.00	-.39	
					Other	.00	.00	.00		
25	0-25	.33	.34	.43	A	.00	.02	.00	-.50	
					B	.33	.16	.50	.43	*
					C	.62	.81	.45	-.40	
					D	.04	.02	.05	.02	
					Other	.00	.00	.00		
26	0-26	.50	.27	.31	A	.50	.42	.69	.31	*
					B	.47	.50	.31	-.24	
					C	.02	.05	.00	-.33	
					D	.01	.03	.00	-.30	
					Other	.00	.00	.00		
27	0-27	.38	.08	.10	A	.04	.06	.00	-.32	
					B	.52	.50	.56	.03	
					C	.38	.33	.40	.10	*
					D	.06	.11	.02	-.30	
					Other	.00	.00	.00	.52	
28	0-28	.65	.10	.11	A	.09	.13	.02	-.22	
					B	.15	.09	.18	.12	
					C	.65	.64	.74	.11	*
					D	.11	.14	.06	-.17	
					Other	.00	.00	.00		

Seq. No.	Item Statistics				Alternative Statistics					Key
	Scale -Item	Prop. Correct	Disc. Index	Biser.	Alt.	Prop. Total	Endorsing Low	Endorsing High	Biser.	
29	0-29	.19	.09	.14	A	.33	.41	.31	-.06	
					B	.40	.39	.35	-.07	
					C	.19	.16	.24	.14	*
					D	.08	.05	.10	.05	
					Other	.00	.00	.00		
30	0-30	.67	.47	.52	A	.28	.48	.03	-.52	
					B	.01	.02	.00	-.18	
					C	.67	.48	.95	.52	*
					D	.04	.02	.02	-.05	
					Other	.00	.00	.00		
31	0-31	.95	.11	.42	A	.02	.06	.00	-.58	
					B	.02	.05	.00	-.39	
					C	.95	.88	.98	.42	*
					D	.01	.02	.02	-.01	
					Other	.00	.00	.00		
32	0-32	.51	.41	.41	A	.09	.17	.06	-.24	
					B	.03	.02	.03	.07	
					C	.37	.50	.18	-.35	
					D	.51	.31	.73	.41	*
					Other	.00	.00	.00		
33	0-33	.16	.04	.09	A	.07	.05	.08	.08	
					B	.28	.45	.18	-.31	
					C	.16	.13	.16	.09	*

CHECK THE KEY

					Anexo 5					
C was specified, D works better					D	.48	.38	.55	.15	?
					Other	.01	.00	.00	.50	
34	0-34	.08	.19	.54	A	.33	.27	.35	.15	
					B	.06	.13	.02	-.37	
					C	.53	.58	.40	-.24	
					D	.08	.03	.23	.54	*
					Other	.00	.00	.00		
35	0-35	.58	.46	.49	A	.10	.13	.06	-.18	
					B	.25	.44	.08	-.42	
					C	.58	.34	.81	.49	*
					D	.06	.08	.05	-.17	
					Other	.00	.00	.00	-.45	

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics					
		Prop. Correct	Disc. Index	Biser.	Alt.	Prop. Total	Endorsing		Biser.	Key
36	0-36	.25	-.01	-.01	A	.17	.11	.18	.12	?
CHECK THE KEY					B	.22	.19	.21	.02	
C was specified, A works better					C	.25	.28	.27	-.01	*
					D	.34	.42	.32	-.12	
					Other	.00	.00	.00	.52	
37	0-37	.31	.44	.50	A	.16	.17	.05	-.20	
					B	.39	.48	.27	-.24	
					C	.31	.17	.61	.50	*
					D	.14	.17	.06	-.17	
					Other	.00	.00	.00		
38	0-38	.22	.20	.32	A	.46	.39	.44	-.00	
					B	.12	.19	.15	-.08	
					C	.22	.13	.32	.32	*
					D	.20	.30	.10	-.27	
					Other	.00	.00	.00		
39	0-39	.62	.43	.42	A	.62	.44	.87	.42	*
					B	.23	.31	.06	-.32	
					C	.07	.16	.02	-.35	
					D	.07	.09	.03	-.16	
					Other	.00	.00	.00	.52	
40	0-40	.20	.28	.37	A	.20	.09	.37	.37	*
					B	.34	.34	.39	.09	
					C	.13	.13	.10	-.14	
					D	.33	.44	.15	-.30	
					Other	.00	.00	.00		
41	0-41	.86	.06	.15	A	.05	.06	.03	-.13	
					B	.04	.02	.05	.06	
					C	.86	.81	.87	.15	*
					D	.06	.11	.05	-.21	
					Other	.00	.00	.00		
42	0-42	.35	.03	.07	A	.28	.17	.40	.24	?
					B	.34	.42	.24	-.20	
CHECK THE KEY					C	.35	.33	.35	.07	*
C was specified, A works better					D	.03	.08	.00	-.52	
					Other	.00	.00	.00		

Anexo 5

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics					
		Prop. Correct	Disc. Index	Biser.	Alt.	Prop. Total	Endorsing Low	Endorsing High	Biser.	Key
43	0-43	.45	.24	.24	A	.30	.38	.31	-.03	
					B	.17	.17	.08	-.19	
					C	.08	.09	.02	-.26	
					D	.45	.36	.60	.24	*
					Other	.00	.00	.00		
44	0-44	.31	.10	.06	A	.21	.30	.13	-.21	
					B	.31	.25	.35	.06	*
					C	.35	.28	.37	.12	?
					D	.13	.17	.13	-.07	
					Other	.00	.00	.00	.52	
CHECK THE KEY B was specified, C works better										
45	0-45	.06	.10	.37	A	.58	.53	.50	-.03	
					B	.16	.17	.11	-.10	
					C	.21	.27	.26	-.03	
					D	.06	.03	.13	.37	*
					Other	.00	.00	.00		
46	0-46	.22	-.04	-.04	A	.14	.19	.15	-.05	
					B	.22	.22	.18	-.04	*
					C	.33	.27	.44	.20	?
					D	.32	.33	.24	-.14	
					Other	.00	.00	.00		
CHECK THE KEY B was specified, C works better										
47	0-47	.45	.09	.11	A	.45	.39	.48	.11	*
					B	.34	.36	.40	.05	
					C	.10	.13	.06	-.14	
					D	.10	.11	.05	-.17	
					Other	.00	.00	.00	-.45	
48	0-48	.38	.17	.23	A	.38	.31	.48	.23	*
					B	.16	.17	.11	-.14	
					C	.24	.30	.19	-.14	
					D	.22	.22	.21	-.03	
					Other	.00	.00	.00		
49	0-49	.24	.37	.47	A	.24	.08	.45	.47	*
					B	.22	.31	.15	-.21	
					C	.41	.47	.23	-.24	
					D	.12	.13	.16	.04	
					Other	.01	.00	.00	.04	

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics					
		Prop. Correct	Disc. Index	Biser.	Alt.	Prop. Total	Endorsing Low	Endorsing High	Biser.	Key
50	0-50	.63	.21	.19	A	.11	.19	.10	-.17	
					B	.63	.50	.71	.19	*
					C	.24	.30	.18	-.13	
					D	.01	.00	.02	.17	
					Other	.00	.00	.00	-.45	
51	0-51	.56	.24	.20	A	.56	.45	.69	.20	*
					B	.08	.05	.05	-.02	
					C	.05	.08	.03	-.25	
					D	.31	.41	.23	-.14	
					Other	.00	.00	.00	-.34	
52	0-52	.78	.15	.18	A	.17	.20	.13	-.08	
					B	.04	.08	.00	-.37	

					Anexo 5					
					C	.78	.70	.85	.18	*
					D	.01	.02	.02	.01	
					Other	.00	.00	.00		
53	0-53	.63	.29	.37	A	.05	.05	.00	-.25	
					B	.20	.28	.18	-.22	
					C	.13	.23	.10	-.27	
					D	.63	.44	.73	.37	*
					Other	.00	.00	.00		
54	0-54	.77	.15	.23	A	.04	.08	.05	-.15	
					B	.02	.03	.00	-.36	
					C	.17	.23	.15	-.16	
					D	.77	.66	.81	.23	*
					Other	.00	.00	.00		
55	0-55	.57	.24	.31	A	.57	.45	.69	.31	*
					B	.15	.20	.10	-.20	
					C	.17	.16	.18	-.03	
					D	.10	.19	.03	-.38	
					Other	.00	.00	.00		
56	0-56	.17	.23	.42	A	.19	.20	.23	-.01	
					B	.51	.64	.40	-.22	
					C	.11	.05	.05	-.09	
					D	.17	.08	.31	.42	*
					Other	.01	.00	.00	.01	

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics					
		Prop. Correct	Disc. Index	Biser.	Alt.	Prop. Total	Endorsing Low	Endorsing High	Biser.	Key
57	0-57	.66	.45	.51	A	.09	.14	.03	-.35	
					B	.66	.41	.85	.51	*
					C	.03	.09	.00	-.52	
					D	.21	.33	.11	-.29	
					Other	.01	.00	.00	-.40	
58	0-58	.54	.30	.30	A	.08	.09	.05	-.08	
					B	.54	.41	.71	.30	*
					C	.11	.20	.02	-.34	
					D	.27	.30	.23	-.12	
					Other	.00	.00	.00		
59	0-59	.52	.30	.31	A	.52	.41	.71	.31	*
					B	.30	.31	.19	-.15	
					C	.07	.11	.08	-.14	
					D	.10	.16	.02	-.26	
					Other	.00	.00	.00	-.45	
60	0-60	.27	.31	.33	A	.27	.14	.45	.33	*
					B	.50	.56	.39	-.15	
					C	.13	.20	.08	-.16	
					D	.10	.09	.08	-.08	
					Other	.00	.00	.00		
61	0-61	.79	.17	.28	A	.06	.08	.02	-.22	
					B	.06	.11	.03	-.34	
					C	.09	.11	.08	-.10	
					D	.79	.70	.87	.28	*
					Other	.00	.00	.00		
62	0-62	.62	.34	.38	A	.11	.16	.10	-.15	
					B	.19	.30	.08	-.36	
					C	.62	.45	.79	.38	*

Anexo 5

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics					
		Prop. Correct	Disc. Index	Biser.	Alt.	Prop. Total	Endorsing		Biser.	Key
							Low	High		
					D	.06	.08	.03	-.12	
					Other	.00	.00	.00	-.34	
63	0-63	.63	.18	.19	A	.33	.39	.29	-.11	
					B	.63	.52	.69	.19	*
					C	.03	.08	.00	-.46	
					D	.01	.02	.02	-.13	
					Other	.00	.00	.00		
64	0-64	.44	.40	.42	A	.23	.31	.16	-.22	
					B	.44	.28	.68	.42	*
					C	.16	.16	.10	-.12	
					D	.17	.25	.06	-.28	
					Other	.00	.00	.00		
65	0-65	.40	.49	.54	A	.13	.25	.10	-.21	
					B	.40	.14	.63	.54	*
					C	.20	.33	.05	-.41	
					D	.27	.28	.23	-.15	
					Other	.00	.00	.00		
66	0-66	.48	-.03	.01	A	.06	.06	.06	-.05	
					B	.48	.53	.50	.01	*
					C	.32	.27	.37	.08	?
					D	.14	.13	.06	-.10	
					Other	.00	.00	.00	-.45	
					CHECK THE KEY B was specified, C works better					
67	0-67	.60	.11	.16	A	.60	.52	.63	.16	*
					B	.27	.25	.29	-.03	
					C	.06	.08	.06	-.05	
					D	.06	.13	.02	-.35	
					Other	.01	.00	.00	-.26	
68	0-68	.57	.05	.12	A	.08	.08	.05	-.10	
					B	.04	.06	.02	-.19	
					C	.29	.31	.37	.00	
					D	.57	.50	.55	.12	*
					Other	.02	.00	.00	-.31	

There were 218 examinees in the data file.

Scale Statistics

Scale:	0

N of Items	68
N of Examinees	218
Mean	34.798
Variance	45.776
Std. Dev.	6.766
Skew	0.337
Kurtosis	-0.365
Minimum	21.000
Maximum	55.000
Median	34.000
Alpha	0.715
SEM	3.611
Mean P	0.512

Anexo 5

Mean Item-Tot. 0.223
 Mean Biserial 0.305
 Max Score (Low) 30
 N (Low Group) 64
 Min Score (High) 39
 N (High Group) 62

SCALE # 0 Score Distribution Table

Number Correct	Freq- uency	Cum Freq	PR	PCT	
. . . No examinees below this score . .					
20	0	0	1	0	+
21	4	4	2	2	##
22	0	4	2	0	
23	2	6	3	1	#
24	6	12	6	3	###
25	1	13	6	0	+
26	7	20	9	3	###
27	10	30	14	5	#####
28	12	42	19	6	#####
29	8	50	23	4	####
30	14	64	29	6	+#####
31	13	77	35	6	#####
32	13	90	41	6	#####
33	9	99	45	4	####
34	12	111	51	6	#####
35	16	127	58	7	+#####
36	10	137	63	5	#####
37	7	144	66	3	###
38	12	156	72	6	#####
39	11	167	77	5	#####
40	6	173	79	3	+###
41	5	178	82	2	##
42	6	184	84	3	###
43	8	192	88	4	####
44	4	196	90	2	##
45	6	202	93	3	+###
46	6	208	95	3	###
47	1	209	96	0	
48	2	211	97	1	#
49	3	214	98	1	#
50	2	216	99	1	+#
51	1	217	99	0	
52	0	217	99	0	
53	0	217	99	0	
54	0	217	99	0	
55	1	218	99	0	+
56	0	218	99	0	
57	0	218	99	0	
. . . No examinees above this score . . .					

-----+-----+-----+-----+-----+
 5 10 15 20 25
 Percentage of Examinees

Anexo 6

Resultados puntuales por ítem de la evaluación 2010-1 del ETAM

Los ítems que se presentan a continuación son un ejemplo muestra de los ítems que se presentaron en la evaluación del área metodológica. En cada una de los ítems que se encuentran en la siguiente tabla pueden observarse varios recuadros con los siguientes datos:

- Codificación de los contenidos evaluados por asignatura y eje curricular
- Contenidos de formación red curricular.
- Ítem muestra con su respectiva respuesta correcta (*).
- Posición del ítem en el contexto de la prueba.
- Porcentaje de estudiantes que contestaron correctamente el ítem.
- Niveles de logro al que corresponde el ítem:

Por debajo del nivel básico
 Básico
 Medio
 Avanzado

Codificación de los contenidos evaluados por asignatura y eje curricular (Unidad temática).	
I.- Métodos en psicología I	II.- Métodos en psicología II
I.1.- Unidad 1: La investigación en psicología	II.1.- Unidad 1: El diseño de investigación cuantitativa y cualitativa
I.2.- Unidad 2: El inicio del proceso de investigación	II.2.- Unidad 2. Recolección y análisis de los datos
I.3.- Unidad 3: los requisitos protocolarios del proceso de investigación	

Contenidos oficiales a evaluados	Red curricular de pertenencia	Ítem muestra del contenido evaluado	Posición del ítem	% de estudiantes que acertaron el ítem	Niveles de logro
I.1.1.2.3. Los enfoques de investigación (98%)	Lógica de la investigación (57%)	5. Enfoque de investigación que se fundamenta en analizar datos de tipo numérico: a) Etnográfico b) Cuantitativo* c) Ideográfico d) Cualitativo	5	98	Avanzado
I.2.2.1.1. Los alcances de la investigación (51%)	Planteamiento o del problema (56%)	31. Tipo de alcance que busca determinar la vinculación entre dos o más variables. a) Exploratorio b) Explicativo c) Correlacional* d) Descriptivo	31	95	Avanzado

Contenidos oficiales a evaluados	Red curricular de pertenencia	Ítem muestra del contenido evaluado	Posición del ítem	% de estudiantes que acertaron el ítem	Niveles de logro
I.2.1.4. El planteamiento del problema de investigación (57%)	Planteamiento del problema (56%)	24. En el campo científico, se presenta como la posible respuesta a la pregunta de investigación. a) Objetivos de investigación b) Hipótesis de investigación* c) Variables, lógica y diseño d) Sujetos, marco teórico y variables	24	90	Avanzado
I.1.1.1.1. Corrientes epistemológicas (69%)	Lógica de la investigación (57%)	19. Las características representativas del enfoque cuantitativo son: 1. Busca la generalización y comprobación de los resultados* 2. Se enfoca en conocer a profundidad a la unidad de análisis 3. Se centra en la noción ideográfica de la realidad 4. El constructivismo es su paradigma representativo	19	86	Avanzado
I.3.2.2. Trabajo de gabinete (fuentes de información y redacción de referencias) (73%)	Redacción académica (68%)	41. Tipo de citas en las que se retoman las ideas de otro autor, pero con palabras propias. a) Textual b) Por mención a trabajo de otro autor c) De paráfrasis d) De cita - nota	41	86	Avanzado
II.2.1.2. Desarrollo y piloteo de las herramientas de recolección de datos (64%)	Herramientas de recolección de datos (65%)	23. Indica el proceso para elaboración de un instrumento de medición: a) Identificación del dominio de la variable, construcción del instrumento, prueba piloto, construcción de la versión final del instrumento y revisión de la literatura sobre la variable de interés. b) Identificación del dominio de la variable, construcción del instrumento, construcción de la versión final del instrumento, prueba piloto, revisión de la literatura sobre la variable de interés. c) Revisión de la literatura sobre la variable de interés, identificación del dominio de la variable, construcción del instrumento, prueba piloto, análisis de la calidad técnica del instrumento, construcción de la versión final del instrumento. d) Construcción del instrumento, identificación del dominio de la variable, revisión de la literatura sobre la variable de interés, análisis de la calidad técnica del instrumento, prueba piloto, construcción de la versión final del instrumento.	23	85	Avanzado

Contenidos oficiales a evaluados	Red curricular de pertenencia	Ítem muestra del contenido evaluado	Posición del ítem	% de estudiantes que acertaron el ítem	Niveles de logro
I.1.1.1.3. La institucionalización de la actividad investigativa en psicología (84%)	Historia y ciencia en psicología (84%)	20. ¿Cuál de los siguientes eventos históricos fue primordial para que se reconociera a la psicología como ciencia? a) Los experimentos realizados por el fisiólogo Ivan Pavlov. b) El trabajo de Sigmund Freud, fundador del psicoanálisis. c) La creación del primer laboratorio de psicología por Wilhem Wunt. d) La creación de un laboratorio antropométrico por Sir Francis Galton	20	84	Avanzado
I.2.1.1. Procesos de investigación (63%)	Proceso de investigación (59%)	16. Elige la opción que represente en forma correcta la secuencia del proceso de la investigación cuantitativa: a) Generación de la idea, planteamiento del problema, elaboración del marco teórico, elección del diseño, recolección y análisis de los datos, y reporte de resultados de la investigación b) Generación de la idea, planteamiento del problema, elaboración del marco teórico, recolección y análisis de los datos, elección del diseño, y reporte de resultados de la investigación c) Generación de la idea, elaboración del marco teórico, elección del diseño, planteamiento del problema, recolección y análisis de los datos, y reporte de los resultados de la investigación d) Generación de la idea, planteamiento del problema, elaboración del marco teórico, y reporte y socialización de resultados de la investigación.	16	84	Avanzado
II.1.2.4. Población y selección de la muestra (61%)	Técnicas de muestreo (62%)	61. Subgrupo de población que puede ser de tipo probabilístico o no probabilístico. a) Muestra dirigida b) Cohorte c) Unidad de análisis d) Muestra	61	79	Avanzado
II.2.1.1. Tipos de herramientas utilizadas en la investigación científica (65%)	Herramientas de recolección de datos (65%)	52. Herramienta de recolección de datos que requiere de estandarización, confiabilidad y evidencias de validez. a) Entrevista estructurada b) Test proyectivos c) Test psicométricos d) Observación cuantitativa	52	78	Avanzado

Contenidos oficiales a evaluados	Red curricular de pertenencia	Ítem muestra del contenido evaluado	Posición del ítem	% de estudiantes que acertaron el ítem	Niveles de logro
II.1.2.4. Población y selección de la muestra (61%)	Técnicas de muestreo (62%)	54. Subgrupo de población en donde todos los elementos tienen la misma posibilidad de ser elegidos. a) Marco muestral b) Muestreo dirigido c) Muestreo al azar por marcado telefónico d) Muestreo probabilístico	54	77	Avanzado
I.2.2.2.3. Clasificación de los diseños cualitativos (76%)	Diseño de investigación (34%)	10. Son diseños y técnicas representativas de la investigación cualitativa. a) Los cuasi-experimentos b) Los grupos focales y casos de estudio c) Los experimentos puros d) La observación estandarizada	10	76	Avanzado
I.1.2.2.1. Investigación cuantitativa (55%)	Lógica de la investigación (57%)	22. Tipo de investigación que busca la objetividad y la generalización. a) Investigación ideográfica b) Investigación cualitativa c) Investigación cuantitativa d) Investigación - acción	22	73	Medio
I.2.1.4. El planteamiento del problema de investigación (57%)	Planteamiento del problema (56%)	8. ¿Cuál de las siguientes opciones presenta un punto esencial para justificar cualquier trabajo de investigación? a) Interés individual. b) Identificar áreas del conocimiento no exploradas c) Los problemas socioeconómicos d) La relevancia social	8	73	Medio
I.1.2.1.2. Método deductivo (71%)	Lógica de la investigación (57%)	6. Método que parte de categorías generales para hacer afirmaciones sobre casos particulares. a) Cualitativo b) Analógico c) Inductivo d) Deductivo	6	71	Medio

Contenidos oficiales a evaluados	Red curricular de pertenencia	Ítem muestra del contenido evaluado	Posición del ítem	% de estudiantes que acertaron el ítem	Niveles de logro
I.1.2.2.1. Investigación cuantitativa (55%)	Lógica de la investigación (57%)	<p>Indicaciones (estudio ficticio): Lee con cuidado el siguiente texto, una vez leído, elige la opción correcta según correspondan a las preguntas 11, 12, 13 y 14.</p> <p>Un psicólogo del desarrollo está interesado en la cantidad de conducta verbal que niños muy pequeños mostraban dependiendo de quién estaba presente. Los niños en el estudio tienen 3 años de edad. Los mismos fueron observados en un ambiente de laboratorio por un periodo de 30 minutos. La mitad de los niños fue asignada a la condición en la cual otro niño estaba presente durante la sesión de observación. El psicólogo midió: el número, la duración y la complejidad de las verbalizaciones de cada uno de los niños observados.</p> <p>11. ¿Cuál es el enfoque investigativo en el que se fundamenta el estudio?</p> <p>a) Complementario b) Cualitativo c) Cuantitativo d) Constructivista</p>	11	71	Medio
I.2.1.4. El planteamiento del problema de investigación (57%)	Planteamiento del problema (56%)	<p>30. Algunos autores especializados en metodología de la investigación hacen énfasis en que la mejor manera de plantear un problema es en forma:</p> <p>a) Hipotética b) Aleatoria c) De pregunta d) Metodológica</p>	30	67	Medio
II.1.2.4. Población y selección de la muestra (61%)	Técnicas de muestreo (62%)	<p>57. Subgrupo de la población en la que la elección de los elementos no depende de la probabilidad sino de las características de la investigación o de quien hace la muestra.</p> <p>a) Muestreo por racimos b) Muestreo no probabilístico c) Muestreo probabilístico d) Muestreo estratificado</p>	57	66	Medio

Contenidos oficiales a evaluados	Red curricular de pertenencia	Ítem muestra del contenido evaluado	Posición del ítem	% de estudiantes que acertaron el ítem	Niveles de logro
I.2.1.3. Las etapas de la investigación (56%)	Proceso de investigación (59%)	<p>28. Indica los puntos esenciales para la adecuada elección del tema.</p> <p>1. El Interés particular de un grupo social, evaluación de los recursos económicos y la conveniencia del investigador</p> <p>2. La evaluación de los recursos cognitivos, la conveniencia del investigador y la facilidad del estudio.</p> <p>3. La evaluación de los recursos cognitivos y económicos, la trascendencia social y las limitaciones posibles del estudio</p> <p>4. El interés particular del investigador, la relevancia social y la planeación operativa</p> <p>a) 1, 2 y 3 b) 3 y 1 c) 3 y 4 d) 2, 3 y 4</p>	28	65	Medio
II.1.2.3. Características específicas de los diseños cualitativos (42%)	Diseños de investigación (34%)	<p>50. A los diseños de investigación en donde el investigador recolecta datos sobre las historias de vida y experiencias de ciertas personas para describirlas y analizarlas, se les conoce como:</p> <p>a) Micro-etnográficos b) Narrativos c) Etnográficos d) Teoría fundamentada</p>	50	63	Medio
II.1.2.4. Población y selección de la muestra (61%)	Técnicas de muestreo (62%)	<p>53. Permite identificar, enumerar y seleccionar los elementos muestrales de una población.</p> <p>a) Población b) Diseño de investigación c) Muestreo aleatorio d) Marco muestral</p>	53	63	Medio
II.2.1.1. Tipos de herramientas utilizadas en la investigación científica (65%)	Herramientas de recolección de datos (65%)	<p>63. Técnica en donde el investigador utiliza primordialmente registros no estandarizados, y se sumerge a profundidad en situaciones sociales manteniendo un papel activo y reflexivo.</p> <p>a) Entrevista clínica b) Observación cualitativa c) Test psicométricos d) Observación cuantitativa</p>	63	63	Medio

Contenidos oficiales a evaluados	Red curricular de pertenencia	Ítem muestra del contenido evaluado	Posición del ítem	% de estudiantes que acertaron el ítem	Niveles de logro
II.2.1.1. Tipos de herramientas utilizadas en la investigación científica (65%)	Herramientas de recolección de datos (65%)	62. De la siguiente lista elige las principales escalas de actitudes: 1. Thurstone 2. Frostig 3. Guttman 4. Bender 5. Likert 6. Holland a) 1, 2, 3 y 5 b) 1, 2 y 5 c) 1, 3 y 5 d) 3, 4, 5 y 6	62	62	Medio
II.2.3.2. Elección del análisis estadístico (37%)	Análisis de los datos (37%)	39. Nivel de medición más elemental que utiliza símbolos para clasificar o tipificar una variable: a) Nominal b) Ordinal c) Razón d) Intervalos	39	61	Medio
I.3.2.2. Trabajo de gabinete (fuentes de información y redacción de referencias) (73%)	Redacción académica (68%)	67. Libros, artículos científicos de publicaciones periódicas, tesis y disertaciones, son ejemplos de fuentes de información de tipo: a) Primarias b) Secundarias c) Terciarias d) Generales	67	60	Medio
I.1.2.2.1. Investigación cuantitativa (55%)	Lógica de la investigación (57%)	Indicaciones (estudio ficticio): Lee con cuidado el siguiente texto, una vez leído, elige la opción correcta según correspondan a las preguntas 11, 12, 13 y 14. Un psicólogo del desarrollo está interesado en la cantidad de conducta verbal que niños muy pequeños mostraban dependiendo de quién estaba presente. Los niños en el estudio tienen 3 años de edad. Los mismos fueron observados en un ambiente de laboratorio por un periodo de 30 minutos. La mitad de los niños fue asignada a la condición en la cual otro niño estaba presente durante la sesión de observación. El psicólogo midió: el número, la duración y la complejidad de las verbalizaciones de cada uno de los niños observados. 14. ¿Cuáles son las variables dependientes manejadas en el estudio?	14	58	Medio

Contenidos oficiales a evaluados	Red curricular de pertenencia	Ítem muestra del contenido evaluado	Posición del ítem	% de estudiantes que acertaron el ítem	Niveles de logro
		a) Numero, duración y complejidad de las verbalizaciones de cada uno de los niños b) Adulto presente, número y duración de las verbalizaciones de cada uno de los niños c) La presencia de un adulto o un niño durante la sesión de observación d) Niños pequeños dentro de la sesión de observación			
I.2.3.1. Desarrollo de operaciones básicas (57%)	Diseño de investigación (34%)	68. Indica el proceso para elaboración del marco teórico: a) Revisión de la literatura, detección y consulta de las fuentes, construcción del marco teórico, y adopción de una teoría. b) Detección y consulta de las fuentes, construcción del marco teórico, recopilación de la información, y revisión de la literatura. c) Detección y consulta de las fuentes, adopción y desarrollo de una teoría, recopilación de la información, y construcción del marco teórico. d) Revisión de la literatura, detección y consulta de las fuentes, recopilación de la información, adopción y desarrollo de una teoría, y construcción del marco teórico.	68	57	Medio
I.3.1.1.3. Los apartados del anteproyecto y tesis (57%)	Redacción académica (68%)	55. Sección de la tesis en donde se comentan las teorías que se manejan y los estudios previos que fueron relacionados con el planteamiento del problema. a) Marco teórico b) Resumen c) Introducción d) Conclusiones	55	57	Medio
II.2.1.1. Tipos de herramientas utilizadas en la investigación científica (65%)	Herramientas de recolección de datos (65%)	51. Al tipo de entrevista más utilizada en estudios exploratorios, se le conoce como: a) No estructurada b) De diagnóstico c) De grupo focal d) Estructurada	51	56	Medio

Contenidos oficiales a evaluados	Red curricular de pertenencia	Ítem muestra del contenido evaluado	Posición del ítem	% de estudiantes que acertaron el ítem	Niveles de logro
II.1.2.4. Población y selección de la muestra (61%)	Técnicas de muestreo (62%)	58. Procedimientos de selección utilizados en el muestreo probabilístico: 1. Tómbola 2. Casos típicos 3. Números aleatorios 4. Selección sistemática 5. Por cuotas a) 1,4 y 5 b) 1,3 y 4 c) 2,3 y 4	58	54	Medio
I.1.1.2.1. La articulación de la actividad investigativa (50%)	Lógica de la investigación (57%)	2. Principales objetivos de investigación científica en psicología. a) Identificar las necesidades de la investigación formal y su relevancia social b) Generar conocimiento de tipo teórico y solucionar problemas en contextos específicos c) Analizar de los datos obtenidos de la realidad y socializar los resultados d) Comprobar hipótesis y consolidar la investigación	2	53	Medio
I.1.1.1.1. Corrientes epistemológicas (69%)	Lógica de la investigación (57%)	4. Corriente filosófica que enfatiza el papel de la experiencia a través de la percepción sensorial y justifica la posibilidad de validar las teorías desde la experimentación. a) Fenomenología b) Materialismo c) Positivismo d) Empirismo	4	52	Medio
II.1.2.4. Población y selección de la muestra (61%)	Técnicas de muestreo (62%)	59. Son muestras clasificadas como no probabilísticas. a) Por decisión y de cuotas b) Estratificadas c) Por azar y heurística d) Por azar y de cuotas	59	52	Medio
II.1.1.3. Validez interna y externa de los diseños (36%)	Diseños de investigación	32. Posibilidad de generalizar los resultados de un experimento a situaciones no experimentales, y a otras poblaciones. a) Noción nomotética b) Validez interna c) Generalización d) Validez externa	32	51	Básico

Contenidos oficiales a evaluados	Red curricular de pertenencia	Ítem muestra del contenido evaluado	Posición del ítem	% de estudiantes que acertaron el ítem	Niveles de logro
I.2.2.1.2. Características generales de los diseños cuantitativos (25%)	Diseño de investigación (34%)	26. Es el factor que manipula el investigador dentro de un experimento. a) V. Independiente b) V. Dependiente c) Variable extraña d) Constructo hipotético	26	50	Básico
I.1.1.2.1. La articulación de la actividad investigativa.	Lógica de la investigación (57%)	3. Conjunto de procesos sistemáticos y empíricos que se aplican al estudio de un fenómeno real. a) Investigación cualitativa b) Investigación acción c) Investigación científica d) Investigación experimental	3	48	Básico
II.2.3.2. Elección del análisis estadístico (37%)	Análisis de los datos (37%)	66. Tipo de análisis que utiliza las pruebas del coeficiente de correlación de Pearson, la prueba t, y el análisis de varianza factorial (ANOVA), entre otras. a) No paramétrico b) Paramétrico c) Multivariado d) Descriptivo	66	48	Básico
I.2.1.3. Las etapas de la investigación (56%)	Proceso de investigación (59%)	7. Actividad o etapa del proceso de investigación en donde se define el plan para la obtención de los datos. a) La elección del alcance de investigación b) La elección del diseño de investigación c) La construcción de Instrumentos de medición d) La elección del análisis estadístico	7	47	Básico
II.1.2.2. Características específicas de los diseños cuantitativos (27%)	Diseños de investigación (34%)	43. Tipo de diseño experimental que se caracterizan por un bajo nivel de control y, por lo tanto, baja validez interna y externa. a) Pre-experimentales b) Experimentos puros c) Transeccional d) Cuasi-experimentos	43	45	Básico
II.1.3.1. Ética e investigación (41%)	Ética de la investigación (42%)	47. Juicio ético para determinar si la investigación debe ser llevada a cabo o no. a) Evaluación sobre la relevancia de la investigación b) Evaluación subjetiva de riesgos y beneficios c) Evaluación del engaño en la investigación d) Evaluación de los créditos de publicación y el plagio	47	45	Básico

Contenidos oficiales a evaluados	Red curricular de pertenencia	Ítem muestra del contenido evaluado	Posición del ítem	% de estudiantes que acertaron el ítem	Niveles de logro
II.2.1.2. Desarrollo y piloteo de las herramientas de recolección de datos (64%)	Herramientas de recolección de datos (65%)	64. Asignación de valores numéricos que dan orden a un conjunto de datos para su posterior análisis estadístico. a) Categorización b) Codificación c) Nominación d) Enumeración	64	44	Básico
I.2.1.1. Procesos de investigación (63%)	Proceso de investigación (59%)	21. ¿Cuál es el proceso de investigación en espiral, en donde las etapas interactúan entre sí, sin seguir una secuencia rigurosa? a) Proceso cuantitativo b) Proceso de delimitación del tema c) Proceso cualitativo d) Proceso de búsqueda de la información	21	42	Básico
I.1.2.2.1. Investigación cuantitativa (55%)	Lógica de la investigación (57%)	Indicaciones (estudio ficticio): Lee con cuidado el siguiente texto, una vez leído, elige la opción correcta según correspondan a las preguntas 11, 12, 13 y 14. Un psicólogo del desarrollo está interesado en la cantidad de conducta verbal que niños muy pequeños mostraban dependiendo de quién estaba presente. Los niños en el estudio tienen 3 años de edad. Los mismos fueron observados en un ambiente de laboratorio por un periodo de 30 minutos. La mitad de los niños fue asignada a la condición en la cual otro niño estaba presente durante la sesión de observación. El psicólogo midió: el número, la duración y la complejidad de las verbalizaciones de cada uno de los niños observados. 13. ¿Cuáles son las variables independientes manejadas en el estudio? a) Numero, duración y complejidad de las verbalizaciones de cada uno de los niños b) Niños pequeños dentro de la sesión de observación c) La presencia de un adulto o un niño durante la sesión de observación d) Número y duración de las verbalizaciones de cada uno de los adultos presentes	13	41	Básico

Contenidos oficiales a evaluados	Red curricular de pertenencia	Ítem muestra del contenido evaluado	Posición del ítem	% de estudiantes que acertaron el ítem	Niveles de logro
II.1.2.4. Población y selección de la muestra (61%)	Técnicas de muestreo (62%)	1. Subgrupo de población que se divide en segmentos, seleccionando una muestra para cada uno de ellos. a) Muestreo simple b) Muestreo estratificado c) Muestreo por racimos d) Muestreo dirigido	1	40	Básico
II.2.3.2. Elección del análisis estadístico (37%)	Análisis de los datos (37%)	65. Tipo de escala de medición en donde los valores se encuentran ordenados jerárquicamente y se desconoce la distancia entre ellos. a) Intervalar b) Ordinal c) Nominal d) Escalar o de razón	65	40	Básico
I.2.1.4. El planteamiento o del problema de investigación (57%)	Planteamiento del problema (56%)	27. El enunciado: "La desintegración del matrimonio provoca baja autoestima en los hijos e hijas", es un ejemplo de hipótesis: a) Comparativa b) Correlacional c) Explicativa d) Exploratoria	27	38	Básico
II.1.3.1. Ética e investigación (41%)	Ética de la investigación (42%)	48. Los psicólogos pueden prescindir del consentimiento informado si: a) se asume que la investigación no creará cierto daño o tensión b) los participantes están inhabilitados en la toma de decisiones c) es permitido por el reglamento de la institución científica d) el problema a estudiar tiene una alta relevancia social	48	38	Básico
I.1.2.2.2. Investigación cualitativa (37%)	Lógica de la investigación (57%)	9. Enfoque de investigación que trata de reconstruir la realidad desde la perspectiva del sujeto de estudio: a) Estudios correlacionales b) Estudios cuantitativos c) Estudios causales d) Estudios cualitativos	9	37	Básico

Contenidos oficiales a evaluados	Red curricular de pertenencia	Ítem muestra del contenido evaluado	Posición del ítem	% de estudiantes que acertaron el ítem	Niveles de logro
II.2.3.2. Elección del análisis estadístico (37%)	Análisis de los datos (37%)	17. Prueba estadística para evaluar si dos grupos difieren entre sí de manera significativa respecto a sus medias. a) Puntuaciones Z b) Coeficientes de correlación de Spearman c) Coeficiente de correlación de Pearson d) Prueba t de student	17	36	Por debajo del básico
II.1.1.3. Validez interna y externa de los diseños (36%)	Diseños de investigación (34%)	42. Cuando decimos que la "artificialidad" de las condiciones puede hacer que el contexto experimental resulte atípico, respecto a la manera en que se aplica regularmente el tratamiento, nos referimos a una fuente de invalidación: a) Interna b) Contextual c) Externa d) Mixta	42	35	Por debajo del básico
II.2.3.2. Elección del análisis estadístico (37%)	Análisis de los datos (37%)	18. Tipo de prueba estadística que analiza la relación entre dos variables, las cuales son medidas en nivel de intervalo o de razón. a) Chi cuadrada X^2 , b) Coeficientes de correlación de Spearman c) Coeficiente de correlación de Pearson d) Coeficientes de correlación de Kendall	18	35	Por debajo del básico
II.1.1.3. Validez interna y externa de los diseños (36%)	Diseños de investigación (34%)	25. Grado de confianza de que los resultados del experimento se interpreten adecuadamente y de que se obtienen con un alto grado de control. a) Equivalencia experimental b) Validez interna c) Confiabilidad d) Validez externa	25	33	Por debajo del básico
I.1.2.2.1. Investigación cuantitativa (55%)	Lógica de la investigación (57%)	12. ¿Cuál es el alcance del estudio? a) Explicativo b) Predictivo c) Correlacional d) Descriptivo	12	32	Por debajo del básico

Contenidos oficiales a evaluados	Red curricular de pertenencia	Ítem muestra del contenido evaluado	Posición del ítem	% de estudiantes que acertaron el ítem	Niveles de logro
I.2.2.1.3. Clasificación de los diseños cuantitativos (31%)	Diseño de investigación (34%)	37. Representan dos de los más importantes diseños de investigación de corte cualitativo. a) Análisis de contenido y cuasi-experimentos b) Transversales y longitudinales c) Narrativos e investigación acción d) No experimentales y pre-experimentales	37	31	Por debajo del básico
II.2.3.2. Elección del análisis estadístico (37%)	Análisis de los datos (37%)	44. Indica el procedimiento que se sigue para analizar los datos cuantitativos: a) Seleccionar y ejecutar el programa estadístico, analizar las hipótesis planteadas, explorar descriptivamente los datos, evaluar la confiabilidad y validez del instrumento, análisis adicional de los datos y preparación de los resultados para su presentación. b) Seleccionar y ejecutar el programa estadístico, explorar descriptivamente los datos, evaluar la confiabilidad y validez del instrumento, analizar las hipótesis planteadas, análisis adicional de los datos y preparación de los resultados para su presentación. c) Seleccionar y ejecutar el programa estadístico, evaluar la confiabilidad y validez del instrumento, explorar descriptivamente los datos, analizar las hipótesis planteadas, análisis adicional de los datos y preparación de los resultados para su presentación. d) Seleccionar y ejecutar el programa estadístico, evaluar la confiabilidad y validez del instrumento, explorar descriptivamente los datos, analizar las hipótesis planteadas, análisis adicional de los datos y preparación de los resultados para su presentación.	44	31	Por debajo del básico
II.1.1.3. Validez interna y externa de los diseños (36%)	Diseños de investigación (34%)	60. Fuente de invalidación interna que trata sobre acontecimientos que ocurren durante el desarrollo del experimento afectando a la variable dependiente. a) Maduración b) Inestabilidad c) Instrumentación d) Historia	60	27	Por debajo del básico

Contenidos oficiales a evaluados	Red curricular de pertenencia	Ítem muestra del contenido evaluado	Posición del ítem	% de estudiantes que acertaron el ítem	Niveles de logro
I.2.2.1.2. Características generales de los diseños cuantitativos (25%)	Diseño de investigación (34%)	36. Indica los requisitos para establecer la causalidad de un fenómeno. a) Validez externa e interna del diseño de investigación y la posibilidad de replicar los estudios. b) Medición del efecto de la variable Independiente en la variable dependiente, validez externa del diseño de investigación y la posibilidad de replicar los estudios. c) Medición del efecto de la variable Independiente en la variable dependiente, validez interna del diseño de investigación y la manipulación intencional de la variable independiente. d) Medición del efecto de la variable Independiente en la variable dependiente, validez interna del diseño de investigación y la posibilidad de replicar los estudios.	36	25	Por debajo del básico
II.1.2.2. Características específicas de los diseños cuantitativos (27%)	Diseños de investigación (34%)	35. A los diseños de investigación que pueden analizar en un momento determinado la relación entre autoestima y el temor al logro, se les conoce como: a) Experimentos puros b) Longitudinales c) Transeccionales d) Pre-experimentos	35	25	Por debajo del básico
II.1.2.3. Características específicas de los diseños cualitativos (42%)	Diseños de investigación (34%)	49. Diseño de investigación que tiene como rasgo principal que los datos se categorizan con codificación abierta, luego el investigador organiza las categorías resultantes en un modelo conceptual de interrelaciones. a) Teoría fundamentada b) Estudio de caso c) Estudio transeccional d) Experimento	49	24	Por debajo del básico
I.2.2.2.2. Características de los diseños cualitativos (22%)	Diseño de investigación (34%)	38. Los diseños de investigación en donde los procedimientos no son estandarizados, se les conoce como: a) Cuasiexperimentos b) Preexperimentos c) Cualitativos d) Exploratorios	38	22	Por debajo del básico

Contenidos oficiales a evaluados	Red curricular de pertenencia	Ítem muestra del contenido evaluado	Posición del ítem	% de estudiantes que acertaron el ítem	Niveles de logro
II.1.2.2. Características específicas de los diseños cuantitativos (27%)	Diseños de investigación (34%)	46. A los estudios longitudinales en donde la población es la misma, pero son diferentes los sujetos de las muestras, se les conoce como: a) Panel b) Tendencia c) Cohortes d) Evolución de grupos	46	22	Por debajo del básico
II.1.2.2. Características específicas de los diseños cuantitativos (27%)	Diseños de investigación (34%)	40. Tipo de diseño investigación que pertenece a los experimentos puros. a) Grupos de Salomón b) Test-retest con un solo grupo c) Cuasi-experimentos d) Estudio de caso con una sola medición	40	20	Por debajo del básico
I.2.1.4. El planteamiento del problema de investigación (57%)	Planteamiento del problema (56%)	29. Indica los puntos esenciales en la delimitación del tema de investigación: a) Justificar el tema de investigación, señalar los límites teóricos y temporales, y definir el alcance de estudio. b) Justificar el tema de investigación, definir el alcance de estudio, identificar el diseño de investigación y señalar los límites técnico-metodológicos. c) Señalar límites teóricos, temporales y espaciales, definir unidades de análisis y situar el problema en el contexto sociopolítico. d) Señalar límites teóricos, definir las unidades de análisis y el alcance de estudio, y situar el problema en el contexto sociopolítico,	29	19	Por debajo del básico
I.2.2.1.2. Características generales de los diseños cuantitativos (25%)	Diseño de investigación (34%)	56. Símbolo que representa "tratamiento", "estímulo" o "condición experimental" en un diseño de investigación: a) R b) E c) O d) X	56	17	Por debajo del básico
I.1.2.1.1. Método inductivo (16%)	Lógica de la investigación (57%)	33. Método lógico dentro de un proceso de investigación, en donde no necesariamente se cuenta con hipótesis. a) Metodología b) Lógica deductiva c) Lógica inductiva d) Método dialéctico	33	16	Por debajo del básico

Contenidos oficiales a evaluados	Red curricular de pertenencia	Ítem muestra del contenido evaluado	Posición del ítem	% de estudiantes que acertaron el ítem	Niveles de logro
I.2.2.1.2. Características generales de los diseños cuantitativos (25%)	Diseño de investigación (34%)	15. Símbolo que indica la medición de la variable dependiente en un diseño de investigación. a) R b) X_1 c) O d) X	15	10	Por debajo del básico
I.2.2.1.1. Los alcances de la investigación (51%)	Planteamiento del problema (56%)	34. Identifica el tipo de alcance de investigación que plantea la siguiente pregunta: ¿cuántas de las víctimas por secuestro en Tijuana presentan síntomas o trastorno de estrés posttraumático? a) Exploratorio b) Transaccional c) Correlacional-causal d) Descriptivo	34	8	Por debajo del básico
II.2.3.2. Elección del análisis estadístico (37%)	Análisis de los datos (37%)	45. Tipo de análisis que utiliza las pruebas de chi cuadrada X^2 y los coeficientes de correlación de Spearman y Kendall, entre otras. a) Paramétrico b) Descriptivo c) Multivariado d) No paramétrico	45	6	Por debajo del básico