VNIVERSIDAD NACIONAL
AVTNOMA DE
MEXICO

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE PSICOLOGÍA

"ADAPTACION PSICOMETRICA DEL INVENTARIO DE DEPRESION DE BECK-II EN RESIDENTES DE LA CIUDAD DE MEXICO"

TESIS

QUE PARA OBTENER EL TÍTULO DE LICENCIADA EN PSICOLOGÍA

PRESENTA:
ALINA HERNANDEZ POMAR

DIRECTOR DE TESIS: DR. SAMUEL JURADO CÁRDENAS

ESTA TESIS SE REALIZÓ CON AYUDA DEL CONACYT PROYECTO: 090369

MÉXICO, D. F.
Febrero, 2010

Facultad de Psicología

Universidad Nacional Autónoma de México
UNAM
Biblioteca Central
Dirección General de Bibliotecas de la UNAM

UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

AGRADECIMIENTOS

Antes que nada quiero agradecerle a Dios por haberme dado la oportunidad de llegar a la culminación de uno de los proyectos más importantes de toda mi vida: mi titulación.

Mi segundo agradecimiento se dirige a la Máxima Casa de Estudios de mi país, la Universidad Nacional Autónoma de México (UNAM), que me abrió sus puertas desde mi educación media superior y me ha seguido brindando la instrucción necesaria para el mejor desempeño como la profesionista en la que hoy me he convertido.

Al Dr. Samuel Jurado Cárdenas, por haberme dirigido la tesis que hoy presento, por el apoyo incondicional que siempre recibí de él y, sobre todo, por sus grandes enseñanzas que contribuyeron en gran medida a la realización de la misma.

A mi mesa de sinodales: la Mtra. Alma Mireia López-Arce Coria, a la Dra. Lilian Klein Stern, al Dr. José de Jesús González Núñez y al Mtro. Haroldo Elorza, porque de ellos no tan sólo tuve su aceptación para fungir como tales, sino -lo más importante- por su dedicación y empeño como mis profesores en la carrera.

A mi madre, por ser mi mejor amiga, la persona que ha seguido cada uno de mis pasos y que me ha apoyado en cada una de las decisiones que he tomado en el camino recorrido; por ser siempre mi confidente y

ayudarme a levantar cada vez que he caído, por estar conmigo en los momentos más difíciles, por compartir mis alegrías, mis tristezas, mis sueños y, hoy, un triunfo que es de ella también por desvelarse junto conmigo cuando las tareas se complicaban, por estar incondicionalmente para mí cuando más la necesité, pero sobre todas las cosas, por ser la mejor mamá que alguien pudiera tener.

A mi padre, por haberme dado siempre las bases necesarias para ser la persona que ahora soy, por su gran ejemplo del sentido de la responsabilidad ante las cosas que algún día me harían crecer como persona, por sus palabras de aliento cuando las cosas se volvían complicadas, por esa gran confianza que día a día me brinda para darme una mano e ir junto a mi en los momentos más importantes de mi vida, en pocas palabras, por ser mi padre.

A mi hermano, por haber sido siempre un ejemplo a seguir, porque fue una de mis más grandes inspiraciones para poder llegar hasta donde hoy me encuentro, porque siempre está dispuesto a ayudarme en lo que necesito, aún en tiempos difíciles, por cada uno de los consejos que siempre me brindó y sobre todo por cuidarme y estar siempre al pendiente de mi.

A mi hermana, por ser incondicionalmente mi amiga y por escucharme cuando más lo he necesitado, por sus palabras exactas en el momento

preciso y por estar allí, aún sin saber que su presencia siempre me hizo más fuerte.

A toda mi familia, por ser una de las bases más importantes de mi vida, por compartir conmigo este gran logro y por todos los buenos momentos que hemos compartido.

A mis mejores amigas: Berenice García y Monserrat Farfán, por tantos años de compartir sueños, metas, ilusiones, y muy especialmente a Araceli Carpinteiro, por ser uno de los pilares en mi vida, por estar presente en cada momento, por su apoyo en la realización de esta tesis y por cada uno de los años que hemos pasado juntas.

A Ximena Arciniega, por haber formado parte de mi vida durante 4 años de formación como psicólogas, por las noches de desvelo haciendo trabajos o estudiando para algún examen, por ser una de mis mejores amigas y, sobre todo, por sus valiosos consejos que me han ayudado a ser una mejor persona.

A Viridiana Muñiz porque antes de ser mi jefa, es mi amiga, por ser mi confidente, por sus palabras en cada una de las situaciones acontecidas y porque siempre me apoyó en mis estudios, a todo el equipo Plastilina y Krayolas por formar parte de mi vida durante 7 años y por cada uno de los momentos que hemos compartido juntos.

A Gymboree por la oportunidad de desempeñarme como psicóloga y brindarme las facilidades necesarias para cumplir uno de mis sueños al trabajar con niños pequeños.

Y, finalmente, a todos y cada uno de mis amigos, por formar parte en vida y dejar en mí una huella imborrable.

ÍNDICE

AGRADECIMIENTOS 2

CAPÍTULO 1

Trastornos del Estado de Ánimo 7

- Etiología 7
- Semiología 10
- Trastorno Depresivo Recurrente 18
- Árbol de Diagnóstico de los Trastornos del Estado de Ánimo 23

CAPÍTULO 2

Psicometría 24

- Historia de la Medición en Psicología 24
- Definición de los Tests Psicológicos 32
- Reactivos o Ítems que Conforman un Test 34
- Características de los Tests 37
- Adaptación de Tests para su Uso en Diferentes Idiomas y Culturas 50

CAPÍTULO 3

Desarrollo del Inventario de Depresión de Beck (BDI) 62

- Historia 64
- Desarrollo de los Ítems 70
- Propiedades Psicométricas del BDI 71

Planteamiento del problema 75
Justificación 76
Método 76
Resultados 79
Discusión y Conclusiones 83
Referencias 86

CAPITULO 1

TRASTORNOS DEL ESTADO DE ÁNIMO

Trastornos del Estado de Ánimo

Los trastornos del estado de ánimo son un conjunto de enfermedades en las que la principal manifestación es una alteración en el estado de **ánimo** y del **afecto**, que puede ser bajo (ánimo depresivo) o de un ánimo elevado (la euforia que caracteriza a los cuadros maniacos).

Es importante señalar las diferencias entre estado de ánimo que describe el estado emocional subjetivo de un individuo, y de afecto que es la parte observable u objetiva del estado de ánimo.

Se trata de un grupo de trastornos que poseen una enorme relevancia, dada su frecuencia y capacidad de generar discapacidades.

- **Etiología**

La causa precisa de los trastornos del ánimo no es conocida, pero existe abundante información que muestra que en su génesis participan factores biológicos y psicosociales. Existen claras evidencias de la interacción entre ellos, los factores biológicos, psicológicos, ambientales y culturales se conjugan con una ponderación variable y propia de cada sujeto afectado, condicionando la aparición de episodios

anímicos específicos en diferentes momentos de la vida. Así, un individuo con gran carga genética podría manifestar la enfermedad, incluso, con una modesta contribución psicosocial.

Aspectos biológicos

Factores genéticos:

El trastorno **depresivo mayor** es 1 a 3 veces más frecuente en los familiares de primer grado de las personas con este trastorno, que en la población general.

También en estos familiares existe un riesgo aumentado de trastorno dístimico, dependencia al alcohol (familiares adultos) y de trastorno por déficit de atención con hiperactividad (hijos)

La tasa de concordancia para trastorno bipolar es aproximadamente un 19% en mellizos dicigóticos, pero alcanza un 79% en gemelos monocigóticos. La probable heterogeneidad genética de los trastornos del ánimo, sugiere un modo complejo de herencia y la posibilidad de que varios genes contribuyan a la vulnerabilidad individual para sufrirlos. Análisis de hermanos afectados por trastorno bipolar y estudios que incluyen numerosas familias con alta prevalecía de trastornos del ánimo, encontraron evidencias de la existencia de un gen de susceptibilidad ubicado en el cromosoma 18.

Un vasto número de estudios han comunicado varias anormalidades en productos metabólicos de la serotonina, dopamina y norepinefrina, en sangre, orina y liquido cefalorraquídeo de pacientes con trastornos del ánimo.

A pesar de los significativos avances en nuestro entendimiento de los mecanismos fisiológicos y moleculares que gobiernan la neuroquímca cerebral, hasta la fecha no se dispone de un procedimiento de laboratorio que permita diagnosticar o subtipificar clínicamente los trastornos del animo.

Aspectos psicosociales

Los trastornos del ánimo. Se ha observado una cierta vulnerabilidad a desarrollar depresión clínica en individuos con alteraciones de la personalidad, particularmente aquéllos con un estilo muy perfeccionista, inflexible o rígido, y emocionalmente son muy dependientes.

Los trastornos suponen una crisis en el funcionamiento psicobiológico de las personas con una cierta vulnerabilidad (disposición o características de personalidad) ante determinados estresores (situaciones que generan tensión y necesidades de adaptación). La vulnerabilidad incluye predisposiciones determinadas genéticamente que, a su vez,

influyen determinados niveles de sustancias bioquímicas en el cerebro del individuo.

También incluyen procesos y habilidades psicológicas, como los estilos de pensamiento personales y las habilidades de conducta para afrontar determinadas situaciones.

Existen otros factores psicosociales, como la tensión familiar, escaso apoyo social y falta de estrategias de afrontamiento.

- **Semiología**

  Signos y Síntomas

  <u>Episodio depresivo:</u>

  - Disminución de la atención y concentración.

  - Pérdida de la confianza en sí mismo y sentimientos de inferioridad.

  - Ideas de culpa y de ser inútil.

  - Perspectiva sombría del futuro.

  - Pensamientos y actos suicidas o de auto agresiones.

- Trastornos del sueño.

- Pérdida del apetito.

- Tristeza.

- Pérdida de interés.

- Fatigabilidad (disminución de la actividad).

Episodio maniaco:

- Aumento de actividad o inquietud psicomotriz.

- Aumento notable de la locuacidad (logorrea).

- Fuga de ideas o experiencia subjetiva de pensamiento acelerado.

- Pérdida de las inhibiciones sociales que da lugar a comportamientos inadecuados para las circunstancias y para el carácter del individuo.

- Disminución de las necesidades de sueño.

- Aumento exagerado de la estima de sí mismo o ideas de grandeza.

- Facilidad para distraerse o cambios constantes de actividades o de planes.

- Comportamientos temerarios o imprudentes que implican riesgos no reconocidos como tales por el individuo, por ejemplo, grandes gastos de dinero, proyectos insensatos o conducción imprudente.

- Marcado aumento en el vigor sexual o indiscreciones sexuales.

Episodios depresivos

Esto es un panorama general de todos los síntomas involucrados en los episodios de la depresión y se han clasificado en cuatro grandes grupos, síntomas emocionales, cognitivos, volitivos y somáticos:

1.- Los síntomas emocionales son la característica fundamental de la depresión e incluyen cambios en el estado de ánimo como:

-Tristeza.

-Pesimismo.

-Falta de ilusión.

-Desesperanza.

-Tendencia al llanto.

-Anhedonia (incapacidad para disfrutar con las cosas) que en los casos graves puede ser total, con falta de reacción incluso a estímulos placenteros.

2.-Síntomas cognitivos, derivan de la modulación del ánimo sobre la valoración negativa que el individuo hace de sí mismo del mundo y del futuro.

-Expresa ideas de minusvalía e inutilidad.

-Incapaz de realizar tareas que antes hacia sin dificultad.

-Ideas de muerte también llamado “tedio vital”.

-Ideas autolíticas (cuando la planificación suicida ya es clara).

-Disminución de la atención y concentración: asociada a un aparente deterioro de memoria, se da en pacientes de edad avanzada y da lugar a la llamada “seudodemencia depresiva”.

3.-Síntomas volitivos

*Abulia.

*Apatía.

*Falta de iniciativa y capacidad de decisión.

*Ánimo bajo y la energía.

*Conlleva al deterioro y abandono de la actividad laboral / doméstica.

4.-Síntomas somáticos, afectan a la diversidad de ritmos biológicos:

*Falta de energía o impulso.

*Bradicinesia (lentitud de los movimientos).

*Trastornos del sueño y apetito.

*El sueño se afecta en la mayoría de los casos mostrando un insomnio de conciliación, sueño fragmentario o despertar precoz.

*Pueden aparecer otras alteraciones como pesadillas o mioclonias nocturnas

*El apetito disminuido de forma típica, con falta de gusto por las ingestas y en ocasiones disminución de peso.

*Disminución del deseo sexual.

* Otros tipos de somatización con cefaleas o lumbalgias.

Una primera clasificación para cualquier episodio depresivo es la especificación de la gravedad. Tanto el CIE-10 como el DSM-IV consideran que se puede clasificar en ***leve, moderada y grave.***

La distinción se realiza sobre la base del número total de síntomas que presenta el paciente. Una implicación fundamental será el grado de incapacidad funcional producido por cada tipo.

Pautas para su diagnóstico:

Deben cumplirse dos de los criterios del primer grupo y dos del segundo durante al menos dos semanas.

Al menos dos de los siguientes:

*Tristeza (humor depresivo).

*Pérdida de interés.

*Fatigabilidad (disminución de vitalidad).

Y al menos dos de los siguientes:

*Disminución de atención y concentración.

*Pérdida de confianza en sí mismo y sentimientos de inferioridad.

*Ideas de culpa y de ser inútil.

*Perspectiva sombría del futuro.

*Pensamientos de auto agresiones o actos suicidas.

*Tx del sueño.

*Pérdida de apetito.

Presentando dos síntomas de cada grupo, debe diagnosticarse **episodio depresivo leve.**

Para diagnosticar un **episodio depresivo moderado,** se deben tener dos del primer grupo y al menos tres o cuatro del segundo grupo.

Para diagnosticar un **episodio depresivo grave,** deben estar presentes los tres primeros síntomas y al menos cuatro del segundo grupo, además de una intensidad grave.

Además de cumplir los criterios generales, la depresión se puede clasificar, de acuerdo con la clínica predominante, en diversos subtipos y son:

***Síntomas psicóticos:**

- Delirios o alucinaciones.
- Delirios de ruina y culpa.
- Hipocondríacos.

***Síntomas atípicos:**

- Aumento del apetito y peso.
- Aumento en las horas de sueño.
- Sensación de intensa pesadez, “flojera”, en brazos y piernas.
- Marcada sensibilidad al rechazo personal.

***Síntomas melancólicos:**

- Anhedonia completa con falta de reactividad a estímulos agradables, predominio matutino de los síntomas.
- Despertar precoz (en vez de insomnio de conciliación).
- Pérdida de peso.
- Enlentecimiento psicomotor.
- Mutismo.
- Negativismo o inmovilidad.

Depresión con patrón estacional:

- Episodios depresivos en la época otoño- invierno, mejoría en primavera.

- **Trastorno Depresivo Recurrente**

Este trastorno se caracteriza por la aparición de períodos repetidos de depresión, que pueden presentar los rasgos de episodio depresivo leve, moderado o grave, pero sin antecedentes de episodios aislados de exaltación del estado de ánimo o aumento de la vitalidad suficientes para satisfacer las pautas de manía.

Los criterios para el diagnóstico de Trastorno Depresivo Recurrente son:

- Presencia de dos o más episodios depresivos. Tiene que haber un intervalo de al menos dos meses seguidos en los que no se presenten episodios de exaltación muy grandes como para parecerse a la manía.

- Nunca se ha producido un episodio maniaco, un episodio mixto o un episodio hipomaniaco.

El trastorno depresivo recurrente puede subdividirse como se indica a continuación, especificando primero el tipo de episodio actual y después el tipo que predomina en todos los episodios.

Trastorno Depresivo Recurrente, episodio actual leve

Pautas para el diagnóstico:

Se deben satisfacer todas las pautas para el trastorno depresivo recurrente y el episodio actual debe satisfacer las pautas para episodio depresivo leve. Además, por lo menos dos episodios deben haber durado un mínimo de dos semanas, y entre ellos debe haber un intervalo libre de alteración significativa del humor de varios meses de depuración.

Puede presentarse con síndrome somático o sin él.

Trastorno Depresivo Recurrente, episodio actual moderado

Pautas para el diagnóstico

Se deben satisfacer todas las pautas del trastorno depresivo recurrente y el episodio actual debe satisfacer las pautas para episodio depresivo moderado. Además, por lo menos dos episodios deben haber durado un mínimo de dos semanas, y deben haber estado separados por un periodo de varios meses libres de alteración significativa del humor.

Puede presentarse con síndrome somático o sin él.

Trastorno Depresivo Recurrente, episodio actual grave sin síntomas psicóticos

Pautas para el diagnóstico

Se deben satisfacer todas las pautas del trastorno depresivo recurrente y el episodio actual debe satisfacer las pautas para episodio depresivo grave sin síntomas psicóticos. Además, por lo menos dos episodios deben haber durado un mínimo de dos semanas y deben haber estado separados por un periodo de varios meses libres de alteración significativa del humor.

Curso y Pronóstico

El Trastorno Depresivo Recurrente puede empezar a cualquier edad; la edad promedio de inicio es la mitad de la tercera década de la vida. El curso es variable, algunas personas tienen episodios aislados separados por muchos años sin síntomas depresivos, en tanto que otras tienen episodios agrupados, e incluso otras tienen episodios cada vez más frecuentes a medida que se hacen mayores. Los períodos de remisión duran más en las fases tempranas del curso del trastorno.

La mayor parte de los estudios hallan que el 70-80% de los pacientes se recuperan antes de dos años, pero si ya no mejoran en ese tiempo es improbable que lo hagan. Un 20% de pacientes se pueden cronificar. Los factores que predisponen a esta cronificación son un bajo nivel intelectual previo, rasgos de neuroticismo, histeria, hipocondría, dependencia o paranoidismo e historia de alcoholismo.

El trastorno depresivo recurrente se asocia a una alta mortalidad. Las tasas de suicidio llegan al 15% y las tasas de mortalidad son más altas en comparación con pacientes de la misma edad sin un trastorno depresivo recurrente.

Trastorno Depresivo Recurrente, episodio actual grave con síntomas psicóticos

Pautas para el diagnóstico

Se deben satisfacer todas las pautas del trastorno depresivo recurrente y el episodio actual debe satisfacer las pautas para episodio depresivo grave con síntomas psicóticos. Además, por lo menos dos episodios deben haber durado un mínimo de dos semanas y deben haber estado separados por un periodo de varios meses libres de alteración significativa del humor.

Trastorno Depresivo Recurrente, episodio actual en remisión.

Pautas para el diagnóstico

En el pasado debieron satisfacerse todas las pautas de trastorno depresivo recurrente, pero el estado actual no satisface las pautas para episodio depresivo de cualquier gravedad o para cualquier otro trastorno. Además, por lo menos dos episodios deben haber durado un mínimo de dos semanas y deben haber estado separados por un período de varios meses libres de alteración significativa del humor.

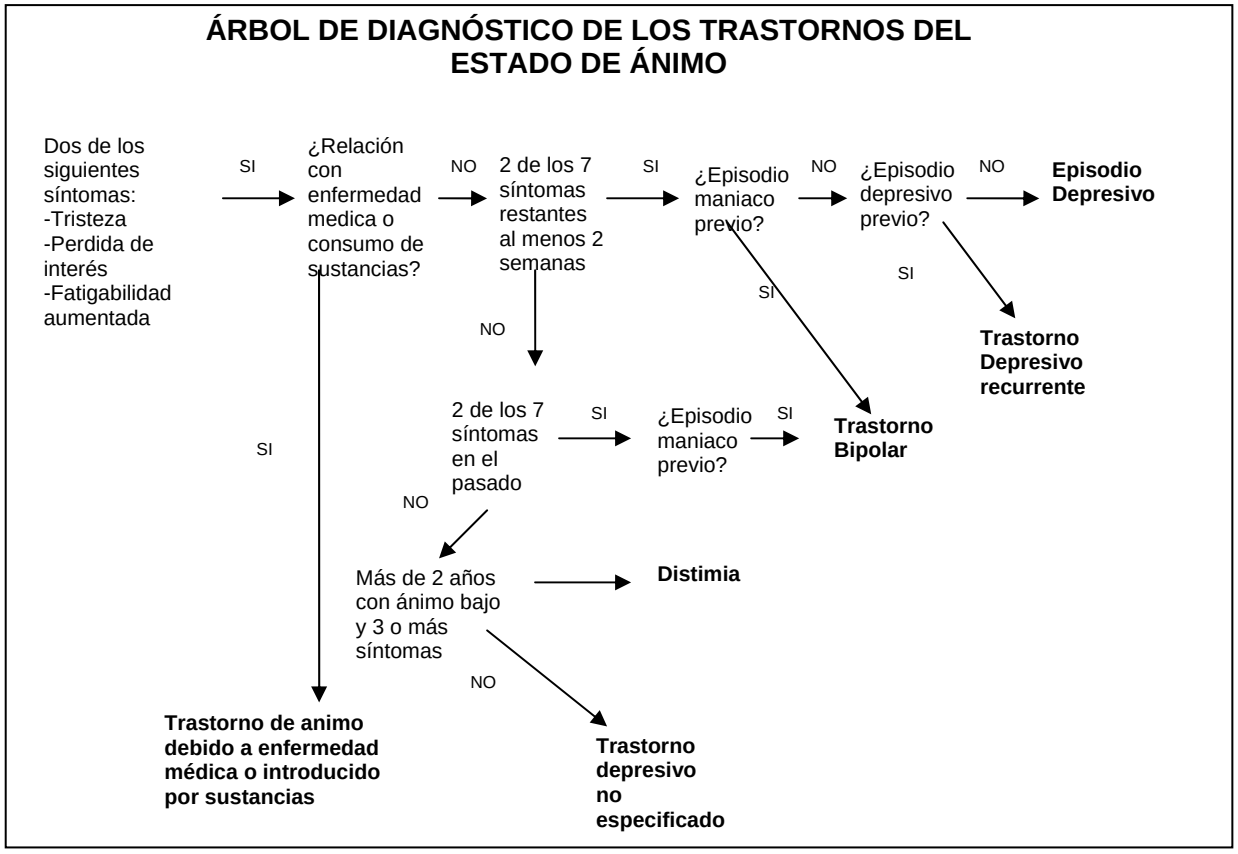

ÁRBOL DE DIAGNÓSTICO DE LOS TRASTORNOS DEL ESTADO DE ÁNIMO
Dos de los siguientes síntomas:
-Tristeza
-Perdida de interés
-Fatigabilidad aumentada
SI
¿Relación con enfermedad medica o consumo de sustancias?
NO
2 de los 7 síntomas restantes al menos 2 semanas
SI
¿Episodio maniaco previo?
NO
¿Episodio depresivo previo?
NO
Episodio Depresivo
SI
Trastorno Depresivo recurrente
SI
NO
2 de los 7 síntomas en el pasado
SI
¿Episodio maniaco previo?
SI
Trastorno Bipolar
SI
NO
Más de 2 años con ánimo bajo y 3 o más síntomas
Distimia
NO
Trastorno de animo debido a enfermedad médica o introducido por sustancias
Trastorno depresivo no especificado

Figura 1.

CAPITULO 2

PSICOMETRÍA

- **Historia de la Medición en Psicología**

El origen de los tests psicológicos surgió debido al interés por conocer las diferencias humanas, por un lado, y para poder tomar decisiones a partir de los datos obtenidos (Nunnally, 1970; Pichot, 1976; Anastasi, 1988; Cerdá, 1984) Sin embargo, se tienen datos sobre los primeros intentos de estudiar a los seres humanos en sus diferencias y semejanzas desde la época de los mandarines chinos que implementaron un servicio civil para determinar si sus funcionarios eran aptos para desempeñar sus deberes gubernamentales; por otro lado, tanto Platón como Aristóteles ya hablaban hace 2 500 años sobre las diferencias individuales (Aiken, 1995; Gregory, 1993)

En el intento por conocer las diferencias humanas se realizaron diversas investigaciones que van desde el estudio de las sensaciones del ser humano, hasta el nivel intelectual y las emociones de cada persona.

Estas investigaciones tuvieron gran importancia dentro de la psicología, ya que llevaron a la creación de las pruebas o tests psicológicos que

hasta la fecha permiten conocer diversas características de los seres humanos.

Entre los primeros estudios de las diferencias humanas, se encuentra el trabajo del español J. Huarte de San Juan quien, en el siglo XVI, daba normas para descubrir habilidades e ingenios y establecer diferencias mediante su *Examen de ingenios.*

En el siglo XIX en Francia, ante el trato inadecuado que sufrían los enfermos mentales, surge el interés por estudiarlos y hacer clasificaciones que permitieran poder darles un tratamiento adecuado. En el año de 1838, el psiquiatra Esquirol publicó: "*Des maladies mentales considerees sous les rapportes medical hygienique et medico-legal",* donde realizó la primera clasificación de los enfermos mentales, la cual se basó en el uso del lenguaje que podía tener un sujeto, ya que consideraba que éste era un criterio adecuado para determinar el nivel intelectual (Cerdá, 1984)

Es importante mencionar que el trabajo que realizó Esquirol sirvió como referencia para la construcción de las primeras escalas de inteligencia y desarrollo, como las de Binet-Simon y Gessell.

Wundt, en el año de 1879, crea el primer laboratorio de psicología experimental, que tenía como finalidad formular leyes sobre la

conducta humana; en este laboratorio se trabajó con diversos sujetos, observando que no todos reaccionaban de la misma manera, por lo que no se podían realizar generalizaciones, por lo tanto, fue necesario que todos trabajaran bajo las mismas circunstancias y se estudiaran entonces las diferencias individuales. Los experimentos que ahí se realizaban tenían influencia médica y fisiológica, ya que se hacían estudios sobre las reacciones de los sentidos ente diversos estímulos.

Una de las aportaciones de Galton fue el uso de procedimientos estadísticos en los datos que reportaban el uso de los tests; dichos procedimientos los desarrolló con su discípulo K. Pearson, quien estableció la fórmula del coeficiente de correlación, la correlación parcial, correlación múltiple y el análisis factorial.

En 1904, Spearman establece el uso del análisis factorial al utilizar el coeficiente de correlación de Pearson en los resultados de diferentes tests.

James McKeen Cattell, un psicólogo norteamericano que fue colaborador de Galton y discípulo de Wundt, siguió los estudios de las diferencias individuales y el uso de los tests en los Estados Unidos. En el año de 1890, publicó un artículo llamado *Mental Test and Measurements,* en la revista *Mind,* (Cerdá, 1984), donde utilizó por primera vez el término test mental para designar una serie de pruebas o

instrumentos psicológicos que se empleaban para conocer las diferencias individuales y la inteligencia. Cattell consideraba, al igual que Galton, que la inteligencia se podía medir a través de la discriminación sensorial y el tiempo de reacción.

Krapelin, quien es considerado el padre de la Psiquiatría, en el año de 1895 crea tests para explorar funciones más complejas como la elaboración de conceptos, enjuiciamiento práctico, fatiga mental, etcétera.

En 1897, Ebbinghaus elabora instrumentos para medir memoria y cálculo aritmético.

Los italianos Guicciardi y Ferrari continuaron el trabajo de Krapelin y diseñaron pruebas destinadas a la exploración de los enfermos mentales.

A finales del siglo XIX, A. Binet y Herni realizan un trabajo llamado *La psychologie individuelle,* donde se oponen totalmente a la teoría de Locke y sus seguidores, que sostenía que la inteligencia se podía medir a través de los sentidos, argumentando que ésta no tomaba en cuenta las diferencias individuales. Propusieron entonces diversos tests que incluían la medición de la memoria, la comprensión, la imaginación y la atención.

Binet y Simon publican, en 1905, en *L´Anné Psychologique,* un artículo llamado *"Nuevos métodos para el diagnóstico del nivel intelectual de los anormales"* (Pichot, 1976), donde se incluía el primer test mental práctico. Esta escala contaba con problemas ordenados en grado de dificultad y de acuerdo al ambiente cultural. Sus subtests medían el juicio, la comprensión y el razonamiento, dentro de sus reactivos se incluía el nombrar objetos, completar frases o comprender preguntas. Este test se presentó como un instrumento provisional y no contaba con ningún método para su calificación.

En 1908, apareció la primera revisión de la Escala Binet-simon: se eliminaron los tests que no eran satisfactorios de la versión anterior y se agregaron nuevos tests. Además, se realizó una clasificación de los tests de acuerdo a diversos niveles de edad. En esta escala se utilizó por primera vez el término "*edad mental"*, que correspondía a los niños cuya ejecución igualaba a la edad cronológica (Anastasi, 1988).

La escala Binet-Simon despertó un gran interés en diversas partes del mundo, se realizaron diversas traducciones y adaptaciones, una de las más importantes fue la que llevó a cabo L.M. Terman en la Universidad de Stanford, en 1916, conocida como *Escala Stanford-Binet,* en la cual Stern sugirió el término *"Coeficiente Intelectual",* que consideraba la relación entre la edad mental y la edad cronológica (Anastasi, 1988).

En 1917, cuando el gobierno de los Estados Unidos decide participar en la Primera Guerra Mundial, se enfrenta ante el problema de seleccionar rápidamente a un millón y medio de personas para el ejército y elegir su cargo, por lo que *El Ministerio del Ejército* forma un comité dirigido por R.K. Yarkes para investigar acerca de todos los procedimientos disponibles para elaborar dicha selección.

Durante esa época, Arthur S. Otis había creado un test colectivo de inteligencia, formado con preguntas de información general, razonamiento y aritmética, que presentó ante el comité, en donde se revisó y adaptó para poderse emplear en la selección de los reclutas. A este test se le conoció como *"Test Alfa" ("Army Alpha Test")* (Cerdá, 1984). Se construyó además un test complementario para personas analfabetas y con poco dominio del idioma inglés, que incluía rompecabezas geométricos y análisis de ilustraciones, el *Army Beta.* Estos tests fueron cedidos para posterior uso civil.

En 1939, aparece la primera Escala Wechsler-Bellevue, que sirve para medir inteligencia, siendo ésta la primer escala para adultos en la que se llevó a cabo una estandarización en función de su edad cronológica.

Anastasi (1988) menciona que el desarrollo de los tests psicológicos, en un principio se orientó principalmente a medir diversas funciones y, a partir de éstas, determinar el nivel de inteligencia; estos instrumentos medían en su mayoría aptitud verbal y, en menor grado, aptitudes para manejar relaciones numéricas, abstractas y simbólicas. Posteriormente, se descubrió que estas medidas proporcionaban poca información de cada función, por lo que fue necesario crear tests que midieran funciones específicas.

Al mismo tiempo en que se descubría la necesidad de elaborar tests para medir las diversas aptitudes de los individuos, se estudiaba la organización de los rasgos, por lo que se facilitó la elaboración de éstos. En esta área, Spearman realizó una de las mayores aportaciones con sus investigaciones estadísticas en estudios de intercorrelaciones de la naturaleza de la inteligencia y las puntuaciones en otras pruebas; y sus estudios sobre el análisis factorial, en donde se descubrió que en las puntuaciones de una prueba están presentes diversos factores o rasgos independientes.

Por otro lado, en el ámbito escolar se experimentaban nuevas técnicas, dentro de las cuales se encontraba la aplicación de pruebas tipificadas de rendimiento, siendo éstas las primeras que consideraban importante que todos los sujetos se sometieran a las mismas circunstancias. En estas pruebas se trataba de abarcar contenidos más amplios y eliminar la

posibilidad de favoritismo por parte del examinador durante el proceso de evaluación (Mann, citado en Anastasi, 1988).

Dentro del estudio de la personalidad, se crearon tests para evaluar los intereses, actitudes y opiniones; dentro de los primeros que se elaboraron, se encuentra el que fue creado en el Seminario para graduados celebrado en el Instituto de Tecnología Carnegie 1919-1920 (Fryer, 1931, citado en Anastasi, 1988), esta prueba se denominaba Strong Vocational Interest Blank (SVIB). El grupo del Carnegie creó diversos inventarios de intereses, los cuales presentaban preferencias o aversiones de los sujetos para una variedad de actividades específicas, objetos o personas que se encuentran comúnmente en la vida diaria (Anastasi, 1988). Posteriormente, se crearon otros inventarios, dentro de los cuales destaca la elaboración del MMPI (1939) (Inventario Multifásico de la Personalidad Minnessota) y el CPI (California Psychological Inventory).

Anastasi (1988) señala que para complementar el SVIB, se elaboró un inventario similar de intereses para diversas ocupaciones especializadas y semiespecializadas. Este inventario se llamó *Minnessota Vocational Interest Inventory,* formado por 21 escalas para ocupaciones tales como panaderos, conductores, reparadores de televisión, entre otros.

Otro de los inventarios más conocidos para evaluar intereses son las diversas formas y versiones del Kuder, dentro de las cuales se encuentra el *Kuder Preference,* el *Kuder Occupational Interest Survey* y el *Kuder General Interest Survey.*

En lo referente a la historia de la medición en psicología, es importante señalar que aquí solamente se han citado algunos de los tests que marcaron un paso importante en la historia, y que el desarrollo de los tests continúa hasta la fecha.

- **Definición de los Tests Psicológicos**

La conceptualización y definición de las pruebas o tests psicológicos varía de acuerdo a cada autor, a continuación se presenta una definición breve de los autores más relevantes:

- Tyler, L. (1972) señala que: "*Una prueba puede definirse como una situación estándar diseñada para tomar una muestra del comportamiento de un individuo*".
- Pichot, P. (1976) da la siguiente definición: "*Un test es una situación experimental estandarizada, sirviendo de estímulo a un comportamiento*".
- Anastasi (1988) considera que un test tiene como finalidad medir las diferencias que existen entre diferentes individuos o las reacciones del mismo individuo en diferentes ocasiones. Por lo

tanto, Anastasi define: "*un test psicológico constituye una medida objetiva y tipificada de una muestra de conducta*".

- Brown (1980) define a una prueba como el: *"Procedimiento sistemático para medir una muestra de conducta", "El procedimiento sistemático se refiere a que una prueba se construye, se administra y se aplica, según las reglas preestablecidas".*

"Test" es una palabra inglesa que significa "prueba" y que se deriva del latín "testis". Esta misma raíz figura en palabras españolas tales como testimonio o testigo. La palabra "test" se utiliza sin traducir en todos los países del mundo y sirve para designar una modalidad de exploraciones muy extendida hoy en día en diversos campos científicos y técnicos, entre ellos el psicológico (Cerdá, 1978).

Cerdá (1978) menciona la definición que fue aceptada en la Antigua Asociación Internacional de la Psicotécnica de H. Pieron que dice: *"Test es una prueba definida, fijando una tarea a realizar, idéntica para todos los sujetos examinados. Debe disponer de una técnica precisa, que permita discriminar entre soluciones satisfactorias y erróneas, o bien de un sistema numérico que permita puntuar el resultado. La tarea puede consistir en poner de manifiesto conocimientos adquiridos (tests pedagógicos) o bien funciones sensoriomotoras o mentales (tests psicológicos)".* Cerdá señala que esta definición sólo incluye tests de

inteligencia, aptitudes y conocimientos, y no toma en cuenta los instrumentos de personalidad.

Considerando la definición de los diversos autores, se puede concluir que una prueba o test psicológico es una situación estandarizada, que permite tener una medida objetiva de la conducta de un individuo.

- **Reactivos o Ítems que Conforman un Test**

En el proceso de la elaboración de una prueba o test, se deben tomar en cuenta, como elemento básico de éstos, los llamados *reactivos o ítems* que lo conforman. Para ello, autores como Aiken (1995), Kline (1986), Redil (1990) hacen las siguientes sugerencias para la elaboración de los reactivos o ítems:

- Se deberán usar palabras sencillas, que sean conocidas por todos los respondientes potenciales.
- Hacer la pregunta lo más concisa posible, ésta no deberá tener más de 20 palabras.
- Formular la pregunta de manera que produzca únicamente la información deseada, evitando reactivos que no estén relacionados con la variable.
- Evitar preguntas ambiguas o que puedan ser interpretadas de diferente manera.
- Evitar preguntas que sugieran la respuesta.

- Cada reactivo o pregunta debe tener sólo una idea, oraciones simples.
- La sintaxis debe ser sencilla y clara: sujeto, verbo y complemento.
- Se debe tener cuidado de no emplear frases que puedan afectar el prestigio del respondiente.
- Decidir si se van a personalizar algunos de los reactivos o ítems, o si se redactarán en tercera persona o en forma impersonal.
- Permitir que se den todas las posibles respuestas, principalmente en los reactivos de opción múltiple.
- Las alternativas de los reactivos de opción múltiple deben ser realistas.
- Emplear con cuidado o casi nunca palabras universales como “siempre”, “ninguno” (sólo para definir los extremos continuos) o aquéllas como: “sólo”, “apenas”, “casi”, “a veces”, porque introducen ambigüedad ya que tienen diferentes significados para las personas.
- Se debe evitar el usar dobles negativos, ya que también producen ambigüedad.
- Tratar de mantener al mínimo la cantidad de escritura por parte del que responde, pidiéndole tan sólo que escriba SI o NO; escriba un número o una letra, ponga una cruz (x), subraye la respuesta que desee, circule la respuesta escogida, etc.
- Evitar en lo posible términos dentro del reactivo que reflejen sentimientos.

- En los reactivos de opción múltiple coloque las opciones en formato (párrafo) de grupo en lugar de una tras otra; utilice números para designar los reactivos y letras para las opciones.

Tipos de reactivos o ítems

Los reactivos que conforman un instrumento representan procedimientos para obtener información sobre los individuos, pero la cantidad y la clase de información varía con la naturaleza de las tareas que implican los distintos tipos de reactivos (Aiken, 1995). Si por ejemplo, le pedimos a un sujeto que compare la batalla de la Noche Triste con la batalla de la toma de Tenochtitlán, esto exige una clase de respuesta diferente, que aquélla que se obtiene si le pedimos al sujeto que indique una serie de eventos que ocurrieron en cada batalla. El primer reactivo requiere capacidades complejas de integración y organización, en el segundo sólo se necesita memoria cognoscitiva. Ahora bien, existen reactivos de ensayo o abiertos y reactivos objetivos que pueden ser de respuesta corta o completamiento, de falso y verdadero, de aparejamiento y de opción múltiple.

- **Características de los Tests Psicológicos**

Toda prueba, test o instrumento de medición psicológica, debe de cumplir con características específicas para que los datos que ésta proporcione sean objetivos, de utilidad y confiables. De ahí que un test debe ser *sensible, objetivo, confiable y válido. Sensible* ya que los

reactivos que componen el instrumento pueden diferenciar entre aquellos sujetos que tienen el rasgo de aquellos que no. *Objetivo* se refiere a que dos aplicadores o más obtengan el mismo puntaje al calificar el instrumento. *Confiable,* que sea lo más preciso en el rasgo que está midiendo. *Válido*, que mida lo que pretende medir.

- **Grado de dificultad y poder discriminativo de los reactivos o ítems**

Una de estas características es el *grado de dificultad* (Anastasi, 1988) o la *sensibilidad de discriminación* (Pichot, 1976) que deben tener los reactivos o ítems que forman una prueba o test psicológico. Esta característica permite conocer las diferencias individuales, y se logra mediante un procedimiento que se lleva a cabo en el momento de construir el test. El procedimiento consiste en agregar *reactivos o ítems* que sólo puedan ser contestados por todos los sujetos y *reactivos o ítems* que sólo puedan ser contestados por aquellos más aptos. Generalmente, en las pruebas psicológicas que evalúan el rendimiento académico, inteligencia y aptitudes, los reactivos son ordenados en grado de dificultad, colocando primero los reactivos más fáciles y posteriormente los de mayor dificultad, esto se hace con el fin de que los sujetos sientan confianza al empezar la prueba y permitir que los sujetos con menor capacidad puedan dedicar mayor tiempo a aquellos reactivos que pueden resolver correctamente.

El poder discriminativo describe el grado en que la probabilidad de la respuesta alfa (aprobar un reactivo, calificando como correcto contra incorrecto, responder un reactivo en dirección clave) se correlaciona con el atributo. Un reactivo que no discrimina debe ser eliminado del instrumento. El índice clásico más común de la discriminación es la correlación con personas entre la respuesta alfa y el puntaje total de la prueba, la correlación reactivo total. (Nunnally, 1995).

- **Estandarización**

La *tipificación o estandarización* es otra de las características que debe incluir una prueba o test psicométrico, ésta consiste en utilizar los mismos procedimientos para aplicar y calificar los instrumentos en todos los sujetos. La tipificación abarca, pues, materiales, límites de tiempo, instrucciones, demostraciones previas, formas de resolver dudas a los sujetos durante la aplicación. El hecho de tener bajo las mismas condiciones a los sujetos permitirá, posteriormente, generar normas que permitan hacer comparaciones entre éstos, así como el tratar de minimizar al máximo errores durante la ejecución.

Una de las características más importantes que debe cubrir una prueba o test psicométrico, es que éste tenga un grado aceptable de confiabilidad y validez. A continuación se presenta una breve definición

de ambos términos y, posteriormente, se revisará en detalle cada uno de éstos.

La *confiabilidad* de una prueba o test "*se refiere a la consistencia o estabilidad de las puntuaciones obtenidas de un mismo sujeto en diferentes ocasiones o con conjuntos distintos de elementos equivalentes"* (Anastasi, 1988).

La validez *"es el grado en que un test psicológico mide lo que realmente pretende medir"* (Anastasi, 1988), es decir, los elementos que conformen la prueba o test psicométrico deben ser representativos de la conducta que se desea medir.

- **Confiabilidad**

Al realizar diversas mediciones sobre un mismo objeto, es muy probable que se cometan errores entre una y otra medición, por lo que éstas pueden no ser exactas o verse alteradas. Al medir atributos psicológicos, nos encontramos con que muchos de éstos no son estables, y a veces, aunque se utilice el mismo instrumento en distintas ocasiones, las puntuaciones obtenidas de un mismo sujeto pueden variar. Por tal motivo, es importante determinar si las variaciones en las puntuaciones de los sujetos se deben a un cambio real en el atributo medido o a un error cometido durante el proceso de medición.

Para tal efecto, la teoría psicométrica contempla el estudio del *error de medición* al calcular la confiabilidad de una prueba o test. Se considera *error de medición* a todas aquellas condiciones ajenas o imprevistas que se presentan durante el proceso de medición y que no fueron contempladas cuando se diseñó la prueba o test psicométrico.

Nunnally (1995) menciona que el *error de medición* es una combinación de procesos sistemáticos y procesos aleatorios. Los procesos sistemáticos son las fallas en el procedimiento, que incluyen desde la elaboración, aplicación y calificación de la prueba, hasta la captura y el análisis de los datos. Los procesos aleatorios son los que quedan fuera de control del proceso de la medición, como los cambios propios del elemento medido.

Si el *error de medición* es sistemático y altera todas las mediciones por igual, es considerado como un *error constante,* que no tendrá mayor efecto en la medición puesto que no alterará las comparaciones de los grupos. Pero cuando el error sólo afecta algunas de las mediciones, se considera que hay un sesgo en la medición.

El *error aleatorio* tiene mayor importancia en las mediciones psicológicas, puesto que los puntajes obtenidos se ven afectados por diversas situaciones tales como la elección adecuada de los reactivos de acuerdo a la muestra de sujetos, es decir, determinada prueba

puede arrojar puntuaciones elevadas en un grupo y bajas en otro, probablemente porque uno de estos tuvo mayor acceso a información que le fue útil al contestar la prueba y el otro grupo no; otra situación que puede variar las puntuaciones de un mismo sujeto, es el empleo del azar al contestar los reactivos de la prueba o que no tenga la misma disposición en las diversas mediciones. Este tipo de error es casi imposible de eliminar, por lo que se debe de tratar de reducir al mínimo siempre que sea posible.

Debido a la cantidad de errores que se pueden cometer en las mediciones es importante conocer la confiabilidad de una prueba. Anastasi (1988) define *la confiabilidad* de una prueba o test como "*la consistencia de las puntuaciones obtenidas por los mismos individuos cuando son examinados con el mismo test en diferentes ocasiones, con conjuntos distintos de elementos equivalentes o bajo otras condiciones variables del examen".*

Para obtener la confiabilidad de una prueba o test psicométrico, se obtiene un *coeficiente de correlación*, que indica el grado de concordancia que hay entre dos conjuntos de calificaciones. Para estimar el *coeficiente de correlación* se emplean algunos métodos estadísticos, como la fórmula de correlación de Pearson o Spearman, entre otros.

En seguida se describirá brevemente en qué consiste cada uno de los métodos para obtener la confiabilidad, así como el tipo de coeficiente de correlación que aporta y el tipo de varianza de error que contempla.

Métodos para obtener la confiabilidad

1. **Método de confiabilidad test-retest**

Este método consiste en aplicar una prueba y después de un intervalo de tiempo, volver a aplicar la misma. Posteriormente, se calcula el coeficiente de confiabilidad con las calificaciones obtenidas en la primera y segunda aplicación de cada sujeto. El coeficiente de confiabilidad que se obtiene con este método se denomina *coeficiente de estabilidad temporal*. La varianza de error se debe a la variación que se puede producir en el transcurso del tiempo de una aplicación a otra, o a las condiciones de aplicación en que se administró la prueba. Es importante señalar que este tipo de confiabilidad se podría ver alterado con rasgos variables de la personalidad, como el estado de ánimo.

2. **Método de confiabilidad con formas equivalentes o paralelas**

Consiste en elaborar dos pruebas o instrumentos parecidos que midan el mismo atributo; la construcción de pruebas paralelas consiste en crear dos formas de la prueba que contengan el mismo número de reactivos y el mismo grado de dificultad. Ambas pruebas deben tener formato, instrucciones y límite de tiempo en la administración que sean iguales, para posteriormente ser comparadas. Una vez construidas ambas, se

aplica una forma y con un periodo mínimo de tiempo se aplica la segunda forma al mismo grupo de sujetos. Para estimar la confiabilidad, se correlacionan las calificaciones obtenidas en las dos formas y el resultado de ésta se denomina *coeficiente de equivalencia.* La varianza de error muestra las variaciones en la actuación de los sujetos en ambas formas. Este método permite, además, conocer la magnitud del error de medida debido al muestreo de contenido.

3. Método de confiabilidad de división por mitades

En este método se aplica sólo una forma de la prueba o test. Después se divide el instrumento en dos mitades equivalentes que se califican por separado para obtener dos calificaciones distintas para cada sujeto. Una de las formas más comunes para dividir la prueba, es separar los reactivos pares e impares y obtener una calificación sumando todos los reactivos pares y otra con los reactivos impares, en esta división se debe tener en cuenta que los reactivos pares e impares sean similares en el grado de dificultad. Posteriormente, para obtener la confiabilidad de la prueba estas dos puntuaciones se correlacionan y se obtiene un *coeficiente de consistencia interna*, se le llama así puesto que sólo se emplea una forma de la prueba. Este método proporciona sólo la confiabilidad de la mitad de la prueba, pero entre mayor sea el número de reactivos o ítems que contiene el instrumento, mayor será la confiabilidad.

4. **Método de confiabilidad de Kuder Richardson**

Este método consiste en evaluar la consistencia interna de una prueba, para lo cual se requiere una sola aplicación y la confiabilidad se estima a partir de las respuestas de los sujetos a todos los reactivos de la prueba; la forma más común de evaluar este tipo de consistencia es utilizando la fórmula Kuder-Richardson, que hace un examen de la ejecución de cada elemento. Las fuentes de error que influyen en este método son el muestreo de contenido y la heterogeneidad de lo que se pretenda medir.

- **Validez**

En los tests psicológicos no sólo es importante que haya confiabilidad en las mediciones, puesto que ésta sólo indica la consistencia en las puntuaciones, un problema más importante que enfrenta una prueba psicológica es comprobar empíricamente que ésta mide el rasgo o atributo para el que fue construida.

Magnusson (1973) menciona que *"la validez de un método es la exactitud con que pueden hacerse medidas significativas y adecuadas con él, en el sentido que midan los rasgos que se pretenden medir".* Anastasi (1988) señala que, para poder determinar la validez, es necesario contar con criterios específicos y fuentes objetivas de información que definan el rasgo que mide la prueba. Aiken (1995)

define la validez como *"el grado en el cual ésta mide aquello para lo que se diseñó".*

Una prueba psicológica no tiene una sola validez, sino varias, que están en función del uso para el que ésta se destine y el grupo en el que se utilice.

La validez se estima mediante un coeficiente de correlación, denominado *coeficiente de validez,* que indica la relación existente entre los datos obtenidos de la prueba psicológica y los datos que se utilizan como índices para los puntajes del individuo en la variable de criterio (Magnusson, 1973).

Brown (1980) menciona que los diversos procedimientos para conocer la validez de una prueba se clasifican en tres categorías principales, según los *Standards for Educational and Psychological Tests:* validez de contenido, validez relacionada con el criterio y la validez de construcción.

a. **Validez de Contenido:** este tipo de validez supone un adecuado muestreo de contenido, es decir, una selección cuidadosa de los reactivos que conforman la prueba psicológica, para lo cual se realiza un examen sistemático del contenido de los reactivos que conforman el

instrumento. En el muestreo de contenido se debe especificar la amplitud que fue considerada, y ésta debe tratar de abarcar todas las áreas posibles. Este análisis permitirá determinar si éstos comprenden una muestra representativa del atributo o rasgo que se pretende medir.

b. **Validez empírica o de criterio:** Los tests psicológicos se emplean en algunas situaciones específicas. En este caso, se requiere comparar los resultados de la prueba con un criterio específico. El *criterio* es una medida directa e independiente de lo que mide la prueba (Anastasi, 1988). Las medidas de criterio para obtener la validez de una prueba se pueden obtener mediante dos formas, al mismo tiempo en que se aplica la prueba o después de un intervalo de tiempo.

La validez de criterio se puede clasificar en *validez predictiva y* ***validez concurrente***, la primera se emplea para predecir la actuación o ejecución de un individuo ante determinada situación, y la medida del criterio se obtiene después de un determinado intervalo de tiempo; este tipo de validez se emplea comúnmente en los instrumentos empleados en la selección y clasificación de personal. La validez concurrente se emplea cuando ya se tienen datos de los individuos (criterio) y éstos se pretenden substituir con

los resultados obtenidos en la prueba, esto es con la finalidad de ahorrar tiempo en determinados procesos, esta validez se emplea generalmente en instrumentos de diagnóstico.

c. **Validez de construcción o validez estructural:** Este tipo de validez se emplea para comprobar un constructo que se supone mide la prueba o test: *"En la medida en que una variable es abstracta y latente, más que concreta y observable se le denomina de constructo. Es algo que los científicos construyen y que no existe como dimensión observable de la conducta. Un constructo refleja la hipótesis de que una variedad de conductas se correlacionan entre sí, en estudios de diferencias individuales serán afectados de manera semejante por manipulaciones experimentales"* (Nunnally, 1995).

- **Normalización**

Un punto más a considerar en el trabajo con los tests o pruebas psicológicas, es la interpretación de los datos que éstas proporcionan, ya que la calificación aislada de una prueba solo proporciona puntajes crudos que no se pueden interpretar en situaciones prácticas. Para lograr una adecuada interpretación, es necesario hacer una normalización de las calificaciones.

La calificación que obtiene un sujeto en una prueba o test depende principalmente de su ejecución en ella, y en algunas pruebas del grado de dificultad que ésta presenta; para poder interpretar esta calificación es necesario recurrir a normas. *"Las normas son cualesquiera datos estadísticos que proporcionen un marco de referencia para interpretar los puntajes de un individuo en relación con los puntajes de otros"* (Nunnally, 1995).

Magnusson (1973) señala que *"la normalización es una transformación de la distribución de los puntajes originales a una distribución normal que permita tener escalas con diferentes niveles de dificultad".* De esta manera, las calificaciones de una prueba podrán interpretarse con mayor precisión.

Anastasi (1988) señala que los puntajes normalizados tienen una doble finalidad, indicar la situación relativa de un sujeto en relación al grupo normativo que permite valorar su ejecución con relación a otras personas; y proporcionar medidas que hagan posible la comparación directa de la actuación del sujeto en distintos instrumentos.

- **Adaptación de Tests para su Uso en Diferentes Idiomas y Culturas**

Existen bastantes datos indicadores de que está creciendo el interés por los estudios internacionales comparativos sobre rendimiento académico y por la investigación intercultural. Con este crecimiento aparece la necesidad de adaptar (habitualmente denominada “traducir”) tests de rendimiento y aptitudes, así como otros instrumentos psicológicos, para su uso en múltiples culturas e idiomas.

Hay muchas razones para adaptar tests. Una razón es que adaptar un nuevo test es generalmente más barato y rápido que preparar uno nuevo para un segundo idioma. Por ejemplo, el cuestionario de ansiedad-rasgo de Speilberger ha sido adaptado para ser utilizado en más de cincuenta países. Una segunda razón es que en el segundo idioma no siempre existe la experiencia técnica para producir un nuevo test validado. Esto es especialmente cierto en los países de tercer mundo. Los investigadores pueden sentirse más seguros utilizando una versión adaptada de un test que usando uno nuevo. Una tercera razón es que para llevar a cabo estudios comparativos inter países, interidiomas, y/o interétnicos, se requieren tests adaptados. Estos estudios se han vuelto muy populares en los últimos años, dado que muchos países se esfuerzan por establecer estándares educativos internacionales y desean comparar sus progresos con los de otros países. Los tests preparados en cada país, incluso utilizando los mismos presupuestos, serían suficientemente diferentes como para que las comparaciones entre países no fueran válidas.

Finalmente, los tests se adaptan a veces para mejorar la imparcialidad de la evaluación, permitiendo a las personas utilizar los tests y las escalas psicológicas en su idioma preferido.

Mientras que las muchas razones para la adaptación de los tests están claras, los métodos para llevarlas a cabo y establecer las equivalencias entre las puntuaciones no lo están tanto (Hambleton, 1993, 1994). La adaptación de un test incluye todo tipo de tareas, desde decidir si mide o no el mismo constructo en idiomas y culturas diferentes, hasta el problema de la selección de traductores, la decisión sobre los ajustes necesarios que hay que hacer para usar el test en un segundo idioma, o la propia adaptación del test y la comprobación de la equivalencia de la forma adaptada.

Fuentes de error o invalidez en la adaptación de los tests

Los estándares sobre los tests psicológicos y educativos elaborados en 1985 por las asociaciones americanas *American Educational Research Association* (AERA), *American Psychological Association* (APA) y *National Council on Measurement in Eduaction* (NCME), proporcionan cuidadosas directrices para aquellos psicólogos que seleccionan, construyen, aplican y usan tests psicológicos y educativos. Tres de los

estándares son especialmente relevantes en el contexto de la adaptación y uso de los tests:

Estándar 6.2. Cuando se hacen cambios sustanciales en el formato del test, modo de aplicación, instrucciones, idioma o contenido, el usuario debería de revalidar el test para las nuevas condiciones, o tener argumentos que apoyen que no es necesario o posible una validación adicional.

Estándar 13.4. Cuando se traduce un test de un idioma o dialecto a otro, debe establecerse su fiabilidad y validez en los nuevos grupos lingüísticos a los que se aplique.

Estándar 13.6. Cuando se pretende que las dos versiones de un test en idiomas distintos sean comparables, hay que aportar pruebas acerca de la comparabilidad de los tests.

Estos estándares proporcionan un marco para analizar las fuentes de error o invalidez que pueden surgir en la adaptación de los tests. Las fuentes de error o invalidez presentes en la adaptación de los tests pueden organizarse en tres amplias categorías:

- Diferencias culturales y de idioma
- Aspectos técnicos y métodos
- Interpretación de los resultados

Si no se atienden adecuadamente todos los aspectos incluidos en estas categorías, podemos terminar con un test adaptado que no es equivalente en los dos idiomas o grupos culturales. Los tests no equivalentes conducen necesariamente a errores en la interpretación y a conclusiones defectuosas acerca de los grupos implicados. Un buen ejemplo de las interpretaciones erróneas que puede seguirse de adaptaciones deficientes, es el siguiente (el ejemplo fue facilitado por Richard Wolf del *Columbia Teachers College.)* En un reciente estudio comparativo internacional de lectura, se presentaban a los estudiantes americanos pares de palabras y se les pedía que dijeran si tenían un significado similar o distinto. *Sanguine-pessimistic* fue uno de los pares de palabras en los que la ejecución de los estudiantes americanos resultó sólo ligeramente mejor que lo esperado por mero azar. En el país que quedó el primero, el 98% de los estudiantes respondieron esta cuestión correctamente. En el proceso de intentar comprender la razón de esta gran diferencia en la ejecución, se descubrió que *sanguine* no tenía una palabra equivalente en el idioma del país que había quedado el primero, por lo que se usó el equivalente al término inglés *optimistic.* Esta situación hizo la pregunta considerablemente más fácil y hubiese sido contestada también correctamente por un alto porcentaje de los estudiantes americanos si se les hubiese presentado el par *optimistic-pessimistic.* El objetivo de este ejemplo no es otro que subrayar el peligro de extraer conclusiones de estudios internacionales

sin tener una fuerte evidencia de que la adaptación del test genera dos instrumentos realmente equivalentes.

Directrices prácticas para la adaptación de los tests

La literatura técnica acerca del proceso de adaptación de los tests es incompleta desde el punto de vista de la medición, y aparece diseminada en una plétora de revistas internacionales, informes técnicos y libros. En pocas palabras, no existe una única y completa fuente que los profesionales puedan consultar, ni se han elaborado hasta hace poco un conjunto de directrices sobre la adaptación de los tests. Además, los métodos de medición más complejos (por ejemplo, los modelos de respuesta a los ítems y ecuaciones estructurales) que resultan muy útiles en el establecimiento de la equivalencia formal entre las puntuaciones obtenidas en tests adaptados para varios idiomas y culturas, no son bien conocidas por los investigadores que llevan a cabo las adaptaciones (Hulin, 1967).

A la vista del hecho de que el mayor interés se asocia a menudo con los resultados de estudios comparativos interculturales o internacionales sobre rendimiento educativo, parece clara la necesidad de disponer de unas directrices desarrolladas y validadas profesionalmente para la adaptación de los tests y el establecimiento de las equivalencias entre las puntuaciones. En muchos países ya se dispone de estándares técnicos o directrices relativos a la construcción de los tests, la fiabilidad

de las mediciones, la validez y baremación; sin embargo, se ha prestado poca atención a la preparación de directrices para la adaptación de los tests y el establecimiento de equivalencias entre las puntuaciones. Por ejemplo, en los ampliamente utilizados Estándares de la APA, AERA y NCME, sólo tres estándares se ocupan de la adaptación de los tests. En Canadá, que es un país bilingüe, sólo tres directrices de los estándares sobre los tests elaborados por la Asociación de Psicólogos Canadienses, se dedican a la adaptación de los tests. La Comisión Internacional de Tests (ITC) se ha ocupado de este problema y ha elaborado un conjunto de directrices prácticas para la adaptación de los tests.

Las directrices de la ITC para la adaptación de tests psicológicos y educativos se organizan dentro de cuatro secciones: Contexto, Construcción del test y adaptación, Aplicación y Documentación/ interpretación de las puntuaciones. Se consideró que las directrices serían más fáciles de usar si estaban organizadas en categorías significativas. Las directrices pertenecientes a la categoría *contexto* abordan aspectos relativos a la equivalencia de los constructos en los grupos idiomáticos de interés. De hecho, las adaptaciones no tienen ningún valor si no se puede establecer la equivalencia de los constructos. La categoría *construcción del test y adaptación* incluye directrices que emergen del proceso de adaptación del test, todo aquello que va desde la elección de los traductores hasta los métodos

estadísticos para analizar los datos empíricos, pasando por la investigación de la equivalencia de las puntuaciones. La tercera categoría, *aplicación,* se ocupa de directrices que tienen que ver con las distintas formas de aplicar los tests a grupos con diferentes idiomas, incluyendo desde la selección de aplicadores, hasta la elección de formatos de los ítems, o el establecimiento de los tiempos límite. La cuarta categoría de directrices tiene que ver con la *documentación y la interpretación de las puntuaciones.* Habitualmente, los investigadores proporcionan muy poca documentación acerca del proceso de adaptación para establecer la validez de un test adaptado, por lo que son frecuentes las interpretaciones erróneas de las puntuaciones de los tests en múltiples idiomas. Las directrices abordan asuntos incluidos en estas áreas.

El comité de la ITC adoptó la siguiente definición de *directriz* para la adaptación de un test:

Una directriz para la adaptación de un test es una práctica que se considera importante para realizar y evaluar la adaptación o construcción simultánea de tests psicológicos y educativos para su uso en diferentes poblaciones.

Considérese el siguiente ejemplo: una persona encargada de elaborar tests construyó un nuevo inventario de personalidad, lo estandarizó, validó sus

principales usos en una serie secuencial de estudios interrelacionados, y lo vendió exitosamente entre psicólogos practicantes de los Estados Unidos. Dicha persona viajó por todo el país para dictar conferencias a psicólogos, instruirlos sobre el modelo que subyace en el instrumento, y capacitarlos para utilizar el inventario de forma adecuada. Después de varios años, se realizaron investigaciones importantes sobre el instrumento y se llevaron a cabo conferencias nacionales para discutir su uso.

Claramente se observó que el instrumento referido era válido en su construcción y útil desde el punto de vista clínico. El único aspecto negativo del inventario fue una controversia menor sobre la generalización del instrumento en varios grupos minoritarios de baja representación del país. No obstante, el instrumento se empleaba tanto, que su editor y algunos psicólogos internacionales que realizaron sus trabajos de graduación en los Estados Unidos y estaban capacitados en el uso del instrumento, comenzaron a presionar al autor del test para que lo tradujera a otros idiomas.

¿Qué aspectos deben considerarse en la adaptación de un instrumento a diferentes países, culturas y personas? Si somos capaces de identificar los problemas inherentes en la adaptación de instrumentos a nuevas culturas, ¿no es pertinente también pensar en algunos de estos mismos problemas al emplear un instrumento en varios subgrupos o poblaciones secundarias dentro de un país o una población determinada? En otras palabras, ¿existen

similitudes de procedimiento entre la adaptación de un inventario de personalidad para ser empleado con individuos de España y la adaptación del mismo inventario para ser empleado con Latinos del los Estados Unidos?

Aunque la literatura está llena de ejemplos de tests traducidos del inglés a otros idiomas, son incipientes aún los procedimientos para comprometerse en estas actividades, y muchos de los mejores procedimientos aún no se han puesto en práctica plenamente.

La adaptación de instrumentos de valoración para nuevas poblaciones-objetivo, normalmente se requiere cuando la nueva población-objetivo difiere notablemente de la población original con la que se utiliza el instrumento de valoración, desde el punto de vista cultural o de los antecedentes culturales, el país y el idioma. La mayoría de las adaptaciones culturales cruzadas de los instrumentos de valoración, incluye la traducción de un instrumento de un idioma a otro.

En algunos casos, sin embargo, es necesaria la adaptación de instrumentos de valoración, aunque el idioma continúe siendo el mismo, ya que son diferentes la cultura o las experiencias de la vida de los que hablan ese mismo idioma. (Se han adaptado numerosos instrumentos para su uso en grupos que difieren en experiencias de aquéllos con quienes se usó originalmente el instrumento. Por ejemplo, el Inventario Minnesota de Personalidad de Fases Múltiples – Adolescente (MMPI-A, por sus siglas en

inglés), es una adaptación del MMPI-2, el cual se desarrolló para ser empleado con adolescentes cuya experiencia en la vida sistemáticamente difiere de la de los adultos. Se han desarrollado otras medidas con diferentes versiones para hombres y mujeres, por ejemplo, con base en las experiencias en la vida, generalmente diferentes, y las orientaciones resultantes de hombres y mujeres.)

La traducción y adaptación de las medidas psicológicas del inglés a otro idioma, actualmente es una actividad ampliamente difundida. Estos esfuerzos requieren tanto el idioma como la sensibilidad cultural; en los últimos años, algunas medidas sólo se tradujeron de manera lingüística. Se tuvo poco cuidado de adaptar medidas con base en las diferencias culturales. Deben darse dichos pasos, pues el problema de la adaptación cultural puede dificultar la traducción directa y la utilización de ítems de algunas medidas y, en algunos casos, la traducción y la utilización de ítems de medidas enteras. Los pasos recomendados para adaptar medidas permitirán asegurar el uso adecuado de medidas adaptadas.

Aunque la teoría psicométrica observa un retraso en esta área, se cuenta con procedimientos psicométricos para estudiar medidas adaptadas e intentar determinar si éstas valoran las mismas estructuras en la forma adaptada, si los resultados pueden utilizarse con una validez equivalente, si está presente la inclinación al nivel del ítem o con respecto a todo el instrumento, y si los resultados son igualmente importantes en ambas formas.

Las normas y las prácticas profesionales requieren la realización de pruebas piloto y de campo subsecuentes -a larga escala- sobre las medidas adaptadas.

En la práctica, podría aplicarse la noción de que una versión revisada de un instrumento mide lo que se propone, y que lo que mide en la versión original de idioma o cultura es información esencial que debe reunirse antes de adaptar medidas. Quienes adaptan medidas, también deben ser sensibles a las necesidades de capacitación y documentación de los usuarios del instrumento de valoración. Si se siguen dichas prácticas experimentales, es probable que el uso de las medidas resultantes genere efectos benéficos. Cuando se pasan por alto los pasos en la construcción, traducción y validación de las medidas traducidas, es menos probable lograr una utilización responsable de pruebas.

CAPITULO 3

DESARROLLO DEL INVENTARIO DE DEPRESIÓN DE BECK (BDI)

Los datos internacionales indican un aumento importante de los trastornos depresivos, por lo que se empieza a considerar al siglo XXI como la era de la melancolía. Cada nueva generación, desde principios del siglo XX, ha tenido un riesgo mayor que la generación de sus padres de sufrir una depresión importante en el curso de su vida. Este trastorno, que adopta diversas formas en función de características sociales y culturales, no se trata simplemente de la presencia de tristeza, sino que implica desinterés, desaliento y una abrumadora desesperanza vital (CNCG, 1993). A ello se añade el descenso en la edad de inicio de aparición de este trastorno.

Desde hace tiempo se ha constatado un aumento del diagnóstico de depresión en personas jóvenes. Además, estudios realizados en todo el mundo muestran que en épocas de crisis sociales y políticas aumentan los índices de los trastornos depresivos y de ansiedad (Patel y Kleinman, 2003). Estas tendencias han promovido el estudio de la depresión tanto en investigaciones psicológicas (básicas y aplicadas), como sociológicas y biológicas.

La primera versión del BDI (Beck, 1961) se realizó sobre la base de análisis de los contenidos más frecuentes en las sesiones terapéuticas con pacientes deprimidos. Lo conformaban 21 ítems que se presentaban en tres fases de formulación gradual y que referían a criterios diagnósticos básicos de la depresión (humor, pesimismo, culpa, preocupaciones somáticas, pérdida de libido, etc.).

Luego de una revisión realizada en 1987 (BDI-A), en 1994 se inició el estudio piloto para el diseño del Inventario de Depresión de Beck-II (BDI-II), cuyo objetivo principal fue adecuar el contenido de los ítems a los criterios diagnósticos para la depresión considerados por el *Manual Diagnóstico y Estadístico de los Trastornos Mentales IV (DSM-IV).*

Por sus características –características psicométricas, facilidad y rapidez en su aplicación y justeza en la identificación de síntomas depresivos según criterios del DSM-IV-, se trata de una prueba que sirve al psicólogo, al psiquiatra y al neurólogo para discriminar rápidamente los posibles casos de depresión; es un instrumento de autoinforme compuesto por 21 ítems, cuya finalidad es medir la severidad de la depresión en adultos y adolescentes, a partir de los 13 años.

Después de 35 años de investigación y experiencia con el BDI, se consideró propicio revisar y modificar el Beck Depression Inventory (BDI-IA; Beck, Rush, Shaw y Emery, 1979) y modernizarlo. El BDI-IA remplazó al

instrumento original (BDI) que había sido desarrollado por Beck, Ward, Mendelson, Mock y Erbaugh (1961). En la nueva versión, BDI-II, cuatro ítems (Pérdida de Peso, Cambio de Imagen Corporal, Preocupación Somática y Dificultad para Trabajar) fueron reemplazados por otros nuevos (Agitación, Desvalorización, Dificultad de Concentración y Pérdida de Energía) con el objetivo de identificar síntomas típicos de la depresión severa o de la depresión que requiere hospitalización.

Se cambiaron dos ítems para incluir los cambios en el apetito y en el sueño (tanto los aumentos como las disminuciones). Además, se reformularon algunos enunciados (o alternativas) utilizados para evaluar otros síntomas. A diferencia del BDI-IA, el BDI-II constituye una revisión sustantiva del BDI original.

- **Historia**

En los últimos años, el BDI se ha convertido en uno de los instrumentos más aceptados para evaluar la severidad de la depresión en pacientes diagnosticados y para detectar casos de depresión en poblaciones normales (Sherry y Séller, 1985; Archer, Maruish, Imhof y Piotrowski, 1991; Piotrowski y Séller, 1992).

En dos revisiones exhaustivas sobre las aplicaciones y propiedades psicométricas del BDI, se ha informado de su alta confiabilidad tanto en

poblaciones clínicas como en no clínicas (Steer, Beck y Garrison, 1986; Beck, Steer y Garbin, 1988).

Por ejemplo, el coeficiente promedio del BDI para pacientes psiquiátricos es superior a 0.80. De igual forma, se ha establecido la validez concurrente y de constructo del BDI con respecto a una variedad de mediciones psicológicas (Beck y Steer, 1987). Además, el BDI diferenció pacientes con depresión clínica de pacientes psiquiátricos no depresivos (Steer y otros, 1986).

El BDI original se basó en los enunciados descriptivos típicos de aquellos síntomas que habían sido reportados con frecuencia por pacientes psiquiátricos con depresión, y sólo algunas veces por pacientes psiquiátricos no deprimidos (Beck y otros, 1961). Las observaciones clínicas y las descripciones de los pacientes fueron consolidadas sistemáticamente en 21 ítems que incluyeron síntomas y actitudes representativos. Estos ítems fueron organizados de acuerdo con la severidad del contenido de los enunciados alternativos. A cada ítem se le dio un valor según una escala de cuatro puntos, de 0 a 3, en términos de severidad. La versión original se diseñó para ser administrada por entrevistadores entrenados que leían en voz alta los enunciados de síntomas a los pacientes. Luego, los pacientes seleccionaban el enunciado de cada ítem que mejor se ajustaba a sus estados de ánimo del momento.

Los entrevistadores generalmente tardaban entre 10 y 15 minutos para administrar el instrumento y se obtenían las puntuaciones mediante la suma de las elecciones que los pacientes hacían en cada uno de los 21 ítems.

Los 21 síntomas y actitudes depresivos elegidos por Beck y otros (1961) para ser incluidos en el BDI, se basaron en las descripciones verbales de pacientes, y no fueron seleccionados para reflejar ninguna teoría de depresión en particular. Estos ítems fueron: (1) Estado de Ánimo, (2) Pesimismo, (3) Sensación de Fracaso, (4) Insatisfacción (anhedonia), (5) culpa, (6) Castigo, (7) Disgusto Consigo Mismo, (8) Acusaciones a Sí Mismo, (9) Ideas Suicidas, (10) Llanto, (11) Irritabilidad, (12) Aislamiento Social, (13) Indecisión, (14) Cambio de Imagen Corporal, (15) Dificultad en el Trabajo, (16) Insomnio, (17) Fatiga, (18) Pérdida de Apetito, (19) Pérdida de Peso, (20) Preocupación Somática y (21) Pérdida de Libido.

A principios de 1971, Beck y sus asociados en el Centro de Filadelfia para Terapia Cognoscitiva de la Universidad de Pensilvania, comenzaron a emplear una versión modificada del BDI que eliminaba palabras y frases alternativas para los mismos síntomas y evitaba la doble negación (Beck y otros, 1979). El número de respuestas opcionales se limitó estrictamente a cuatro alternativas para cada ítem. incluyendo

la opción nula, y se realizaron modificaciones mínimas en la formulación de 15 ítems.

Seis ítems (Llanto, Irritabilidad, Fatiga, Falta de Apetito, Pérdida de Peso y Pérdida de Libido) no fueron corregidos. Se realizaron varias versiones piloto y la revisión final fue registrada en 1978 y publicada en 1979. Aparentemente, muchos investigadores desconocían la existencia de esta versión actualizada, ya que el origen de 1961 continuó siendo el más citado en la mayoría de los estudios que incluían la administración del BDI (Steer y Beck, 1985).

En 1987, se publicó un manual técnico para el BDI-IA (Beck y Steer, 1987). La edición de 1993 de este manual incluyó leves revisiones de puntuación recomendadas para interpretar el nivel de severidad de síntomas depresivos.

A lo largo de los años se han realizado varios ajustes en la definición de depresión, especialmente con el advenimiento del Manual Diagnóstico y Estadístico de los Trastornos Mentales Tercera Edición (DSM-III, 1980) de la American Pshychiatric Association. Moran y Lambert (1983) compararon el contenido del BDI con los criterios de Trastornos Afectivos del DSM-III y llegaron a la conclusión de que de los nueve criterios establecidos en el DSM-III, el BDI coincidía sólo con seis, parcialmente sólo con dos, y un criterio directamente no se incluía. El BDI se refería a la

pérdida de apetito, pero no a su aumento. Sin embargo, algunas personas con depresión, especialmente aquéllas con depresiones "atípicas", experimentan incremento del apetito. Existe una situación similar con respecto a la variación del sueño, ya que algunos pacientes con depresión duermen considerablemente más de lo habitual. Moran y Lambert (1983) también notaron que el BDI carecía de ítems explícitos referidos a la agitación y la actividad psicomotriz, a pesar de que los criterios del DSM-III incluyen a ambas en el diagnóstico de trastorno depresivo mayor.

Vredenburg, Krames y Flett (1985) plantearon preguntas similares con respecto a si el BDI se correspondía con lo criterios del DSM-III. Steer y Beck (1985) indicaron que las preguntas sobre el incremento del sueño y apetito fueron omitidas en el BDI porque Beck (1967, p.34) había encontrado que 72% de los pacientes con depresión severa se quejaban por la pérdida de apetito y 87% reportaba disminución de sueño. Además, Steer y Beck (1985) expresaron que el incremento de apetito y sueño ocurría en adultos normales y que la inclusión de estos dos síntomas en el BDI podría producir un número elevado de falsos positivos; es decir, que podría identificar a individuos como depresivos, cuando en realidad no lo estaban (Steer y Beck, 1985).

Los lanzamientos consecutivos del Manual Diagnóstico y Estadístico de los Trastornos Mentales Tercera Edición Revisada (DSM-III-R; 1987) y el

Manual Diagnóstico y Estadístico de los Trastornos Mentales Cuarta Edición (DSM-IV, 1994) de la American Pshychiatric Association, destacaron la necesidad de una nueva medición psicológica de depresión que evaluara síntomas conforme a los criterios del DSM-III-R y el DSM-IV. Además, algunos de los síntomas originales del BDI que habían sido típicamente observados en pacientes hospitalizados por mucho tiempo, en particular, Cambio de Imagen Corporal, Pérdida de Peso y Preocupación Somática, fueron gradualmente perdiendo utilidad con el paso de los años para la evaluación de la severidad de la depresión en pacientes externos y en pacientes hospitalizados por poco tiempo.

El ítem Pérdida de Peso del BDI, por ejemplo, resultó ser menos efectivo que en 1961 para estimar la depresión. Esta disminución de efectividad se debe a que el grado de pérdida de peso depende de la duración de la depresión y, hoy en día, la depresión esta siendo diagnosticada y tratada en estadios más tempranos que en 1960, cuando Beck realizó el BDI original.

- **Desarrollo de los Ítems**

Beck y sus colaboradores comenzaron el trabajo piloto para el BDI-II en 1994. Los ítems específicos que se referían a los criterios de depresión del DSM-IV y que no habían sido incluidos en el BDI-IA, tales como

Agitación, Desvalorización, Dificultad de Concentración y Pérdida de Energía, fueron construidos y testeados junto con otros ítems del BDI-IA, cuyas opciones de respuesta fueron parcialmente reformuladas para hacerlas más entendibles. Los ítems Insomnio y Pérdida de Apetito fueron reformulados para reflejar tanto los aumentos como las disminuciones de sueño y apetito. El ítem Pérdida de Peso no fue incluido en el estudio piloto.

A fin de reducir los 27 ítems del estudio piloto a la versión final de 21 ítems, Beck y sus colaboradores realizaron una serie de análisis factoriales y de análisis de ítems con las respuestas de 193 pacientes psiquiátricos externos diagnosticados con varios trastornos psiquiátricos.

Sólo tres ítems del BDI-IA no fueron reformulados: Castigo, Ideas Suicidas y Pérdida de Libido. En algunas instancias, para la reformulación de ítems también se cambiaron los títulos del BDI-IA. Por ejemplo, Autoacusación fue cambiado por Autocrítica, ya que representa mejor los enunciados alternativos. Como se mencionó, Cambio de Imagen Corporal, Pérdida de Peso, Preocupación Somática y Dificultad para Trabajar, se reemplazaron por Agitación, Desvalorización, Dificultad de Concentración y Pérdida de Energía. El ítem Insomnio ahora indaga tanto sobre pérdida como sobre incremento de sueño, y fue renombrado como Cambios en los Hábitos de Sueño. El ítem Pérdida de Apetito ahora evalúa tanto los aumentos como las disminuciones de

apetito, y fue renombrado como Cambios en el Apetito. También se agregaron a algunos ítems implicaciones clínicas adicionales. Por ejemplo, el ítem Exclusión Social del BDI-IA, que se refería sólo a las relaciones interpersonales, ahora en el BDI-II incluye a personas y actividades en el ítem Pérdida de Interés.

- **Propiedades Psicométricas del BDI**

Las propiedades psicométricas del BDI han sido reportadas por Beck, Steer y Garbin en un detallado artículo que integra los estudios de investigación realizados con este inventario durante 25 años. Dicho trabajo incluye datos sobre aspectos tales como las versiones que se han desarrollado, la forma de aplicarlo, la confiabilidad, la validez discriminativa, de constructo y factorial, así como las normas tanto para muestras clínicas como no clínicas.

Como resultado del minucioso análisis comparativo realizado de la bibliografía de estos autores, concluyen que después de un meta-análisis de los valores de consistencia interna reportados, el valor promedio de coeficiente alfa para muestras clínicas es de 0.86 y para muestras no clínicas es de 0.81. La estabilidad temporal test-retest ha sido probada en intervalos de una semana de aplicación o más tiempo. El valor promedio de correlación, validez concurrente, con la Escala de Hamilton, la Escala de Zung y la Escala de Depresión del MMPI, oscila

entre 0.72 y 0.73, respectivamente. El análisis factorial muestra que el número de factores obtenidos, dependiendo del tipo de la muestra utilizada, oscila entre 3 y 7. No obstante lo anterior, la evidencia apunta hacia el hecho de que el BDI representa un síndrome general de depresión que se puede dividir en tres factores altamente intercorrelacionados: *actitudes negativas hacia sí mismo, deterioro del rendimiento y alteración somática.*

Igualmente, concluyen que es crucial la correcta identificación de la composición factorial del BDI debido a que se utiliza para detectar y evaluar la depresión en pacientes con problemas médicos y en pacientes psiquiátricos cuyos síntomas pueden ser atribuidos al abuso de sustancias o a la enfermedad propiamente dicha, por lo que la confirmación de la composición factorial permite excluir los síntomas somáticos y el deterioro del rendimiento característico de dichos problemas. Por último, señalan que existe evidencia de que el BDI discrimina entre subtipos de depresión y claramente diferencia la depresión de la ansiedad. Además de los estudios anteriores, enfocados propiamente a los aspectos psicométricos, otros se han interesado en evaluar si hay diferencias entre género inter e intra culturas.

En México, al igual que en otros países, el BDI ha sido traducido y ha ganado una gran popularidad tanto en la práctica clínica, como en la investigación realizada por psicólogos o psiquiatras. Sin embargo, hasta

donde la información tácitamente reportada en revistas científicas mexicanas lo demuestra, su uso no ha estado cimentado sobre la base de una investigación sistemática de las propiedades psicométricas de la escala que lleve, entre otros aspectos, a obtener normas para la población mexicana. Es muy posible que en alguna escuela, centro, facultad o instituto se cuente con datos de algún trabajo psicométrico escrupuloso, pero desafortunadamente el no darlos a conocer en los medios especializados imposibilita el avance en las investigaciones y sólo los convierte, en el mejor de los casos, en instrumentos de autoconsumo.

La revisión, lo más exhaustiva posible, de las revistas especializadas de circulación nacional, muestra que en México se han realizado trabajos psicométricos con el BDI (11, 13, 15). Desafortunadamente, si se hace una evaluación de sus resultados y de la forma en que cumplen, tanto con el rigor psicométrico, como con los criterios internacionales de adaptación ya señalados, sus hallazgos son definitivamente cuestionables respecto a su validez, debido a que no pueden garantizar el haber medido lo que pretendían medir.

PLANTEAMIENTO DEL PROBLEMA

Las pruebas psicológicas, validadas y confiables, son una herramienta de vital importancia para la evaluación y diagnóstico tanto experimental como clínico. Para la Psicología Clínica y para la Psiquiatría, la evaluación psicométrica de la depresión es fundamental en el proceso de evaluación y tratamiento de dicho trastorno. Dado que en la Cuidad de México no se contaba con una prueba con validez y confiabilidad psicométrica, Jurado y Col, 1998, publicó la estandarización del Inventario de Depresión de Beck. A partir de su publicación en castellano, el Inventario de Depresión de Beck ha sido utilizado ampliamente en diferentes trabajos de Tesis de Licenciatura y Posgrado en la Facultad de Psicología de la UNAM, en otras escuelas de Universidades Privadas, así como en investigaciones publicadas en revistas arbitradas.

En apego a las normas psicométricas establecidas internacionalmente, se debe revisar cada diez años la validez y confiabilidad de una prueba psicológica, así como su capacidad de medir/ identificar el fenómeno para el que fue construido, atendiendo a los cambios del paso del tiempo (específicamente causados por aspectos culturales o teóricos derivados de las investigaciones pertinentes) que permitan tener un constructo vigente y actualizado. Por lo anterior, la presente investigación tiene como objetivo, realizar la adaptación psicométrica

del Inventario de Depresión de Beck II en residentes de la Cuidad de México. Lo anterior permitirá contar con un instrumento revisado y actualizado que facilitará la medición objetiva de la depresión en escenarios académicos y no académicos.

JUSTIFICACIÓN

La depresión existe y hay muchos instrumentos psicométricos para identificarla; es de todos sabido que el Inventario de Depresión de Beck fue construido originalmente en los Estados Unidos y ha sido utilizado a nivel internacional y traducido al español, chino, francés, sueco, árabe, persa y, recientemente, al búlgaro. Sin embargo, en apego a las normas psicométricas, cada diez años se debe hacer una revisión de los instrumentos para tener un constructo vigente y actualizado.

MÉTODO

- **Objetivo**

Evaluar la validez concurrente entre el BDI-II y la Escala de Zung en una muestra de 300 sujetos residentes de la Cuidad de México (100 estudiantes de bachillerato de la Escuela Nacional Preparatoria #6 “Antonio Caso”, 100 estudiantes de licenciatura de la Universidad Autónoma Metropolitana y 100 adultos, empleados del Centro Nacional

de Prevención de Desastres (CENAPRED), y la confiabilidad test-retest en un tiempo establecido de 1 mes.

- **Sujetos**

Fueron seleccionados sujetos de ambos sexos al azar, en diferentes escenarios de la Ciudad de México; se les entregó primero el Inventario de Depresión de Beck Segunda Edición y enseguida la Escala de Zung.

- **Procedimiento**

Se realizaron cartas dirigidas a los encargados de cada una de las Instituciones para solicitar el permiso para aplicar el instrumento, así como la propuesta de posibles fechas para su realización, comprometiéndome a entregar resultados generales de las comparaciones efectuadas entre los diversos grupos antes mencionados.

Se reunió a los participantes en grupos de 15 o 20 personas para brindarles una atención personalizada en caso de que hubiera surgido alguna duda, explicándoles que con su participación estaban colaborando en un proyecto para la actualización del Inventario de Depresión de Beck-II y que si alguien requería sus resultados individuales, podrían pedirlos de manera anónima.

Al momento de la aplicación se les pidió leer con cuidado las instrucciones que estaban escritas en cada instrumento, acto seguido, se les indicó que procedieran a contestarlos. Al terminar, se les agradeció su participación y colaboración.

- **Consideraciones Éticas**

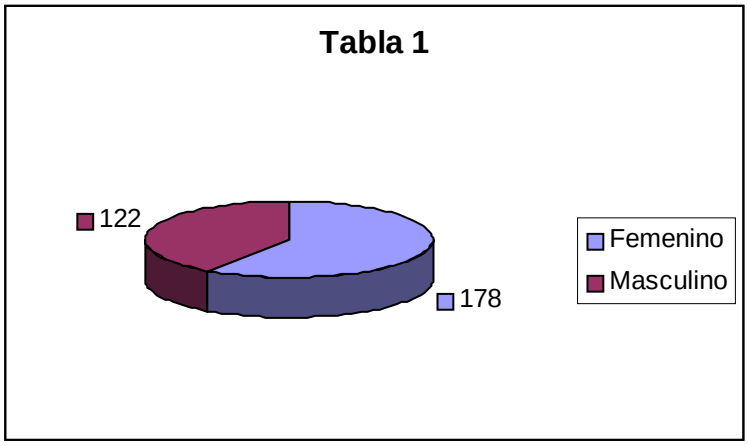
Se garantizó la confidencialidad, ya que sólo se les pidió que escribieran un sobrenombre o alguna clave con la que se pudiera identificar fácilmente el test y el re-test de cada sujeto, y se les dijo que los datos sólo serían manejados por personal calificado. Asimismo, se les informó que no era obligatorio contestarlo y que no tendrían ninguna consecuencia en caso de no aceptar participar en el estudio, que la presente investigación no se realizaría con fines diagnósticos ni clínicos y que si alguien tuviera alguna duda en cuanto a su situación en particular, sería canalizado al Centro de Servicios Psicológicos (CSP) de la Facultad de Psicología de la UNAM.

RESULTADOS

A continuación se describen las características de la muestra:

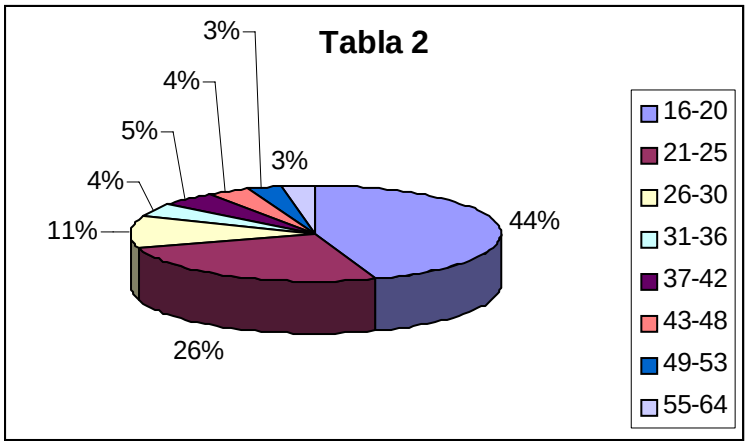
- **Sexo**

Con respecto al sexo, la muestra se integró por 300 sujetos, de los cuales 122 fueron hombres y 178 mujeres, como se puede observar en la Tabla 1.



- **Edad**

Con respecto a la edad, la muestra se integró por 134 sujetos de entre 16 y 20 años, 77 sujetos de entre 21 y 25 años, 32 sujetos de entre 26 y 30 años, 13 sujetos de entre 31 y 36 años, 15 sujetos de entre 37 y 42 años, 11 sujetos de entre 43 y 48 años, 9 sujetos de entre 49 y 53 años, y 9 sujetos de entre 55 y 64 años, como se puede observar en la Tabla 2.



- **Estado Civil**

Del total de la muestra, participaron 242 sujetos solteros, 40 casados, 7 en unión libre, 2 viudos, 4 separados y 5 divorciados, como lo muestra la Tabla 3.

Tabla 3

Estado Civil

	Frecuencia	**%**
Soltero	242	80.7
Casado	40	13.3
Unión Libre	7	2.3
Viudo	2	0.7
Separado	4	1.3
Divorciado	5	1.7
Total	300	100

- **Escolaridad**

Del total de la muestra, participaron 100 estudiantes de preparatoria, 110 estudiantes de licenciatura y 90 trabajadores, como lo indica la Tabla 4.

Tabla 4

Escolaridad

	Frecuencia	**%**
Estudiantes de preparatoria	100	33.3
Estudiantes de licenciatura	110	33.3
Trabajadores	100	33.3
Total	300	100

Dado que el objetivo de este estudio era evaluar si existía validez de criterio, así como la validez concurrente entre el Inventario de Depresión de Beck-II y la Escala de Zung, las respuestas a los inventarios se codificaron y procesaron estadísticamente con el paquete SPSS para Windows, versión 12.0. Los resultados encontrados arrojaron los siguientes coeficientes de correlación de Pearson:

- Entre la Escala Zung Pre y Post Test se obtuvo una correlación de R= 0.469, P<.000

- Entre el BDI-II Pre y Post Test se obtuvo una correlación de R= 0.7, P<.000
- Entre la Escala Zung Pre Test y el BDI-II Pre Test se obtuvo una correlación de R= –0.245, P<.000
- Entre la Escala Zung Post Test y el BDI-II Post Test se obtuvo una correlación de R= –0.244, P<.000

DISCUSIÓN Y CONCLUSIONES

El nivel significativo de correlación obtenido entre la Escala de Beck y la Escala de Zung, permite afirmar que ambos instrumentos cuentan con validez, cumpliéndose así el objetivo del presente trabajo, el cual era obtener la validez de criterio y la validez concurrente del Inventario de Depresión de Beck-II.

Asimismo, se cumplió el objetivo principal, ya que sí hay una alta confiabilidad test-re test en el BDI-II.

La escala de Zung requeriría ser revisada y actualizada para identificar los síntomas de depresión, para ello, sugiero que se podría correlacionar con la Escala 2 del MMPI-2.

Debemos tener en cuenta que la psicometría es una herramienta sólida, capaz de medir objetivamente manifestaciones del comportamiento humano y el omitir pasos en la construcción y/o estandarización de instrumentos, no permite promover el uso ético, responsable y profesional de los mismos, por lo cual se resalta la importancia del trabajo psicométrico realizado en torno al BDI-II en población mexicana, trabajo que permite afirmar que es un instrumento validado y útil.

Por otro lado, cumpliendo con las consideraciones éticas, a los 30 participantes que pidieron sus resultados individuales, se les entregó un

informe donde se indicaba su puntaje y, en caso de requerirlo, fueron canalizados al Centro de Servicios Psicológicos de la Facultad de Psicología de la UNAM.

La validez concurrente busca encontrar constancia y estabilidad de una prueba, es por ello que se recomienda continuar con esta investigación, extendiendo el uso y la aplicación del BDI-II en diferentes escenarios, y evaluando sus cualidades psicométricas.

La principal limitación encontrada es que hubo mucha gente que no estaba dispuesta a cooperar.

Tomando en cuenta que la muestra está integrada por un mayor número de adolescentes que de adultos, considero que podría haber un sesgo, ya que es justamente en la adolescencia donde existen: cambios constantes en el estado de ánimo, mayor inadaptación al tratar con un ambiente social al que consideran difícil, incapacidad para experimentar placer en actividades previamente gratificantes, pérdida del sentido de autovalía, constantes relaciones interpersonales problemáticas que pueden llegar a tener o la misma pérdida de una relación importante; todo esto puede hacer que se experimenten, en mayor o menor medida, síntomas de depresión que pueden llegar a ocasionar pensamientos o incluso intento de suicidio, más comúnmente que en los adultos.

La presente investigación aporta un instrumento de medición que va a cuantificar de manera objetiva un fenómenos natural como lo es la depresión; mi trabajo fue la continuación de lo realizado por Pozos L. (2009), quien realizó la primera versión confiable de la Escala BDI-II, la cual se tradujo al español y obtuvo la confiabilidad por consistencia interna. Con este trabajo se consiguió la confiabilidad test-re test de dicho instrumento, que nos indica la estabilidad de las respuestas, a través del tiempo.

Sin embargo, este tipo de medida no puede ser usada en personas diagnosticadas previamente con depresión, ya que al estar en tratamiento se esperaría que los puntajes tiendan a bajar.

Se debe considerar que sería antiético aplicar el Inventario de Depresión de Beck-II en dichos pacientes, dejarlos un mes sin tratamiento y, posteriormente, aplicar el re test; ya validado el instrumento, se podría hacer para ver el cambio que existe entre un momento y otro.

REFERENCIAS

A. Harris C. , D´Eon J.L. (2007) Psychometric properties of the Beck Depression Inventory-Second Edition (BDI-II) in individuals with chronic pain. Pain 1-14.

Azocar F., Areán P., Miranda J., Muñoz R.F. (2001) Differential Item Functioning in a Spanish Translation of the Beck Inventory Depression. Journal of Clinical Psychology 57, 355-365.

Beck A., Steer R. y Brown G. 2006 BDI-II Inventario de Depresión de Beck Segunda Edición. Buenos Aires. Editorial Piados.

Caballo, V. (et. al.) 1995 Manual de psicopatología y trastornos psiquiátricos. España. 1ª ed. Siglo Veintiuno Editores. Cáp. 17

CIE-10, Trastornos mentales y del comportamiento. Descripciones clínicas y pautas para el diagnóstico (1992) Organización Mundial de la Salud. Editorial Meditor. España. Págs. 141-166.

Cohen A. (2008) The underlying structure of the Beck Depression Inventory II: A multidimensional scaling approach. Journal of Research in Personality 42 779-786.

D. Theoms B., Ziegelstein R. C., Beck C.A., Pilote L. (2008) A general factor model for the Beck Depression Inventory-II: Validation in a sample of patients hospitalized with acute myocardial infarction. Journal of Psychosomatic Research 1-7.

DSM-IV-TR, Manual diagnóstico y estadístico de los trastornos mentales. (2002). Editorial Masson. 1ª ed. España. Págs. 387-476.

Dum M., Pickren J., Sobell L.C., Sobell M.B. (2008) Comparing the BDI-II and the PHQ-9 with outpatient substance abusers. Addictive Behaviors 33, 381-387.

F. Geisinger K. Valoración Normativa y Cultural Cruzada: Problemas de Traducción y Adaptación que Influencian la Interpretación Normativa de los Instrumentos de Valoración

Hambleton K. R. (1995). Adaptación de tests para su uso en diferentes idiomas y culturas: fuentes de error, posibles soluciones y directrices prácticas. Capítulo 6

J. Moran P. and Mohr D.C. (2004) The Validity of Beck Depression Inventory and Hamilton Rating Scale for Depression Items in the Assessment of Depression Among Patients with Multiple Sclerosis. Springer Science, 35-41.

Jurado C. S., Villegas H. E., Méndez S. L., Rodríguez S. F., Loperena S. V., y Varela D. R. W. (1998) La estandarización del Inventario de Depresión de Beck para los residentes de la Cuidad de México. Revista Salud Mental. Vol. 21, Núm. 3, Págs. 26-31.

Robles R., Varela R., Jurado S. y Páez F. (2001) Versión Mexicana del Inventario de Ansiedad de Beck: Propiedades Psicométricas. Revista Mexicana de Psicología 18, 211-218

Ruca B. M. 1999. Trastornos del Humor. España, 1ª ed. Panamericana.

Seignourel P.J. , Green C. , Schmitz J.M. (2008) Factor structure and diagnostic efficiency of the BDI-II in treatment-seeking substance users. Drug and alcohol Dependence 93, 271-278.

Vázquez B. J.L. 1998. Psiquiatría en Atención Primaria. España. Libros Princeps, Biblioteca Aula médica. Pág. 231-258.

Yin P. and Fan X. (2000) Assesing the Reliability of Beck Depression Inventory Scores: Reliability Generalization Across Studies. Educational and Psychological Measurement 60, 201-223.