

01921  
56



**UNIVERSIDAD NACIONAL AUTÓNOMA  
DE MÉXICO**

**FACULTAD DE PSICOLOGÍA  
DIVISIÓN DE ESTUDIOS PROFESIONALES**

**"CONFIABILIDAD Y VALIDEZ DE LA ESCALA DE HABILIDAD GENERAL  
PARA ADULTOS (GAMA) EN UNA MUESTRA DE ESTUDIANTES  
UNIVERSITARIOS MEXICANOS"**

**T E S I S**  
QUE PARA OBTENER EL TÍTULO DE:  
**LICENCIADO EN PSICOLOGÍA**  
P R E S E N T A :  
**GENARO DE LA CUADRA CASTILLO**

DIRECTOR DE TESIS:  
LIC. RUBÉN W. VARELA DOMÍNGUEZ

REVISOR: DR. SAMUEL JURADO

SINODALES: MTRA. ROCÍO AVENDAÑO S.  
LIC. ESTELA CORDERO B.  
LIC. RICARDO DÍAZ G.

TESIS CON  
FALLA DE ORIGEN



**FACULTAD  
DE PSICOLOGÍA**

CIUDAD UNIVERSITARIA

JUNIO DEL 2003

A



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**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.

## **DEDICATORIAS**

**Dedico esta tesis a mis padres, a mi profesores,  
a mis familiares y amigos.**

**Una especial dedicatoria a ti madre,  
por enseñarme a amar la vida, como siempre lo hiciste,  
y por que serás siempre un ejemplo a seguir en mi vida.**

**Papá, gracias por tu consejo y apoyo incondicional,  
por creer en mi y en mis sueños, muchas gracias.**

**Hermanos y familia, los amo a todos,  
gracias por escucharme y apoyarme,  
siempre serán mis aliados consentidos.**

**Pily, sigo diciéndome afortunado por estar a tu lado,  
un incansable GRACIAS por tu amor, paciencia y apoyo,  
por creer en mis proyectos y sueños.**

**Te amo.**

**A todos los compañeros de generación,  
que durante más de cuatro años vivieron  
día con día este proyecto de vida.**

TESIS CON  
FALLA DE ORIGEN

**“Con todo mi cariño y esfuerzo dedico esta tesis a todos  
los hombres y mujeres que tiene fe en ellos mismos  
para lograr sus objetivos en la vida”**

## **AGRADECIMIENTOS**

**Gracias a Dios, a mis padres, familiares, hermanos, maestros y amigos que me brindaron una enseñanza, un aliento y consejo en este proceso tan importante en mi vida.**

**A mi asesor de Tesis Rubén Varela:**

**“Gracias por tu apoyo incondicional, tu enseñanza y guía, por tu esfuerzo para hacer de esta ciencia, una forma de vida”**

**Un agradecimiento especial a la Mtra. Enedina Villegas, asesora de Tesis y excelente persona.**

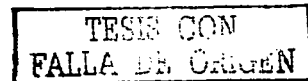
**Al comité de Tesis integrado por la Mtra. Rocío Avendaño S., Lic. Estela Cordero B., Dr. Samuel Jurado C. y Lic. Ricardo Díaz G., por el interés y apoyo mostrado en esta Tesis.**

**A todos aquellos profesores y alumnos, Escuelas y Facultades que participaron e hicieron posible el desarrollo de esta Tesis.**

**“Por mi raza hablará el espíritu”**

**Atentamente**

**Genaro de la Cuadra C.**



## INDICE

**RESUMEN.....V**

**INTRODUCCIÓN..... VI**

### **CAPITULO I "INTELIGENCIA (HABILIDADES)"**

1.1. INTODUCCION.....	9
1.2. HISTORIA DEL CONCEPTO DE INTELIGENCIA .....	9
1.2.1. HABILIDADES.....	12
1.3. CONCEPTOS Y TEORÍAS DE INTELIGENCIA (HABILIDADES) .....	15
1.3.1. CONCEPTOS .....	15
1.3.2. TEORIAS .....	19
1.4. MEDICIÓN DE INTELIGENCIA (HABILIDADES) .....	31

### **CAPITULO II "CARACTERISTICAS DE LOS TESTS, CONFIABILIDAD, VALIDEZ, ESTANDARIZACIÓN"**

2.1. CONFIABILIDAD .....	37
2.1.1. COEFICIENTE DE CONFIABILIDAD .....	40
2.2. TIPOS DE CONFIABILIDAD .....	40
2.2.1. CONFIABILIDAD TEST – RETEST .....	40
2.2.2. CONFIABILIDAD DE FORMAS ALTERNAS .....	41
2.2.3. CONFIABILIDAD DE DIVISIÓN POR MITADES .....	43
2.2.4. CONFIABILIDAD DE KUDER-RICHARDSON Y COEFICIENTE ALFA .....	44
2.2.5. CONFIABILIDAD ENTRE CALIFICADORES .....	46
2.3. EL CONCEPTO DE VALIDEZ Y SU EVOLUCIÓN HISTÓRICA .....	47
2.4. TIPOS DE VALIDEZ .....	55
2.4.1. VALIDEZ DE CONTENIDO .....	56

2.4.2. VALILDEZ CONCURRENTE Y VALIDEZ PREDICTIVA (EN RELACION CON UN CRITERIO .....	57
2.5. NORMAS Y SIGNIFICADO DE LAS PUNTUACIONES DE LOS TEST .....	61
2.6. CONCEPTOS ESTADÍSTICOS .....	62
2.7. NORMAS .....	65
2.7.1. PROCESO DE NORMALIZACIÓN DE UN TEST PSICOLÓGICO .....	67
2.7.2. LA RELACIÓN DE LAS NORMAS DE UN TEST TIPIFICADO CON LA POBLACIÓN PARA LA CUAL SE ELABORA ESE TEST .....	68
2.8. TIPOS DE NORMAS .....	70
2.8.1. NORMAS DE DESARROLLO .....	70
2.8.2. NORMAS INTRAGRUPPO .....	70
2.9. RELATIVIDAD DE LAS NORMAS .....	75
2.10. TESTS PSICOLOGICOS Y SU ADAPTACIÓN A OTRAS CULTURAS .....	79
2.11. TESTS PSICOLÓGICOS Y SU APLICACIÓN EN LA PSICOLOGÍA .....	82

## **CAPITULO III "ESCALA DE HABILIDAD GENERAL PARA ADULTOS (GAMA)"**

3.1. INFORMACIÓN DESCRIPTIVA .....	88
3.2. OBJETIVOS DEL GAMA .....	90
3.3. ESTRUCTURA DEL GAMA .....	90
3.4. USOS DEL GAMA .....	92
3.5. ADMINISTRACIÓN Y CALIFICACIÓN .....	93
3.6. DESARROLLO DEL TEST Y ESTANDARIZACIÓN .....	93
3.6.1. DESARROLLO DE LA PRUEBA .....	94
3.6.1.1. DESARROLLO DE LOS ITEMS .....	94
3.6.1.2. LEGIBILIDAD DE LAS INSTRUCCIONES DEL GAMA .....	94
3.6.1.3. TRYOUT Y SELECCIÓN DE ITEMS .....	95
3.6.1.4. RECOLECCIÓN DE LA BASE DE DATOS NACIONAL PARA LA ESTANDARIZACIÓN .....	96
3.6.1.5. CONTROL DE CALIDAD .....	97
3.6.2. ESTANDARIZACIÓN .....	97

3.6.2.1.	CARACTERÍSTICAS DEMOGRÁFICAS DE LA MUESTRA DE ESTANDARIZACIÓN .....	97
3.6.2.2.	DERIVACIÓN DE LAS CALIFICACIONES DE LA PRUEBA .....	100
3.6.2.3.	CALIFICACIONES DEL SUBTEST .....	100
3.6.2.4.	PUNTUACIÓN DEL C.I. GAMA .....	101
3.6.2.5.	DESARROLLO DE LOS PERCENTILES E INTERVALOS DE CONFIANZA DEL C.I. GAMA .....	101
3.7.	PROPIEDADES ESTADÍSTICAS .....	101
3.7.1.	CONFIABILIDAD .....	101
3.7.2.	CONSISTENCIA INTERNA .....	102
3.7.3.	ERROR ESTÁNDAR DE MEDICIÓN .....	103
3.7.4.	CONFIABILIDAD TEST – RETEST .....	104
3.7.5.	VALIDEZ .....	105
3.7.5.1.	DESARROLLO DEL TEST .....	106
3.7.6.	CORRELACIÓN CON OTROS TESTS DE HABILIDADES GENERALES .....	108

## **CAPITULO IV “MÉTODO”**

4.1.	PLANTEAMIENTO Y JUSTIFICACIÓN DEL PROBLEMA .....	112
4.2.	PREGUNTAS DE INVESTIGACIÓN .....	112
4.3.	HIPÓTESIS DE INVESTIGACIÓN .....	112
4.4.	VARIABLES .....	113
4.4.1.	VARIABLE INDEPENDIENTE .....	113
4.4.2.	VARIABLE DEPENDIENTE .....	113
4.5.	DEFINICIÓN CONCEPTUAL DE LA VARIABLE .....	113
4.6.	DEFINICIÓN OPERACIONAL DE LA VARIABLE .....	113
4.7.	POBLACIÓN .....	114
4.8.	MUESTRA .....	114
4.9.	ESCENARIO .....	114
4.10.	INSTRUMENTO .....	114
4.11.	DISEÑO .....	115
4.12.	TIPO DE ESTUDIO .....	115
4.13.	PROCEDIMIENTO .....	116
4.14.	PROCEDIMIENTO ESTADÍSTICO .....	118

## **CAPITULO V "RESULTADOS"**

5.1. ANÁLISIS ESTADÍSTICO .....	119
5.2. DESCRIPCIÓN DE LA MUESTRA .....	119
5.3. PODER DISCRIMINATIVO Y CONFIABILIDAD .....	120
5.4. VALIDEZ .....	124
5.5. DISCUSIÓN .....	127
5.6. CONCLUSIONES Y APORTACIONES .....	128

<b>REFERENCIAS DOCUMENTALES.....</b>	<b>133</b>
--------------------------------------	------------

<b>APENDICE.....</b>	<b>137</b>
----------------------	------------



## **RESUMEN**

Esta investigación tuvo como objetivo el obtener datos de Confiabilidad y Validez de la Escala de Habilidades Generales para Adultos (GAMA) al ser administrada a una muestra de estudiantes universitarios de la ciudad de México, esto con el fin de decidir si dicho instrumento es útil como herramienta de evaluación para la población mexicana. Este instrumento de tipo no verbal, de lápiz y papel, se administró a 670 estudiantes de nivel universitario de la ciudad de México; todos y cada uno de los 66 reactivos que componen el GAMA discriminaron adecuadamente con un análisis de consistencia interna que arrojó coeficientes de correlación ítem-Total superiores a .20; la confiabilidad general del instrumento se obtuvo a través del método de Alpha de Cronbach arrojando un coeficiente de correlación de .85. Referente a la Validez de constructo, esta se comprobó por medio de un análisis factorial de componentes principales con rotación varimax que arrojó 1 factor general y 4 secundarios, comprobándose que no existe diferencia entre la estructura original de la prueba y la de la versión mexicana. Se concluye a partir de los datos obtenidos que la Escala de Habilidades Generales para Adultos (GAMA) es útil para utilizarse como herramienta de evaluación psicológica en población mexicana adulta.

# **CONFIABILIDAD Y VALIDEZ DE LA ESCALA DE HABILIDAD GENERAL PARA ADULTOS (GAMA) EN UNA MUESTRA DE ESTUDIANTES UNIVERSITARIOS MEXICANOS**

## **INTRODUCCIÓN**

**E**ste trabajo surge a partir de la creciente necesidad ética y el compromiso de un correcto manejo y aplicación de instrumentos psicométricos, los cuales puedan contar con el sustento teórico adecuado para su aplicación en los diversos grupos y escenarios laborales. Encontramos un mal manejo de pruebas en el proceso de reclutamiento y selección, promoción y evaluación de personal aplicados en la empresa Mexicana. Diversos estudios e investigaciones arrojan datos alarmantes del notable decremento de la importancia en la aplicación de las pruebas psicológicas, esto debido en gran medida a la falta de conocimiento de los aspectos psicométricos y a su vez, el uso indiscriminado de estas. Evidentemente el prestigio y certeza que alcanza una disciplina científica está determinada por el rigor con que utiliza medidas precisas y validas en la recopilación, análisis e interpretación de los datos que estudia, esto nos lleva al hecho de que al no ser aplicada una prueba psicológica de manera adecuada para lo que fue construida y dirigida a la población para la cual fue diseñada, los resultados en primera instancia serán incorrectos y esta situación podría acarrear consecuencias verdaderamente lamentables para los examinados, y por ende, para la empresa en sí.

Es importante anotar que antes de poder utilizar una prueba con cierta seguridad como instrumento de medición preciso de la condición psicológica que debe medir y para el fin que se le dé, debe obtenerse información con respecto de su

confiabilidad y validez. Además, resulta útil para los propósitos de interpretación de las calificaciones tener disponible datos sobre el desempeño de un grupo extenso de personas como aquel con las que va a emplearse el instrumento. A fin de lograr esta tarea, deben estandarizarse la prueba, el inventario, la escala de calificaciones u otro instrumento psicométrico.

En los instrumentos psicológicos, las puntuaciones suelen interpretarse haciendo referencia a normas que representan el desempeño de la muestra de estandarización en la prueba; es decir, las normas se establecen empíricamente al determinar lo que hacen en la prueba los miembros de un grupo representativo. La puntuación transformada de cualquier individuo se refiere entonces a la distribución de las puntuaciones obtenidas por la muestra de estandarización para descubrir qué lugar ocupa en esa distribución (Anastasi, 1998).

Es claro que cualquier norma se restringe a la población normativa particular de la que se tomó la muestra. El usuario no debe perder de vista la forma en que se establecen las normas, pues no son, de manera alguna, absolutas, universales o permanentes, sino que simplemente representan la ejecución en la prueba de las personas que formaron parte de la muestra de estandarización. Al elegir dicha muestra, suelen hacerse esfuerzos por obtener una que sea representativa de la población a la que va dirigida la prueba. Siempre hay que considerar y verificar nuestra muestra con respecto a la distribución geográfica, el nivel socioeconómico, la composición étnica y otras características pertinentes para asegurar que representan de verdad a la población definida (Anastasi, 1998).

**S**e pretende realizar un trabajo trascendente y funcional para la psicología general, ya que se ha detectado una problemática en materia psicométrica, la cual incluye una falta de instrumentos que cuenten con datos de estandarización y normas adecuadas para sujetos mexicanos. Se ha observado que muy pocas empresas consideran estas variables para la aplicación de pruebas psicométricas. Esta realidad conlleva una serie de problemas internos y externos. En primer lugar es trascendental el siguiente cuestionamiento: ¿Cuál es el papel del psicólogo general con respecto al uso de pruebas psicológicas? Partiendo de que el psicólogo debe estar formado bajo principios éticos y profesionales, cuestiones como estas no se presentarían en ningún campo de aplicación en psicología en el cual se requiera el uso de pruebas psicológicas, y la segunda pregunta obligada sería: ¿Cuáles son las consecuencias de un mal manejo de las pruebas psicológicas en el diagnóstico? En este sentido son muchas las desventajas que se presentan, por ejemplo, que las interpretaciones de pruebas no estandarizadas y normadas para poblaciones específicas ciertamente serían erróneas. Una consecuencia de esto puede crear un impacto importante en el desarrollo de profesionales, que llevaría a que los procesos de selección de personal, ascensos y promociones no tengan un sustento válido ni confiable, lo cual puede repercutir en el desarrollo y en el beneficio económico de la empresa. Dificultades como estas son causadas por un error que pudiera parecer insignificante y que en realidad afecta el curso para lograr los objetivos planteados. Este breve análisis nos orienta a poner énfasis en la importancia de realizar investigaciones y trabajos comprometidos en materia de psicometría aplicada a cualquier área de la psicología.

# **CAPITULO I "INTELIGENCIA (HABILIDADES)"**

## **1.1. INTRODUCCIÓN**

En la actualidad la inteligencia tiene múltiples acepciones ya que ha sido definida por diversos autores que se dedicaron a hacer investigación en este campo de diversas maneras. Dichas definiciones van desde lo operacional hasta lo abstracto. Aquí es importante hacer notar que la inteligencia no es una entidad sino que más bien es un constructo hipotético usado para referirse a una gama amplia de conductas etiquetadas como "inteligentes". La conducta inteligente se caracteriza por ser adaptativa, además de ser indispensable en la solución de problemas, el razonamiento, la autocrítica y la productividad (Arce y Arce, 2001).

El término de "inteligencia" o "habilidades" se utilizan con una gran cantidad de significados, no solo por el público en general sino por miembros de diferentes disciplinas, como la biología, filosofía, pedagogía, etc., Y en la psicología difiere el significado dependiendo de la orientación teórica.

Para el público en general el C.I. no se identifica con un tipo particular de puntaje de un test, mas bien, a menudo se designa como inteligencia.

## **1.2. HISTORIA DEL CONCEPTO DE INTELIGENCIA**

El concepto de inteligencia se deriva del latín "intelligere" que significa "recolectar de entre" y abarca los conceptos de percepción, discernimiento, selección y establecimiento de relaciones.

Binet (1953), señala que el término inteligencia se remonta a Aristóteles, quien distingue **orexis**, las funciones emocionales y mentales, de **dianoia**, las funciones cognoscitivas e intelectuales (Vernon, 1982).

# PAGINACIÓN DISCONTINUA

Filósofos griegos como Platón y Aristóteles, y escolásticos medievales como Tomas Aquino (1225 - 1274), estudiaron el concepto "inteligencia" y concluyeron que ésta era una "facultad mental"; Sin embargo, el éxito para medir dicha variable se logró en este siglo. Aquino describió esta facultad general como la habilidad para detectar similitudes y desigualdades entre cosas. Se piensa que el concepto de inteligencia nace de la observación de las personas que intentan resolver problemas o aprender cosas difíciles y que exigen esfuerzo como las matemáticas, las lenguas o la historia.

Platón distinguió tres aspectos básicos a los que llamó inteligencia, emoción y voluntad; la primera, guía y dirige; las segundas suministran la fuerza motriz. Aristóteles simplificó esta división: contrapuso la capacidad intelectual o cognitiva a la orectiva o apetitiva que abarca a la emoción y voluntad. Asimismo, realizó otra aportación fundamental al contrastar la actividad o conducta **observada**, con la aptitud (capacidad hipotética **latente**); postuló que la inteligencia es una aptitud que puede mostrarse o no en la práctica y que tiene que deducirse de la conducta observada mediante el uso de reglas científicas propias de todo procedimiento experimental. Por lo tanto, las **aptitudes** (en relación con la función cognitiva), los **rasgos** (en relación con la personalidad), y las **actitudes** (en relación con las posturas y opiniones sociales), son conceptos de estructura latente muy importantes (Eysenck, 1983).

Varios autores opinan que Cicerón tradujo el concepto platónico y aristotélico de capacidad cognitiva o intelectual por la palabra intelligentia (inter-dentro, leer-reunir, escoger, discriminar); así nació el concepto de inteligencia.

El término inteligencia (actualmente popular), era desconocido y raramente utilizado hace un siglo. A finales del siglo XIX, el filósofo Herbert Spencer y el científico Francis Galton, motivados por los escritos de Charles Darwin, acerca del origen de las diferencias entre las especies, formularon una teoría, la cual sostiene

la comprensión de un doble proceso, analítico o discriminativo por una parte, y sintético o integrativo por otra. Su función esencial consiste en capacitar al organismo para que se adapte a un medio complejo y cambiante. Spencer volvió a introducir el término latín *inteligencia* para referirse a las diferencias individuales en la capacidad mental, es decir, para designar las características básicas de toda manifestación y diferenciación cognitivas, agregando elementos biológicos. Estos autores y sus seguidores creían que en los seres humanos existía un grado innato de inteligencia general la cual se distinguía de las capacidades especiales.

A diferencia de Spencer, Galton, además de investigar sobre la inteligencia, intentaba demostrar sus fundamentos hereditarios (Eysenck, 1983).

Existe una definición fisiológica de la inteligencia considerada como "la aptitud para establecer relaciones entre las percepciones sensoriales o para abstraer y asociar conceptos". Esta definición establece que de la comparación entre el comportamiento de los diversos animales se detectan claras diferencias entre el hombre y otras especies, como son: el sentido de orientación; representación espacial, y el lenguaje ligado a la noción de espacio, la capacidad de aprendizaje para explorar objetos precisos que le permite estar abierto a todo lo que hay a su alrededor; curiosidad explorativa para poder adoptar gran variedad de opciones; capacidad para ordenar el proceso de experimentación entre la práctica y el conocimiento. Esta última capacidad, resumen y culminación de las anteriores, se localiza fisiológicamente en el centro práctico y fásico del cerebro, que es la base del lenguaje; su manifestación diferencia a la inteligencia animal de la inteligencia humana considerando esta última como mucho más evolucionada y perfecta.

En el estudio de la inteligencia se tienen los enfoques psicométricos y los empíricos-teóricos; los primeros estudian los conceptos abstractos de tipo cuantitativo como la inteligencia, la dificultad de un problema, los factores y vectores, el espacio de  $n$  dimensiones. En cambio, los segundos se centran sobre



el número de neuronas, las conexiones sinápticas, la organización estructural del ARN, etc.; todo ello, susceptible de estudios empíricos y de la observación directa. Ambos enfoques son importantes y deben llevarse a la práctica si se pretende formar una idea real de lo que es la inteligencia (Vernon, 1982).

Diversos autores poseen diferentes enfoques en lo que se refiere al pensamiento abstracto (por ejemplo, Terman); la resolución de problemas, la capacidad de planeación (por ejemplo, Porteus); la atención, la adaptabilidad, la capacidad de educación o aprendizaje, el discernimiento y la captación de las relaciones (por ejemplo, Spearman). Binet consideró a la inteligencia como un conjunto de cualidades, incluyendo: 1) la apreciación de un problema y la dirección de la mente hacia su ejecución, 2) la capacidad para realizar las adaptaciones necesarias para alcanzar un fin determinado y 3) el poder de autocrítica. En otros términos, Binet señala que la cualidad fundamental es "el juicio denominado también buen sentido, sentido práctico, iniciativa y la facultad de adaptarse a las circunstancias. El juzgar, el comprender y el razonar bien son (Binet y Simon, 1905), las actividades esenciales de la inteligencia" (Vernon, 1982).

Binet se refiere tanto a los rasgos motivacionales como a los cognoscitivos, y en esto Wechsler (1958), está de acuerdo, pues definió la inteligencia como el conjunto o capacidad global del individuo para actuar con un fin, pensar de modo racional y enfrentarse con eficacia a su ambiente (Vernon, 1982).

### **1.2.1. HABILIDADES**

Para ayudar a las personas a alcanzar su nivel óptimo de funcionamiento necesitamos comenzar por su punto de partida, evaluar sus fortalezas y debilidades y hacer un plan en función de ello. Los tests psicológicos en general, y particularmente los de inteligencia, nos ayudan a elaborar dicho plan; pero los tests no deben emplearse como una simple etiqueta para clasificar a los

individuos, sino como un instrumento que nos ayude a comprenderlos. Así mismo, el CI es sencillamente la expresión del nivel de habilidad que, de acuerdo con las normas de edad disponibles, muestra un individuo en cierto momento.

Muchas de las habilidades desarrolladas durante la etapa escolar, como la lectura y las operaciones aritméticas, pueden aplicarse luego a una gran variedad de situaciones de aprendizaje. De la misma manera, las técnicas eficaces y sistemáticas de resolución pueden aplicarse para resolver nuevos problemas. Las diferencias individuales en la adquisición de esas habilidades se verán reflejadas en la ejecución de un gran número de tareas diferentes. Los factores motivacionales, la personalidad del individuo y la experiencia influyen en las habilidades y su medición. La inteligencia humana comprende la combinación de habilidades cognoscitivas y de conocimiento que demanda, fomenta y recompensa el contexto de experiencias en el que funciona el individuo (Anastasi, 1986).

En un esfuerzo por evitar el significado excesivo que se ha asociado a los términos "aptitud" y "aprovechamiento", cada vez es más frecuente el uso del término "habilidad" para designar las mediciones de la conducta cognoscitiva. Cualquier prueba cognoscitiva, independientemente de cómo se llamara antes, ofrece una muestra de lo que el individuo sabe en el momento en que se le probó y mide el nivel de desarrollo obtenido en una o más habilidades. Los resultados obtenidos con estos tests deben ser interpretados tomando en cuenta los antecedentes y las experiencias de cada individuo. En este sentido, cada puntuación tiene un pasado que debe ser explorado para la adecuada comprensión del sujeto que la obtuvo; pero también tiene un futuro en la medida en que permite cierta predicción de lo que hará en otras situaciones, ajenas a la prueba, así como en algún tiempo futuro. Los tests de habilidades desarrollados se han diseñado como tests de inteligencia general, baterías de aptitudes múltiples, tests de aptitudes especiales o pruebas de aprovechamiento (Anastasi, 1998).

En las escuelas se emplea casi cualquier prueba disponible; en el repertorio del consejero y de los psicólogos escolares pueden encontrarse tests de inteligencia, de aptitudes especiales y de aptitudes múltiples para medir las habilidades de los educandos. La aplicación de pruebas de aprovechamiento bien elaboradas puede facilitar el aprendizaje en los estudiantes; ya que estas pruebas proporcionan un medio para adaptar la instrucción a las necesidades individuales, pueden emplearse como apoyos en la evaluación y el mejoramiento de la enseñanza, así como en la formulación de las metas educativas además de proporcionar información sobre lo adecuado de la enseñanza de las habilidades y contenido esencial.

En la empresa, es costumbre que las pruebas psicológicas se empleen como apoyos en la toma de decisiones ocupacionales, que incluyen tanto la consejería individual como las decisiones institucionales concernientes a la selección y clasificación de personal en función de sus habilidades. Desde el punto de vista del empleado y el empleador, es obvia la importancia de que los individuos se coloquen en los trabajos para los cuales están calificados. La colocación adecuada también implica que los rasgos que sean irrelevantes para los requisitos del trabajo no afecten (favorable o desfavorablemente) las decisiones de selección. Además, para hacer un análisis de puestos adecuado, capaz de identificar los requisitos que lo distinguen de otros, se deben tomar en cuenta manuales de entrenamiento y operación, registros de desempeño, a expertos en la materia, etc., y concentrarse en los aspectos del desempeño que hacen una buena diferenciación entre los mejores y peores empleados. Lo anteriormente expuesto nos permite predecir hasta cierto punto el desempeño laboral del solicitante.

En el campo clínico y en la consejería vocacional resulta común el uso de instrumentos de medición diversos, entre los que se encuentran los tests que miden habilidades. La práctica de la evaluación en la elección de las carreras supone ayudar al individuo a decidir cuál es la carrera apropiada para él, teniendo

en cuenta sus habilidades, intereses, metas valores y temperamento, así como los requisitos de la ocupación. Pocas áreas de la vida son tan importantes para la gente como su ocupación, no solo por la gran cantidad de tiempo que la mayoría dedica a trabajar, sino también porque el trabajo proporciona oportunidades para recibir muchas recompensas intrínsecas y extrínsecas (Super y Sverko; Citados por Anastasi, 1998).

### **1.3. CONCEPTOS Y TEORÍAS DE INTELIGENCIA** **(HABILIDADES)**

#### **1.3.1. CONCEPTOS**

- \* Inteligencia es la capacidad académica general o capacidad educativa general (Aiken, 2003).
- \* Inteligencia sirve para designar una cierta categoría de actos distinguidos de las actividades automáticas o instintivas (Howard, 1998).
- \* Inteligencia se emplea para definir la facultad de conocer o comprender. Significa el rendimiento general del mecanismo mental. Capacidad de ciertos organismos para enfrentarse con una situación nueva improvisando una reacción de adaptación nueva también (Merani, 1976).
- \* Inteligencia es la aptitud para enfrentarse a situaciones nuevas con rapidez y éxito (Merani, 1976).
- \* Éxito medido en ejecuciones de tareas comúnmente llamadas intelectuales (Merani, 1976).
- \* Inteligencia consta de tres conjuntos de aptitudes básicas:
  - resolución de problemas prácticos

- capacidad de expresión
- desempeño social

La resolución de problemas prácticos incluye aptitudes tales como ser capaz de razonar con lógica, identificar relaciones entre ideas, considerar todos los aspectos de un problema y tener criterio amplio.

La capacidad de expresión incluye aptitudes como expresarse claramente, saber conversar, tener conocimientos en un campo particular, ser estudioso, leer mucho y tener un vocabulario extenso.

El desempeño social incluye una serie de características como ser capaz de aceptar a otros tal como son y admitir sus defectos, tener conciencia social y ser sensible ante las necesidades y los deseos de los demás (Darley, 1990).

\* Inteligencia practica: consiste en aptitudes como ser capaz de manejar cualquier situación; saber como alcanzar un objetivo y mostrar interés en el mundo en general (Darley, 1990).

\* La inteligencia consta de siete aptitudes mentales primarias, las cuales se median de diferente manera (Darley, 1990). :

1. Comprensión verbal: estimada mediante pruebas de vocabulario.
2. Expresión fluida: estimada por la rapidez con que una persona puede producir palabras con alguna letra específica.
3. Número: estimada mediante problemas de razonamiento aritmético.
4. Visualización del espacio: reflejada en la capacidad de realizar tareas de tipo de comparación de patrones.
5. Memoria: medida por el desempeño de tareas de recuerdo.

6. Razonamiento: estimada por el rendimiento en problemas de analogía.
7. Velocidad para percibir: medida mediante pruebas que exigen velocidad.  
(1938; Thurstone y Thurstone, 1962, citados por Gregory 2001)

Otras definiciones por autores son (Arce y Arce, 2001):

- \* Spearman (1904,1923): Habilidad general, la cual incluye principalmente la enseñanza y la capacidad para establecer relaciones desde las más simples hasta las más complejas.
- \* Binet y Simon (1905): Habilidad para juzgar bien, comprender bien y razonar bien.
- \* Terman (1916): Capacidad para formar conceptos y comprender sus significados.
- \* Pintner (1921): Habilidad para adaptarse adecuadamente a situaciones nuevas.
- \* Thorndike (1921): Poder de responder de manera adecuada desde la verdad y los hechos.
- \* Thurstone (1921): Capacidad para inhibir adaptaciones instintivas y realizar modificaciones.
- \* Wechsler (1939): Capacidad global para pensar racionalmente y desarrollarse efectivamente en el medio ambiente.
- \* Humphreys (1972): Repertorio de habilidades adquiridas, conocimiento, aprendizaje y generalización de dichas tendencias que se encuentran disponibles en cualquier momento dado y que permiten adaptarse al medio.

\* Sternberg (1985, 1986): Capacidad mental para procesar de manera automática la información y emitirla contextualmente a conductas apropiadas; la inteligencia también incluye componentes tales como las ejecuciones y la adquisición de conocimientos.

\* Eysenck (1986): Transmisión de la información a través de la corteza. Considera la eficiencia del procesamiento neuronal como la base de la inteligencia biológica humana.

\* Gardner (1986): Habilidad o destreza para resolver problemas en cualquier cultura; es decir, es un conjunto de capacidades distintas que actúan independientemente a la demanda de las exigencias.

\* Creci (1994): Habilidades múltiples innatas, las cuales sirven como un rango de posibilidades; el desarrollo de éstas depende de la motivación y de las experiencias educativas.

Pueyo (1996), realiza una recopilación de otras definiciones con respecto al término de inteligencia por autores (Arce y Arce, 2001):

\* Boring: Es lo que miden los tests de inteligencia.

\* Terman: Capacidad para realizar pensamientos abstractos.

\* Woodrow: Capacidad para adquirir habilidades.

\* Anastasi: Cualidad de la conducta adaptativa.

\* Berry, J. B.: Concepto social que opera en el ámbito académico, técnico y práctico.

\* Glaser, R.: Habilidad adquirida tras ejecutar conductas eficaces, que determina el rendimiento de los individuos.

\* Jensen, R.: Es el factor "g" descrito por Spearman.

\* Horn, J.: Es la rectificación de una unidad funcional del procesamiento de la información.

\* Detterman, D. K.: Es un conjunto de aptitudes cognitivas básicas distintas inmersas en un sistema complejo de relaciones que constituye la arquitectura cognitiva humana.

### 1.3.2. TEORÍAS

La psicología diferencial considera tres concepciones de la inteligencia (Arce y Arce 2001):

- a) es una propiedad del cerebro
- b) es una cualidad de la conducta
- c) es una amalgama o conglomerado de capacidades (aptitudes)

\* **Francis Galton.** Las primeras teorías de la inteligencia, propuestas aproximadamente en 1800, resaltaron las áreas sensoriales. Francis Galton y J. Mckeen Cattell, creyeron que la inteligencia estaba relacionada con las habilidades sensoriales. Galton pensaba que los seres humanos varían respecto a su grado de "aptitud natural" (natural ability), lo cual confirmó al realizar diversos estudios con familias, en los que concluyó que la aptitud natural era una propiedad hereditaria (López, 1989).

\* **Cattell r. Y Hebb** han distinguido dos clases de inteligencia, llamándolas "inteligencia a e inteligencia b". O "inteligencia fluida y cristalizada". Las distinciones hechas por estos autores son bastante parecidas. En ambos casos a



una clase de inteligencia se le considera como potencialidad genética y a la otra clase principalmente como el resultado de la experiencia, el aprendizaje y los factores ambientales (Butcher, 1995; citado por Gregory 2001).

\* **Binnet** sostenía que la inteligencia se concentraba en complicados procesos intelectuales, como el juicio, el razonamiento, la memoria y la comprensión, él creía que la inteligencia no era estática, sino que crece y se transforma a lo largo de la vida (Grace, 1997).

\* **Charles Spearman** desarrolla la primer teoría de la inteligencia basada en un análisis estadístico de las puntuaciones de una prueba fue la teoría bifactorial del psicólogo inglés Charles Spearman. Cerca de 1900, en su planteamiento original, Spearman sostenía que todas las actividades intelectuales comparten un único factor común, llamado **factor general** o **G**. Postulaba además la existencia de numerosos **factores específicos** o **s**, cada uno de los cuales se dirigía concretamente a una actividad. Es decir, propuso que la inteligencia estaba compuesta por dos tipos de factores: un factor general (**g**) para la ejecución de tareas cognitivas, y numerosos factores específicos (**s**) que son grupos de redes neuronales que interactúan en una actividad o subtest específico para la ejecución de tareas particulares. De acuerdo con la teoría bifactorial, el propósito de la evaluación debería ser la medición de la cantidad de **g** que posee cada individuo, ya que al encontrarse en todas las habilidades, es la única base para predecir la ejecución del individuo en distintas situaciones. Sería inútil medir los factores específicos ya que, por definición, cada uno opera en una actividad única (Anastasi, 1998).

En consecuencia, Spearman propuso sustituir la heterogénea colección de reactivos que se encuentran en los tests de inteligencia con una sola prueba que tuviera una elevada saturación de **g**. Dos tests que miden el factor **g** de Spearman son las Matrices Progresivas de Raven y el Test de Inteligencia justo para la

Cultura. Cuando se comparan actividades similares, parte de la correlación entre ellas puede estar por encima de lo atribuible al factor **g**, lo que implica que, además de los factores general y específicos, debe existir otra clase intermedia que no es tan universal como **g** ni tan específica como **s**. Dicho factor, común a un grupo de actividades pero no a todas, se le llamó factor de grupo. En la primera formulación de su teoría, Spearman admitía la posibilidad de que existieran factores de grupo muy pequeños e insignificantes; posteriormente, y de acuerdo con investigaciones de varios de sus estudiantes, incluyó factores de grupo mucho más amplios, como las habilidades aritmética, mecánica y lingüística (Anastasi, 1998).

Este autor contribuyó a la invención del análisis de factores para apoyar sus investigaciones sobre inteligencia. El análisis factorial, es una técnica estadística que consiste en un sistema de procesos matemáticos que permite determinar la dimensionalidad de un conjunto de mediciones (número de factores que contiene un instrumento), y ha sido aplicado al análisis de los resultados de los tests de inteligencia por muchos años. Esta teoría sigue vigente salvo que en ocasiones cuestionan el número de factores específicos que pueden existir, puesto que algunas veces han encontrado factores "**s**" que sirven para un determinado test.

\* **Jean Piaget** señala que la inteligencia no es una facultad causal o distinta de la mente, sino una extensión de los productos biológicos de adaptación, que se observa en la evolución de los animales; además, enuncia que la conducta se vuelve progresivamente más inteligente, cuanto más compleja sea la interacción entre el organismo y el ambiente, y cuanto más amplias y lógicas sean las concepciones que tienen los niños del mundo y sus procesos de pensamiento.

Jean Piaget propuso dos procesos para el desarrollo de la cognición: la asimilación y la acomodación en el mundo externo. Al interactuar con el medio ambiente, el

crecimiento del niño desarrolla proyectos o estructuras mentales que sirven como exploración de mapas y guías para su conducta. En el desarrollo de la cognición, los niños normalmente maduran en su intelecto y pasan por sus diferentes estadios: sensoriomotor, preoperacional, concreto operacional y operacional formal. Piaget creyó que el crecimiento de la inteligencia se daba alrededor de los 15 años, pero varios investigadores dudan de esta conclusión. Él sugirió que los proyectos-modelos organizados de conducta o estructuras mentales que dirigen al conocimiento -, encaminaban hacia la madurez; a dicho proceso lo denominó equilibrio. En la teoría de Piaget, la asimilación es la aplicación de un proyecto con un objeto, persona o evento, si este proyecto trabaja surgirá un estado de equilibrio, si no, el resultado será un desequilibrio (estado de tensión dinámica). Por último, la persona debe ajustarse al proyecto y trabajarlo, este proceso lo denominó acomodación (Arce y Arce, 2001).

\* **L. L. Thurstone** fue uno de los principales exponentes de la teoría de los factores múltiples. Thurstone a diferencia de Spearman y el factor "g", realizó un estudio para identificar las dimensiones básicas e independientes de la inteligencia que consistían de aproximadamente una docena de factores de grupo más que de un factor general (de los cuales solo siete se han corroborado frecuentemente). Dichos factores los denominó como **habilidades mentales primarias** y fueron los siguientes:

Factor espacial (E), numérico (N), velocidad perceptiva (P), fluidez verbal (W), memoria asociativa (M), razonamiento inductivo (I) y comprensión verbal (V). Finalmente, Thurstone desconoció la inteligencia de g como un factor superior (Gregory, 2001; Anastasi, 1998).

\* **J. R. Guilford** propuso la teoría de la estructura del intelecto y fue quien describió el número más grande de factores cognitivos; es decir, enfatizó que la inteligencia es multifacética. A través de la aplicación de baterías de inteligencia, Guilford encontró correlaciones con respecto al tipo de procesos mentales que

intervienen, tipos de procesamiento de información, o la forma que toman los ítems en cuanto a la información se refiere. Dichas correlaciones lo llevaron a proponer su modelo. La teoría de Guilford es una extensión de la teoría de las habilidades mentales de Thurstone. Asimismo, calificó su modelo como "morfológico o taxonómico". El modelo original de la estructura del intelecto contenía 120 factores diferentes, los cuales eran independientes, y para cada uno de ellos Guilford y sus colaboradores designaron un test (Gregory, 2001).

En una modificación de su modelo original, este autor clasificó las habilidades intelectuales en tres dimensiones llamadas:

**Operaciones:** Corresponde a lo que hace el individuo. Incluye la cognición, la memoria (actualmente el registro y la retención en memoria), el pensamiento divergente (que es importante en la creatividad), el pensamiento convergente y la evaluación.

**Contenidos:** Se refiere a la naturaleza de los materiales o de la información presentada al examinado sobre las que se realizan las operaciones. Incluye contenidos visuales, auditivos, simbólicos (letras, números), semánticos (palabras) y conductuales (información acerca de la conducta, actitudes y necesidades de otra persona).

**Productos:** Corresponde a la forma en que el individuo procesa la información. Se refiere a los diferentes tipos de estructuras mentales que deben intervenir para dar la respuesta correcta. Los productos se clasifican en unidades, clases, relaciones, sistemas, transformaciones e implicaciones (Gregory, 2001; Anastasi, 1998).

Para 1967 Guilford encontró 5 tipos de operaciones, 5 tipos de contenido y 6 tipos de productos. Esto condujo a la existencia de  $5 \times 5 \times 6 = 150$  posibles tareas intelectuales compuestas por la estructura del intelecto. El modelo de la estructura del intelecto de Guilford es representado en forma de cubo. Inicialmente, Guilford,

asumió que estos 150 factores eran independientes uno del otro, más adelante esto falló (Keiderman, Mellenberg y Elshout, 1981; citados por Gregory, 2001). Por lo tanto, reemplazó el modificado modelo por habilidades y por jerarquías consistentes de 150 factores de primer orden, 85 factores de segundo orden y 16 factores de tercer orden (citado por Gregory, 2001). Sin embargo, este nuevo modelo no ha sido investigado y varios autores lo califican como complejo ya que no se puede verificar empíricamente porque los factores que deberían ser independientes suelen salir correlacionados.

Como resultado de investigaciones posteriores de Guilford el modelo actual consta de 180 factores ( $6 \times 5 \times 6 = 180$ ) ya que divide la faceta de contenido figurativo en componentes visual y auditivo, y la faceta de memoria de operación en los componentes de registro y retención (Guilford, 1967 y 1988; citado por Anastasi, 1998).

\* **Wechsler** (1939), define el concepto de inteligencia como "el agregado o capacidad global de individuo para actuar propositivamente, para pensar racionalmente y para conducirse adecuada y eficientemente en su medio ambiente". Para este autor, difiriendo con Spearman, el factor "g", no puede considerarse de importancia primaria. Es mas bien, un factor de integración entre el conjunto de aptitudes propias del ser humano (Butcher, 1995; citado por Gregory, 2001).

\* **Burt** (1955) desarrolla la teoría jerárquica. Es casi el único que mantiene firmemente su opinión definiendo a la inteligencia como " aptitud cognoscitiva general innata ". Identifica varios niveles de bifurcación: en el primero y más alto nivel existen relaciones, en el segundo asociaciones, en el tercero percepciones y en el cuarto sensaciones (Butcher, 1995; citado por Gregory, 2001).

\* **Guilford** (1967, 1982) sostiene que la llamada inteligencia general no existe, así que podemos ser muy buenos o muy malos en diferentes cosas, y no es necesario que exista una relación entre ellas. Una persona puede ser apta para manejar ciertos contenidos, como los visuales y simbólicos, pero inepta para otros, por ejemplo los auditivos. De la misma manera, una persona puede ser muy hábil en tareas que requieren de la memoria pero muy torpes para tareas que exijan la evaluación de materiales. Por ello Guilford propuso 120 aptitudes mentales, todas ellas teóricamente independientes una de otra (Darley, 1990).

\* **Luria**, desarrolla la teoría del Procesamiento Simultáneo y Sucesivo que surgió de sus investigaciones neuropsicológicas. Luria distinguió 2 formas básicas de la actividad integrativa de la corteza cerebral. La primera de ellas asegura que la integración de la información llega dentro del cerebro primeramente en grupos espaciales simultáneos, y la segunda forma de integración de la información llega consecutivamente al cerebro dentro de organizaciones temporales, series sucesivas.

El "Procesamiento Simultáneo" de la información se caracteriza por la ejecución de diferentes operaciones mentales simultáneamente. Las formas de pensar y percibir que requieren del análisis espacial necesitan procesar la información al mismo tiempo. Luria descubrió que este tipo de procesamiento simultáneo de la información está asociado a los lóbulos parietal y occipital del cerebro.

El "Procesamiento Sucesivo" de la información se necesita para actividades mentales en donde es necesaria una secuencia apropiada de operaciones que deben ser seguidas. A diferencia del Procesamiento Simultáneo de la información en donde la secuencia no es importante, aquí es indispensable. Luria localizó el procesamiento sucesivo en el lóbulo temporal y las regiones frontales adyacentes a él (Gregory, 2001).

Cabe aclarar que muchas formas de procesamiento de la información requieren interactuar con mecanismos simultáneos y sucesivos al mismo tiempo.

\* **Vernon (1982)**, indica que existe un consenso entre los psicólogos acerca de los procesos cognoscitivos que reciben el nombre de inteligencia. Sin embargo, ésta comprende un conjunto de habilidades variadas y no una entidad claramente definida. Para este autor, la inteligencia y los logros dependen del potencial genético y el estímulo ambiental; por lo que es difícil clasificar ciertas habilidades como dependientes de uno u otro factor.

Vernon (1960; citado por Gregory, 2001) clasifica las descripciones de la inteligencia como biológica, psicológica y operativa. Freeman (1962; citado por Gregory, 2001) las divide según que acentúen a) el poder de adaptación al medio, b) la capacidad para aprender y c) la aptitud para el pensamiento abstracto (Butcher, 1995; citado por Gregory, 2001).

No obstante, para Vernon, la inteligencia se refiere a las habilidades más generalizadas, las estrategias del pensamiento y el nivel conceptual general; que se aplican en actividades cognoscitivas o en nuevos aprendizajes, y que se constituyen principalmente mediante la interacción con las experiencias ambientales cotidianas y de manera secundaria mediante la estimulación de la escuela. Además, los logros son más específicos y dependen primordialmente de la instrucción proporcionada en la escuela y del interés que el individuo tenga en ella o en la motivación para aprender el tema de que se trate. El cociente intelectual ayuda a predecir las realizaciones futuras; sin embargo, la inteligencia superior ya no es considerada como causa del éxito en la escuela, ni que la baja inteligencia sea causa de los fracasos (Vernon, 1982).

Vernon jerarquiza la inteligencia en varios niveles: en la parte superior de la jerarquía se encuentra el factor general (g) de Spearman, en el siguiente nivel se localizan dos principales factores de grupo que corresponden a las aptitudes: verbal-educativa (v:ed) y la práctica-mecánica (p:m), que a su vez, en el siguiente nivel pueden dividirse en factores más finos; por ejemplo, el factor verbal-educativo puede descomponerse en los subfactores (factores de grupo menores) verbal y numérico, mientras que el factor práctico-mecánico puede dividirse en subfactores (factores de grupo menores) como habilidades de información mecánica, espacial y psicomotora. Es posible identificar subfactores aún más finos mediante el análisis subsecuente de, digamos, tareas verbales. En el nivel inferior de la jerarquía se encuentran los factores específicos (Anastasi, 1998).

En una elaboración posterior del modelo, Vernon incluyó algunas relaciones más complejas y contribuciones cruzadas de factores de tercer nivel, especialmente los relacionados con el aprovechamiento educativo y vocacional. Así, las habilidades científicas y técnicas están asociadas con la habilidad espacial y la información mecánica; las habilidades matemáticas están asociadas con la habilidad espacial y numérica y más directamente al factor g mediante un factor de inducción (Anastasi, 1998).

En resumen, Vernon piensa que la inteligencia es la velocidad de procesamiento de información característica del sistema cognitivo, cuya idea es compatible con la teoría de Spearman, aunque relega las habilidades mentales primarias de Thurstone (Vernon, 1982).

\* **H. Gardner** (1983) ha propuesto una teoría de inteligencias múltiples basadas en el estudio del cerebro, su conducta y sus relaciones. El autor enfatizó que existen inteligencias humanas relativamente independientes al mismo tiempo que admite que la naturaleza exacta, extensión y número de inteligencias no se han podido establecer con precisión.



El esquema de la inteligencia propuesta por Gardner consta de 6 inteligencias múltiples que son: inteligencia lingüística (verbal), musical, lógico-matemática, espacial, cinestésica-corporal, e inteligencias personales. Tres de los seis tipos de inteligencia son bien conocidos (inteligencia lingüística (verbal), lógico-matemática y espacial).

La "Inteligencia cinestésica-corporal" incluye el tipo de habilidades usadas por atletas, bailarines, mecanógrafas, artistas mímicos o cazadores primitivos.

Las "Inteligencias Personales" incluyen la capacidad tener acceso a los sentimientos de la vida y la habilidad para hacer distinciones acerca de los estados de ánimo, temperamento, motivaciones y las intenciones de los otros. Está relacionada con aspectos intra e interpersonales.

La "Inteligencia Musical" está relacionada con la facilidad para aprender, usar un instrumento o escribir composiciones. Ahí se incluyen aspectos estructurales de la melodía, el ritmo y el timbre.

El fenómeno de la deficiencia mental o discapacidad provee de argumentos sólidos a Gardner para afirmar la existencia de inteligencias separadas, ya que se han reportado casos en los que personas con deficiencia mental desarrollaron increíbles talentos en áreas específicas como el arte, la memoria, la música (Aiken, 1996; Gregory, 2001).

\* **R. Sternberg** (1985) conceptualizó al cerebro como un órgano de información (y resolucionador de problemas), el cual contiene procesos para resolver problemas. Por lo tanto, propuso una teoría triple de inteligencia, comúnmente llamada Triárquica, ya que afirma que la inteligencia está integrada por 3 aspectos que son: Inteligencia Componencial, Inteligencia Experiencial e Inteligencia Contextual.

La Inteligencia Componencial consiste en mecanismos mentales internos que son responsables del comportamiento inteligente, es decir, elementos relacionados con el aprendizaje y la práctica. Este tipo de inteligencia tiene 3 diferentes funciones:

a) Metacomponentes: Son los procesos ejecutivos que dirigen las actividades de todos los otros componentes de la inteligencia. Son responsables de determinar la naturaleza del problema intelectual, seleccionar la estrategia para resolverlo, y estar seguro de que el problema se haya resuelto. Los metacomponentes reciben constantemente retroalimentación sobre qué cosas están ocurriendo a la hora de ir resolviendo el problema. Al respecto, Sternberg encontró que la inteligencia alta o elevada está asociada con utilizar relativamente mayor tiempo en global o planear en un orden más elevado (para formar una macroestrategia) y necesitar relativamente menor tiempo para planear en un orden más bajo.

b) Componentes de Ejecución: Son los procesos mentales correctamente entrelazados que se deben utilizar para ejecutar un ejercicio o para resolver un problema.

c) Componentes de la Adquisición de Conocimientos: Son los procesos usados para aprender. Sternberg afirma que el hecho de que algunas personas sean más hábiles que otras es debido a su capacidad creciente para adquirir conocimientos en el primer momento. Así mismo asegura que si el conocimiento nuevo se integra a un contexto dado el sujeto podrá hacer un mejor uso del mismo).

La Inteligencia Experiencial está compuesta por la habilidad para abordar situaciones novedosas y procesar automáticamente la información.

La Inteligencia Contextual es la actividad mental encargada de la adaptación, formación y selección del entorno real del mundo que para el individuo es relevante en su vida; es decir, elementos relacionados con la adaptación al medio ambiente inmediato. Consta de 3 partes: Adaptación (desarrollo de habilidades requeridas para un entorno en particular; una adaptación exitosa será diferente de una cultura a otra), Selección (habilidad para salir del entorno en el que estamos y seleccionar un entorno diferente que sea mejor para nuestros talentos y necesidades) y Formación.

En resumen; el primer componente es necesario para identificar el tipo de problema y decidir como resolverlo; el segundo, para procesar la información y dar una respuesta, y el tercero se usa para adquirir selectivamente nuevos conocimientos e integrarlos en un todo coherente (Darley, 1990).

Sternberg ha tratado de combinar los conceptos de inteligencia y personalidad (Aiken, 1996; Gregory, 2001).

En la actualidad se considera que los modelos presentados (con excepción del de Guilford), no difieren entre sí, ya que la mayoría de ellos concuerdan en la agrupación de diversas capacidades.

\* De las definiciones, descripciones y teorías particulares propuestas, algunas han llegado a ser muy conocidas, tales como " la aptitud para llevar a cabo pensamientos abstractos " **de Woodrow** y " el poder para construir buenas respuestas desde el punto de vista de la verdad y de los hechos " de **Thorndike**. Por ello, este concluye que la inteligencia puede describirse como la capacidad para vivir una existencia de tanteo (ensayo y error) con alternativas que son todavía solo conductas incompletas (Butcher, 1995).

\* Las ideas actuales sobre la inteligencia se inclinan hacia el enfoque de varias aptitudes diferentes más que una inteligencia de una sola cualidad. Como **Thurstone y Gardner** (1983) proponen que la inteligencia consiste en siete cualidades diferentes (Darley, 1990):

- Inteligencia lingüística (muy obvia en escritores y poetas)
- inteligencia lógica - matemática (esencial en las ciencias, las matemáticas y la filosofía) - inteligencia espacial (se manifiesta en el arte, la navegación y el ajedrez a ciegas)
- inteligencia corporal - cenestésica (danza, atletismo, utilización de instrumentos)
- y dos inteligencias " personales " sensibilidad hacia otras personas y conocimiento de sí mismo. Según este autor, estas siete inteligencias son independientes, aunque "interactúan una con otra y se apoyan una a otra desde el principio de la vida ".

#### **1.4. MEDICIÓN DE INTELIGENCIA (HABILIDADES)**

En los primeros intentos por evaluar la inteligencia (habilidad), se utilizaron distintas pruebas sensoriomotrices, pero la primera medida de inteligencia practica fue la que desarrollaron Alfred Binet y Theodore Simon durante la primera década del siglo XX. La escala de Binet - Simon, una serie de tareas relacionadas con la escuela organizada en orden de dificultad ascendente, producía una calificación de la edad mental para cada sujeto.

Más tarde se llevaron a cabo revisiones de dicha escala por lo que la escala de inteligencia de Stanford-Binet, cuyo autor fue Levis Terman, se volvió la más popular; originalmente se publicó en 1916 y se revisó en 1937, 1960 y 1986. La prueba produjo un coeficiente intelectual promedio, que se definió como CI=100 (EM/EC), a pesar de que también se pudo calcular un CI de desviación de la revisión de 1960 (Aiken, 2003).

La prueba de Binet-Simon fue adoptada en otros países y en 1916, L. M. Terman, de la Universidad de Stanford, publicó una revisión que cubría casi toda la gama de la inteligencia desde los tres años hasta los niveles adultos. La prueba de inteligencia individual más utilizada fue la de Stanford-Binet, hasta que fue reemplazada por las pruebas de Terman-Merrill en 1937 (Vernon, 1982).

La cuarta edición de la escala de inteligencia Stanford-Binet comprendió una teoría psicométrica y procedimientos más avanzados; además, proporcionó calificaciones separadas en quince pruebas. Esta cuarta edición identificó el CI de los individuos y proporcionó información para el diagnóstico de las causas específicas de las dificultades en el aprendizaje; además, adopta un modelo de jerarquización de inteligencia. El factor general está medido por puntos. Los subcomponentes del factor general incluye un factor de habilidades cristalizadas (razonamiento verbal y razonamiento cuantitativo), un factor de habilidades fluidas-analíticas y un factor de memoria a corto plazo. El SB:FE consiste de 15 subtests, pero se administran según las edades. Dicha prueba tiene una excelente estandarización (Gregory, 2001). De las distintas traducciones y revisiones de la escala de Binet - Simon, la escala de inteligencia de Stanford - Binet se volvió la más popular.

Posteriormente las escalas de Wechsler ganaron mayor popularidad. David Wechsler fue quien realizó el Test de inteligencia denominado Wechsler-Bellevue, publicado en 1939 y adaptado en 1946. Otros tests y sus revisiones son: Wechsler para niños y la Escala Primaria de Inteligencia Revisada (1967, 1989), la Escala de Inteligencia Wechsler para Niños-III (1949,1974, 1991) y la Escala de Inteligencia Wechsler para Adultos Revisada (1955, 1981) (Gregory, 2001).

Todas las escalas Wechsler usaron el mismo formato: de 10 a 13 subtests divididos en parte verbal y parte ejecutiva; se obtiene el CI con un estándar de 100 y

desviación de 15; una característica común de los subtests, es que los examinados pueden fácilmente transferir sus habilidades de una escala a otra (Gregory, 2001).

La Escala de Inteligencia Wechsler para Adultos Revisada (WAIS-R), es la más utilizada. El test tiene excelente confiabilidad y validez. La escala de Inteligencia Wechsler para Niños III (WISC-III) para niños con edades de 6 a 12), contiene cuatro áreas: comprensión verbal (subtests verbales), organización perceptual (subtests de ejecución), distracción (aritmética y dígitos) y procesamiento (códigos y símbolos) (Gregory, 2001).

Las escalas Wechsler han sido más populares debido a que sus subpruebas se agrupan en diez categorías de acuerdo a su contenido y a que proporcionan tres clases de coeficientes intelectuales de desviación: verbal, de ejecución y la escala completa. Por otro lado, las subpruebas en el WAIS-R ofrece información clínica para el diagnóstico del daño cerebral orgánico y ciertos trastornos mentales.

Entre las muchas pruebas de inteligencia con propósitos especiales que se aplican en forma individual, se encuentran las pruebas como la prueba de vocabulario con ilustraciones de Peabody y la prueba de madurez mental de Columbia, así como otras pruebas de una sola tarea como los laberintos de Porteus y los diseños con bloques de Kohs. Otras pruebas no verbales para individuos con incapacidades del lenguaje o físicas incluyen las baterías de pruebas de desempeño, como la escala internacional de desempeño de Leiter. Las pruebas de inteligencia que se aplican en grupo, y que pueden presentar varias personas a la vez, se utiliza con mayor frecuencia que las pruebas individuales. Estas se derivan de la primera guerra mundial, basados en los primeros trabajos de Otis y otros psicólogos. Algunos ejemplos actuales de pruebas de inteligencia de grupos actuales son la prueba de capacidades escolares de Otis - Lennon, la de habilidades cognoscitivas y la de capacidades cognoscitivas. Ciertas pruebas de grupo de capacidad académica, sobre todo la prueba de aptitud académica (sat). También están disponibles

pruebas de inteligencia de grupo no verbales para personas con incapacidades de lenguaje como la prueba de dibujo de Goodenough - Harris.

Los tests de inteligencia son descendientes directos de las escalas originales de Binet, aunque han sido diseñados para emplearse en diversos contextos y para diversos fines. Como muchos de los tests de inteligencia se validan con medidas de aprovechamiento académico, es común que se les conozca como pruebas de aptitud escolar o de inteligencia académica.

La inteligencia evaluada no debe considerarse como una explicación, sino como una descripción. El CI es sencillamente la expresión del nivel de habilidad que, de acuerdo con las normas de edad disponibles, muestra el individuo en cierto momento, pero no existe un test de inteligencia que pueda explicar las causas de dicho desempeño. Por ende, atribuir el fracaso en un instrumento o en la vida cotidiana una "inteligencia inadecuada" es una tautología que en nada contribuye a comprender los problemas del individuo y sí, en cambio, nos hace postergar los esfuerzos por explorar en su historia personal las verdaderas causas de su desventaja. Otro punto que debe considerarse es el que la inteligencia no es una capacidad unitaria, sino un compuesto de varias funciones o habilidades. El término suele emplearse para cubrir la combinación de habilidades que una cultura requiere para sobrevivir y progresar. De lo anterior se deduce que las habilidades específicas incluidas en ese compuesto, así como sus relativos, varían según el tiempo y el lugar; por consecuencia, puede esperarse que las condiciones requeridas para un aprovechamiento exitoso difieran en culturas distintas, en distintos periodos históricos de la misma cultura e incluso en el ciclo vital del individuo, de la infancia a la etapa adulta. Una habilidad relativa tiende a aumentar o disminuir con la edad en aquellas funciones cuyo valor destaca o menoscaba la cultura o subcultura a la que se pertenece (Anastasi, 1998).

Los esfuerzos por comprender lo que miden los test de inteligencia no solo comprenden los procedimientos estadísticos de validación de constructo como el análisis factorial, sino también la aplicación de las técnicas de procesamiento de información a las tareas presentadas en dichos instrumentos. Esta última aproximación no se limita si la respuesta del sujeto fue correcta, sino que pretende identificar los procesos elementales por los que el examinado llega a ella. Es de esperar que este análisis contribuya de manera importante a utilización de diagnósticos de los instrumentos y ala elaboración de programas de capacitación que permitan satisfacer las necesidades del individuo.

El campo de los tests psicológicos ha crecido con rapidez desde la década de los 20s, en la actualidad, se producen y distribuyen comercialmente cientos de tests. Muchos de estos son de habilidades y personalidad que se crearon en Estados Unidos y se tradujeron a otros idiomas. Además de los test estandarizados que se publican, están disponibles cientos de estos sin publicar. Estos tests que se mencionan en revistas y libros profesionales o especializados, se utilizan en Estados Unidos, Europa y muchos otros países. Generalmente se emplean como instrumentos de clasificación preliminar a cuya aplicación debe seguir la medición de aptitudes especiales, una práctica frecuente al examinar a los adolescentes y adultos normales con fines de consejería educativa, ocupacional, selección de personal y propósitos similares.

Existen bastantes datos indicadores de que está creciendo el interés por los estudios internacionales comparativos sobre inteligencia, rendimiento académico y por la investigación intercultural. Con este crecimiento aparece la necesidad de adaptar (habitualmente denominado "traducir") test de rendimiento, inteligencia, así como otros instrumentos psicológicos, para su uso en múltiples culturas e idiomas (Muñiz, 1996).



La información sobre los tests psicológicos y otros instrumentos de evaluación se pueden encontrar en los catálogos de las compañías que los distribuyen o, con más detalle en los manuales que vienen con los tests, también están disponibles muchos libros que se refieren a los tests.

## **CAPITULO II "CARACTERISTICAS DE LOS TESTS, CONFIABILIDAD, VALIDEZ, ESTANDARIZACION"**

### **2.1. CONFIABILIDAD**

El Término "confiabilidad" nos indica Anastasi (1998) en su libro "Test Psicológicos" se refiere a la consistencia de las puntuaciones obtenidas por las mismas personas cuando se les examina en distintas ocasiones con el mismo instrumento, con conjuntos equivalentes de reactivos o en otras condiciones de examinación. El concepto fundamenta el cálculo del error de medición de un solo resultado, con el que podemos predecir la probable fluctuación en la calificación de un solo individuo debida a factores aleatorios irrelevantes o desconocidos. El concepto de confiabilidad se ha dirigido a varios aspectos de la consistencia de las calificaciones. En un sentido más amplio, la confiabilidad de una prueba indica la medida en que las diferencias individuales en los resultados pueden atribuirse a "verdaderas diferencias" en las características consideradas y el grado en que pueden deberse a errores fortuitos. En términos más técnicos, la medición de la confiabilidad de una prueba permite estimar qué proporción de la varianza total de las puntuaciones se debe a la varianza de error. No se trata de errores, comenta Anastasi (1998), indica que en el sentido de que hubieran podido evitarse o corregirse con una mejor metodología, el sustantivo error es una herencia de las primeras épocas de la psicología, cuando estaba interesada en encontrar las leyes generales de la conducta y en evaluar lo que se suponía que eran rasgos básicos fijos. En la actualidad, los psicólogos reconocen la variabilidad como una propiedad intrínseca del comportamiento y procuran investigar e identificar sus muchas fuentes.

Con respecto a la confiabilidad de las puntuaciones, el problema estriba en la definición de la varianza de error. Factores que para este propósito podrían considerarse varianza de error para otros serían clasificados bajo el rubro de varianza verdadera. En esencia, cualquier condición que sea irrelevante para el propósito de la prueba responde lo que se conoce como varianza de error. Así, cuando los examinadores tratan de mantener uniformes las condiciones de aplicación mediante el control del entorno, las instrucciones, los límites de tiempo, el rapport y otros factores similares, reducen la varianza de error y hacen más confiables las puntuaciones de la prueba; sin embargo, aun cuando se cuente con condiciones óptimas para su aplicación, ninguna prueba es un instrumento perfectamente confiable, de ahí que todas deban ser acompañadas por un reporte de su confiabilidad. También es necesario especificar las características de la muestra normativa, ya que la confiabilidad declarada caracteriza a la prueba cuando se aplica en condiciones normales a personas similares a las que conformaron la muestra normativa u original.

Es evidente que hay tantas variedades de confiabilidad como condiciones que afecten a los resultados de la prueba, ya que cada una de éstas puede ser irrelevante para cierto propósito y por consecuencia clasificada como varianza de error. Pese a ello, en la práctica se calculan relativamente pocas clases de confiabilidad (Anastasi, 1998). Es preciso contemplar que existen diversas técnicas para medir la confiabilidad de las puntuaciones, las cuales cuentan con varianzas de error específicas.

Todas las categorías de confiabilidad pueden expresarse en términos de un coeficiente de correlación porque todas las atañe el grado de consistencia o acuerdo entre dos conjuntos de puntuaciones obtenidas independientemente; por lo que podemos definir al coeficiente de correlación ( $r$ ) como el que expresa el grado de correspondencia, o relación, entre dos conjuntos de puntuaciones. Si el individuo que obtiene la mayor calificación en la variable 1 consigue el mejor

resultado en la variable 2, el individuo con la segunda mejor calificación en la variable 1 también logra el segundo mejor resultado en la variable 2, y así va disminuyendo hasta llegar al individuo con la menor calificación del grupo, entonces existiría una correlación perfecta entre las variables 1 y 2 y sería igual a +1.00. Por el contrario una correlación de cero indica la ausencia absoluta de relación, en general los coeficientes que se encuentran en la práctica se localizan entre estos extremos y dan valores superiores a cero pero menores a 1.00. Las correlaciones entre las medidas de habilidades son casi siempre positivas, aunque a menudo son bajas. Cuando se obtienen una correlación negativa entre dos de las variables, por lo general se debe a la forma de expresar los resultados; por ejemplo, si se correlacionan calificaciones de tiempo con calificaciones de cantidad es probable que se obtenga una correlación negativa.

Anastasi (1998) menciona que los coeficientes de correlación se calculan de diversas maneras, dependiendo de la naturaleza de los datos. El más común es el coeficiente de correlación producto – momento de Pearson, que no sólo toma en consideración la posición de la persona en el grupo, sino también la cantidad de su desviación por encima o por debajo de la media del grupo. Hay que recordar que cuando la posición de cada sujeto se expresa en términos de calificaciones estándares, quienes caen por encima del promedio reciben calificaciones positivas mientras que los que caen por debajo obtienen calificaciones negativas. Si ahora multiplicamos la calificación estándar de cada individuo en la variable 1 por su calificación estándar en la variable 2, todos los productos serán positivos si cada cual cae del mismo lado de la media en ambas variables. El coeficiente de correlación de Pearson es simplemente la media de esos productos: será alta y positiva cuando las calificaciones estándares correspondientes sean de igual signo y aproximadamente iguales en ambas variables. Cuando los sujetos estén por encima del promedio en una variable y por debajo de la otra, los productos cruzados correspondientes serán negativos. Cuando algunos productos son positivos y otros negativos, la correlación será cercana a cero.

Durante muchos años, los niveles de significancia han sido la forma tradicional de evaluar las correlaciones, pero cada vez es mayor la conciencia de la insuficiencia y los defectos de este procedimiento. Demostrar que un coeficiente de confiabilidad (o cualquier correlación) es significativamente mayor a cero proporciona pocos conocimientos para propósitos teóricos o prácticos. Ni siquiera una correlación elevada logra pasar la "prueba de significancia" cuando se obtiene de una muestra pequeña. Una posible alternativa consiste en considerar la magnitud real de la correlación obtenida y estimar el intervalo de confianza dentro del cual es probable que caiga la correlación de la población a un nivel especificado de confianza. Esta tendencia al uso de los intervalos de confianza, como complemento si no es que como sustituto de las pruebas de significancia, anuncian un cambio importante en el análisis de los coeficientes de correlación en los siguientes años (Anastasi, 1998).

### **2.1.1. COEFICIENTE DE CONFIABILIDAD**

Los coeficientes de correlación tienen muchos usos en el análisis de los datos psicométricos, entre ellos la medición de la confiabilidad de una prueba. Este se puede obtener a través del método de producto – momento de Pearson. Se considera una correlación positivo elevada cuando el resultado cae entre .80 a .90.

## **2.2. TIPOS DE CONFIABILIDAD**

### **2.2.1. CONFIABILIDAD TEST – RETEST**

El método más inteligente para encontrar la confiabilidad de las puntuaciones de una prueba consiste en aplicar el mismo instrumento por segunda ocasión. En este caso, el coeficiente de confiabilidad ( $r_n$ ) es simplemente la relación entre los resultados de las mismas personas en las dos aplicaciones de la prueba. La

varianza de error corresponde a las fluctuaciones aleatorias de la ejecución de una sesión a otra, que pueden provenir en parte de condiciones no controladas de la aplicación, como cambios climáticos externos, ruidos repentinos o la punta rota del lápiz, sin embargo, hasta cierto punto surgen de los cambios en la condición de los propios examinados, por ejemplo, una enfermedad, fatiga, etc. La confiabilidad del retest muestra el grado en el que los resultados de una prueba pueden generalizarse en otras ocasiones; entre mayor sea la confiabilidad menos susceptibles serán los resultados a los cambios fortuitos en la condición cotidiana de los examinados o en el entorno en el que se aplica la prueba. Es importante resaltar que cuando se informa de la confiabilidad test – retest en el manual de la prueba, siempre debe especificarse el intervalo en el que se midió. Como las correlaciones test – retest disminuyen progresivamente conforme aumenta el intervalo, no hay uno sino un número infinito de coeficientes de confiabilidad test – retest para cualquier prueba. Aunque en apariencia es simple y directa, la técnica test – retest presenta dificultades al aplicarse a la mayor parte de las pruebas psicológicas. Sólo las pruebas que no son afectadas de manera apreciable por la repetición son adecuadas para la técnica test – retest, en esta categoría se encuentran una serie de pruebas motoras y de discriminación sensorial, sin embargo, en la mayor parte de los casos no es apropiado repetir el examen con el mismo instrumento para encontrar el coeficiente de confiabilidad (Anastasi, 1998).

### **2.2.2. CONFIABILIDAD DE FORMAS ALTERNAS**

El uso de formas alternas de la prueba es una manera de evitar las dificultades de la confiabilidad test – retest. Las mismas personas pueden ser evaluadas con una forma en la primera ocasión y con otra equivalente en la segunda. La correlación entre las puntuaciones de las dos formas representa el coeficiente de confiabilidad de la prueba, que no sólo mide la estabilidad temporal, sino también la consistencia de las respuestas a diferentes muestras de reactivos (o formas de la prueba), lo que permite combinar dos tipos de confiabilidad. Como

ambas son importantes para casi todos los propósitos de la examinación, la confiabilidad de formas alternas proporciona una medida útil para evaluar muchas pruebas. Es conveniente examinar de cero el concepto de muestreo de reactivos, o muestreos de contenido, ya que no sólo fundamenta la confiabilidad de formas alternas, sino también otras clases de confiabilidad.

Como la confiabilidad test – retest, la confiabilidad de formas alternas siempre deben ser acompañadas por el informe de la duración del intervalo entre las aplicaciones de la prueba y de una descripción de las experiencias intermedias relevantes. Si las dos formas se aplican en sucesión inmediata, la correlación muestra la confiabilidad únicamente entre las dos formas, no entre ocasiones. En este caso, la varianza de error representa fluctuaciones en el desempeño de un conjunto de reactivos a otro, pero no fluctuaciones temporales.

Es evidente que en la preparación de formas alternas deben tomarse las precauciones necesarias para asegurar que en realidad son paralelas. En principio, deben hacerse independientemente y diseñarse de modo que cumplan las mismas especificaciones; han de contener el mismo número de reactivos, y éstos tienen que expresarse de la misma forma, cubriendo el mismo contenido y ser de igual alcance y nivel de dificultad. Además, es necesario supervisar la equivalencia de las instrucciones, límites de tiempo, ejemplos ilustrativos, formato y todos los otros aspectos de la prueba. Tener dispuestas formas de la prueba no sólo es deseable por su conveniencia para la determinación de la confiabilidad de la prueba, puesto que también son útiles en los estudios de seguimiento o para investigar los efectos de algún factor experimental que interviene en el desempeño de la prueba. El uso de varias formas alternas brinda asimismo un medio para reducir la posibilidad de preparación o de fraudes (Anastasi, 1998).

Es importante señalar que cuenta con sus limitaciones, en primer lugar, si las funciones de conducta consideradas son muy susceptibles al efecto de la práctica

el uso de formas alternas reduce pero no elimina dicho efecto. Otra cuestión relacionada atañe el grado en el que la naturaleza de la prueba cambia con la repetición; en estos casos no basta con cambiar el contenido de los reactivos de la segunda forma para eliminar este acarreo de la primera. Por último, dadas las dificultades prácticas para elaborar formas verdaderamente equivalentes, muchas pruebas no disponen de formas alternas. Por todo lo anterior, a menudo se requieren otras técnicas para estimar la confiabilidad de la prueba.

### **2.2.3. CONFIABILIDAD DE DIVISIÓN POR MITADES**

Mediante diversos procedimientos de división por mitades es posible obtener una medida de confiabilidad a partir de una única aplicación de una forma de la prueba, ya que al dividirla en mitades equivalentes se obtienen dos puntuaciones de cada persona. Es obvio que la confiabilidad de división por mitades proporciona una medida de la consistencia del contenido muestreado, pero no de la estabilidad temporal de las puntuaciones porque el procedimiento consiste en una única sesión de prueba. Este tipo de coeficiente de confiabilidad se conoce como coeficiente de consistencia interna, ya que requiere una sola aplicación de una única forma.

Puesto que cualquiera prueba puede dividirse de muchas maneras distintas, el primer problema es cómo dividirla para obtener las mitades más equivalentes para encontrar la confiabilidad de división por mitades. En la mayor parte de los casos, es posible que las mitades no sean equivalentes por diferencias en la naturaleza y el nivel de dificultad de los reactivos, por los efectos acumulativos de calentamiento, prácticas, fatiga, aburrimiento y por cualquier otro factor que varíe progresivamente desde el inicio hasta el final de la prueba. Un procedimiento adecuado para casi todos los propósitos es separar las puntuaciones de los reactivos nones y pares. Si los reactivos están dispuestos en un orden aproximado de dificultad, la división producirá calificaciones casi equivalentes de ambas mitades. Una precaución que debe tomarse al hacer la división entre nones y pares



tiene que ver con los grupos de reactivos que abordan un solo problema, como las preguntas que se refieren a un diagrama mecánico particular o a ciertos pasajes en una prueba de lectura. En este caso, el grupo entero de reactivos debe asignarse intacto a una u otra mitad. Si los reactivos de dicho grupo se colocaran en mitades diferentes, la similitud de los resultados estaría influida de manera espuria, por que cualquier error en la comprensión del problema afectaría a los reactivos de ambas mitades.

Una vez que se tienen las puntuaciones de ambas mitades de cada sujeto, pueden correlacionarse por el método usual, pero no hay que perder de vista que esta correlación sólo proporciona la confiabilidad de media prueba. Mientras más extensa sea la prueba más confiable resultará. Es razonable esperar que con una muestra mayor de conducta pueda obtenerse una medida más adecuada y consistente. El efecto que tendrá alargar o acortar la prueba sobre el coeficiente puede estimarse por medio de la formula de Spearman Brown (Anastasi, 1998).

#### **2.2.4. CONFIABILIDAD DE KUDER-RICHARDSON Y COEFICIENTE ALFA**

Este método requiere de una sola aplicación de una sola forma, se basa en la consistencia de las puntuaciones a todos los reactivos de la prueba. Esta consistencia entre reactivos está influida por dos fuentes de varianza de error (1) el muestreo de contenido (como en la confiabilidad de formas alternas y de división por mitades), y (2) la heterogeneidad del área de conducta muestreada. Entre más homogénea sea el área, mayor será la consistencia entre reactivos. Es evidente que las puntuaciones de la prueba serán ambiguas cuando se deriven de instrumentos relativamente homogéneos. Una cuestión de gran importancia es si el criterio que la prueba pretende predecir es en sí mismo homogéneo o heterogéneo. Aunque se prefieren las pruebas homogéneas porque sus resultados permiten una interpretación poco ambigua, es evidente que un solo instrumento

homogéneo no es un predictor adecuado de un criterio muy heterogéneo. Más aun, en la predicción de éste, la heterogeneidad de los reactivos no representa por fuerza una varianza de error. Los tests de inteligencia son un buen ejemplo de un instrumento heterogéneo para predecir criterios relativamente homogéneos; sin embargo, en un caso así puede ser deseable formular varias pruebas relativamente homogéneas, de las que cada una mida una fase diferente del criterio heterogéneo. De este modo puede combinarse una interpretación no ambigua de los resultados de la prueba con una adecuada cobertura del criterio.

Anastasi (1998) indica que el procedimiento más común para encontrar la consistencia entre reactivos se debe a Kuder y Richardson (1937). Como en los métodos de división por mitades, la consistencia entre reactivos se encuentra en una sola aplicación de una única prueba; sin embargo, en lugar de requerir las puntuaciones de las dos mitades, la técnica se basa en el examen del desempeño en cada reactivo. De las diversas fórmulas derivadas del artículo original, la de mayor aplicación, conocida comúnmente como la "formula 20 de Kuder-Richardson" es:

$$KR_{20} = \frac{K}{K-1} \frac{(S^2 t - \sum pq)}{S^2 t}$$

En la  $KR_{20}$  es el coeficiente de confiabilidad de toda la prueba,  $K$  el número de reactivos y  $S^2 t$  la desviación estándar de la puntuación total de la prueba. El único término nuevo de la fórmula,  $(\sum pq)$ , se encuentra calculando la proporción de personas que aciertan ( $p$ ) y la de personas que no aciertan ( $q$ ) en cada reactivo. Para obtener  $(\sum pq)$  se calcula para cada reactivo el producto de  $p$  y  $q$ , y luego se suman los productos de todos los reactivos. Este método requiere de pocos cálculos adicionales porque al elaborar el instrumento rutinariamente se registra  $p$  para encontrar el grado de dificultad de cada reactivo.

La fórmula de Kuder-Richardson puede aplicarse a pruebas cuyos reactivos se califiquen como aciertos o errores, o de acuerdo con algún otro sistema de todo o nada. Ahora bien, cuando las respuestas son ordenadas en diferentes puntuaciones, se deriva la fórmula del coeficiente alfa, en la que la suma de las varianzas de las puntuaciones del reactivo, sustituye a  $(\sum pq)$ . El procedimiento consiste en encontrar la varianza de todas las puntuaciones individuales de cada reactivo y sumar las varianzas de todos los reactivos.

### **2.2.5. CONFIABILIDAD ENTRE CALIFICADORES**

Anastasi (1998) menciona que es evidente que las distintas formas de confiabilidad difieren en los factores que incluyen bajo la varianza de error. En un caso, ésta cubre las fluctuaciones temporales, en otro se refiere a la diferencia entre conjuntos de reactivos paralelos y en otro más comprende cualquier inconsistencia entre reactivos. Por otra parte, los factores excluidos de las medias de la varianza de error son, de manera general, de dos clases: a) los factores cuya varianza debe permanecer en las puntuaciones porque forma parte de las diferencias consideradas, y b) los factores irrelevantes que pueden controlarse experimentalmente. La mayor parte de las pruebas proporciona procedimientos tan estandarizados para la administración y calificación que la varianza de error atribuible a esos factores es insignificante. Una varianza de error que puede supervisarse de manera sencilla es la varianza del calificador. Algunas categorías de pruebas (sobre todo las de creatividad y las proyectivas de personalidad) dejan mucho al juicio del calificador. La confiabilidad del calificador se consigue con una muestra de pruebas calificadas independientemente por dos o más examinadores. Los dos resultados se correlacionan de acuerdo con la forma común y el coeficiente de correlación es la media de la confiabilidad del calificador. Esta clase de confiabilidad suele calcularse cuando se emplean en la investigación instrumentos de calificación subjetiva.

En resumen las clases de coeficiente de confiabilidad se demuestran en la siguiente tabla, clasificando las operaciones seguidas para obtener cada tipo de confiabilidad en relación con el número de formas de prueba y de sesiones de examinación que se requieren, así como las fuentes de varianza de error con relación a los coeficientes de confiabilidad.

<b>Tipos de coeficiente de confiabilidad</b>	<b>Sesiones de aplicación requeridas</b>	<b>Formas de pruebas requeridas</b>	<b>Varianza de error</b>
✓ Test-Retest	Dos	Una	Muestreo de tiempo
✓ Formas alternas (inmediata)	Una	Dos	Muestreo de contenido
✓ Formas alternas (demorada)	Dos	Dos	Muestreo de tiempo y de contenido
✓ División por mitades	Una	Una	Muestreo de contenido
✓ Kuder-Richardson y coeficiente alfa	Una	Una	Muestreo de contenido y heterogeneidad de contenido
✓ Del calificador	Una	Una	Diferencias entre calificadores

Anastasi, A. (1998) "Test psicológicos"

### **2.3. EL CONCEPTO DE VALIDEZ Y SU EVOLUCIÓN HISTÓRICA**

La validez siempre ha sido considerada el aspecto central de la psicometría, pero podríamos decir sin riesgo de equivocarnos que es también el más difícil de definir, lo que ha llevado a una gran proliferación de nombres o etiquetas diferentes para representarlo. Como con otros muchos conceptos de la psicología, una breve revisión de la evolución del término puede ayudar a aclarar el significado del concepto, que cambiado considerablemente en los últimos 40 años (Agnoff, 1988).

Convencionalmente la validez se define como "el grado en que un test mide lo que pretende medir", definición que no ha cambiado si la comparamos con una de las

definiciones antiguas, como la de Garrett en 1937: "la validez se refiere al grado en que un test o un conjunto de operaciones mide lo que dice medir". El problema se encuentra a la hora de operacionalizar dicho grado de relación y es ahí donde ha habido una gran evolución histórica. Durante muchos años primó una aproximación pragmática, empírica, e incluso ateorica, motivada por la expectativa utilitaria de las aplicaciones de los tests y el operacionalismo dominante en la filosofía de las ciencias. Una breve revisión de las principales definiciones de validez dadas antes de 1950, nos confirmará esta impresión. Así, Bingham (1937) define la validez en términos puramente operacionales como "la correlación entre puntuaciones de un test y alguna otra medida objetiva de lo que el test intenta medir". Guilford (1946) la define de un modo muy similar: "de una forma muy general, un test es válido para aquello con lo que correlaciona". Cureton (1950) también define la validez como la correlación entre las puntuaciones observadas del test con las verdaderas del criterio, pero ya introduce algunos matices teóricos, distinguiéndola del poder predictivo del test. Al que considera como correlación entre puntuaciones verdaderas y observables del test y del criterio; diferenciado éste a su vez de la relevancia, a la que define como correlación entre puntuaciones verdaderas de ambos. En la práctica, y tal como señala Angoff (1988), fue el uso de la validez en su sentido predictivo la que denominó la escena psicométrica hasta los años cuarenta. Durante este mismo periodo se fue introduciendo otro tipo de validación por criterios, la validez concurrente (APA; AERA, 1954), considerada como la correlación del test con criterio, difiriendo de la concepción predictiva anterior en que la recogida de las puntuaciones de ambos se realizaba simultáneamente. Desde este punto de vista, se aceptaba como evidencia de la validez de un test la correlación de éste con alguna medida aceptable del rasgo en cuestión.

Otra característica distintiva de la aproximación de la validez ligada a criterios externos resultaba demasiado restrictiva, ya que, aunque directamente aplicable en contextos de selección con finalidad esencialmente predictiva, no servía para

otros muchos tests en los que ellos mismos constituyen su propio criterio, como es el caso de los tests de rendimiento, competencias, trastornos de conducta, etc. Esto llevó a una primera ampliación del concepto de validez, definida ahora además como "el grado en que su contenido representa una muestra satisfactoria del dominio", apareciendo el concepto de validez de contenido.

Otro tipo de validez que fue popular durante los años cuarenta fue el de validez aparente citado por Mosier en 1947, aunque nunca se le concedió mucho interés teórico en la psicometría; se refiere a la imagen, la presentación y lo que a simple vista se ve bien del instrumento empleado.

Un cambio importante fue el introducido en los Standards for Psychological and Educational Test (APA, AERA, 1954), en los que se habla de aspectos de la validez: contenido, predictiva, concurrente y de constructo, introduciendo esta última por primera vez como nuevo tipo. Los Standards de 1964 y 1971 combinaron la concurrente y la predictiva en validez referida al criterio.

No fue casual que los Standards hablasen por primera vez de la validez de constructo, ya que L.J. Cronbach fue el presidente del Comité y Meehl uno de sus miembros y al año siguiente apareció el importantísimo artículo de ambos autores (Cronbach y Meehl, 1955) sobre la validez de constructo y comienza a percibirse ésta como el aspecto fundamental e inclusivo de los restantes aspectos de la validez. La concepción teórica del constructo dicta la naturaleza de los datos a recoger para la validación y la interpretación de las puntuaciones. A su vez, los datos resultantes del estudio de validación se usarán para validar, revisar o rechazar la teoría misma. En este sentido, ya Gulliksen (1950) hablaba de la validez intrínseca, según la cual se postula existencia de constructos alternativos y se examinan los datos de varias medidas, estudiando sus patrones de covariación. Desde el punto de vista de Cronbach y Meehl, todos los datos que provienen de la teoría son útiles para la validez de constructo: distribuciones de grupo, matrices de

correlaciones internas (de ítems) y externas (de tests), criterios externos, datos de desarrollo y cambio, análisis de contenidos, etc. Dentro de esta nueva concepción, parece obvio que la validez no puede expresarse por un coeficiente único, sino que requiere muchas líneas de evidencia, no todas cuantitativas. Todas las medidas que proporcionen datos que puedan considerarse expresiones comportamentales del constructo, incluyendo tanto tests como criterios, pueden usarse en la validación de constructo.

La validez de constructo tuvo un importante avance metodológico pocos años después con la aparición del artículo de Campbell y Fiske en 1959, en el que ofrecieron pruebas contundentes y empíricas para la valoración de constructo, basadas en medidas correlacionales, y diferenciando dos importantes tipos de evidencia de la validez de constructo: validez convergente y validez discriminante. La triple clasificación de la validez tuvo consecuencias teóricas no deseables ya que fue interpretada por los autores y usuarios de los tests en el sentido de que los tests podían validarse por medio de uno o más de los tres procedimientos generales. Además comienzan a proliferar las denominaciones de validez con nombres ligados a los procedimientos empíricos utilizados en su determinación. Se hacía necesario de nuevo la revisión del concepto desde una nueva perspectiva integradora y unificada.

Aunque no podemos olvidar los esfuerzos de otros autores en este sentido, un autor con una significación especial en la integración del concepto de validez fue Meesick que integra las diferentes aproximaciones a la validez dentro de la validez de constructo.

La conclusión de todos los autores es que hay una única validez, la validez de constructo, tal como se refiere en las palabras de Meesick: "la validez de constructo es el concepto unificador de validez que integra consideraciones de contenido y de criterio, en un marco general para probar hipótesis racionales

acerca de relaciones teóricamente relevantes” (1980). Añade además que la validez representa la interpretación de los tests basados en la evidencia proporcionada por sus puntuaciones. Messick señala en el mismo artículo que la validez de contenido mostraría la relevancia de los contenidos y las validades predictiva y concurrente, serían utilidad predictiva y utilidad en el diagnóstico, respectivamente (Martínez, 1995).

Como se puede observar en esta reseña histórica, los trabajos que siguieron al artículo conceptual de Cronbach y Meehl, tuvieron un importante papel clarificador del concepto y de los procedimientos de la validez. Cuando deseamos medir una característica psicológica de los sujetos, en primer lugar se define ésta y a continuación se establecen operaciones diseñadas para proporcionar descripciones cuantitativas del grado en que un individuo la posee o manifiesta. Una cuestión fundamental que debemos plantearnos es el grado en que las propiedades medidas por dicho conjunto de operaciones se corresponden de hecho a las características definidas y son apropiadas para las inferencias y acciones basadas en las puntuaciones de los tests; es decir, antes de aceptar un conjunto de operaciones para la medida de un rasgo, es necesario saber si mide lo que pretende medir y cuáles son las consecuencias potenciales de la interpretación de las puntuaciones.

Es importante destacar que la validez es una cuestión de grado y que no es definitiva, pudiendo ser modificada por nuevas evidencias, consecuencias sociales potenciales, etc. **En concreto la validación de un test es un proceso continuo, nunca acabado.**

Esta perspectiva unificadora de la validez aparece recogida en los Standards For Educational and Psychological Testing (APA, AERA, 1985) que dan la siguiente definición de la validez: “ **la validez se refiere a la adecuación, significación y utilidad de las inferencias específicas hechas a partir de las**



**puntuaciones de los tests... una gran variedad de inferencias pueden hacerse a partir de las puntuaciones de un determinado test y hay muchas formas de acumular evidencia que soporten una inferencia particular. La validez, no obstante, es un concepto unitario y siempre se refiere al grado en que la evidencia soporta las inferencias hechas desde las puntuaciones de los tests. Se validan las inferencias para propósitos especiales, no el test mismo... Una validación ideal incluye varios tipos de evidencias, que comprenden los tres tradicionales (contenido, constructo y criterio)... Los juicios profesionales guiarán las decisiones respecto a las evidencias más importantes a la luz del uso pretendido del test”.**

Esta definición de la validez tiene importantes implicaciones, primero, debemos destacar el énfasis en las puntuaciones y no en los tests o instrumentos, ya que éstos no tienen fiabilidad ni validez, únicamente las tienen las puntuaciones derivadas de ellos. En segundo lugar, y como sucede con la fiabilidad que no es de un test “per se”, sino de unas medidas que representan a las puntuaciones verdaderas, y que en las perspectivas de la TG depende del objeto y de las facetas, los tests o instrumentos tampoco tienen una validez, sino que utilizamos las puntuaciones de los tests para un propósito particular, para hacer inferencias concretas; por tanto, diferentes tipos de inferencias de las puntuaciones de los tests pueden requerir distintas evidencias extraídas de las puntuaciones. Es aquí donde encontramos una ventaja a la definición, refiriéndose a que liga la validez a unas necesidades específicas a las que el psicólogo pretende responder usando el instrumento de medida.

En tercer lugar, mientras que la fiabilidad de una puntuación observada derivada de un procedimiento de medida, puede describirse numéricamente por medio de un coeficiente de fiabilidad, error típico de medida o una función de información, la validez no puede resumirse de forma adecuada por un índice numérico; no hay un

único coeficiente de validez. Los diversos estadísticos denominados coeficientes de validez descritos en los manuales de los tests, trabajos de investigación, etc. , son descripciones numéricas de la fuerza o grado de algunas de las evidencias empíricas de la validez. Finalmente, puesto que la validez es el grado de adecuación de las puntuaciones de un test para un uso particular, la validación de los tests es el proceso de acumulación de pruebas y evidencias que soporten dichos usos e inferencias. El proceso de validación puede estar formado por muchos subprocesos independientes. En resumen y como señala Cronbach (1984) "la meta final de la validación es la explicación y la comprensión y por tanto esto nos lleva a considerar que toda validación es validación de constructo". Además señala que las tradicionales tres formas de validez no pueden considerarse alternativas, sino estrategias de investigación diferentes (Martínez, 1996)

Otra consideración importante que tiene que ver con la validez es la de las consecuencias sociales del test, ya que el uso de los tests está generalizado en todos los ámbitos de la vida cotidiana y en este sentido, Cronbach (1988) señala que "...debe ligar conceptos, evidencias, consecuencias sociales y personales y valores... y los validadores tienen la obligación de revisar si una práctica tiene o no las consecuencias apropiadas para individuos e instituciones, y especialmente evitar las consecuencias adversas". Estas últimas consideraciones se han llevado a considerar como parte del proceso de validación de otros aspectos hoy muy importantes en la psicometría, tales como el sesgo de los ítems y de los tests.

La validación de los tests para el trabajo nos resulta familiar, pues el simple hecho psicométrico de que es necesario averiguar la validez del instrumento para los usos particulares de la prueba, aunque ha adquirido nueva importancia por la preocupación generalizada acerca de colocación laboral de minorías que sufren de desventajas culturales y educativas. Una prueba no válida o una que incluya elementos que no están relacionados con el grupo de trabajo considerado puede excluir de manera injusta a los miembros de grupos minoritarios que podrían

realizar ese trabajo de manera satisfactoria. Otra preocupación pertinente, tanto para las organizaciones como para la sociedad en su conjunto, surge de la relación demostrada entre la productividad en el trabajo y la validez de los instrumentos de selección. Las ganancias y pérdidas estimadas en la productividad que se asocian con los aumentos y las disminuciones de la validez de los procedimientos de selección de personal son considerables; en las instituciones que contratan a muchos empleados, como las dependencias gubernamentales, el valor acumulado de estas ganancias y pérdidas es tan grande que merece una estrecha atención (Anastasi, 1998).

Podemos decir entonces que el proceso de interpretación de los constructos sitúa a las puntuaciones de las pruebas tanto en un contexto teórico de las relaciones implicadas con el constructo y otros constructos, como en un contexto de valores. Las pruebas empíricas del primer tipo de relaciones contribuyen a poner las bases evidenciales para la interpretación de las pruebas; las valoraciones del segundo tipo proporcionan las bases consecuenciales para la interpretación de las pruebas. Así pues, diríamos que la validación de las pruebas radica sobre cuatro bases:

	Interpretación del Test	Uso del test
Bases de Evidencia	Validez de Constructo	Validez de constructo + Relevancia/Utilidad
Bases de Consecuencias	Implicaciones de Valor	Consecuencias Sociales

(Martínez , 1996)

Estas cuatro bases son: Resumen inductivo de la evidencia convergente y discriminante de que las puntuaciones del test tienen una interpretación de constructo.

- 1) Evaluación de las aplicaciones de valor de la interpretación del test.
- 2) Evidencia de la relevancia del constructo y de su utilidad en aplicaciones particulares.
- 3) Estimación de las consecuencias sociales del uso propuesto del test y de las reales cuando se usa.

## **2.4. TIPOS DE VALIDEZ**

Hay dos formas de establecer la validez de un examen psicológico. En primer lugar, el proceso de validación tomando en consideración la totalidad del instrumento, en segundo lugar, procesos, tomando en consideración solamente las partes del examen o los ítems individuales que el examen contiene; esto se conoce como el Proceso de Análisis de los Ítems.

Como indica Herrans (2000), cuando se toma en consideración la totalidad del examen psicológico para determinar su validez, se encuentran dos tipos de validez, la subjetiva y la objetiva. La validez objetiva se divide a su vez en cuatro subclases, a saber: validez de contenido, validez concurrente, validez predictiva y validez de construcción lógica. La validez subjetiva o aparente de un examen psicológico es el grado hasta el cual el examen parece medir lo que dice que esta midiendo. Este tipo de validez se refiere únicamente a que los usuarios del instrumento y los expertos en la disciplina estimen por inspección, que el contenido del examen se relaciona con el rasgo o el tipo de conducta que el examen dice que está midiendo. Es necesario que los exámenes psicológicos tengan validez aparente pero esto no basta; en la medición psicológica se exige que se establezca además la validez objetiva para estimar así la utilidad de un instrumento psicológico.

La validez aparente se determina por la inspección del contenido de un examen. El examinando o el examinador lo inspecciona para determinar si le sirve para el

propósito que tiene en mente, o si en efecto mide aquello que a través de su nombre o título, proclama medir.

En síntesis, la validez aparente de un examen determina su aceptación inicial entre los usuarios de pruebas psicológicas.

### **2.4.1. VALIDEZ DE CONTENIDO**

Se refiere a sí el contenido de la prueba produce un rango de respuestas que presenta un área o universo completos de habilidad, comprensiones y otros comportamientos que se supone la prueba debe medir. La validez de contenido se refiere primordialmente a que los ítems que contenga la prueba concuerden con los objetivos de la misma.

Herrans (2000) indica que existen los siguientes principios que rigen la determinación de validez de contenido:

1. La prueba debe medir una sola función.
2. La sustancia de la prueba debe ser representativa de la sustancia y objetivo de la instrucción.
3. La prueba debe contener suficientes preguntas de manera que se obtenga una muestra representativa de la sustancia y los objetivos de la instrucción.
4. Las preguntas que contenga la prueba deben estar bien formuladas. Esto es, que el constructo de la prueba debe asegurarse de que las preguntas que incluye no tengan indicadores que faciliten la adivinanza al asar de las respuestas correctas.

Ese universo de comportamientos se debe especificar con detenimiento o sea el examen sistemático de la prueba para determinar su cubre una muestra representativa del área de conducta que debe medirse. Así, si la prueba está bien diseñada, las respuestas que dan los sujetos a los reactivos serán representativas de

las que serían sus respuestas al universo de situaciones de las que la prueba es una muestra. Puede parecer que basta con la simple inspección del contenido de la prueba para establecer su validez para dicho propósito; pero la solución no es tan sencilla. Nos indica Anastasi (1998) que un problema es el muestreo adecuado del universo de reactivos. El área de conducta por examinar debe analizarse sistemáticamente para garantizar que los reactivos cubran todos los aspectos importantes y en la proporción correcta. Es fácil cargar en exceso las pruebas con los aspectos del campo que más se presentan para la preparación de reactivos objetivos. Por ello, resulta conveniente describir de antemano toda el área considerada en lugar de hacerlo después de que la prueba esta lista. En consecuencia, hay que definir el contenido de manera amplia para que, además del conocimiento real incluya objetivos importantes como la aplicación de principios y la interpretación de datos. Más aun, la validez depende más de la relevancia que las respuestas del individuo tenga para el área conductual considerada que la importancia aparente del contenido del reactivo. La simple inspección de la prueba no siempre basta para revelar los procesos empleados por los examinados por los empleados. En relación con el área muestreada por la prueba, es importante prevenir cualquier tendencia ala sobregeneralización.

#### **2.4.2. VALILDEZ CONCURRENTENTE Y VALIDEZ PREDICTIVA (EN RELACION CON UN CRITERIO)**

La validez concurrente y la predictiva de las pruebas psicológicas son formas de validez relacionadas con un criterio, en donde entendemos por criterio como una medida externa, ajena a la prueba que se propone validar y que existe antes de la construcción de dicha prueba. La medida externa que se usará como criterio tiene que ser una que los expertos en la disciplina consideren satisfactoria como medida del rasgo que se quiere medir con la nueva prueba. La validez concurrente de una prueba psicológica se refiere al grado hasta el cual la prueba mide lo mismo que miden otras

pruebas psicológicas ya existentes, consideradas como medidas adecuadas del mismo rasgo. También se tiene validez concurrente de una prueba cuando esta correlacionada con algún criterio ya existente que mida lo mismo que la prueba. La validez predictiva es el grado hasta el cual la ejecución de una prueba sirve para predecir eficientemente la conducta futura de las personas que se han examinado.

Los procedimientos para establecer ambos tipos de validez son semejantes en cuanto a los procesos estadísticos que se usan para calcularlos. La diferencia básica entre ambas formas de validez radica en el tiempo que transcurre entre la administración de la nueva prueba que se pretende validar y la aplicación del criterio que se utiliza para validarlo. Si el lapso entre la administración de la nueva prueba por validarse y la aplicación del criterio es corto, se está bregando con validez concurrente. Si el lapso entre la administración de la nueva prueba por validarse y la aplicación del criterio que se utiliza para la validación, es largo, entonces se está bregando con validez predictiva. Ambos tipos de validez se miden computando un coeficiente de correlación entre los resultados de ambas administraciones; estos se conocen como coeficientes de validez. Además del tiempo entre la administración de la prueba y la aplicación del criterio, otra diferencia esencial entre la validez concurrente y la predictiva radica en el propósito para el cual se efectúa el proceso de validación. En la validez concurrente el propósito es constatar que la nueva prueba mide lo mismo que miden las pruebas anteriores (criterio); en la validez predictiva el propósito es la predicción del comportamiento futuro de los sujetos que son examinados con la nueva prueba. Por tanto, los criterios por usarse en el proceso de validez predictiva son mucho más variados; incluyen pero no se limitan, a la ejecución de los sujetos en otra prueba semejante a la que se está validando.

Por ejemplo, si el coeficiente de correlación que se obtiene (coeficiente de validez concurrente) es positivo y alto, se dice que en efecto, ambas pruebas están midiendo lo mismo. Si el coeficiente de correlación que se obtiene entre pruebas de aptitud y el índice académico es alto y positivo, se dice que la nueva prueba de aptitud predice eficientemente la ejecución futura.

La validez concurrente de una prueba psicológica no indica nada acerca de la capacidad predictiva de ese instrumento. Esta solo señala que el nuevo instrumento que se ha desarrollado para medir un rasgo, es tan eficiente o es tan válido para medir ese rasgo como lo era el criterio que existía previamente.

Aiken (1996) en su libro "test psicológicos y evaluación" indica que existen factores que afectan la validez con relación a los criterios, estos incluyen diferencias de grupo, extensión de la prueba, contaminación de criterios y calificación de base. El incremento de validez de una prueba también debe tomarse en cuenta al decidir si se va a utilizar una prueba para propósitos de selección y ubicación.

Diferencias de grupo: las características del grupo de personas para quienes es válida una prueba incluyen variables como sexo, edad y rasgos de personalidad. Estos factores, que denominaremos variables moderadoras, pueden afectar la correlación entre una prueba y una medida de criterio. Evidentemente el coeficiente de validez se vera influenciados por el grado de heterogeneidad del grupo de validación en lo que mide la prueba. Los coeficientes de validez tienden a ser más reducidos en los grupos más homogéneos; es decir, en los grupos que contienen un rango más limitado de calificaciones de prueba. Ya que el tamaño de un coeficiente de correlación es una función de dos variables, la reducción del rango de calificaciones ya sea en la variable de predicción o de criterio tenderán a disminuir el coeficiente de validez predictiva.

Debido a que la magnitud de un coeficiente de validez varía con la naturaleza del grupo sometido a prueba, una prueba recién creada que se descubre como indicador válido de una variable de criterio particular en un grupo de sujetos debe tener validez cruzada. Está comprende la aplicación de la prueba a una segunda muestra de personas para determinar si conserva su validez en distintas muestras.



Puede comprender también una prueba diferente, sea el caso por lo regular existe cierta reducción en el coeficiente de validez en la validación cruzada.

**Extensión de la prueba:** al igual que la confiabilidad, la validez varía en forma directa con la extensión de la prueba y la heterogeneidad del grupo de personas que la presenta. Hasta cierto punto, las calificaciones en una prueba más extensa y aquellas en una prueba que se aplica a un grupo de individuos que varían en gran medida en las características a medir tienen varianzas mayores. Por otra parte, las calificaciones obtenidas en pruebas cortas o breves o en aquellas que se aplican a grupos de personas más homogéneas tienen varianzas más bajas.

**Contaminación de criterios:** la validez de una prueba esta limitada no sólo por la confiabilidad y el criterio, sino también por la validez del criterio mismo como una medida de la variable de interés particular. En ocasiones, el criterio de hace menos válido o se contamina, por el método particular para determinar las calificaciones del criterio. Esa contaminación del criterio puede controlarse por medio del análisis a ciegas; es decir, sin comunicar, a quien realiza el diagnostico, ninguna información sobre los sujetos aparte de sus calificaciones en la prueba.

**Índice de base:** otro factor que puede afectar la magnitud de un coeficiente de validez con relación a los criterios es el índice de base de la característica o comportamiento de interés en la población a la que se dirige la prueba. El índice de base se refiere a la proporción de personas en la población que muestran la característica o comportamiento de interés. Una prueba que esta diseñada para predecir un tipo particular de comportamiento es más efectiva cuando el índice de base es de 50% y menos efectiva cuando el índice de base es muy alto o muy bajo.

## **2.5. NORMAS Y SIGNIFICADO DE LAS PUNTUACIONES DE LOS TESTS**

Para poder entender las pruebas psicológicas y para su adecuada interpretación de sus resultados se requiere que la misma cuente con normas, confiabilidad, validez, análisis de los reactivos y el conocer el diseño del test.

En ausencia de datos interpretativos adicionales, la sola puntuación cruda de cualquier prueba psicológica carece de significado. Tampoco los porcentajes brindan una solución satisfactoria al problema de interpretar las puntuaciones de los instrumentos; obviamente, el grado de dificultad de los reactivos que componen cada prueba determinará el significado del resultado. Al igual que las puntuaciones crudas, los porcentajes solo pueden interpretarse en un marco de referencia claramente definido y uniforme. En los instrumentos psicológicos, las puntuaciones suelen interpretarse haciendo referencia a **normas** que representan el desempeño de la muestra de estandarización en la prueba; es decir, las normas se establecen empíricamente al determinar lo que hacen en la prueba los miembros de un grupo representativo. La puntuación transformada de cualquier individuo se refiere entonces a la distribución de las puntuaciones obtenidas por la muestra de estandarización para descubrir qué lugar ocupa en esa distribución. ¿Coincide el resultado con el desempeño promedio del grupo de estandarización?, ¿Está ligeramente por debajo del promedio o se encuentra cerca del extremo superior de la distribución?. Para averiguar la posición exacta del individuo en relación con la muestra de estandarización, la puntuación cruda se convierte en una medida relativa. Estas puntuaciones transformadas cumplen un doble propósito; primero, indican la posición del individuo en relación con la muestra normativa, lo que permite la evaluación de su desempeño en comparación directa del desempeño del individuo en pruebas diferentes. Como las puntuaciones crudas de diferentes pruebas suelen expresarse en unidades distintas resulta

imposible su comparación directa. El nivel de dificultad de cada prueba también afecta esta comparación. Por otro lado, las puntuaciones derivadas pueden expresarse en las mismas unidades y referirlas a muestras normativas iguales o muy similares en distintas pruebas lo que permite comparar la ejecución relativa en funciones muy diferentes. Aunque hay varias formas de convertir las puntuaciones crudas para satisfacer los objetivos planteados, las puntuaciones derivadas se expresan esencialmente en dos formas principales: (1) el nivel de desarrollo alcanzado, o (2) la posición relativa dentro de un grupo específico (Anastasi, 1998)

## **2.6. CONCEPTOS ESTADÍSTICOS**

Nos indica Anastasi (1998): "Un propósito fundamental de la estadística es organizar y resumir los datos cuantitativos para facilitar su comprensión".

Anastasi en su libro "Test Psicológicos" nos comenta que basados en esta razón el primer paso para poner orden en semejante caos de datos es tabular las puntuaciones en una distribución de frecuencias, en donde se agrupan las puntuaciones en intervalos de clase, y se marca cada una en el intervalo adecuado, cuando se han anotado todas las puntuaciones se cuentan las marcas para encontrar la frecuencia o el número de casos en cada intervalo de clase. La suma de esas frecuencias será igual a  $N$ , el número total de casos en el grupo.

La información que proporciona una distribución de frecuencia también puede representarse en forma gráfica en una curva de distribución, en donde en la línea base el eje horizontal representa las puntuaciones agrupadas en intervalos de clase; en el eje vertical se localizan las frecuencias o números de casos que caen en cada intervalo. Estas gráficas pueden ser representadas en histogramas o polígonos de frecuencia. En la mayoría de las representaciones gráficas de estas puntuaciones surge una figura conocida como curva normal, esta curva tiene importantes propiedades matemáticas y brinda la base para muchos análisis

estadísticos, sin embargo, para nuestros propósitos actuales sólo mencionaremos algunos de esos rasgos. En esencia, la curva indica que el mayor número de casos se agrupa en el centro de la distribución, y que al acercarse a los extremos va disminuyendo gradualmente en ambas direcciones. La curva es simétrica y tiene un pico en el centro. Casi todas las distribuciones de rasgos humanos, estatura, aptitudes y características de personalidad aproximan a la curva normal, y en general, entre más grande sea la muestra o grupo más parecida resultara la distribución a la curva teórica normal.

Un grupo de puntuaciones también puede describirse en términos de alguna medida de tendencia central, que ofrece la puntuación más común o representativa del desempeño de todo el grupo. La más conocida de esas medidas es el promedio, que técnicamente se llama media ( $M$ ) y se obtiene al sumar todas las puntuaciones y dividir la suma entre el número de casos ( $N$ ). Otra medida es la moda o puntuaciones más frecuentes. En una distribución de frecuencias, la moda es punto central del intervalo de clase con la mayor frecuencia. Hay que observar que esta puntuación corresponde al punto más alto en la curva de distribución. Otra medida de tendencia central es la mediana, el punto más cercano al centro de una distribución de puntuaciones ordenadas por tamaño. La mediana es el punto que divide la distribución por encima y por debajo del cual la mitad de los casos.

Las medidas de variabilidad, de la extensión de las diferencias individuales alrededor de la tendencia central, proporcionan otra descripción del conjunto de puntuaciones de un test. La forma más evidente y familiar de mostrar la variabilidad es el rango entre el mayor y la menor puntuación; sin embargo, se trata de una medida muy cruda e inestable, toda vez que la determinan apenas dos puntuaciones, lo cual supone que una sola puntuación inusualmente elevada o baja puede afectar mucho su tamaño. Un método más preciso para medir la

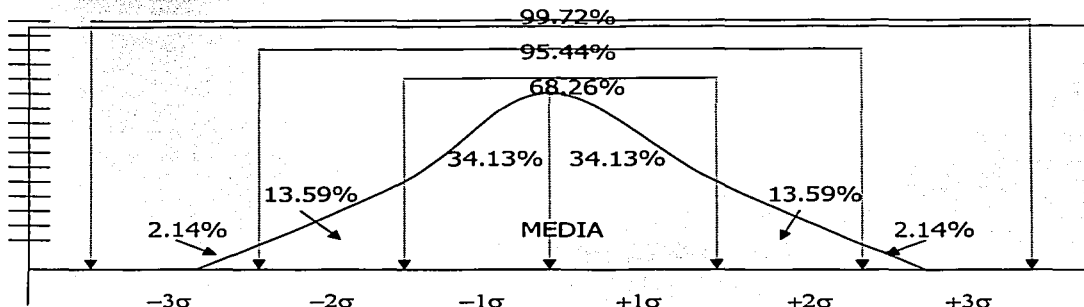
variabilidad se basa en la diferencia entre la puntuación de cada individuo y la media del grupo.

De manera convencional, designamos las puntuaciones crudas originales con una X mayúscula, la x minúscula se emplea para referirse a las desviaciones de cada puntuación respecto a la media del grupo y la letra griega sigma significa "la suma de".

La desviación estándar (simbolizada por DE) es una medida de variabilidad mucho más útil en la que los signos negativos se eliminan al elevar al cuadrado cada desviación, se conoce como varianza o media de los cuadros de las desviaciones a la suma de esta columna dividida entre el número de casos. La varianza ha demostrado ser de gran utilidad para separar la contribución de diferentes factores a las diferencias individuales en el desempeño de una prueba; sin embargo, para los propósitos presentes lo que más nos interesa es la DE que es la raíz cuadrada de la varianza. Esta medida suele emplearse al compara la variabilidad de distintos grupos. La distribución con la mayor variabilidad individual produce una DE más grande que la que tiene menores diferencias individuales. Al evaluar la ejecución relativa de dos grupos, deben compararse tanto las DE CÓMO las medias. Si se encuentra que la variabilidad de los grupos difiere, puede suponerse que son disímiles en la proporción de puntuaciones altas, bajas o en ambas, independientemente de las diferencias entre las medias. En la actualidad se dispone de procedimientos que permiten combinar los efectos de las diferencias entre las medias y las diferencias en la DE. La DE también proporciona la base para expresar en términos de normas las puntuaciones del individuo en diferentes pruebas. La interpretación de la DE resulta especialmente clara al aplicarse a una curva de distribución normal o aproximadamente normal, en la que hay una relación exacta entre la DE y la proporción de casos.

Ejemplo: en la línea horizontal de la curva se ha marcado las distancias que representan una, dos y tres desviaciones normales por encima y por debajo de la

media. En la curva normal el porcentaje de casos que cae entre la media y  $+1\sigma$  es 34.13 %. Como la curva es simétrica también se encuentra 34.13% de los casos entre la media y  $-1\sigma$ , de modo que entre  $+1\sigma$  y  $-1\sigma$  ambos lados de la media se encuentra el 68.16 % de los casos. Casi todos los casos (99.72%) caen dentro de  $\pm 3\sigma$  de la media. Estas relaciones son particularmente importantes en la interpretación de las puntuaciones estándares y los percentiles (Anastasi, 1998).



## **2.7. NORMAS**

Recordemos que un examen psicológico es un instrumento de medición para el cual se debe establecer normas. Las normas de un examen psicológico se refieren a la ejecución típica realizada en ese examen por la muestra de estandarización. Específicamente el promedio aritmético y la desviación estándar derivados de la muestra de estandarización son las normas del examen que permiten efectuar la conversión de las puntuaciones brutas de la prueba a una escala de puntuaciones derivadas ( z, percentilas, T, etc.). Esta distribución de puntuaciones brutas obtenidas por los sujetos de la muestra de estandarización y convertidas a puntuaciones derivadas es lo que generalmente se conoce como las normas del examen (Herrans, 2000).

TESIS CON  
 FALLA DE ORIGEN

La muestra de estandarización es una porción representativa de los miembros de la población específica para la cual se desarrolla un instrumento de medición. Una muestra representativa de una población contiene todos los factores y características de esta población que son de interés principal para el constructor del examen. Los estadísticos dicen que la posibilidad de lograr representatividad de la muestra de estandarización se maximiza mediante la selección al azar de los sujetos que forman esa muestra de hecho de acuerdo con Guilford, las estadísticas de muestreo (incluyendo el concepto de error estándar) se basan en la premisa de que las muestras han sido seleccionadas al azar (Guilford, 1965). La selección al azar presupone que cada miembro de la población ha tenido igual oportunidad de ser seleccionado para inclusión en la muestra. También presupone que la selección de cada miembro de la muestra se hizo independientemente de la selección de cualquier otro miembro incluido. Si esas dos condiciones no se dan, la muestra está sesgada o contaminada y por tanto no se podrá utilizar para establecer normas para un instrumento de medición psicológica.

Las normas del examen permiten establecer la posición relativa de un sujeto en comparación con la ejecución típica de la muestra de estandarización. Es importante recordar que esta es la única función de las normas. Usualmente las normas de los instrumentos psicológicos proporcionan una escala de medición de intervalos. Nunca señalan de un rasgo posee un individuo; por ejemplo, nunca indican en términos absolutos cuánta inteligencia tiene un individuo. Al comparar con la ejecución de un individuo con la norma del examen, lo único que se sabe cuán distante del promedio aritmético de esa muestra está situada esa persona (Herrans, 2000).

### **2.7.1. PROCESO DE NORMALIZACIÓN DE UN TEST PSICOLÓGICO**

La normalización de un test psicológico es un proceso riguroso que incluye varias fases. La primera fase, la estandarización del examen consiste en la construcción de la forma final del test y el desarrollo de un manual de procedimientos uniformes para administrar y valorar los resultados de este test. La segunda fase es la selección de una muestra representativa de la población para la cual se construye el test. La tercera fase es la administración del test a la muestra de estandarización (la muestra representativa) siguiendo los procedimientos uniformes estipulados en el manual de instrucciones del test. La cuarta fase es el establecimiento de normas: la derivación del promedio aritmético y la desviación estándar de los puntajes obtenidos por la muestra de estandarización que permita luego la conversión de puntajes brutos a puntajes estándar. La quinta y última fase es el proceso de conversión de los puntajes brutos (total de las respuestas correctas obtenidas por cada sujeto en el test) a puntajes a escala.

Un puntaje bruto en un test no tiene significado a menos que se le compare con una vara o con un criterio específico. Una vez que se conocen esos datos, se puede derivar algún significado del puntaje bruto que se obtuvo en el test. Este proceso de darle significado a los puntajes brutos es parte del proceso de conversión de puntuaciones brutas a puntuaciones a escala. Las puntuaciones a escala se conocen también como puntuaciones típicas y puntuaciones derivadas. No se debe menosvaluar la importancia de cada una de las fases del proceso de normalización de los tests psicológicos. Cada una de esas fases tiene un propósito que luego determina la interpretación que se hace de los resultados que se obtienen al administrar la prueba una vez que esta ha sido tipificada.



Se puede hacer la analogía entre el uso de test estandarizados y el control que se debe tener al realizar un experimento en el laboratorio. La primera fase (procedimientos uniformes para administrar y valorar los resultados) equivale al control experimental. Si en un experimento no se controlan las variables extrañas que pueden influir en los resultados del experimento esos resultados no se pueden atribuir al efecto que pueda tener la variable independiente en la variable dependiente del experimento. De igual modo, si en la administración de una prueba ya tipificada si por alguna razón el examinador altera en alguna forma los procedimientos uniformes, los resultados que se derivan no se pueden comparar con las normas de esa prueba. Eso equivaldría atribuir los resultados del test a cualquier otro factor que no sea la variable que se esta midiendo en ese momento con ese test. La alteración de los procedimientos uniformes en la administración, valoración o interpretación de los resultados de un test que ha sido estandarizado equivalen a invalidar los resultados que se obtienen. La violación de esos procedimientos uniformes afecta directamente a la confiabilidad de los resultados obtenidos(Herrans, 2000).

### **2.7.2. LA RELACIÓN DE LAS NORMAS DE UN TEST TIPIFICADO CON LA POBLACIÓN PARA LA CUAL SE ELABORA ESE TEST.**

Herrans (2000) en su libro "Psicología y medición" nos indica que es de suma importancia señalar que las normas de un test psicológico son aplicables única y exclusivamente a los miembros de la población específica de la cual se sustrajo la muestra de estandarización. No se debe usar normas de un test que han sido desarrolladas en una población específica con personas que pertenecen a poblaciones diferentes. Las normas de un test son la ejecución típica de la muestra de estandarización, que es un grupo representativo de una población dada. En esa

población prevalecen condiciones particulares que pueden ser distintas a las condiciones particulares que prevalezcan en otra población.

La práctica de aplicar las normas de una población a otra población distinta es frecuente en países latinoamericanos o del caribe. A menudo, las normas de los tests psicológicos desarrollados para la población de los estados unidos se usan con las poblaciones de estos países que socioculturalmente son diferentes. Dichas normas no se deben aplicar. También en Estados Unidos, muchos psicólogos clínicos y escolares, así como psicólogos de consejería, utilizan pruebas preparadas para la población norteamericana blanca y de clase media y con esas normas ajenas miden sectores de su población (hispanohablantes, negros, indios, asiáticos) que no estuvieron representados en las muestras de estandarización de esos tests. La crítica y el ataque continuo en los Estados Unidos al uso de las pruebas psicológicas que están dirigidas en gran medida a esa practica equivocada de aplicar normas preparadas para poblaciones o subgrupos específicos de la nación norteamericana, a los llamados grupos minoritarios. El negro, el puertorriqueño, el mexicano-americano, el asiático y el indio al ser medidos con una norma del blanco norteamericano de clase media, podrían puntuar por debajo o por arriba de la norma. Si la situación se invirtiera y se midiera el blanco norteamericano de clase media con la norma de un test desarrollado para hispanohablantes, para negros, para asiáticos o para indios norteamericanos, sería de esperar, que el blanco norteamericano puntuara por debajo de la norma.

La relación estrecha que existe entre las normas de un test psicológico y la población para la cual se desarrolla este test es de vital importancia y afecta tanto al aspecto de confiabilidad de ese instrumento como al aspecto de su validez.

## **2.8. TIPOS DE NORMAS**

### **2.8.1. NORMAS DE DESARROLLO**

Las puntuaciones de las pruebas adquieren significado si indican qué tanto ha progresado el individuo en el patrón de desarrollo normal. Las puntuaciones basadas en normas de desarrollo tienden a ser burdas en términos de la psicometría, y no se presentan a un tratamiento estadístico preciso; no obstante, resultan muy atractivas para propósitos descriptivos, en especial en el estudio clínico intensivo de individuos y para ciertos propósitos de investigación.

### **2.8.2. NORMAS INTRAGRUPU**

Casi todos los tests estandarizados proporcionan alguna forma de normas intragrupo, con las que la ejecución del individuo se evalúa en términos del desempeño del grupo de estandarización más próximo. Las puntuaciones intragrupo tienen un significado cuantitativo uniforme y definido y pueden emplearse adecuadamente en casi todos los análisis estadísticos( Anastasi, 1998).

Percentiles. Las calificaciones percentiles se expresan en términos del porcentaje de sujetos de la muestra de estandarización que caen bajo determinada puntuación cruda. El percentil indica la posición relativa del individuo en la muestra de estandarización. Los percentiles también pueden verse como rasgos en un grupo de 100, excepto porque en éstos se acostumbra empezar a contar por la parte superior, y el mejor individuo del grupo recibe un rango de 1, por su parte, con los percentiles se comienza por el extremo inferior, de modo que entre más bajo sea el percentil más pobre es la posición del individuo.

El percentil 50 ( $P_{50}$ ) corresponde a la mediana, ya estudiada como medida de tendencia central. Los percentiles por encima de 50 representan una ejecución por

encima de la mediana; los que están por debajo de 50 significan una ejecución inferior. Los percentiles 25 y 75 se conocen como el primer y el tercer puntos cuartiles ( $Q_1$  y  $Q_3$ ), por que separan los cuartos inferior y superior de la distribución. Como la mediana, proporcionan referencias convenientes para describir una distribución de puntuaciones y compararla con otras distribuciones. Los percentiles no deben confundirse con los porcentajes, que son puntuaciones crudas que se expresan en términos del porcentaje de los reactivos correctos y el total; los percentiles son puntuaciones transformadas que se expresan en términos del porcentaje del individuo. Una puntuación cruda inferior a cualquiera de las calificaciones obtenidas en la muestra de estandarización tiene un rango percentil de cero ( $P_0$ ); una superior a cualquier puntuación de la muestra de estandarización tiene un rango percentil de 100 ( $P_{100}$ ); sin embargo, los percentiles no implican una puntuación cruda de cero ni una puntuación cruda perfecta.

Los rangos percentiles tienen varias ventajas: son sencillos de calcular y de comprender, incluso por personas sin capacitación técnica. Más aún, son de aplicación universal, pueden usarse lo mismo con adultos como con niños y son adecuados para cualquier prueba, sea que mida variables de aptitud o de personalidad.

Su principal desventaja está en la marcada disparidad de sus unidades, en especial en los extremos de la distribución. Si la distribución de las puntuaciones crudas se aproximan a la curva normal (como sucede con los resultados de la mayor parte de las pruebas), las diferencias en las puntuaciones crudas cercanas del centro de la distribución (o mediana) aparecen exageradas al transformarse a percentiles; por el contrario, la diferencias en las puntuaciones crudas próximas a los extremos quedan considerablemente acortadas. Como recordara en la curva normal los casos se agrupan cerca del centro y su dispersión aumenta al aproximarse a los extremos; en consecuencia, cualquier porcentaje de casos cercanos al centro

cubre una distancia más corta en la línea horizontal que el mismo porcentaje cerca de los extremos de la distribución (Anastasi, 1998).

Resulta evidente que los percentiles muestran la posición relativa de cada individuo en la muestra normativa pero no la magnitud de la diferencia entre puntuaciones; sin embargo, si se grafican en una carta de probabilidad aritmética, las puntuaciones percentiles también brindan una imagen correcta de las diferencias entre las puntuaciones. La carta de probabilidad aritmética es una gráfica con secciones transversales en la que las líneas verticales muestran el mismo el mismo espaciamiento que los puntos percentilares de una distribución normal, mientras que el espaciamiento entre líneas horizontales es uniforme, o a la inversa. Las gráficas de percentiles normales pueden usarse para mostrar las puntuaciones de diferentes individuos en la misma prueba o los resultados de la misma persona en diferentes pruebas. En cualquier caso, la diferencia real entre puntuaciones aparecerá representada correctamente. Muchas baterías de aptitudes y de aprovechamiento utilizan esta técnica en sus perfiles de calificaciones, que muestran el desempeño del individuo en cada prueba (Anastasi, 1998).

Calificaciones estándares. Los instrumentos actuales hacen un uso creciente de las calificaciones estándares, que desde cualquier punto de vista constituyen el tipo más satisfactorio de puntuación derivada. Estas calificaciones expresan la distancia del individuo de la media en términos de la desviación normal de la distribución normal. Las calificaciones estándares pueden obtenerse por medio de transformaciones lineales o no lineales de las puntuaciones crudas originales. Las calificaciones estándares obtenidas por medio de una transformación lineal conservan las relaciones numéricas exactas de la puntuación originales gracias a que para calcularlas se resta una constante de cada puntuación cruda y se divide el resultado entre otra constante. La magnitud relativa de las diferencias entre las calificaciones estándares así derivadas corresponde exactamente a la que existe entre las puntuaciones crudas. Todas las propiedades de la distribución original de

las puntuaciones crudas se duplican en la distribución de calificaciones estándares, lo cual implica que en las dos clases de puntuaciones es posible realizar el mismo cálculo sin distorsionar de los resultados.

Las puntuaciones estándares derivadas linealmente suelen conocerse simplemente como "calificaciones estándares" o como "calificaciones z" para calcular una calificación z, se encuentra la diferencia entre la calificación individual cruda y la media del grupo normativo, y luego se divide la diferencia entre la DE del grupo normativo. Cualquier puntuación no procesada que sea exactamente igual a la medida es equivalente a una puntuación z de cero. Es evidente que dicho procedimiento producirá puntuaciones derivadas con signo negativo para todos los que caigan por debajo de la media. Además, como el rango total de la mayor parte de los grupos no se extiende más allá de tres DE por encima y por debajo de la media, las puntuaciones estándares deben expresar con al menos un decimal para permitir la suficiente diferenciación (Anastasi, 1998).

Calificaciones estándares normalizadas. Estas son calificaciones estándares expresadas en términos de una distribución que ha sido transformada para ajustarse a una curva normal. Dichas calificaciones pueden calcularse recurriendo a tablas que muestran el porcentaje de casos que cae a diferentes distancias DE de la media de una curva normal, para ello, primero se encuentra el porcentaje de individuos de la muestra de estandarización que cae en o por encima de cada puntuación cruda, este porcentaje se localiza luego en la tabla de frecuencia de la curva normal y se obtiene la correspondiente calificación estándar normalizada. Las calificaciones estándares normalizadas se expresan de la misma forma que las calificaciones estándares derivadas linealmente, es decir, con una medida igual a cero y una DE igual a 1. Así, una calificación normalizada de cero indica que el individuo cae en la medida de la curva normal, que supera al 50 % del grupo. Una calificación de  $-1$  significa que sobrepasa aproximadamente a 16 % del grupo y una calificación de  $+1$  que aventaja al 84%.

Igual que las calificaciones estándares derivadas linealmente, las calificaciones estándares normalizadas pueden adoptar cualquier forma conveniente. Si la calificación estándar normalizada se multiplica por 10 y se suma o resta de 50 se convierte en una calificación T, una puntuación que propuso primero W.A. McCall (1922) y en la que una puntuación de 50 corresponde a la media, una de 60 a una DE por encima de la media, etc. Otra transformación conocida es la de la escala estanina, elaborada por la aviación estadounidense durante la Segunda Guerra Mundial, que proporciona un sistema de calificaciones de un solo dígito con una media de 5 y una DE aproximadamente igual a 2. La restricción de las calificaciones a números de un solo dígito tiene ciertas ventajas para el cálculo, ya que cada calificación requiere una sola columna en la computadora (Anastasi, 1998).

Anastasi señala que aunque las calificaciones estándares normalizadas son el tipo de puntuación más adecuado para la mayor parte de los propósitos, hay ciertas objeciones técnicas a la normalización rutinaria de todas las distribuciones. Dicha transformación sólo debería llevarse a cabo cuando la muestra es grande y representativa y cuando existen razones para creer que la desviación de la normalidad resulta de defectos en la prueba más que de las características de la muestra y otros factores que afecten la conducta considerada. También debe observarse que cuando la distribución original de puntuaciones crudas se aproxima a la normalidad, las calificaciones estándares derivadas linealmente y las calificaciones estándares normalizadas resultan muy similares a pesar de las diferencias entre los métodos empleados para derivarlas. Como es obvio, el proceso de normalizar una distribución que de hecho es prácticamente normal producirá poco o ningún cambio. Siempre que sea factible suele ser más deseable obtener una distribución normal de la prueba más que por la normalización subsecuente de una distribución marcadamente no normal. Con una distribución más o menos normal de puntuaciones crudas, las calificaciones estándares

linealmente derivadas cumplirán el mismo propósito que las calificaciones estándares normalizadas.

## **2.9. RELATIVIDAD DE LAS NORMAS**

Comparaciones entre reactivos. El CI, o cualquier otra calificación, debe acompañarla siempre el nombre del test del que se obtuvo. Los resultados de las pruebas no pueden interpretarse en abstracto, tiene que ser referidos a pruebas particulares. Del mismo modo, la posición relativa de un individuo en diferentes funciones puede ser muy mal representada por la falta de comparabilidad de las normas de una prueba. Supongamos que a una estudiante se le aplica una prueba de comprensión verbal y un test de aptitud especial para determinar su ubicación relativa en ambos campos. Si la prueba de habilidad verbal se estandarizó con una muestra aleatoria de estudiantes de secundaria mientras que la prueba especial se estandarizó en un grupo selecto de estudiantes a curso para elección de carrera, el examinador podría llegar a la conclusión errónea de que la capacidad verbal de la chica es mucho mayor que la especial, cuando podía ser el caso contrario (Anastasi, 1998).

Hay tres razones principales para explicar las variaciones sistemáticas entre las calificaciones obtenidas por los mismos individuos en pruebas distintas. Primero, a pesar de la similitud en el nombre, el contenido de las pruebas puede diferir, como sucede con muchos de los llamados test de inteligencia. Aunque lleven el mismo nombre, una de las pruebas puede abarcar únicamente contenido verbal, otra ante todo aptitudes espaciales y la tercera puede cubrir contenido verbal, numérico y espacial en proporciones iguales. Segundo, las unidades de la escala pueden no ser comparables. Como ya explicamos, si en una prueba el CI tiene una DE de 12 y en otra de 18, entonces es más probable que un individuo que en la primera prueba obtuvo un CI de 112 en la segunda reciba un C.I. de 118. Tercero, puede haber diferencias en la composición de las muestras de estandarización



empleadas para establecer las normas de las diferentes pruebas. Es obvio que aparecerá que los mismos individuos tuvieron un mejor rendimiento al compararlos con un grupo menos capaz, que cuando se hace con uno de mayor capacidad. Las faltas de compatibilidad del contenido de la prueba o a su manual. Sin embargo, es más probable que pasen por alto las diferencias en las respectivas muestras normativas, diferencias que pueden ser la explicación de muchas otras discrepancias en los resultados que no pueden entenderse de otra manera (Anastasi, 1998).

La muestra normativa. Independientemente de cómo se exprese, cualquier norma se restringe a la población normativa particular de la que se tomó la muestra. El usuario no debe perder de vista la forma en que se establecen las normas, pues no son, de manera alguna, absoluta, universal o permanentes, sino que simplemente representan la ejecución en la prueba de las personas que formaron parte de la muestra de estandarización. Al elegir dicha muestra, suelen hacerse esfuerzos por obtener una que sea representativa de la población a la que va dirigida la prueba.

En la terminología estadística se distingue entre muestra y población. La primera se refiere al grupo de individuos realmente examinado, mientras que la población designa al grupo mayor, pero de similar constitución, de donde se extrajo la muestra. Hay que verificar la muestra en lo que atañe a la distribución geográfica, el nivel socioeconómico, la composición étnica y otras características pertinentes para asegurar que representa la verdad a la población definida.

En el desarrollo y aplicación de las normas debe dedicarse mucha atención a la muestra de estandarización. Es evidente que la muestra en la que se basan las normas debe ser lo suficientemente grande como para proporcionar valores estables, lo que quiere decir que no se deben obtener normas muy diferentes al aplicar la prueba a cualquier otra muestra cuidadosamente elegida de la misma

población. Las normas con un error de muestreo grande tendrán desde luego poco valor en la interpretación de los resultados de la prueba. De igual importancia es el requisito de que la muestra sea representativa de la población considerada, por lo que es necesario investigar con cuidado factores selectivos sutiles que pudieran haber disminuidos la representatividad de la muestra. Algunos de estos factores se encuentran en las muestras institucionales que, debido a su tamaño relativamente grande y a su fácil acceso para propósito de examinación, resultan un campo tentador para la acumulación de datos normativos. De cualquier forma, es necesario analizar con cuidado las limitaciones especiales de las muestras.

Relacionada con la cuestión de la representatividad de la muestra, está la necesidad de definir la población específica a la que pueden generalizarse las normas. Como es evidente, una forma de asegurar que una muestra es representativa consiste en restringir la población para que se ajuste a las especificaciones de la muestra disponible. Por supuesto, lo ideal es que la población deseada se defina de antemano en función de los objetivos de la prueba, para formar luego una muestra adecuada. No obstante, los obstáculos prácticos para obtener participantes puede hacer que esta meta resulte inalcanzable. En dicho caso, es mucho mejor definir la población de manera más reducida que señalar normas sobre una población ideal que no está adecuadamente representada en la muestra de estandarización. En la practica real, son muy pocas las pruebas que se estandarizan sobre poblaciones tan amplias como suele suponerse, y ninguna prueba ofrece normas para la especie humana. Además, es dudoso que muchas pruebas brinden normas verdaderamente adecuadas para poblaciones tan amplias como "los europeos adultos", "los niños mexicanos de años", etc. En consecuencia, las muestras obtenidas por diferentes autores de pruebas pueden no ser representativas de las poblaciones alegadas y presentar diversos sesgos, de ahí que las normas resultantes no sean comparables. Al interpretar los resultados de las pruebas, el usuario debe tomar en consideración las influencias específicas que pueden haber actuado sobre la

muestra normativa utilizada al estandarizarla, y que incluyen factores selectivos especiales así como las condiciones sociales prevalecientes en el momento en que fueron obtenidos los datos normativos (Anastasi, 1998).

Normas específicas. Otro acercamiento a la no equivalencia de las normas existentes –que quizá sea el más realista para la mayor parte de los instrumentos– consiste en estandarizarlas sobre poblaciones definidas de manera más restringida y elegidas para adaptarse para los propósitos específicos de cada prueba. En estos casos, los límites de la población normativa debe informarse claramente con las normas. Así puede decirse que las normas se aplican a “los empleados administrativos de las grandes empresas” o a “los estudiantes de primer año de ingeniería”. Para muchos propósitos de información resulta deseable contar con normas muy específicas. Aunque pueda disponerse de normas representativas de una población más amplia, siempre que subgrupos inidentificables obtengan puntuaciones apreciablemente diferentes en una prueba particular resulta útil informar por separado las normas del subgrupo. Los subgrupos pueden formarse de acuerdo con la edad, el grado escolar, el programa de estudios, el sexo, la región geográfica, el ambiente urbano o rural, el nivel socioeconómico y muchas otras variables. El uso que valla hacerse de la prueba determina la diferenciación que resulte más pertinente, así como la conveniencia de contar con normas generales o específicas. Debemos mencionar también el caso de las normas locales, que con frecuencia son establecidas por los propios usuarios en ambientes particulares. Los grupos empleados para derivar dichas normas se definen de manera aun más limitada que los subgrupos considerados antes. Así un empleador puede acumular normas de los solicitantes para un determinado trabajo en una empresa particular o la oficina de admisión de una universidad puede evaluar el rendimiento de los alumnos en términos de su propia distribución de clasificaciones. Esas normas locales son más apropiadas que las normas nacionales para muchos propósitos de examinación como la predicción del desempeño en el empleo o académico, la comparación del aprovechamiento relativo de un niño en

diferentes materias o bien la medición del progreso de un individuo a lo largo del tiempo (Anastasi, 1998).

Grupo fijo de referencia. Aunque la forma en que se calcula la mayor parte de las calificaciones derivadas permite una interpretación normativa inmediata de la ejecución en una prueba, hay algunas excepciones notables. Una escala no normativa utiliza un grupo fijo de referencia para asegura la comparabilidad y continuidad de las calificaciones sin brindar una evaluación normativa del desempeño. Con dicha escala, la interpretación normativa tiene que referirse a normas obtenidas de manera independiente de una población adecuada. Con este propósito suelen emplearse normas locales u otras normas específicas. Las escalas formas a partir de un grupo fijo de referencia son análogas en un aspecto a las escalas empleadas en la medición física.

## **2.10. TESTS PSICOLÓGICOS Y SU ADAPTACIÓN A OTRAS CULTURAS**

La finalidad de la psicometría, es llevar cabo la medición de la conducta tanto en el ser humano como en los animales. Morales (1990), refiere que la medición "es el procedimiento mediante el cual asignamos números (calificaciones, medidas) a las propiedades, atributos o características de los objetos (o de unidades experimentales), estableciendo las reglas específicas sobre las que se fundamentan tales asignaciones". Por lo tanto, un instrumento de medición en psicología, se refiere a aquel procedimiento u operación que permita llegar a obtener, objetivamente, información acerca de la conducta que se está midiendo.

Pierre Pichot establece que una prueba psicológica "es una situación experimental y estandarizada que sirve de estímulo a un comportamiento y que, manipulada

estadísticamente, permite clasificar a los sujetos tipológica o cuantitativamente” (Morales, 1990).

Diversos autores, entre ellos Anastasi (1998), concuerdan en que para determinar si un instrumento psicológico constituye un adecuado diseño de medición, debe cubrir los siguientes criterios:

a) Estandarización: Permite determinar el grado de errores de interpretación que han sido controlados, con el fin de evitar caer en ellos.

b) Confiabilidad: Indica la libertad relativa de error variable; es decir, las puntuaciones obtenidas por la misma persona cuando se le aplica la misma prueba es equivalente.

c) Objetividad: Denota el grado en que los errores personales han sido evitados.

d) Validez: Indica el grado en que la prueba mide aquellos fenómenos para los cuales ha sido construida; por tanto, no está influida por el error constante. La validez indica el grado en que la prueba mide lo que pretende medir.

Las pruebas de inteligencia, se denominan pruebas de potencia o de ejecución máxima debido a que se exige el máximo rendimiento en la tarea o tareas que se le pide al sujeto que ejecute. Tres factores determinantes en las calificaciones de una prueba de potencia son: a) habilidad innata, es decir, la capacidad potencial para ejecutar cualquier tarea que se les presente a los sujetos; b) habilidad adquirida, es decir, el ejercicio y entrenamiento mediante el aprendizaje, que a través de su interacción con el ambiente va obteniendo el individuo, y c) motivación, o sea, el grado de disponibilidad en el que se encuentra el sujeto para ejecutar la tarea.

La naturaleza de la inteligencia se puede analizar a través del estudio de las diferencias individuales en el rendimiento ante determinadas tareas cognitivas. El estudio de las correlaciones entre el rendimiento de distintos individuos ante las mismas tareas nos permite determinar el rango de variabilidad de este rendimiento

y comparar este rango entre diferentes tareas. Si ante distintas tareas los individuos muestran un nivel similar de variabilidad en su rendimiento, es posible suponer que esta variabilidad esté reflejando una misma aptitud o capacidad (Pueyo, 1996).

Por esta razón, la adaptación y estandarización así como sus tablas de normas o grados de calificación para grupos específicos, deben ser tomadas en cuenta en la elección y evaluación del instrumento psicológico.

Cualquier prueba estandarizada tiene instrucciones homologadas para su aplicación y calificación que deben seguir muy de cerca, dejando pocas oportunidades para la interpretación o inclinación personal. Sin embargo, para estandarizar una prueba, es necesario pero no suficiente contar con instrucciones estándar. Además, la estandarización de cualquier instrumento de evaluación psicológica requiere de la aplicación a una extensa muestra de individuos (la muestra de estandarización) que se seleccionen como representativos de la población meta para la que se elabora el instrumento. Es claro entonces que el propósito principal del proceso de estandarización es determinar la distribución de calificaciones brutas en el grupo de estandarización (grupo de norma). Las calificaciones brutas que se obtienen se convierten entonces en alguna forma de calificaciones derivadas, o normas (Aiken, L. 2003).

Las normas de calificación son importantes para comprender los resultados de una prueba, ya que proporcionan los parámetros que nos van a permitir comparar la ejecución de un individuo en dicha prueba dentro de su propio grupo (Morales, 1990).

Las tablas de normas de calificación pueden ser de varios tipos y, por ende, contener diferente información; por ejemplo, normas de edad, normas de calificaciones escolares, etc.

Una tabla de normas debe contener las calificaciones reales y sus correspondientes calificaciones individuales (Morales, 1990).

En términos del tamaño y calidad representativa de la muestra, con frecuencia las pruebas de grupo, y en especial las pruebas de aprovechamiento, se estandarizan de manera más adecuada que las individuales. Las normas para las pruebas de grupo pueden basarse hasta en 100,000 personas, en tanto que el tamaño del grupo de norma para una prueba individual estandarizada en forma detallada puede ser de 2,000 a 4,000. Desde luego, contar con una muestra de estandarización extensa no garantiza que sea representativa de la población de interés. Para ser representativa de esa población meta, es necesario seleccionar la muestra con detenimiento.

## **2.11. TESTS PSICOLÓGICOS Y SU APLICACIÓN EN LA PSICOLOGÍA GENERAL**

Anastasi (1998) cita: Una de las especialidades más antiguas y exitosas de la psicología es la que abarca las pruebas de capacidad cognoscitiva. A pesar del fragoroso debate sobre la confiabilidad de aplicar pruebas a unas minorías, la opinión de los psicólogos es la siguiente: la inteligencia humana es una característica que todas las personas poseen en mayor o menor grado; los psicólogos pueden identificarla y cuantificarla. Además. Los psicólogos creen que la inteligencia general permite predecir, con confiabilidad uniforme en todos los casos, el desempeño de las personas con respecto a una amplia gama de trabajos (Dipboye, 1983). Las pruebas cognoscitivas son instrumentos confiables para predecir el desempeño en todos los tipos de trabajo y en todos los ambientes (Hopkins, 1983). Las pruebas de capacidad cognoscitiva son igualmente confiables

para los grupos mayoritarios que los minoritarios. Los costos laborales pueden reducirse utilizando las pruebas de inteligencia.

#### Importancia de la inteligencia para la selección de personal:

La importancia de una buena selección radica en tener mejores expectativas de un trabajador en el desempeño de sus actividades dentro de una empresa, por consiguiente la evaluación debe realizarse en forma secuencial y precisa considerando los siguientes aspectos:

- Experiencia laboral
- Trayectoria académica
- Aspectos psicológicos

En cuanto a los aspectos psicológicos, es ahí donde adquiere una gran importancia la selección técnica de personal para una empresa, es evaluar los componentes básicos de la personalidad de un sujeto lo que hace posible su óptimo desarrollo dentro de una sociedad y, específicamente en su ámbito laboral, por lo que deben analizarse usando técnicas confiables. Por medio de estas, la selección se relaciona íntimamente con una serie de técnicas encaminadas a encontrar a la persona adecuada para el puesto adecuado.

El primer paso en el que interviene el psicólogo para hacer una buena selección de personal consiste en elaborar el perfil, en estos casos, este transcribirá la información de carácter psicológico y así seleccionara al candidato mejor preparado en base al perfil que resulte de los datos, que permitan elaborar un método mas adecuado de evaluación. Estos datos se refieren básicamente al área intelectual, de personalidad y habilidad.

La estructuración de una batería psicológica debe estar basada directamente en los datos contenidos en el perfil, y este, a su vez, en los que presenta la requisición.



Las características intelectuales estarán determinadas por los factores de capacidad, en términos de análisis, síntesis, nivel de pensamiento, integración, organización, etc. Que el puesto requiera. Las características de habilidades, por las condiciones técnicas determinadas por el tipo de labor que se vaya a desempeñar. Las de personalidad se refieren a los factores internos como motivación, intereses, estabilidad emocional, juicio, etc.

Es muy importante evaluar en términos de inteligencia a los candidatos para algún puesto, ya que los resultados de esta evaluación, significan un valor muy determinante para su desempeño profesional y laboral (Arias, 1978).

Desde el punto de vista del empleador y del empleado, es obvia la importancia de que los individuos se coloquen en los trabajos para los cuales están calificados. La colocación adecuada también implica que los rasgos que sean irrelevantes para los requisitos del trabajo no afecten (favorable o desfavorablemente) las decisiones de la selección. Ya nos resulta familiar el simple hecho psicométrico de que es necesario averiguar la validez del instrumento para los usos particulares de la prueba, aunque a adquirido nueva importancia por la preocupación generalizada acerca de la colocación laboral de minorías que surgen de desventajas culturales y educativas. Una prueba no valida o una que incluye elementos que no están relacionados con el trabajo considerado puede excluir de manera injusta a los miembros de grupos minoritarios que podrían realizar ese trabajo de manera satisfactoria.

Otra preocupación pertinente, tanto para las organizaciones como para la sociedad en su conjunto, surge de la relación demostrada entre la productividad en el trabajo y la validez de los instrumentos utilizados. Las ganancias y pérdidas estimadas en la productividad que se asocian con los aumentos y las disminuciones de la validez de los procedimientos de selección de personal son considerables; en las instituciones que contratan a muchos empleados, con las

dependencias gubernamentales, el valor acumulado de estas ganancias y pérdidas es tan grande que merece una estrecha atención (Anastasi, 1998).

Por este motivo se lleva a cabo la estandarización y normalización de instrumentos en poblaciones específicas más amplias.

Es definitivo que los aspectos culturales e idiomáticos influyen en la medición de la Inteligencia. Las conceptualizaciones de inteligencia dependen de cada cultura y sociedad. Dichas concepciones se ven influenciadas debido a que existen diferentes factores específicos de la cultura en cuestión como son: los sistemas de enseñanza, las políticas educativas, las diferencias curriculares, la riqueza, el nivel de vida, los valores culturales, etc. que pueden ser esenciales para interpretar adecuadamente dicho concepto (Muñiz Fernández, 1996).

Por ejemplo, Westbury (1992) señaló que los resultados obtenidos en el Segundo Estudio Internacional de Matemáticas (SIMS) indican que los estudiantes americanos rendían pobremente en todos los cursos y en todos los aspectos de las matemáticas incluidos en el test. Cuando se compara el rendimiento de los estudiantes japoneses con el de los americanos, se observan grandes diferencias importantes entre los estudiantes de los dos países.

Debido a lo anterior, para llevar a cabo cualquier interpretación de inteligencia con sentido, hay que tener en cuenta las diferentes realidades sociales, políticas y económicas a las que se enfrentan los países, así como la relevancia de las oportunidades educativas a la luz de estas realidades (Olmedo, 1981).

Así mismo, en las Directrices de la Comisión Internacional de Tests (ITC) (Hambleton, 1994) para adaptación de tests psicométricos y educativos se hace especial hincapié en que las diferencias existentes entre varios grupos idiomáticos y culturales son función no solo de las diferentes tradiciones, normas y valores,

sino también de diferentes interpretaciones y visiones del mundo. Por ello es muy posible que el mismo constructo sea interpretado y comprendido de forma completamente diferente por dos grupos diferentes.

Por ejemplo, el concepto de inteligencia existe en casi todas las culturas. Sin embargo, en muchas culturas occidentales este concepto se asocia con ser capaz de dar respuestas muy rápidamente, mientras que para muchas culturas del Este la inteligencia se asocia a menudo con un pensamiento reposado, de reflexión y de decir lo correcto (Lonner, 1990; citado por Gregory, 2001).

Patricia M Greenfield (1997) de la Universidad de California, en su artículo: "You Can't Take It With You Why Ability Assessments Don't Cross Cultures", pone en evidencia la necesidad de llevar a cabo estudios basados en la psicología transcultural, ya que ella argumenta que este tipo de investigaciones permiten proveer a psicólogos dedicados a la investigación y a clínicos con herramientas para detectar, corregir y disminuir las diferencias transculturales que afectan la validez de los tests de habilidades e inteligencia aplicados fuera de la cultura en la que fueron creados. Cuando los psicómetras utilizan tests desarrollados en una cultura específica muy pocas veces toman en cuenta a los evaluados procedentes de otra cultura, generalmente consideran que los evaluados de otra cultura poseen los valores, conocimiento, y comunicación implícita que asumió el test. Es por eso que, los estudios transculturales toman gran importancia en los tests de habilidades y de inteligencia ya que ponen especial atención en las diferencias étnicas de la sociedad.

**Instrumentos de evaluación que utilizan con mayor frecuencia los psicólogos asesores o consejeros en el área industrial.**

<b>Tests</b>	<b>Rngo de uso</b>
• Inventario multifacético de la personalidad de Minnesota.	1
• Escala de inteligencia Wechsler para adultos, revisada.	2
• Inventario de intereses de Strong - Campbell.	3
• Test de frases incompletas.	4
• Test gestaltico visomotor de Bender.	5
• Escala de inteligencia Wechsler para niños.	6
• Test casa - árbol - persona.	7
• Test del dibujo de la persona.	8
• Test de apercepción temática.	9
• Test de manchas de tinta de Rocharch.	10.5
• Cuestionario de los 16 factores de personalidad.	10.5
• Test de rendimiento de rango amplio.	12
• Escala de preferencias personales de Edwards	13.5
• Índice de intereses ocupacionales de Kuder.	13.5
• Inventario psicológico de California.	15
• Escala de Wechsler de memoria.	16
• Escala de inteligencia Wechsler para preescolares y primaria.	17
• Test de aptitudes diferenciales.	18

Adaptación de Watkins, Campbell y McGregor, 1988 (Aiken, 2003)

## **CAPITULO III “ESCALA DE HABILIDAD GENERAL PARA ADULTOS (GAMA)”**

### **FICHA PSICOMÉTRICA**

**Título:** General Ability Measure for Adults (Escala de Habilidad General para Adultos)

**Siglas de la prueba:** GAMA

**Autor:** Naglieri, Jack A. Bardos, Achilles N.

**Materiales de la Prueba:** 1. Manual. 2. Cuadernillo de Aplicación. 3. Cuadernillo de Aplicación a Hispanos

**Año de Publicación:** 1997

**Actualización más reciente:** Abril de 1998

**Número de localización (ETS):** TC020250

### **3.1. INFORMACIÓN DESCRIPTIVA**

Resumen: El GAMA (General Ability Measure of Adults) es una prueba breve, auto administrada y no verbal que mide inteligencia. Esta prueba evalúa sobre todo la habilidad general con ítems que requieren la aplicación del razonamiento y lógica para resolver los problemas que exclusivamente utiliza diseños abstractos y figuras. El instrumento cuenta con 66 ítems y puede ser completada en 25 minutos. Se compone de cuatro subtest: igualdades, analogías, secuencias y construcción.

Ya que se llevo a cabo la aplicación con gente que alcanza el nivel de lectura del tercer año, esto permite que pueda ser leído y comprendido sin mayor problema.

La estandarización de la prueba se realizó a nivel nacional (EU), donde cuidadosamente se aplico en diferentes estratos seleccionados al azar con la intención de acercarse a la población del censo de 1990 de los EUA. La muestra

incluyó nueve grupos de edad con un número similar de sujetos. Cada grupo fue estratificado por género (1079 hombres y 1281 mujeres, muestra total), raza o grupo étnico (12.5% de afroamericanos, 0.7% de indio americano, 2.6% de asiáticos, 7.8% de latinos, 76.3% de blancos, muestra total), nivel educativo y por región geográfica (80 ciudades en 23 estados). La administración fue realizada por psicólogos y profesionales previamente capacitados, con la intención de llevar un estricto control en las instrucciones de aplicación y de tiempo. Se llevó un control de calidad de las aplicaciones y materiales.

La confiabilidad de la puntuación del GAMA CI y de los cuatro subtest, son determinadas por dos métodos. La consistencia interna fue determinada por el método de mitades, y la estabilidad de las puntuaciones sobre el tiempo fue determinado a través de un test – retest.

La consistencia interna de los subtest del GAMA fueron determinados por el método de mitas con la correlación de Spearman Brown. La confiabilidad de los coeficientes promedio a través de los 11 grupos de edades fue de .66 para el subtest de igualdades, .81 para el subtest de analogías, .79 para el subtest de secuencias, .65 para el subtest de construcción. La puntuación promedio de confiabilidad a través de los 11 grupos de edad fue de .90 en donde a rango de 80 y más años obtuvo .79, y el rango de 35 a 44 años de edad fue de .94.

La validez del GAMA fue examinada evaluando una variedad de factores relacionados a las inferencias que se pueden hacer de las puntuaciones del test. Contiene una validez de contenido según su desarrollo, validez concurrente según la correlación con otras habilidades generales de test reconocidos y validez del criterio utilizado según las diferencias de grupos.

### **3.2. OBJETIVOS DEL GAMA**

El test GAMA provee una evaluación no verbal de la habilidad general para profesionales de la salud mental que necesitan medir la inteligencia en una manera eficiente. El GAMA fue diseñado para proveer una medida de habilidad general que:

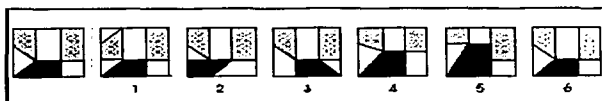
- Es normada en una muestra reciente de estandarización de 2360 individuos entre 18 y 96 años.
- Utiliza un mínimo de contenidos verbales
- Tiene un mínimo de requerimientos motrices (no manipulativos)
- Es breve y fácil de administrar
- Puede ser auto administrada
- Puede ser aplicada individual o grupal mente
- Ofrece un estimado de habilidad mental que es justo para diversos grupos sociales
- Utiliza colores para mantener la atención del sujeto que la realiza
- Utiliza colores que son lo menos problemático para las personas daltónicas
- Puede ser calificada manualmente o por computadora

### **3.3. ESTRUCTURA DEL GAMA**

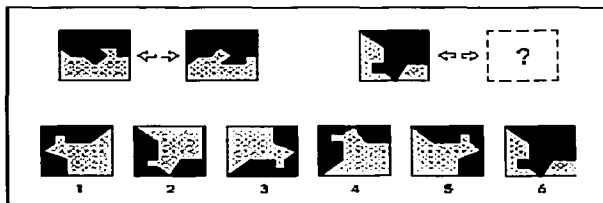
Cuenta con 66 ítems organizados en cuatro tipos de subtest:

- a) Igualdades
- b) Analogías
- c) Secuencias
- d) Construcción.

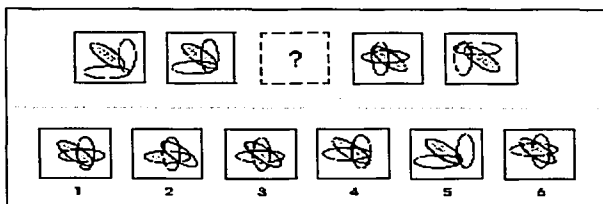
- En el subtest de igualdades se requiere que el examinado determine cuál de las opciones es idéntica al estímulo en color, forma y configuración.



- En el subtest de analogías se requiere que el examinado reconozca la relación entre dos figuras abstractas y después identifique la opción que tenga un par de figuras diferentes con el mismo concepto relacional.



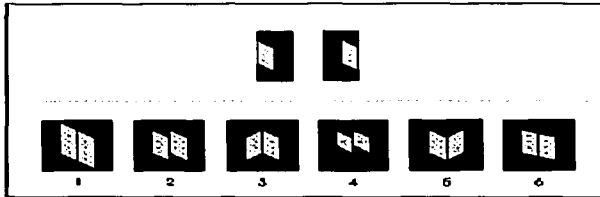
- En el subtest de secuencias, la forma, el color, y la localización de un diseño geométrico cambie en una secuencia lógica, se requiere que el examinado reconozca el patrón de cambio y escoja la opción que concuerde con el patrón.



TESIS CON  
FALLA DE ORIGEN



- En el subtest de construcción se requiere que el examinado determine como se pueden combinar dos figuras para crear una de los diseños que se encuentran como opción.



Los ítems deben ser analizados por el examinado y sintetizar mentalmente la construcción estimulada por dos, tres o cuatro estímulos de varias formas o colores.

### **3.4. USOS DEL GAMA**

Las siguientes sugerencias se basan en la necesidad de medir la inteligencia de forma general.

- Como parte de la evaluación Psicológica
- En escenarios clínicos
- En orientación vocacional
- Para la industria y los negocios
- Para pruebas de gran escala
- Para escuelas publicas y escenarios colegiales

TESIS CON  
FALLA DE ORIGEN

### **3.5. ADMINISTRACIÓN Y CALIFICACIÓN**

#### **MATERIALES**

La prueba incluye un manual, un libro de aplicación y dos tipos de hojas de respuesta.

El manual de la prueba contiene información general acerca del GAMA, una guía específica para la administración y calificación de la prueba, tablas de normas, guía de interpretación y las propiedades psicométricas de la prueba.

El libro de aplicación incluye las instrucciones del test, ejemplos para cada ítem y los 66 ítems que dan forma al GAMA.

La forma de auto evaluación es utilizada para recopilar información demográfica y las respuestas de los ítems. Las instrucciones de llenado las contiene la forma.

La hoja de respuestas escaneable es utilizada para recopilar información demográfica y las respuestas de los ítems, para ser calificados y procesados por computadora, arrojando una calificación computarizada, además el software de reporte permite al examinado resolver el test y producir un reporte de los resultados generado por computadora.

### **3.6. DESARROLLO DEL TEST Y ESTANDARIZACIÓN**

En este punto se describen los métodos usados para desarrollar el GAMA y los procedimientos seguidos en esta estandarización. Debido a que una muestra fiable de estandarización es necesaria para unos datos normativos de alta calidad, se presto una atención cuidadosa a todos los aspectos de identificación de la muestra y a la recolección de los datos.

### **3.6.1. DESARROLLO DE LA PRUEBA**

#### **3.6.1.1. DESARROLLO DE LOS ITEMS**

Aproximadamente 200 ítems fueron creados en el conjunto inicial, de donde se seleccionaron los ítems que componen el conjunto inicial, esto a través de estudios psicométricos. Estos ítems fueron escritos para evaluar la capacidad general usando varias aproximaciones o propuestas, incluyendo la rotación mental, el análisis espacial, el razonamiento analógico y concordancia a los estándares. Los ítems fueron organizados en subpruebas, basados en la manera en que fueron construidos, y en las demandas resultantes de los ítems. Cada uno de los ítems fue examinado apropiadamente para que midieran la habilidad general de manera no verbal.

Tras una revisión inicial de los ítems, un estudio piloto fue conducido para reunir las propiedades psicométricas de los ítems, para comenzar el proceso de selección de ítems. Once tipos de ítems fueron distribuidos a través de cuatro formas de la versión piloto del GAMA. Cada forma contenía 60 ítems (20 ítems para cada uno de los tres tipos, con un tipo repetido en dos folletos). Ciento veintisiete adultos y adolescentes tomaron una de cuatro formas de la prueba. Únicamente ocho tipos de ítems fueron retenidos.

#### **3.6.1.2. LEGIBILIDAD DE LAS INSTRUCCIONES DEL GAMA**

El nivel de lectura requerido para la comprensión de las instrucciones del GAMA fue cuidadosamente monitoreado a lo largo del desarrollo de la prueba. Las instrucciones fueron evaluadas a través del método de Flesch-Kincaid (Flesch, 1948) que da una calificación grado-nivel de lectura de fácil lectura. Los resultados obtenidos para las versiones usadas durante la fase de desarrollo consecuentemente demostraron, que las instrucciones necesitaron un nivel de

lectura elemental. Los resultados para las instrucciones finales del GAMA indicaron que se requería de un grado escolar de 2.4 para su comprensión y lectura.

### 3.6.1.3. TRYOUT (Ejercicio de Prueba) Y SELECCIÓN DE ITEMS

El propósito del tryout fue el de reunir las estadísticas de los ítems de una muestra de adultos para completar el proceso de selección de ítems. Ocho grupos de los diferentes tipos de ítems fueron distribuidos a través de cuatro formas de la prueba. Cada forma de tryout contenía 40 ítems (10 de cada tipo de ítem). Los sujetos de la muestra del tryout fueron seleccionados cuidadosamente para asegurar similitud con las estadísticas del censo de 1990 de los EE.UU. respecto a edad, género, raza o grupo étnico y nivel educacional. Un total de 604 adultos tomaron una de las 4 formas de la prueba. La prueba fue proporcionada con un límite de tiempo de 1 HR. Asegurándose que la mayoría de los sujetos completaran todos los ítems.

Algunos análisis fueron conducidos sobre la base de datos del tryout para evaluar las características psicométricas del GAMA. Estas incluyen la examinación de las calificaciones por edad y género, computando la consistencia interna de cada uno de los 40 ítems del tryout de cada uno de los 4 tipos de ítems. Las estadísticas de los ítems fueron detalladamente examinadas, incluyendo la correlación total de los ítems, dificultad de los ítems y las variables discriminatorias. Todos los ítems fueron evaluados utilizando el procedimiento de Mantel-Haenzel para detectar diferencias del funcionamiento de los ítems. Estos resultados formaron parte para seleccionar los ítems. Los ítems que presentaron diferencias significativas entre hombres y mujeres, personas de raza blanca y personas de otras razas, fueron eliminados a través del procedimiento de Mantel-Haenzel. De los 160 ítems examinados en el tryout, se encontraron 10 que demostraban una diferencia estadísticamente significativa entre hombres y mujeres y otros 3 que mostraban diferencias estadísticamente significativas entre hombres de raza blanca y hombres

de otras razas. Estas y otras depuraciones se realizaron, dando como resultado después de 3 tryouts los 66 ítems en cuatro categorías para la prueba final. Además de probar con diferentes tiempos límite se acordó que 25 minutos era un tiempo apropiado como tiempo límite.

#### 3.6.1.4. RECOLECCIÓN DE LA BASE DE DATOS NACIONAL PARA LA ESTANDARIZACIÓN

El GAMA fue estandarizado utilizando un plan cuidadosamente diseñado de muestreo al azar, estratificando, con el objeto de que la población se acercara lo más posible a la población del censo de los EE.UU. de 1990. La muestra para la estandarización incluyó nueve grupos de edad con la misma cantidad de sujetos, estos eran:

18-19, 20-24, 25-34, 35-44, 45-54, 55-64, 65-69, 70-74, 75-96.

Cada grupo de edad fue estratificado por género, por raza o grupo étnico (Afroamericanos, Indioamericanos, Asiáticos o Isleños del Pacífico, Hispanos y Blancos), por nivel educativo (menos de 9 años completados de estudio, de 9 a 12 años de estudio sin graduarse, nivel secundaria graduados, preparatoria incluyendo programas vocacionales y licenciatura o más), y por región geográfica (norte, sur, este, medio oeste y este). La recolección de los datos se inició en Septiembre de 1995, y se completó en Mayo de 1996. La estandarización Nacional consistió en una muestra de 2360 sujetos entre los 18 y 96 años.

Los administradores de la prueba fueron profesionales en el campo de la Psicología u otras Profesiones con diversos conocimientos comprendiendo clínicos, pasantes y graduados en la Carrera de Psicología de los programas de Psicología Clínica. A todos los administradores de la prueba se les informó con detalle sobre los

lineamientos de la prueba, instrucciones y reglas de tiempo. Los participantes fueron evaluados individual y grupal mente en diversos escenarios.

#### 3.6.1.5. CONTROL DE CALIDAD

Todas las formas del gamma fueron sometidas a una minuciosa inspección visual por personal calificado, quienes se aseguraron que completaron la prueba. Además las características demográficas fueron verificadas en un alto porcentaje para garantizar la veracidad de los datos. Finalmente el análisis de los datos y el desarrollo de las normas fueron sometidos a una serie de procedimientos con garantía de calidad.

### 3.6.2. ESTANDARIZACIÓN

#### 3.6.2.1. CARACTERÍSTICAS DEMOGRÁFICAS DE LA MUESTRA DE ESTANDARIZACIÓN

##### REGIONES GEOGRÁFICAS

La estandarización del Gamma se llevó a cabo en 80 Ciudades y 23 Estados. Esto demuestra que la distribución geográfica de la estandarización se acerca a la muestra que se tomó para el censo de Estados Unidos de 1990.

Región Geográfica												
Grupo de Edad	Noroeste			Sur			Medio oeste			Oeste		
	Muestra GAMA		Población de U.S.	Muestra GAMA		Población de U.S.	Muestra GAMA		Población de U.S.	Muestra GAMA		Población de U.S.
	<i>n</i>	%	%	<i>n</i>	%	%	<i>n</i>	%	%	<i>n</i>	%	%
18-19	56	21.1	19.9	94	35.5	35.5	59	22.3	24.1	56	21.1	20.4
20-24	55	20.3	20.5	93	34.3	34.4	60	22.1	23.1	63	23.2	21.9
25-34	66	21.3	20.3	101	32.6	34.1	77	21.8	22.3	66	21.3	22.4
35-44	72	24	20.5	99	33	33.9	71	23.7	23.5	58	19.3	22.1
45-54	59	21.5	21.1	96	35	34.5	64	23.4	23.8	55	20.1	20.6
55-64	58	22.6	22	90	35	34.5	65	25.3	24.4	44	17.1	19.1
65-69	56	23.5	22.1	81	34	34.7	57	23.9	24.1	44	18.5	19.1
70-74	49	22.4	22.4	70	32	34.5	56	25.6	24.4	44	20.1	18.7
75-n	53	23.5	22.6	59	26.1	34	69	30.5	25.6	45	19.9	17.9
Muestra Total	524	22.2	21	783	33.2	34.3	578	24.5	23.8	475	20.1	20.9

TESIS CON  
FALLA DE ORIGEN

GÉNERO Y EDAD

**LA MUESTRA ES SIGNIFICATIVAMENTE PARECIDA EN PORCENTAJE A LA POBLACIÓN NORTEAMERICANA SEGÚN EL CENSO DE 1990.**

Genero y Edad							
Grupo de Edad	Hombres			Mujeres			Muestra Total
	Muestra Gama		Población de U.S	Muestra Gama		Población de U.S	
	n	%	%	n	%	%	
18-19	135	50.9	51	130	49.1	49	565
20-24	139	51.3	50.8	132	48.7	49.2	21
25-34	151	48.7	49.9	159	51.3	20.1	310
35-44	145	48.3	49.5	155	51.7	50.5	300
45-54	129	47.1	48.8	145	52.9	51.2	274
55-64	116	45.1	47.1	141	54.9	52.9	257
65-69	97	40.8	44.8	141	59.2	55.2	238
70-74	89	40.6	42.7	130	59.4	57.3	219
75-n	78	34.5	34.8	148	65.5	65.2	226
Muestra Total	1079	45.7	47.8	1281	54.3	52.2	2360

EDAD Y RAZA

Los resultados ilustran que la muestra es significativamente parecida en porcentaje a la población norteamericana según el censo de 1990.

Grupo de Edad	Afroamericanos		Indio Americano		Asiático / Islas del pacifico			Hispano / Latinos			Blancos				
	Muestra GAMA		Población de U.S.	Muestra GAMA		Población de U.S.	Muestra GAMA		Población de U.S.	Muestra GAMA		Población de U.S.			
	n	%	%	n	%	%	n	%	%	n	%	%			
18-19	43	16.2	14.5	1	0.4	1.0	14	5.3	3.3	27	10.2	11.0	180	67.9	70.2
20-24	42	15.5	13.3	2	0.7	0.9	12	4.4	3.2	37	13.7	11.7	178	65.7	70.8
25-34	46	14.8	12.4	4	1.3	0.8	14	4.5	3.2	32	10.3	10.0	214	69.0	73.5
35-44	34	11.3	11.1	6	2	0.8	5	1.7	3.3	31	10.3	7.7	224	74.6	77.0
45-54	31	11.3	10.4	3	1.1	0.7	4	4.5	2.9	22	8.0	6.7	214	78.1	79.3
55-64	34	13.2	9.4	1	0.4	0.6	1	0.4	2.2	14	5.4	5.5	207	80.5	82.3
65-69	24	10.1	8.5	0	0.0	0.4	5	2.1	1.7	10	4.2	4.0	199	83.6	85.3
70-74	21	9.6	8.0	0	0.0	0.4	3	1.4	1.5	7	3.2	3.2	188	85.8	86.9
75-n	19	8.4	7.6	1	0.4	0.3	3	1.3	1.1	5	2.2	3.0	198	87.6	87.9
Muestra T	294	12.5	11	18	0.8	0.7	61	2.6	2.8	185	7.8	7.7	1802	76.3	77.8

TESIS CON FALLA DE ORIGEN

### NIVEL EDUCATIVO Y EDAD

Los resultados ilustran que la muestra es significativamente parecida en porcentaje a la población norteamericana según el censo de 1990.

<b>Nivel Educativo y Edad</b>															
Grupo de Edad	8vo grado o menor			9no a 12vo grado con diploma			Secundaria Graduados			Preparatoria o equivalente			Licenciatura o mayor		
	Muestra GAMA		Población de U.S.	Muestra GAMA		Población de U.S.	Muestra GAMA		Población de U.S.	Muestra GAMA		Población de U.S.	Muestra GAMA		Población de U.S.
	n	%	%	n	%	%	n	%	%	n	%	%	n	%	%
18-19	4	1.5	3	101	38.1	35.5	81	30.6	30.4	79	29.8	31.0	0	0	0.1
20-24	6	2.2	3.6	46	17	14	82	30.3	31.2	112	41.3	40.7	25	9.2	10.6
25-34	10	3.2	4	39	12.6	12	98	31.6	30.6	97	31.3	30.7	66	21.3	22.7
35-44	10	3.3	4.7	31	10.3	9.8	91	30.3	28	93	31	30.5	75	25	27.0
45-54	19	6.9	7.9	36	13.1	14.1	91	33.2	31.9	66	24.1	24.4	62	22.6	21.7
55-64	29	11.3	13.9	49	19.1	18.5	89	34.6	32.8	51	19.8	18.8	39	15.2	16
65-69	43	18.1	18	47	19.7	20.2	76	31.9	32.5	38	16	16.8	34	14.3	12.5
70-74	45	20.5	22.5	46	21.0	21.5	68	31.1	30.6	31	14.2	14.7	29	13.2	10.6
75-n	72	31.9	34.3	45	19.9	20.9	48	21.2	22.8	33	14.6	12.5	28	12.4	9.5
Muestra T	238	10.1	9.4	440	18.6	15.2	724	30.7	30.1	600	25.4	26.7	358	15.2	18.5

### NIVEL EDUCATIVO Y RAZA O GRUPO ÉTNICO

**LOS RESULTADOS ILUSTRAN QUE LA MUESTRA ES SIGNIFICATIVAMENTE PARECIDA EN PORCENTAJE A LA POBLACIÓN NORTEAMERICANA SEGÚN EL CENSO DE 1990.**

<b>Nivel Educativo y Raza o Grupo Étnico</b>															
Grupo de Edad	8vo grado o menor			9no a 12vo grado con diploma			Secundaria Graduados			Preparatoria o equivalente			Licenciatura o mayor		
	Muestra GAMA		Pob. de U.S.	Muestra GAMA		Pob. de U.S.	Muestra GAMA		Pob. de U.S.	Muestra GAMA		Pob. de U.S.	Muestra GAMA		Pob. de U.S.
	n	%	%	n	%	%	n	%	%	n	%	%	n	%	%
Afroamericana	43	14.6	11.8	88	29.9	24.2	89	30.3	29.1	59	20.1	25.0	15	5.1	10
Indio americano	2	11.1	12.2	5	27.8	22.7	5	27.8	30	6	33.3	27.2	0	0	8
Asiático/l. del Pacífico	8	13.1	11.2	8	13.1	10.6	8	13.1	19.6	17	27.9	25.9	20	32.8	32.7
Hispano/latino	54	29.2	27.4	54	29.2	22	38	20.5	22.7	32	17.3	20.1	7	3.8	7.8
Blancos	131	7.3	7.2	285	15.8	13.4	584	32.4	31.4	486	27.0	27.7	316	17.5	20.4
Muestra T	238	10.1	9.4	440	18.6	15.2	724	30.7	30.1	600	25.4	26.7	358	15.2	18.5

TESIS CON  
FALLA DE ORIGEN



## LENGUAJE

Los individuos que participaron en la muestra de estandarización hablaban los siguientes idiomas: Chino, inglés, francés, alemán, italiano, japonés, coreano, polaco, español, vietnamita y Tagalog. Únicamente el 6.8% de la muestra reporto que hablaba mas de un idioma.

### 3.6.2.2. DERIVACIÓN DE LAS CALIFICACIONES DE LA PRUEBA

La construcción de las normas del GAMA comenzaron con la examinación de las medias, desviaciones estándar y frecuencias estadísticas para los cuatro subtests. Este análisis se realizó en el total de la muestra y por grupos de edades. La distribución de las calificaciones de los subtest (el número de ítems correctos dentro de un subtest) y el total de las calificaciones (el número de ítems correctos de toda la prueba) fueron examinados a través de los grupos de edad por años. Basados en la progresión de las calificaciones medias, se determinó que la mejor elección era crear once grupos de edad. Por esta razón las calificaciones normativas estándares de los subtest fueron ordenadas dentro de los once grupos de edad.

### 3.6.2.3. CALIFICACIONES DEL LOS SUBTEST

Se obtuvieron las puntuaciones de la distribución de frecuencia acumulativa para los once grupos de edad, esta distribución fue normalizada y las puntuaciones netas fueron convertidas en puntuaciones escalares, de modo que cada grupo de edad tuvo una media de 10 y una desviación estándar de 3. La progresión de las puntuaciones escalares dentro y entre grupos de edad fueron examinadas y pulidas para corregir pequeñas irregularidades en la progresión de las puntuaciones, obteniendo como resultado las tablas de conversión de puntuaciones brutas a escalares.

#### 3.6.2.4. PUNTUACIÓN DEL C.I. GAMA

El primer paso en la puntuación del C.I. gama fue calcular la media y la desviación estándar de la suma de los cuatro subtest por grupos de edad. Un análisis de varianza de la suma de los subtest indicó que no existía una variación significativa por edad. Además se llevó a cabo el test de Barlett para obtener la homogeneidad de la varianza, la cuál reveló que no existían diferencias significativas por edad. Por lo tanto, los grupos de edad fueron combinados para construir la tabla de normas para las puntuaciones de C.I. gama. El C.I. gama fue estandarizado para obtener una media de 100 y una desviación estándar de 15.

#### 3.6.2.5. DESARROLLO DE LOS PERCENTILES E INTERVALOS DE CONFIANZA DEL C.I. GAMA

La tabla de conversión de Puntuaciones del C.I. Gama incluye los rangos percentiles e intervalos de confianza. Los rangos percentiles están asociados con la Teoría de la Curva Normal equivalente de la puntuación estándar. Estos valores están basados en la estimación del error estándar y están centrados en la estimación de puntuaciones obtenidas.

### **3.7. PROPIEDADES ESTADÍSTICAS**

#### **3.7.1. CONFIABILIDAD**

La confiabilidad de la prueba esta definida por Anastasi(1988) como "la consistencia de las calificaciones obtenidas por la misma persona cuando se le reexamina con un mismo instrumento en ocasiones diferentes, o con ítems equivalentes, o bajo otras condiciones de examinación" la confiabilidad de las calificaciones del C.I. GAMA y de sus cuatro subtest fue determinado utilizando dos métodos. La consistencia interna fue determinada por el método de división

por mitades, y la estabilidad de las puntuaciones sobre tiempo fue determinado a través de un test – retest.

### 3.7.2. CONSISTENCIA INTERNA

La consistencia interna para los subtests del GAMA fue determinada utilizando el método de división por mitades con la corrección de Spearman Brown. Los ítems de cada uno de los subtest fueron divididos en dos grupos utilizando el método de pares y nones. Los resultados de este análisis se presentan en la siguiente tabla para cada grupo de edad y para la muestra total de estandarización.

<b>Coefficiente de Confiabilidad para el GAMA y Calificaciones C.I. GAMA</b>						
	<i>Concordancia</i>	<i>Analogías</i>	<i>Secuencias</i>	<i>Construcción</i>	<i>Calificaciones C.I. GAMA</i>	<i>Tamaño de la Muestra</i>
18-19	0.65	0.83	0.81	0.71	0.91	265
20-24	0.63	0.82	0.84	0.73	0.91	271
25-34	0.62	0.85	0.82	0.71	0.9	310
35-44	0.77	0.87	0.86	0.78	0.94	300
45-54	0.74	0.85	0.87	0.67	0.92	274
55-59	0.71	0.87	0.85	0.69	0.92	153
60-64	0.58	0.78	0.76	0.53	0.88	104
65-69	0.68	0.83	0.79	0.72	0.91	238
70-74	0.62	0.79	0.72	0.54	0.86	219
75-79	0.53	0.77	0.69	0.53	0.85	135
80 +	0.6	0.59	0.56	0.39	0.79	91
Promedio	0.66	0.81	0.79	0.65	0.9	2360

NOTA: Todos los coeficientes de confiabilidad se obtuvieron con el método de división por mitades.

El coeficiente de confiabilidad de la calificación C.I. GAMA se basa en una composición lineal calculada a partir de la fórmula  $r_{yy} = 1 - (k - \sum r_{ii}) / \sum r_{ii}$  (Nunnally & Bernstein, 1994).

El Promedio del coeficiente de confiabilidad se basa en la transformación z de Fisher de los coeficientes de confiabilidad de los grupos de edad.

Los promedios de los coeficientes de confiabilidad a través de los 11 grupos de edad en el subtest de concordancia fueron de .66, analogías de .81, secuencias de .79 y construcción de .65.

También se presenta la consistencia interna del coeficiente del C.I. GAMA.

TESIS CON  
FALLA DE ORIGEN

El promedio de confiabilidad del C.I. GAMA a través de los 11 grupos de edad fue de .90 y sus rangos fueron de .79 en el grupo de 80+ a .94 en el grupo de 35 – 44 años. El coeficiente de confiabilidad de los subtest resulta bajo ante las el coeficiente de confiabilidad de las calificaciones del C.I. GAMA por que fueron calculadas en una base de pocos ítems.

### 3.7.3. ERROR ESTANDAR DE MEDICIÓN

El error estándar de medición (EEM) para las Calificaciones del C.I. GAMA y de las calificaciones de los subtests fueron obtenidas por el coeficiente de confiabilidad utilizando la siguiente formula:

$$EEM = DS \times (\text{raíz cuadrada}) 1 - r_{yy}$$

En donde la DS es la desviación estándar de las puntuaciones de los subtests y de las calificaciones del C.I. GAMA, y  $r_{yy}$  es el coeficiente de confiabilidad apropiado.

El promedio de las EEM fue calculado utilizando la siguiente formula:

$$\text{Promedio de EEM} = (\text{raíz cuadrada}) \frac{EEM_1^2 + EEM_2^2 + \dots + EEM_{11}^2}{n}$$

En donde el  $EEM_1$  se refiere al primer grupo de edad, el  $EEM_2$  se refiere al segundo y así sucesivamente;  $n$  representa el total de grupos de edad.

Los EEM para las calificaciones de los subtests y del C.I. GAMA están expresadas en la misma puntuación estándar. La Desviación estándar para los subtests es de 3, mientras que la de las calificaciones del C.I. GAMA es de 15; por lo que los EEM para los subtests se muestran bajos respecto al C.I. GAMA. Los EEM se muestran en la siguiente tabla.

<b>Error Estándar de Medición Para los Subtests y Calificaciones C.I. GAMA</b>						
	<i>Concordancia</i>	<i>Analogías</i>	<i>Secuencias</i>	<i>Construcción</i>	<i>Calificaciones C.I. GAMA</i>	<i>Tamaño de la Muestra</i>
18-19	1.78	1.46	1.44	1.68	4.78	265
20-24	1.81	1.33	1.36	1.7	4.69	271
25-34	1.86	1.14	1.22	1.55	4.29	310
35-44	1.46	1.22	1.21	1.58	4.1	300
45-54	1.58	1.27	1.25	1.72	4.46	274
55-59	1.63	1.07	1.3	1.72	4.36	153
60-64	1.81	1.28	1.44	2.01	4.88	104
65-69	1.86	1.31	1.46	1.66	4.72	238
70-74	1.76	1.31	1.51	1.99	4.92	219
75-79	1.93	1.49	1.59	1.92	5.2	135
80 +	1.83	1.85	1.57	2.56	5.77	91
Promedio	1.76	1.35	1.4	1.85	4.76	2360

NOTA: Todos los errores estándares de medición se basaron en la confiabilidad de división por mitades. El promedio del error estándar de medición fue calculado utilizando la raíz cuadrada de los promedios de la suma de los cuadrados de los errores estándares de cada grupo de edad.

Los promedios de los EEM para todos los grupos de edad en los subtest de concordancia, analogías, secuencias y construcción fueron de 1.76, 1.35, 1.40 y 1.85 respectivamente. El promedio del EEM para las calificaciones del C.I. GAMA callo entre los rangos de 4.10 (grupo de edad de 33-44) a 5.77 (grupo de edad de 80 o más) sobresaliendo el EEM de 4.76. Los EEM reportados en la tabla varían de acuerdo con el coeficiente de confiabilidad; como incrementa la confiabilidad, el EEM decrece.

### **3.7.4. CONFIABILIDAD TEST – RETEST**

La confiabilidad test – retest esta basada en la correlación de puntajes estándar obtenidos cuando la misma persona es examinada dos veces en un periodo de tiempo. La estabilidad de la calificaciones de la prueba fue examinada en un estudio con 86 individuos, a los cuales se les administro el GAMA con un espacio de tiempo de 2 a 6 semanas; permaneciendo como promedio 25 días, con una

desviación estándar de 7.7 . La muestra consistió en 48.8% de hombres y el 51.2% de mujeres. La distribución de raza o grupos étnicos fue como se describe a continuación: 10.5% de Afroamericanos, 1.2% de Asiáticos y de las Islas del Pacífico, 12.8% de hispanos y Latinos, y 75.6% de blancos. El nivel educacional de la muestra fue como se describe a continuación: 10.5% de 8vo grado o menos, 16.3% de 9no a 13vo grado sin título, 30.2% Secundaria graduado, 26.7% Preparatoria o vocacional y 16.3% de licenciatura o mas.

La similitud de las medias del test – retest nos indica una estabilidad temporal. El coeficiente de estabilidad del test- retest es estadísticamente significativo y esto sirve como un índice adicional de la estabilidad de las calificaciones del C.I. GAMA sobre el tiempo.

<b>Estabilidad del desempeño de las calificaciones de los Subtests y calificaciones C.I. GAMA</b>						
	Test		Retest		Recor de Calificaciones	r
	Media	Desviación estándar	Media	Desviación estándar		
Concordancia	10.2	3.0	10.1	3.3	0.1	.55 *
Analogías	9.7	3.0	10.6	3.4	0.9	.65 *
Secuencias	9.7	3.4	10.7	3.8	1.0	.74 *
Construcción	9.7	3.1	10.6	3.5	0.9	.38 *
C.I. GAMA	99.2	15.6	103.2	17.6	4.0	.67 *

NOTA: El intervalo del test - retest fue de 25 días (DS= 7.7). Para todas las comparaciones,  $n = 86$   
\* $p < .001$

### 3.7.5. VALIDEZ

DEACUERDO CON LA "Standards for Educational and Psychological Testing" (1985), la validez se refiere a la calidad de las inferencias hecha para las calificaciones del test y es la más importante consideración en la evaluación de un test. La validez del GAMA fue examinada evaluando una serie de factores relacionados a las inferencias que se pueden hacer para las calificaciones del test. Esta incluye el desarrollo del GAMA (Validez de Contenido), la correlación con

otros test de habilidades generales (Validez Concurrente), y la diferencia entre grupos (Validez de Criterio).

### 3.7.5.1. DESARROLLO DEL TEST

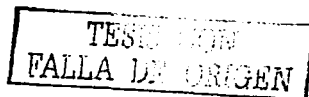
**CREACIÓN DE ITEMS NO VERBALES.** Los ítems del GAMA fueron creados utilizando exclusivamente diseños no verbales para que la inteligencia se pudiera medir sin una confusión potencial, ya sea por las expresiones verbales o la comprensión verbal. Esta característica hace que los ítems sean apropiados para una amplia variedad de individuos con diversos idiomas y lenguajes así como para personas con problemas de comunicación.

**TENDENCIAS DE EDAD.** Se basa en la investigación de Kaufman, Reynolds, & McLean, 1989, la cual indica que la actuación y puntuaciones durante la vida de un adulto van decrementándose. Los grupos de edad del GAMA son parecidos a la investigación mencionada anteriormente. La Media del GAMA cae en el grupo de 20-24 años de edad.

<b>Calificaciones Medias Para los Subtest del GAMA y las Calificaciones Totales de los Grupos de Edad Normativos y la Correlación de las Calificaciones contra la Edad</b>												
<b>Grupos de Edad Normativos</b>												
	18-19	20-24	25-34	35-44	45-54	55-59	60-64	65-69	70-74	75-79	80+	r
	n=26	n=271	n=310	n=300	n=274	n=153	n=104	n=238	n=219	n=135	n=91	N=236
	5											0
Concordancia	8.7	8.7	8.7	8.5	8.0	7.7	7.3	7	6.5	6.1	5.5	-0.43 *
Analogías	11.4	11.4	11.1	10.2	9.2	8.6	7.5	7	6.2	5.6	4.2	-0.56 *
Secuencias	9.1	9.2	8.6	7.7	6.5	6.2	4.8	4.5	3.8	3.1	2.1	-0.56 *
Construcción	8.3	8.2	7.9	7.2	6.5	6.4	5.4	5.4	5.2	4.6	3.9	-0.51 *
Total de calificaciones	37.5	37.6	36.2	33.7	30.1	28.9	25.1	23.8	21.6	19.4	15.7	-0.59 *

Nota: \*p < .001

**CORRELACIÓN TOTAL DE LOS ITEMS.** Una de las técnicas utilizadas durante el desarrollo del GAMA incluye la examinación de la intercorrelación de los ítems y la



relación entre cada ítem y la calificación total. Esta correlación total de los ítems ayudo a asegurarse de que cada ítem estuviera relacionado con el constructo de lo no verbal.

**CORRELACIÓN DE LOS ITEMS DE LOS SUBTESTS.** Examinando la correlación de cada ítem con su respectiva calificación total de su subtest, proveo una media que determino si cada ítem estaba relacionado con cada subtest al cual fuera asignado. Los ítems que correlacionaron significativamente con cada subtest fueron seleccionados para la versión total.

**CORRELACIÓN DE LOS SUBTEST CON LA PRUEBA TOTAL.** La relación de los subtest con el total de la prueba fue examinada para evaluar el límite de relación de cada subtests con los demás. En este análisis se llevo acabo computando la correlación de las puntuaciones de cada subtest sumando las de los tres restantes. Esta intercorrelación, demostró que se mide significativamente un constructo de habilidad general.

<b>Correlación entre las puntuaciones de los subtest del GAMA y la suma de las puntuaciones restantes de las tres escalas restantes, a través de los grupos normativos de edad dentro de la muestra total.</b>												
<b>GRUPOS DE EDAD NORMATIVOS</b>												
	18-19	20-24	25-34	35-44	45-54	55-59	60-64	65-69	70-74	75-79	80+	r
	n=265	n=271	n=310	n=300	n=274	n=153	n=104	n=238	n=219	n=135	n=91	N=2360
Concordancia	0.56	0.52	0.53	0.68	0.60	0.62	0.67	0.61	0.57	0.60	0.46	0.58
Analogías	0.81	0.78	0.76	0.84	0.82	0.79	0.77	0.76	0.70	0.69	0.55	0.77
Secuencias	0.77	0.75	0.70	0.82	0.79	0.77	0.67	0.76	0.63	0.58	0.53	0.74
Construcción	0.69	0.62	0.57	0.69	0.63	0.62	0.61	0.65	0.50	0.57	0.51	0.62
Nota: todas las correlaciones son significativas a partir de $p < .001$												

**INTERCORRELACIÓN DE LOS SUBTEST.** Los subtest del GAMA están organizados de acuerdo a las demandas específicas de los ítems, mejor dicho que representen habilidades separadas. Por eso, todos ellos están dirigidos a medir la inteligencia



general no verbal. La correlación entre los cuatro subtest se muestra en la siguiente tabla. Estas correlaciones son todas altas y significativas y soportan el concepto que los subtest miden un constructo de habilidad general.

<b>Intercorrelación entre las Calificaciones de los Subtest del GAMA a través de los Grupos Normativos de Edad dentro de la Muestra Total</b>												
	<b>Grupos Normativos de Edad</b>											<i>r</i>
	18-19 <i>n</i> =265	20-24 <i>n</i> =271	25-34 <i>n</i> =310	35-44 <i>n</i> =300	45-54 <i>n</i> =274	55-59 <i>n</i> =153	60-64 <i>n</i> =104	65-69 <i>n</i> =238	70-74 <i>n</i> =219	75-79 <i>n</i> =135	80+ <i>n</i> =91	
Concordancia con Analogías	0.58	0.54	0.56	0.68	0.61	0.63	0.66	0.56	0.50	0.54	0.34	0.58
Concordancia con Secuencias	0.49	0.46	0.43	0.63	0.55	0.54	0.53	0.56	0.51	0.41	0.29	0.51
Concordancia con Construcción	0.43	0.37	0.37	0.52	0.42	0.46	0.51	0.48	0.38	0.51	0.43	0.44
Analogías con Secuencias	0.77	0.76	0.72	0.82	0.80	0.77	0.67	0.75	0.65	0.60	0.56	0.74
Analogías con Construcción	0.66	0.59	0.54	0.66	0.61	0.56	0.58	0.61	0.5	0.49	0.39	0.58
Secuencias con Construcción	0.67	0.62	0.56	0.67	0.62	0.61	0.5	0.61	0.36	0.4	0.36	0.58

Nota: todas las correlaciones son significativas a partir de  $p < .01$

### **3.7.6. CORRELACIONES CON OTROS TESTS DE HABILIDADES GENERALES**

El GAMA, WAIS - R y K - BIT (Kaufman Brief Intelligence Test) fueron concurrentemente aplicados a una muestra de 194 individuos, los cuales tenían de entre 25 y 74 años (la media de edad es de 49.4 y la DS es de 16.5). El grupo incluía a 91 hombres (46.9%) y a 103 mujeres (53.1%) y consistía de 171 blancos (88.15%), 22 Afroamericanos (11.3%) y un Asiático (0.5%). Contaban con niveles de educación de 9no a 12vo grado sin graduación (19.1%), con nivel

de secundaria graduados (32.%), con nivel preparatoria o vocacional (25%) y con nivel licenciatura o mayor (23.2%). Los participantes fueron del Noroeste (9.8%, del medio este (54.1%), del este (18.0%) y del sur (18.0%), en donde la mayoría fueron reclutados durante el proceso de estandarización. Examinadores capacitados administraron los tres test a todos los participantes en orden específico.

Las medias y desviaciones estándar para los tres tests se presentan en la siguiente tabla. Es importante destacar que se formaron tres grupos de edad los cuales comprendían de 25 – 34 años, 45 – 54 años y el de 65 – 74 años. Todas las calificaciones cayeron en el rango promedio de la habilidad intelectual. Esto muestra que los tres tests cuentan con calificaciones similares en los tres grupos de edad, soportando el uso del GAMA como una medida de la habilidad cognitiva general.

<b>Medias y Desviaciones Estándar de las calificaciones C.I. Del GAMA, WAIS-R y del K-BIT</b>									
	<b>Grupos de Edad</b>						<b>Muestra Total</b>		
	25 - 34		45 - 54		65 - 74		Media	DS	N
	Media	DS	Media	DS	Media	DS	Media	DS	N
C.I. GAMA	103	14.2	103.5	12.9	107.1	14.1	104.5	13.8	194
WAIS-R PIQa	104	15.2	104.7	16.5	111.3	15	106.7	15.9	194
WAIS-R VIQb	99.6	12.2	103.9	13.2	110.1	13.2	104.6	13.5	194
WAIS-R FSIQc	101.3	13.6	104.5	14.5	111.9	14.8	105.9	14.9	194
K-BIT Matrices	105	11	105	11.5	110.1	9.4	106.7	10.9	189
K_BIT Vocabulario	101.5	10.9	101.2	10.5	111.8	10.1	104.9	11.5	189
K-BIT C.I.	103.6	10.9	103.5	11	112.1	9.1	106.4	11.1	189

Nota: PIQa = C.I. Desempeño(Ejecutivo), VIQb = C.I. Verbal, FSIQc = Escala Completa

La correlación entre las calificaciones del C.I. GAMA y las de los subtest del WZIS-R PIQ (Ejecutivo) y la del K-BIT Matrices fueron computadas y se muestran en la siguiente tabla. La correlación más alta y significativa sugiere que el GAMA mide un constructo de razonamiento no verbal como las otros dos tests. Esto es, que los tres tests miden la habilidad general no verbal, aunque que utilizan métodos diferentes y se administra en tiempos y forma disímil.

**TESIS CON  
FALLA DE ORIGEN**

<b>Correlación entre el C.I. GAMA y WAIS-R y K-BIT</b>					
		<b>Grupos de Edad</b>			
	<b>N</b>	<b>25-34</b>	<b>45-54</b>	<b>65-74</b>	<b>Muestra Total</b>
WAIS-R PIQa	194	0.69	0.8	0.73	0.74
WAIS-R VIQb	194	0.64	0.67	0.63	0.65
WAIS-R FSIQc	194	0.73	0.81	0.72	0.75
K-BIT Matrices	189	0.74	0.66	0.75	0.72
K_BIT Vocabulario	189	0.58	0.63	0.43	0.54
K-BIT C.I.	189	0.73	0.73	0.7	0.7

Nota: Todas las Correlaciones son significativas en  $p < .001$ . PIQa = C.I. Desempeño(Ejecutivo), VIQb = C.I. Verbal, FSIQc = Escala Completa

GAMA, WONDERLIC Y SHIPLEY. La relación entre el Wonderlic Wonderlic Personnel Test, 1992) y el Shipley (Shipley Institute of Living Scale, Zachary, 1991) fue examinada en un grupo de personas que tenían entre 25 a 45 años de edad (media =34.2, Ds=6.2). La muestra incluía a 80 individuos (46.3% hombres y 53.8% mujeres) reclutados del Noroeste (62.5%), Medioeste (20.0%) y del Sur (17.5%). La muestra incluía a 8.8% de Afroamericanos, 3.8% de Indios Americanos, 5.0% de Hispanos y Latinos, y 82.5% de Blancos. Estos individuos describieron su nivel educativo en 12.5% que completaron hasta el 12vo nivel educativo pero que estaban graduados, 28.8% de graduados de secundaria, 40% tenían preparatoria o vocacional y 18.8% contaban con nivel licenciatura o mayor.

Los resultados de este estudio mostró que el GAMA, el Wonderlic y el Shipley muestran calificaciones similares si se miden con la misma métrica ( una media de 100 y una Ds de 15). Adicionalmente se obtuvo una correlación significativa entre las calificaciones del C.I. GAMA y el test de vocabulario de Shipley (.56), Abstracción (.73) y en la calificación final (.72). La relación más fuerte entre las calificaciones del C.I. GAMA y las calificaciones del test de Abstracción de Shipley sugieren que los dos instrumentos miden un mismo constructo general no verbal.

La relación entre las calificaciones del C.I. GAMA y las del Wonderlic fueron altas (.70). Estos resultados indican que las calificaciones obtenidas del GAMA, el Wonderlic y el Shipley están relacionadas con la medición de la habilidad general, pero el GAMA mide la habilidad general utilizando un formato más específico (específicamente contenido no verbal) que no dependen de características de lenguaje y conocimiento académico.

<b>Medias, Desviación Estándar y Correlación entre el C.I. GAMA, Wonderlic y Shipley</b>					
	<b>Escala Original</b>		<b>Métrica con Media de 100 y Ds de 15</b>		
	Media	DS	Media	DS	<i>r</i>
Wonderlic	23.6	7.4	103.8	16.3	0.7
Shipley Vocabulario	48.1	9.4	97.2	14.1	0.56
Shipley Abstracción	56	8.5	109.1	12.8	0.73
Shipley total	53	9.1	104.4	13.7	0.72
C.I. GAMA	104.7	15	104.7	15	

Nota: Todas las correlaciones son significativas cuando  $p < .001$ . Las calificaciones del Wonderlic fueron convertidas a una métrica con media de 100 y una DS de 15 mediante la sustracción de la media normativa de 24.06 y dividiendo por medio de la DS de 7.12 de cada puntaje individual para obtener una calificación estándar. ( Las normas informativas fueron obtenidas del Manual del Wonderlic Personnel Test and Scholastic Level Exam user's Manual, 1992). Las calificaciones del Shipley fueron convertidas de las calificaciones T en forma similar. Para todas las variables la N= 80.

Finalmente se llevaron acabo una serie de estudios extras con sujetos de la misma muestra para estandarización y también población específica para este propósito (Estudiantes de preparatoria, Adultos con problemas de aprendizaje, adultos sordos, adultos con lesiones cerebrales traumáticas, adultos con retraso mental y ancianos). Se les administro el GAMA, el WAIS-R y el K-BIT, obteniendo calificaciones similares en los tres test, los cuales correlacionan significativamente con los subtest de las pruebas que miden lo mismo que el GAMA. La información arrojada de estos estudios soporta la validez del GAMA como un instrumento para medir la habilidad intelectual general.

## **CAPITULO IV "MÉTODO"**

### **4.1. PLANTEAMIENTO Y JUSTIFICACIÓN DEL PROBLEMA**

El objetivo de esta investigación es llevar a cabo la obtención de la validez y confiabilidad del instrumento llamado Escala de Habilidad General para Adultos (GAMA) utilizando población universitaria de escuelas públicas y privadas, dada la falta de trabajo sobre este instrumento, es conveniente que sea elaborado un trabajo que aporte datos sobre el mismo, para que así pueda ser aplicado en población mexicana, en base a este planteamiento se originan las siguientes preguntas de investigación.

### **4.2. PREGUNTAS DE INVESTIGACIÓN**

- ✓ **¿Qué tan sensible (poder discriminativo) será la versión utilizada de la Escala de Habilidades Generales para Adultos "GAMA" en los estudiantes universitarios de ambos sexos?**
- ✓ **¿Cuáles serán las características de Confiabilidad de la Escala de Habilidades Generales para Adultos "GAMA" al ser aplicado en una muestra de estudiantes universitarios?**
- ✓ **¿Cuáles serán las características de Validez de la Escala de Habilidades Generales para Adultos "GAMA" al ser aplicado en una muestra de estudiantes universitarios?**

### **4.3. HIPÓTESIS DE INVESTIGACIÓN**

Hipótesis de Trabajo: Existirán diferencias estadísticamente significativas entre los datos de confiabilidad y validez establecidas por la Escala de Habilidades

Generales para Adultos "GAMA", con los datos de sensibilidad, confiabilidad y validez establecidas en esta investigación.

#### **4.4. VARIABLES**

Variable Dependiente: Inteligencia, es decir los resultados obtenidos de la prueba.  
Variables atributivas: Sexo, edad y escolaridad de los sujetos.

##### **4.4.1. VARIABLE INDEPENDIENTE**

Es una muestra de estudiantes universitarios, de ambos sexos extraídos de las diferentes escuelas y facultades de la UNAM, así como de Universidades particulares (UIC y UVM).

##### **4.4.2. VARIABLE DEPENDIENTE**

Puntajes obtenidos en la aplicación de la Escala de Habilidad General para Adultos a estudiantes universitarios de las diferentes escuelas y facultades de la UNAM, así como de Universidad particulares (UIC y UVM).

#### **4.5. DEFINICIÓN CONCEPTUAL DE LA VARIABLE**

Inteligencia: Se define como la habilidad general, la cual incluye principalmente la enseñanza y la capacidad para establecer relaciones desde las más simples hasta las más complejas (Spearman, 1904,1923).

#### **4.6. DEFINICIÓN OPERACIONAL DE LA VARIABLE**

Inteligencia: Son los puntajes obtenidos en la aplicación la Escala de Habilidad General para Adultos a estudiantes universitarios.

#### **4.7. POBLACIÓN**

Los sujetos fueron 670 estudiantes universitarios (207 hombres y 463 mujeres), con edades entre 18 y 47 años, predominando la edad de 20 años entre los individuos a los que se les aplicó el instrumento. Su principal característica fue la de ser estudiantes pertenecientes al ciclo de licenciatura de las escuelas y facultades de Medicina, Odontología, Arquitectura, Ingeniería, Psicología y Administración de la UNAM en Ciudad Universitaria, así como de Universidad particulares (UIC y UVM) de las licenciaturas de Administración, Contabilidad, Comercio Internacional e Ingeniería.

#### **4.8. MUESTREO**

Es probabilístico ya que la muestra fue proporcional, seleccionada al azar de cada una de las escuelas y facultades de Medicina, Odontología, Arquitectura, Ingeniería, Psicología y Administración de la UNAM en Ciudad Universitaria, así como de Universidad particulares (UIC y UVM campus Sur) de las licenciaturas de Administración, Contabilidad, Comercio Internacional e Ingeniería. Dicha muestra se obtuvo a partir de la Agenda Estadística del año de 2002, que editó la UNAM.

#### **4.9. ESCENARIO**

El lugar donde se realizó la aplicación de los exámenes fue en los salones y aulas de clases, procurando un ambiente adecuado de silencio y sin distracciones con la ayuda del profesor respectivo.

#### **4.10. INSTRUMENTO**

La Escala de Habilidad General para Adultos es un instrumento psicométrico que puede aplicarse individual y colectivamente. Es un instrumento de poder y consta

de 66 reactivos que aumentan progresivamente su grado de dificultad, con un tiempo de aplicación de origen de 25 minutos. La Escala de Habilidades Generales para adultos debe ser utilizada por psicólogos entrenados o capacitados debidamente en conocimientos de psicometría, siendo necesario que esta sea aplicada, interpretada y calificada por los mismos. Su aplicación puede ser individual o colectiva. Cada examinado debe de contar una hoja de respuestas, manual de aplicación y un lápiz con goma. La versión utilizada para esta investigación es la de 1997.

#### **4.11. DISEÑO**

Dado el tipo de estudio (exposfacto), se utilizó un diseño de una muestra con una medición (aplicación de la forma única que compone el material), ya que el interés de esta investigación es obtener datos sobre confiabilidad y validez de la muestra investigada.

#### **4.12. TIPO DE ESTUDIO**

El tipo de estudio es: Exposfacto evaluativo, descriptivo de campo, ya que no se realiza manipulación de variables, solo existe la asignación de los sujetos al azar.

- Descriptivo, porque se describirán las características que presenten los estudiantes en el instrumento.
- De campo, ya que se hará la aplicación en escenarios naturales.



#### **4.13. PROCEDIMIENTO**

Para la aplicación de la Escala de Habilidades Generales para Adultos "GAMA", fue necesario obtener las muestras sobre la población de estudiantes de Ciudad Universitaria en base a la Agenda Estadística que edita la misma universidad, además de aplicar en una muestra significativa de las carreras de las Universidades UIC e UVM Campus Sur. Las muestras debían de tener como característica el que los estudiantes pertenecieran al ciclo de licenciatura, fueran de sexo masculino o femenino.

Una vez obtenidas las muestras, se giraron oficios a los respectivos secretarios generales de cada escuela y facultad, con el fin de obtener su consentimiento, y así, poder llevar a cabo la administración de la Escala de Habilidades Generales para Adultos "GAMA".

La aplicación del instrumento se hizo de manera colectiva dentro de los salones de clase, los cuales contaron con los siguientes materiales: un escritorio, un pizarrón y bancas, dichas aulas tenían buena luminosidad y ventilación. Durante la aplicación se utilizaron también cronómetros, lápices rojos y lápices con goma, hoja de respuestas y cuadernillos del instrumento de medición.

Los estudiantes de licenciatura pasaron a las aulas para la aplicación de la prueba (que se eligió por su fácil administración a nivel grupal), la cual duró los días requeridos para terminar con los grupos. Tales grupos fueron de entre 30 y 40 alumnos en cada sesión con dos personas por grupo como responsables de las aplicaciones.

Es importante señalar que el Manual de la Escala de Habilidades Generales para Adultos "GAMA" especifica que es una prueba contra tiempo, que mide habilidades generales, y que para efectos de obtener la validez y confiabilidad del instrumento

se llevo a cabo la aplicación completa de la prueba, tomando los tiempos de realización por sujeto.

En cada sesión se facilitó el cuadernillo de aplicación a cada estudiante, así como su hoja de respuesta, el responsable del grupo leyó junto con éste las instrucciones y se resolvieron los ejercicios de práctica. Una vez terminados se contestaron las dudas que tenían los evaluados para entonces comenzar con los ejercicios definitivos.

En esta investigación se necesito de una muestra con características similares en cuanto a nivel educativo, edad y sexo se refiere.

La muestra constó de 670 estudiantes (207 hombres y 463 mujeres) a los que se les administró el instrumento completo en una sola sesión. Durante la aplicación del instrumento se tomó en cuenta el tiempo que tardaron los estudiantes en contestarlo en su totalidad, respondiendo con lápiz gris hasta un termino de 25 minutos (los cuales son especificados en el manual original), continuando el resto de la prueba con lápiz rojo (para efectos de control de tiempo limite original y término de la prueba en la nueva población), así como se detectaron cuales reactivos fueron los que se presentaron confusos, poco entendibles o con un alto grado de dificultad para ser contestados.

Una vez finalizada la administración, se calificaron las pruebas para obtener las puntuaciones. Posteriormente, se codificaron los datos para llevar a cabo el análisis estadístico y así obtener los datos de consistencia interna, confiabilidad y validez, de la siguiente manera:

1. Análisis de Frecuencias, medidas de tendencia central y de dispersión.
2. Obtención del grado de sensibilidad de los reactivos.
3. Análisis de Confiabilidad (correlación Ítem-total, KR 20).

#### 4. Obtención de la validez de constructo (análisis factorial).

Cabe aclarar que la codificación de los datos de la muestra se hizo de la siguiente manera: para respuestas correctas se calificó con 1 y 0 para respuestas incorrectas también con el fin de facilitar los análisis estadísticos necesarios.

#### **4.14. PROCEDIMIENTO ESTADÍSTICO**

Inicialmente se llevo a cabo un análisis de frecuencias con el fin de saber como se distribuyeron las respuestas de los sujetos en todos los reactivos que componen el GAMA.

Además se llevó a cabo un análisis de consistencia interna del instrumento, utilizando para ella una correlación Ítem Total, con el fin de saber si los reactivos discriminaban o no; para obtener la confiabilidad se utilizó el método de Alpha de Cronbach para verificar la estabilidad de la misma; y por último para la validez de constructo del GAMA se llevo a cabo un análisis factorial de componentes principales con rotación varimax a fin de obtener la estructura interna del mismo.

Para el análisis estadístico se utilizó el programa estadístico SPSS versión 10.0 para Windows (Statistics Program for Socials Sciencies; Nie, N.Hull, C. Jenkins,2000)

## CAPITULO V "RESULTADOS"

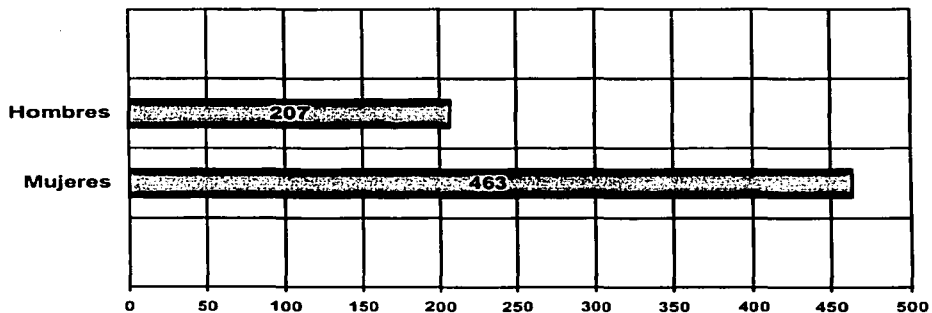
### 5.1. ANÁLISIS ESTADÍSTICO

Se llevó a cabo el siguiente análisis:

- Análisis de frecuencias (para determinar la distribución de los puntajes de los sujetos)
- Análisis de consistencia interna (para determinar la discriminación de los reactivos que componen el instrumento a través de una correlación ítem-total)
- Grado de estabilidad del instrumento con el Alpha de Crombach.
- Validez de constructo (determina que el instrumento mida lo que dice medir, a través de un análisis factorial de componentes principales con rotación varimax)

### 5.2. DESCRIPCIÓN DE LA MUESTRA

Para esta investigación se necesito de una muestra de 670 estudiantes (207 hombres y 463 mujeres) con características similares en cuanto a nivel educativo, edad y sexo.



### **5.3. PODER DISCRIMINATIVO Y CONFIABILIDAD**

Se llevo a cabo un análisis de los reactivos para determinar si estos discriminan adecuadamente mostrándose sensibles y que grado. Estos valores fueron obtenidos por la correlación ítem-total. Y para la confiabilidad se utilizó el método de Alpha de Cronbach que permite verificar la estabilidad del instrumento, Nunnally (1995) afirma que para que un instrumento se considere como confiable es necesario que esté cuenta con valores arriba de .85 de Alpha, y según Magnusson (1978) el criterio para que un instrumento sea confiable se ubica con valores superiores a .51.

Se obtuvieron los siguientes resultados:

1.- Los 66 reactivos que conforman la Escala de Habilidades Generales para Adultos "GAMA", fueron entendidos para la muestra de estudiantes universitarios de escuelas y facultades de Medicina, Odontología, Arquitectura, Ingeniería, Psicología y Administración de la UNAM en Ciudad Universitaria, así como de Universidad particulares (UIC y UVM campus Sur) de las licenciaturas de Administración, Contabilidad, Comercio Internacional e Ingeniería.

2.- El coeficiente de confiabilidad de la Escala de Habilidades Generales para Adultos "GAMA" es de **.8589**, lo que determina su utilidad como instrumento de medición, mostrándose este confiable, entendiendo que cuenta con un grado de precisión adecuado al momento de ser utilizado cuando pretenden medirse las habilidades.

Tabla 5.1

**TABLA DE CORRELACIÓN ITEM-TOTAL (N=670)**

ITEM	VALOR DE CORRELACIÓN	ALPHA
1	.2355	.8594
2	.2687	.8590
3	.2894	.8589
4	.2939	.8590
5	.2389	.8588
6	.2281	.8589
7	.2303	.8592
8	.2235	.8587
9	.3238	.8590
10	.2250	.8583
11	.3821	.8590
12	.2618	.8575
13	.3346	.8587
14	.3491	.8584
15	.2912	.8568
16	.2513	.8673
17	.3462	.8585
18	.2691	.8571
19	.2641	.8572
20	.3994	.8581
21	.3712	.8558
22	.3559	.8561
23	.2028	.8580
24	.2389	.8575
25	.2889	.8567
26	.2740	.8572
27	.3604	.8558
28	.3065	.8564
29	.3230	.8562
30	.3085	.8564
31	.3167	.8562
32	.3129	.8563
33	.2577	.8573
34	.2462	.8575
35	.3282	.8560
36	.3388	.8561
37	.3537	.8557
38	.2824	.8578
39	.4259	.8547
40	.2959	.8585
41	.3834	.8553
42	.4222	.8542
43	.3547	.8555
44	.3747	.8551
45	.4660	.8534
46	.4398	.8538
47	.3238	.8561
48	.3568	.8555
49	.4422	.8538
50	.3558	.8555
51	.3353	.8559
52	.2952	.8586
53	.3280	.8560
54	.3672	.8571
55	.3752	.8570
56	.2458	.8576
57	.4210	.8543
58	.3152	.8563
59	.3409	.8558
60	.2935	.8566
61	.3282	.8560
62	.2779	.8570
63	.2246	.8594
64	.2082	.8595
65	.2625	.8588
66	.2315	.8578

Alpha = .8589

El tiempo promedio para la administración del instrumento en población mexicana a partir de la muestra investigada resulto ser de **40 minutos** de aplicación. (Anexar explicación de que por las condiciones de la población mexicana contrasta con la muestra original)

Tabla 5.2

**TABLA DE CORRELACIÓN ITEM-TOTAL ESCALA DE IGUALDADES  
(N=670)**

ÍTEM	VALOR DE CORRELACIÓN	ALPHA
1	.3464	.7253
2	.4467	.7233
4	.3366	.7939
9	.4617	.7759
13	.3231	.7970
14	.4582	.7217
18	.3264	.7954
20	.3267	.7973
31	.2342	.7170

Alpha = **.7226**

Tabla 5.3

**TABLA DE CORRELACIÓN ITEM-TOTAL ESCALA DE ANALOGÍAS  
(N=670)**

ÍTEM	VALOR DE CORRELACIÓN	ALPHA
3	.3648	.6921
8	.3058	.6929
10	.2175	.6826
17	.4266	.6908
26	.3807	.6721
28	.3556	.6718
29	.3406	.6549
32	.3044	.6624
33	.3294	.6782
36	.4824	.6677
37	.3217	.6588
41	.3980	.6641
51	.3173	.6821
56	.3262	.6801
66	.3330	.6995

Alpha = **.6939**

Tabla 5.4

**TABLA DE CORRELACIÓN ITEM-TOTAL ESCALA DE CONSTRUCCIÓN  
(N=670)**

ÍTEM	VALOR DE CORRELACIÓN	ALPHA
5	.3194	.6345
6	.3994	.6328
7	.4204	.6247
11	.3679	.6331
16	.2946	.6232
30	.2967	.6088
40	.2627	.6300
42	.4061	.5905
43	.2993	.6082
45	.3949	.5929
46	.3950	.5922
48	.3325	.6026
52	.2038	.6239
58	.2979	.6096
59	.3063	.6070
63	.3449	.6300
64	.3110	.6340
65	.2554	.6289

Alpha = **.6318**

Tabla 5.5

**TABLA DE CORRELACIÓN ITEM-TOTAL ESC. DE SECUENCIAS (N=670)**

ÍTEM	VALOR DE CORRELACIÓN	ALPHA
12	.2687	.7336
15	.2952	.7305
21	.3645	.7261
22	.3257	.7288
23	.2168	.7355
25	.2467	.7336
27	.3048	.7295
34	.1936	.7388
39	.4446	.7188
44	.3394	.7263
47	.3188	.7263
49	.4227	.7183
50	.3073	.7294
53	.2858	.7313
54	.2862	.7311
55	.2368	.7360
57	.4054	.7202
60	.2752	.7315
51	.3174	.7285
62	.2815	.7315

Alpha = **.7395**



Las tablas anteriores presentan los datos de consistencia interna y confiabilidad por escala individual que forma la prueba de GAMA, se puede observar que en las cuatro tablas el coeficiente de correlación ítem total en todos y cada uno de los reactivos tuvo valores superiores a .20 positivo, demostrando esto, que los reactivos son sensibles y discriminan adecuadamente.

En lo que respecta a la confiabilidad por escalas, esta se presenta de manera moderada en los cuatro componentes del GAMA con valores que van del .60 al .75, esto desde el punto de vista de Magnusson (1985) las escalas son confiables ya que cualquier instrumento que presente una confiabilidad superior a .51 positivo es considerado como consistente.

#### **5.4. VALIDEZ**

El cálculo de la validez se obtuvo a través de un análisis factorial, de componentes principales, con rotación varimax, teniendo como fin observar como se agrupan los reactivos que componen el GAMA y verificar si la estructura es diferente a la original. Los resultados de esta investigación mostraron 5 factores con valores Eigen superiores a 1.0, que explican 41% de la varianza total del instrumento. Como factores residuales se obtuvieron 5 más con un mínimo de cargas factoriales adecuadas. En el primer factor (Habilidad General) se ubicaron los 66 reactivos, en el segundo factor (Analogías) 17 reactivos, en el tercer factor reactivos (Igualdades) 11, en el cuarto factor (Construcción) 18 reactivos y en quinto factor (Secuencias) 20 reactivos.

Tabla 5.6

**ANÁLISIS FACTORIAL**  
**VALORES EIGEN**  
**EXTRACCIÓN: COMPONENTES PRINCIPALES**

<b>VALOR</b>	<b>VALOR EIGEN</b>	<b>PORCENTAJE DE VARIANZA</b>	<b>VARIANZA ACUMULADA</b>
1	7.029	20.650	20.650
2	2.281	8.455	29.105
3	1.949	5.953	35.058
4	1.870	3.853	38.911
5	1.756	2.661	41.572

Tabla 5.7 RESULTADOS DEL ANÁLISIS FACTORIAL PARA LA VALIDEZ

VARIABLE	FACTOR 1	FACTOR 2 (Analogías)	FACTOR 3 (Igualdades)	FACTOR 4 (Construcción)	FACTOR 5 (Secuencias)
R1	.366		.372		
R2	.720		.513		
R3	.312	.371			
R4	.310		.503		
R5	.310			.636	
R6	.414			.494	
R7	.321			.388	
R8	.415	.433			
R9	.313		.353		
R10	.320	.393			
R11	.310			-.312	
R12	.317				.347
R13	.314		.409		
R14	.319		.375		
R15	.340				.443
R16	.301			-.329	
R17	.419	.323			
R18	.326		.472		
R19	.323	.315			
R20	.243		-.525		
R21	.445				.371
R22	.419				.314
R23	.326				.305
R24	.300	-.343			
R25	.347				
R26	.331	-.332			
R27	.417				-.312
R28	.357	.365			
R29	.374	.387			
R30	.344			.313	
R31	.348		.340		
R32	.357	.314			
R33	.304	-.325			
R34	.428				.311
R35	.358		-.341		
R36	.386	-.321			
R37	.424				
R38	.339		-.376		
R39	.495				.320
R40	.323			-.324	
R41	.445	.359			
R42	.472			.313	
R43	.389			.743	
R44	.432				-.323
R45	.523			.377	
R46	.506			.719	
R47	.378				-.300
R48	.404			.320	
R49	.499				.335
R50	.400				.394
R51	.390	-.317			
R52	.322			-.303	
R53	.367				-.320
R54	.311				.387
R55	.327				-.325
R56	.302	-.388			
R57	.476				.316
R58	.345			.318	
R59	.367			.312	
R60	.335				-.327
R61	.366				.318
R62	.311				-.450
R63	.313			.360	
R64	.312			.319	
R65	.318			.357	
R66	.424	.301			

## **5.5. DISCUSIÓN**

Como en todo método, uno de los aspectos más importantes es poder llegar a obtener el resultado planeado para poder evaluarlo, analizarlo, criticarlo y en un momento dado, establecer parámetros que permitan modificar, adicionando o restringiendo las conclusiones obtenidas.

En este caso, hubieron de llevarse a cabo el procesamiento de datos arrojados por la aplicación del GAMA para analizarlos estadísticamente y obtener la confiabilidad y validez del mismo.

Lo que nos arrojó lo siguiente:

Es una prueba que se puede considerar "culturalmente imparcial", ya que considera una carga verbal mínima, trata con figuras abstractas, las respuestas tienen que ser escritas numéricamente, es una prueba ejecutiva, no verbal y utiliza la solución de problemas novedosos.

Hay que agregar que es un instrumento que supera ciertas limitaciones que surgen en la aplicación de pruebas como es el WAIS-R, ya que el GAMA utiliza el razonamiento y la lógica para resolver problemas exclusivamente de formas y diseños abstractos. Esta conclusión se apoya en que el requerimiento de un adecuado conocimiento del lenguaje (receptivo y expresivo) es indispensable para la evaluación exitosa en pruebas verbales tales como los subtests de Wechsler (Kaufman, 1990, 1994). Los subtests no verbales en el WAIS-R suponen resolver este problema, sin embargo siguen teniendo limitaciones. Por ejemplo, este necesita habilidades motoras finas para la manipulación de los cubos para igualar o diseñar una figura, así como para armar un rompecabezas. Adicionalmente, se requiere rapidez en la acción para puntuar alto.

A diferencia de muchos instrumentos las ventajas que tiene el GAMA como su aplicación individual, colectiva o auto administrada, así como las antes mencionadas brinda la cualidad de poderla aplicar como parte de una evaluación psicológica, en asesoría terapéutica, para consejería vocacional, para el ámbito industrial y de negocios, para investigaciones a gran escala o en colegios y escuelas, y como parte de una batería para diagnóstico neuropsicológico. Además que con una adecuada supervisión e instrucción para la calificación permite obtener calificaciones válidas.

Particularmente, el GAMA nos brinda una alternativa económica en tiempo y costo en el proceso de reclutamiento y selección, ya que permite una evaluación acertada de las habilidades generales de los candidatos a un puesto.

## **5.6. CONCLUSIONES Y APORTACIONES**

Podemos concluir con esta investigación que la aplicación de la Escala de Habilidades Generales para Adultos (GAMA) confirma que es un instrumento válido y confiable para la población mexicana, ya obtuvo una Alpha de . 8589. mismo que puede ser utilizado para investigaciones posteriores.

Hay que considerar que a pesar de la lentitud de las respuestas de los sujetos evaluados, como afirma Díaz Guerrero 1971, citado por Varela (1996) "el psicólogo en desarrollo debe pues, dedicar su atención en su propia cultura, a semejanza de las construcciones empíricas desarrolladas en la cultura angloamericana deberá escudriñar las propias maneras de ser de su propia gente y desarrollar conceptos que les vengán a sus naturalezas específicas e idiosincráticas"; y Díaz Guerrero (1975) que refiere a que los mexicanos son más reflexivos a diferencia de los estadounidenses (se muestran más reactivos ante

estímulos similares), y aunque el tiempo de aplicación se prolongó de 25 a 40 minutos, se observó que es un instrumento sensible, válido y confiable.

Por otro lado, el instrumento resultó novedoso para la muestra estudiada, ya que por lo general se han administrado por años instrumentos como el Beta IIR, el WAIS español o el Raven, los cuales son conocidos ampliamente por la población mexicana.

Hay que agregar que la muestra seleccionada para la presente investigación, fue aleatoria, de escuelas privadas y públicas de nivel licenciatura, en donde se observó que en las escuelas privadas mantenían una reactividad mayor que en las escuelas públicas de educación superior en donde fue aplicado el instrumento. Anotando que sería interesante ampliar el estudio en muestras específicas como en el estudio piloto de Naglieri, en donde involucra muestras con las siguientes variables como raza, idioma, zona geográfica y nivel socioeconómico de la muestra.

Este tipo de estudios favorece la tesis de Varela, Villegas y Arce (1998), que propone la elaboración de estudios de validez y confiabilidad para los instrumentos utilizados en población mexicana, ya que el 95% de los test aplicados en México son extranjeros, mismas que carecen de este tipo de investigaciones.

Al haber realizado este estudio se sugieren nuevas líneas de investigación sobre el mismo instrumento, ya que se debe de estandarizar y normalizar para una población o muestras poblacionales específicas.

Como ya lo he mencionado, las consecuencias de utilizar instrumentos que no están analizados con el debido rigor metodológico, puede acarrear consecuencias que pueden afectar directamente a las personas que son evaluadas con los mismos, por tanto, lo ideal sería que los mexicanos adquirieran una cultura a este respecto, que permita darle la seriedad y asumir la responsabilidad del uso de los instrumentos psicométricos con el único propósito de hacer bien las cosas, que es

la predicción confiable y válida de la conducta. Evidentemente esta responsabilidad es únicamente de los profesionales de la conducta, por lo que es de nuestra competencia llevar a cabo estas investigaciones de forma concienzuda, considerando que es una de las herramientas cotidianas en el trabajo de un psicólogo.

Considero las siguientes aportaciones:

1.- Esta investigación proporciona a la comunidad de psicólogos mexicanos un instrumento nuevo, que viene a renovar el acervo viejo y con errores metodológicos graves que es de uso cotidiano en la psicología nacional, sería conveniente para ello revisar la lista de test psicológicos de medición de habilidades que distribuye la Librería Internacional.

2.- Por otro lado, apoya la línea de investigación de Varela y Villegas que trata de contar con instrumentos novedosos que están actualizando a nivel internacional, logrando poner a la psicometría nacional al mismo nivel de utilización de instrumentos de Estados Unidos y el Reino Unido

3.- Es un instrumento que permite una evaluación rápida de inteligencia con un alto margen de confiabilidad y validez, y que no solamente mide habilidades generales sino cuatro componentes adicionales que contribuyen a una mejor predicción de conducta al momento de resolver problemas.

4.- Tener una prueba como el GAMA permitirá una medición certera de la habilidades en los diversos campos de la psicología, ya sea el clínico, el laboral, el educativo, además del uso del instrumento para la investigación en población mexicana y sus diferencias en otras culturas.

5.- Se aumento el acervo de instrumentos que son susceptibles de ser utilizados en población mexicana y que cuentan con calificaciones validas y confiables.

Realizaré estudios adicionales a partir de este instrumento como son:

- Realizaré estudios de estabilidad temporal aplicado en diferentes muestras poblacionales.
- Elaboraré estudios de validez concurrente utilizando el GAMA con otros instrumentos estandarizados que midan también habilidades como son el WAIS adaptado (Villegas y Varela, 2000) o bien el BETA III (estandarizado de Terrazas, Calas, Villegas y Varela) o el test de personalidad Wonderlic estandarizado de Arce, Arce, Villegas y Varela.
- Realizaré estudios normativos con diferentes niveles de edad, grado escolar, perfil laboral, etc. A fin de lograr una mejor y más calificaciones de sujetos mexicanos a partir de la utilización del GAMA.
- Realizaré estudios en diferentes zonas geográficas de la República Mexicana.
- Aplicaré el GAMA en sujetos analfabetas a fin de ver su utilización en este tipo de sujetos de bajo nivel de estudios.
- Realizaré estudios sobre la validez predictiva del GAMA al ser utilizado en consejería vocacional con sujetos que pretendan elegir carreras relacionadas con arquitectura, ingeniería, diseño industrial; así mismo realizaré estudios también predictivos a fin de observar si el GAMA puede servir como instrumento predictor de la capacitación en sujetos que sean arquitectos, ingenieros, diseñadores industriales, etc.
- Obtendré calificaciones normalizadas para población mexicana con nivel escolar básico, medio y universitario, con rangos de edad de 18 a 60 años con el fin de que sea una norma aplicable a personas en edad laboral de los tres niveles



- Evaluaré los instrumentos con las características de que en su aplicación se utilice el razonamiento y lógica para resolver los problemas que exclusivamente utilizan diseños abstractos y figuras, con el fin de tropicalizar los instrumentos novedosos a México.

Cabe señalar como excepción que en nuestro país durante muchos años se han utilizado los diferentes formatos de la prueba de matrices progresivas de Raven y la prueba del Banco de México, sin embargo no se conoce ninguna publicación en revistas arbitradas sobre las características psicométricas de este instrumento al ser administrado en población mexicana. El uso de la prueba de GAMA, abre toda una línea de investigación para obtener datos actuales sobre como la población mexicana resuelve problemas, y si es viable utilizar instrumentos de características no verbales de diseños geométricos para medir el constructo de inteligencia (habilidades).

## **REFERENCIAS DOCUMENTALES**

Agnoff, W.H. (1988<sup>a</sup>) *Validity: An evolving concept*. En. H. Wainer & H. Braun (EDS.) *Test Validity* Hillsdale, N.J: Lawrence Erlbaum

Agnoff, W.H. (1988<sup>b</sup>) *Proposals for theoretical and applied development in measurement*. *Applied Measurement in Education* Hillsdale, N.J: Lawrence Erlbaum

Aiken, I.R. (2003) *Psychological testing and assessment*, Boston: Allyn and Bacon

American Psychological Association (1954) *Technical Recommendations for Psychological Tests and Diagnostic Techniques* Washington. D.C: Author.

American Psychological Association (1966) *Standars for Educational and Psychological Tests and Manuals* Washington. D.C: Author.

American Psychological Association (APA) American Educational Research Association (AERA) and Nacional Council on Measurement in Education (NCME). (1985) *Standards for Educational and Psychological Testing* Washington. D.C: APA

American Psychological Association (APA) American Educational Research Association (AERA) and Nacional Council on Measurement in Education (NCME). (1974) *Satandards for Educational and Psychological Tests and Manuals* Washington. D.C: APA

Airen, L. R. (1996) *Tests Psicológicos y Evaluación*. México, Prentice may.

- Anastasi, A. (1998) Test Psicológicos. México, Prentice Hall, 7ª Edición.
- Arce, R. D. y Arce, R. V. (2001) Estandarización del Examen de Personal Wonderlic (WPT, Wonderlic Personnel Test) en población universitaria. Tesis de Licenciatura, Facultad de Psicología, UNAM.
- Arias, F. (1978) Desarrollo de Recursos Humanos, México, Diana, p.p. 179 - 181.
- Butcher, H. La Inteligencia Humana. España, Marova, p.p. 40 - 42.
- Díaz Guerrero, R. (1971) La enseñanza de la investigación en psicología en Iberoamérica: un paradigma. Revista latinoamericana de psicología, Vol.3, No. 1, p.p. 5-33
- Díaz Guerrero, R. y cols. (1975) El desarrollo de la personalidad en dos culturas: México y Estados Unidos. México: Trillas
- Cronbach, L.J. & Meehl, P. E. (1955) \_Construct Validity in Psychological Tests. Psychological Bulletin, Vol. 52, p.p. 281-302.
- Cronbach, L.J. (1984) Essentials of Psychological Testing 4ª ed. New York: Harper & Row.
- Cronbach, L.J. (1988) Five Perspective on Validation Argument. En. H. Wainer & H. Braun (Eds.). Test Validity Hillsdale, N.J: Lawrence Erlbaum
- Darley, J. (1990) Psicología México, Prentice Hall, p.p. 364.
- Geinsinger K. (1992) Psychological Testing of Hispanics US: APA

Gregory, R.J. (2001) Evaluación psicológica. Historia, principios y aplicaciones. México, Manual Moderno.

Grace, J. (1997) Desarrollo psicológico, México, Prentice-Hall, p.p. 354-355.

Herrans L. (2000) Psicología y Medición México: McGraw Hill

Howard, C. (1998) Diccionario de Psicología, México, Fondo de Cultura Económica.

López, J. (1989) El Artefacto de la Inteligencia: una reflexión crítica sobre el determinismo biológico de la inteligencia. Barcelona: Anthrapos.

Martínez, A. (1996) Psicometría: Teoría de los Test Psicológicos y Educativos México: Síntesis Psicología.

Merani, A. (1976) Diccionario de Psicología, México, Grijalbo

Muñiz, J. (1996) Psicometría Madrid: Editorial Universitas, S.A.

Naglieri, J. Y Bardos, A. GAMA, Genreal Ability Measure for Adults Manual Mineapolis, National Computer Systems, Inc.

Nash, M. (1985) Como Incrementar la Productividad del Recurso Humano Colombia, Editorial Normal, p.p. 56-57.

Pueyo, A. (1996) Inteligencia y Cognición. Barcelona: Ed. Paidós.

Varela, D.R. (1996) Confiabilidad y Validez del Cornell Index FN2. Tesis de Licenciatura. Facultad de Psicología, UNAM.

**Vernon, E. (1982) Inteligencia, Herencia y Ambiente. México: Ed. Manual Moderno**

El libro de Vernon, E. (1982) Inteligencia, Herencia y Ambiente, México: Ed. Manual Moderno, es una obra que aborda el tema de la inteligencia y su relación con la herencia y el ambiente. El autor, un psicólogo británico, propone un modelo de la inteligencia que se basa en la distinción entre la inteligencia fluida y la inteligencia cristalizada. La inteligencia fluida es aquella que se adquiere a través de la experiencia y el aprendizaje, mientras que la inteligencia cristalizada es aquella que se adquiere a través de la educación y la cultura. Vernon argumenta que la inteligencia fluida es más susceptible a los efectos del ambiente, mientras que la inteligencia cristalizada es más susceptible a los efectos de la herencia. Este modelo ha sido ampliamente aceptado y ha influido en el desarrollo de teorías posteriores de la inteligencia, como la de Spearman y la de Cattell. El libro es una obra fundamental para el estudio de la inteligencia y su relación con la herencia y el ambiente.

## APENDICE

CONTRIBUCIÓN DE GRIEGOS

# Gama

ALFONSO MEXICANA

Jack A. Naglieri, PhD, and Achilles N. Bardas, PhD

TESIS CON  
FALLA DE ORIGEN



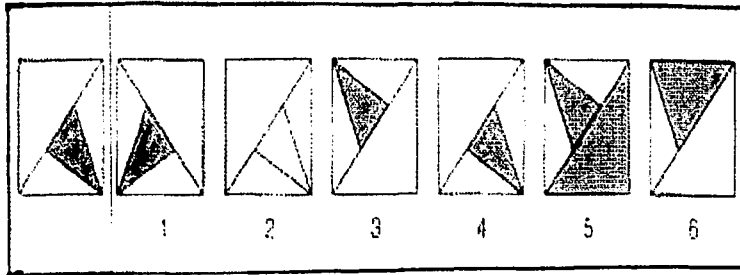
### Instrucciones de ejemplo

Este cuadernillo tiene cuatro tipos de problemas para que los resuelvas. Ejemplos de estos problemas son mostrados en las siguientes páginas. Lee las instrucciones arriba de cada ejemplo y después resuelve el problema. La respuesta correcta esta dada debajo de cada ejemplo. Trata de resolver cada ejemplo cuidadosamente.

Mira el siguiente ejemplo:

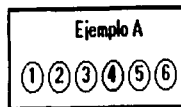
Ejemplo A:

Que respuesta (1,2,3,4,5 ó 6) es igual a la de la primera figura?



La número cuatro es la respuesta correcta porque es la misma que en la primera figura.

Ahora encuentra el Ejemplo A en tu Hoja de Respuestas. Rellena el círculo con el número cuatro como se muestra a continuación.



Si cometes un error, borra la respuesta incorrecta y rellena la nueva respuesta.

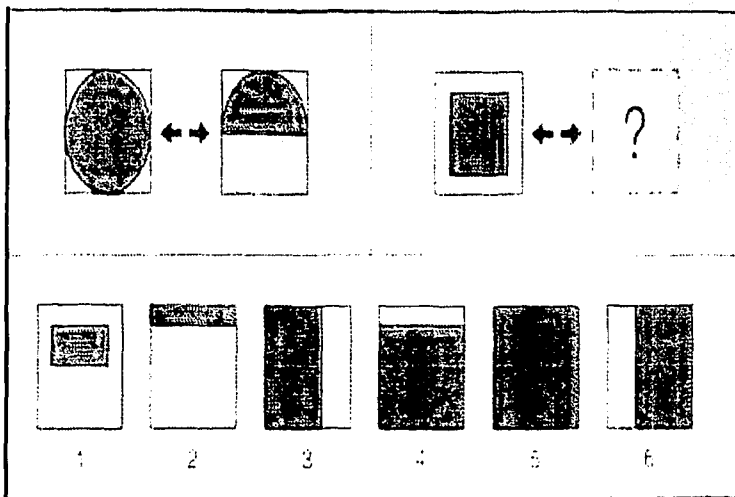
Continúa en la siguiente página. →

TESIS CON  
FALLA DE ORIGEN



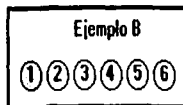
**Ejemplo B:**

Que respuesta (1,2,3,4,5 ó 6) va sobre la incógnita?



La número uno es la respuesta correcta porque es la que completa el patrón.

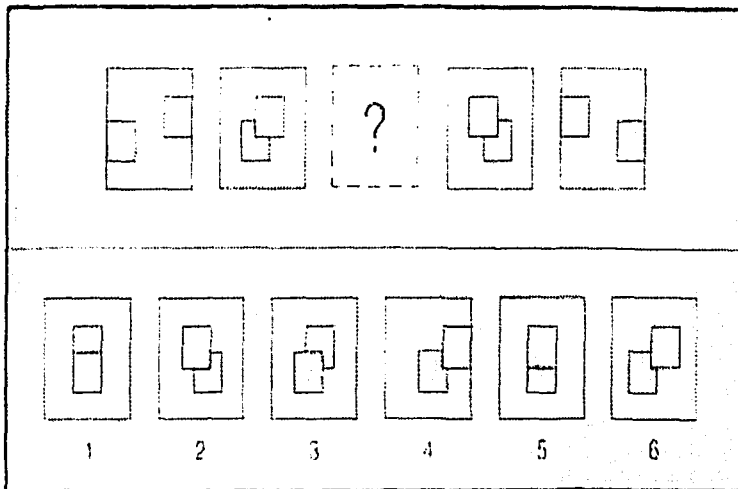
Ahora encuentra el Ejemplo B en tu Hoja de Respuestas. Rellena el círculo con el número uno como se muestra debajo.



Continúa en la siguiente página. →

**Ejemplo C:**

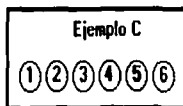
Que respuesta (1,2,3,4,5 ó 6) va sobre la incógnita que completa el patrón?



La número cinco es la respuesta correcta porque es la que completa el patrón.

Ahora encuentra el Ejemplo C en tu Hoja de Respuestas. Rellena el círculo con el número cinco como se muestra debajo.

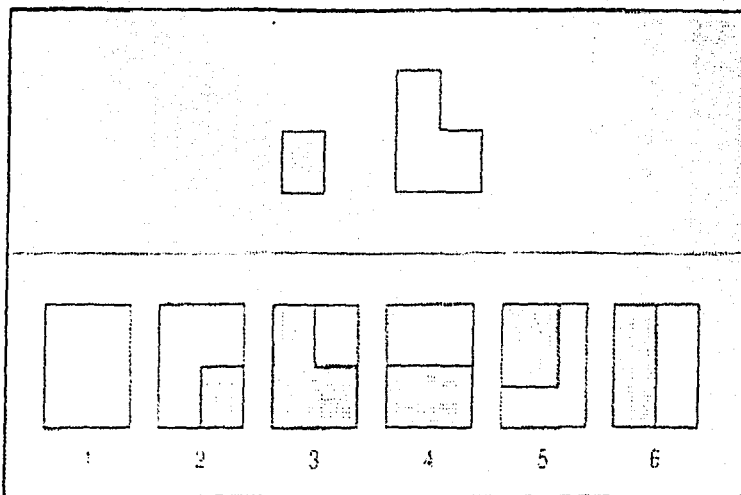
TESIS CON  
FALLA DE ORIGEN



Continúa en la siguiente página. →

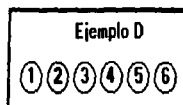
**Ejemplo D:**

Que respuesta (1,2,3,4,5 ó 6) puede ser realizada con las formas en la caja superior?



La número dos es la respuesta correcta porque esta fue hecha con las formas de la caja superior.

Ahora encuentra el Ejemplo D en tu Hoja de Respuestas. Rellena el círculo con el número dos como se muestra debajo.



Continúa en la siguiente página. →

## Instrucciones

No escribas sobre este cuadernillo. Marca tus respuestas en tu Hoja de Respuestas. Si tu cometes un error, borra la respuesta incorrecta y rellena una nueva. Recuerda, existe solamente una respuesta correcta para cada problema. Tu podrás regresar a estas instrucciones y ejemplos cada que lo necesites.

- Deberás trabajar tan cuidadosamente como puedas sobre estos problemas.
- No pierdas mucho tiempo en un problema, pero trata de contestar cada uno.
- Realiza tantos problemas como puedas.

Tienes alguna pregunta?



No des vuelta a esta pagina hasta que se te indique.

Continúa en la siguiente página. →

TESIS CON  
FALLA DE ORIGEN

1. Que respuesta (1,2,3,4,5 ó 6) es la misma a la primera figura?

