



UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO

FACULTAD DE PSICOLOGIA

"CONFIABILIDAD Y VALIDEZ DE LA APLICACION
AUTOMATIZADA DE PRUEBAS PSICOMETRICAS:
INVENTARIO DE INTERESES PROFESIONALES
DE HEREFORD"

T E S I S

QUE PARA OBTENER EL TITULO DE
LICENCIADO EN PSICOLOGIA
P R E S E N T A
HILDA ELENA ESQUIVEL GUILLEN

DIRECTORA DE TESIS: MTRA. CRISTINA HEREDIA ANCONA

MEXICO, D. F.

1995

FALLA DE ORIGEN



Universidad Nacional
Autónoma de México



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 está 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.

Dedico este trabajo:

A mi papá:

*Porque el amor que me brindaste
me permitió llegar a formar los valores que ahora tengo,
por respetar mi decisión profesional,
por el amor que día a día me demuestras,
pero sobre todo porque eres lo que más quiero en esta vida.*

A mi mamá:

*Quien siempre ha estado ahí, apoyándome, consintiéndome,
pero sobre todo queriéndome, mil gracias mamá.*

A Jorgito, mi sobrino:

*El mejor tesoro que tengo en esta vida,
al niño más precioso, con todo mi amor.*

A mis hermanos:

*Alex por la confianza y el respeto que siempre me has brindado,
porque eres el más peque...
y sobre todo por tu cariño.*

**A Hugo por el amor y la paciencia para aguantarme
en los momentos más difíciles (y en los no tan difíciles),
por siempre tener una sonrisa de aliento.**

**A Jorge con mucho cariño,
porque hemos compartido los momentos agradables y
desagradables que esta vida nos ha dado.**

**A mi tío "Peco" por el apoyo brindado
en la conclusión de este trabajo.**

A mi familia, porque a todos ustedes los QUIERO MUCHO !!!

A Jorge Villatoro, porque el estar a tu lado me ha mostrado muchas de las mejores cosas de esta vida, tu compañía e incondicional apoyo me ha permitido salir siempre adelante, por que Te amo

A Laura Loera, por permitirme conocer las posibilidades que me brinda esta vida, y alcanzar mis metas, por tu apoyo infinito gracias.

A Laura Ramos, porque muestra amistad ha ido creciendo día con día, año con año, gracias por estar siempre cuando te necesito.

A Dulce, tu amistad es uno de los mayores tesoros que he recibido, mil gracias a ti y a tus papás, quienes me permitieron ser parte de su familia, por todo lo que hemos convivido, compartido, conocido, etc. ¡Te quiero muchísimo!

Carmen, no encontraría las palabras que describieran todo el agradecimiento y cariño que siento por ti, sólo puedo decir: ¡Gracias por tu amistad!

A ti a quien hace más de seis años has iluminado mi vida, has estado a mi lado (con tus ocasionales desapariciones) regalándome momentos increíbles, apoyándome, escuchándome, aguantándome, por permitirme compartir infinidad de cosas. Con todo mi amor.

**Lulú por todo lo que compartimos
porque siempre estás cuando te necesito,
por la fortaleza que me brindas en todo momento,
con mucho cariño.**

**Paco, gracias por tu bondad y tu enorme corazón,
por tu excepcional amistad.**

**A mis amigos, que hicieron de la carrera
una de las mejores experiencias en mi vida,
a Dulce, Angy, Erika, Carlitos, Paty, Robertits,
Lilla, Diana, Sol, Cynthia, Enrique, Adriana, Betzabel y Silvia.**

**Al grupo de los cuatro, Mily, Alberto y Javier,
por acompañarme en este proceso y
hacerlo tan divertido, gracias por su apoyo y amistad.**

**A Javier Castañeda por tu valiosa amistad, compañía
y por todo lo que hemos compartido.**

AGRADECIMIENTOS:

**A Samuel Jurado, gracias porque la amistad que me brindas es invaluable, tu apoyo, confianza, enseñanzas y asesoría me han enseñado a crecer.
¡¡Gracias por creer en mí, Sam!!**

**A la Mtra. Cristina Heredia, porque gracias a su apoyo incondicional fue posible la finalización de uno de mis principales proyectos de vida, pero sobre todo gracias por la confianza y el cariño que demostró tener en mí.
Con mucho cariño.**

A Carmen Maldonado Navarro y Cuittlahuac Pérez López, por su valiosa amistad y colaboración en el desarrollo de la automatización del Inventario de Intereses Profesionales de Hereford.

A Lourdes Monroy, Miguel Angel Guevara, Benito Ramirez, a la Unidad de Cómputo Informática e Instrumentación de la Facultad de Psicología, por las facilidades brindadas en el desarrollo de esta investigación.

A Rubén Varela porque gracias a su invaluable apoyo fue posible hacer de este trabajo lo que es.

A la Lic. Enedina Villegas por su interés, apoyo, asesoría y facilidades para lograr este trabajo.

***A Rubén Varela, las maestras Enedina Villegas
y Estela Cordero por permitirme acceder a la información,
de su valiosa investigación.***

***A la Dra. Isabel Reyes L. y al Dr. Rogelio Díaz Guerrero
por la información brindada para este trabajo.***

***A los maestros: Carmen Maldonado,
Germán Álvarez, Cristina Heredia,
Fernando Vázquez, Rubén Varela
y Georgina Lozano, pero sobre todo,
a los alumnos que hicieron posible reunir la muestra.***

***A Fernando, Sam, Jorge, Javier,
Sol, Angy y Mily por su valiosa ayuda
en la traducción de algunos de los artículos.***

***A la Lic. Esther Mendoza y María Calzada
por su valiosa amistad y su apoyo
incodicional en el desarrollo de
este trabajo y siempre.***

***A los sinodales:
Mtra. Cristina Heredia Ancona,
Mtro. Samuel Jurado Cárdenas,
Lic. Enedina Villegas Hernández,
Mtra. Fayne Esquivel Ancona y
Lic. Araceli Mendoza Ibarrola
por su valioso tiempo, dedicación y sugerencias
en la revisión de este trabajo.***

Hilda Elena Esquivel Guillén

ÍNDICE

INTRODUCCIÓN	3
CAPÍTULO 1 PRUEBAS PSICOLÓGICAS	
1.1. Historia de la medición en psicología	5
1.2. Definición de las pruebas psicológicas	13
1.3. Características de las pruebas psicológicas	14
1.3.1. Grado de dificultad y poder discriminativo de los reactivos	14
1.3.2. Estandarización	15
1.3.3. Confiabilidad	16
1.3.3.1. Método test-retest	19
1.3.3.2. Método de confiabilidad con formas equivalentes o paralelas	19
1.3.3.3. Método de confiabilidad de división por mitades	20
1.3.3.4. Método de confiabilidad de Kurder Richardson	20
1.3.4. Validez	20
1.3.4.1. Validez de contenido	21
1.3.4.2. Validez empírica o de criterio	23
1.3.4.3. Validez de construcción o estructural	25
1.3.5. Normalización	29
1.4. Clasificación de las pruebas psicológicas	30
1.5. Influencia del examinador en cuanto a la aplicación	33
CAPÍTULO 2 INTERESES	
2.1. Definición de los intereses	39
2.2. Clasificación de los intereses	40
2.3. Factores que determinan los intereses	46
2.4. Medición de los intereses	47
2.5. Instrumentos para evaluar intereses	48

CAPÍTULO 3 AUTOMATIZACIÓN DE PRUEBAS PSICOLÓGICAS

3.1. Definición de la automatización de pruebas psicológicas	54
3.2. Características de una computadora	55
3.3. Estudios realizados	57
3.4. Recomendaciones para elaborar una prueba automatizada	67
3.5. Ventajas	70
3.6. Desventajas	73

CAPÍTULO 4 MÉTODO

4.1. Pregunta de investigación	75
4.2. Hipótesis conceptual	75
4.3. Hipótesis de trabajo	75
4.4. Hipótesis nula	75
4.5. Definición de variables	75
4.6. Población	76
4.7. Sujetos	76
4.8. Muestreo	76
4.9. Diseño	77
4.10. Instrumento y materiales	77
4.11. Procedimiento	79

CAPÍTULO 5 RESULTADOS	81
------------------------------	-----------

CAPÍTULO 6 DISCUSIÓN Y CONCLUSIONES	89
--	-----------

ANEXO	93
--------------	-----------

REFERENCIAS BIBLIOGRÁFICAS	99
-----------------------------------	-----------

INTRODUCCIÓN

INTRODUCCIÓN

Actualmente dentro del campo de la psicología, se ha introducido el uso de las computadoras como una herramienta más para el psicólogo, teniendo éstas diversas aplicaciones comunes, como el uso de procesadores de texto, análisis y graficación de datos. Pero en el campo de la psicología aplicada existen aplicaciones de mayor relación con el trabajo del psicólogo como lo son el procesamiento digital de señales electroencefalográficas, el análisis automatizado de potenciales evocados en humanos (Guevara, y col, 1993), la elaboración de material educativo, mejor conocido como "software educativo" y la automatización de pruebas psicométricas, (Beaumont, 1981; Jäger, 1991; Murphy, 1994; Tallent 1987; Vale 1981; Weiss y Vale 1987); así como su calificación, (Acevedo, M, y Zarabozo, D. 1993) y la interpretación de calificaciones.

Esta investigación está dirigida principalmente a demostrar que un elemento tan útil, como lo es la computadora en diversas áreas, lo puede ser de igual manera en el área de la psicología, en este caso, en las áreas clínica e industrial, donde la computadora permite facilitar el proceso de aplicación y calificación de pruebas psicométricas, empleando el menor tiempo posible y facilitando de gran manera su forma de aplicación y calificación; a este proceso se le denomina "*Automatización de Pruebas Psicométricas*". Este proceso ayuda a evitar errores en el momento de calificar y convertir puntajes brutos a percentiles, permitiendo además una objetividad y exactitud en la calificación; es importante señalar que esto es sólo una herramienta de trabajo, como muchas otras que utiliza el psicólogo, por lo tanto no pretende de ninguna manera substituir el trabajo del psicólogo, ya que se requiere de éste para las demás valoraciones de los sujetos, así como para la interpretación profesional de las pruebas psicométricas.

El proceso de automatización de pruebas consiste en diseñar programas por medio de la computadora que permitan presentar en la pantalla los reactivos y opciones de respuesta de determinada prueba psicométrica, para que el sujeto pueda contestar directamente en la computadora. Así mismo estos programas permiten calificar, graficar y almacenar los resultados obtenidos por uno y/o varios sujetos, así como brindar un diagnóstico en un tiempo mucho más rápido, de esta manera se facilita el trabajo del psicólogo, obteniendo los mismos resultados que en una aplicación de forma tradicional, pero en un menor tiempo.

Por lo anterior el propósito de esta investigación consistió en demostrar que una aplicación automatizada de una prueba psicométrica ofrece resultados confiables y válidos, como los que ofrece una aplicación de forma tradicional, para lo cual se empleó el Inventario de Intereses Profesionales de Hereford en su forma tradicional (lápiz-papel) y en la forma automatizada (por medio de la computadora).

PRUEBAS PSICOLÓGICAS

1.1. Historia de la medición en psicología

El origen de las pruebas psicológicas surgió debido al interés por conocer las diferencias humanas (Nunnally, 1970; Pichot, 1976; Anastasi, 1978; Cerdá, 1984).

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

Estas investigaciones tuvieron gran importancia dentro de la psicología ya que llevaron a la creación de las pruebas o test psicológicos, que hasta la fecha permiten conocer diversas características de los seres humanos.

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

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

CAPÍTULO 1

PRUEBAS PSICOLÓGICAS

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

Durante ese mismo siglo Ernest Weber, en el año de 1840, publica los resultados de sus trabajos sobre relaciones entre estímulos y respuestas sensoriales donde postula diversas leyes, entre las cuales destacan las del "*umbral absoluto* y *umbral diferencial*" (Cerdá, 1984).

Fechner se interesó y continuó los trabajos de Weber; sus investigaciones se dedicaron a medir las sensaciones por medio de los juicios humanos, de esta manera llegó a la conclusión de que la "*respuesta sensorial es proporcional al logaritmo del estímulo*" (Cerdá, 1984).

Wundt, en el año de 1879, crea el primer laboratorio de psicología experimental, que tenía como finalidad formular leyes sobre la conducta humana; en este laboratorio se trabajó con diversos sujetos observando que no todos reaccionaban de la misma manera, por lo que no se podían realizar generalizaciones, por lo tanto fue necesario que todos trabajaran bajo las mismas circunstancias y se estudiaran entonces las diferencias individuales. Los experimentos que ahí se realizaban tenían influencia médica y fisiológica, por lo tanto se hacían estudios sobre las reacciones de los sentidos ante diversos estímulos.

Posteriormente Francis Galton, en Londres, se dedicó a hacer estudios sobre la herencia humana realizando mediciones en personas que tenían o no parentesco. Galton, influenciado por Locke, creó diversos test para medir funciones sensoriomotoras ya que consideraba que el conocimiento entra por los sentidos, y por lo que la medición de estas funciones le permitiría conocer la inteligencia de cada ser humano. En el año de 1884 estableció un laboratorio antropométrico, en el Museo de

South Kensington, donde a los visitantes se les aplicaba una serie de pruebas que les permitía conocer su altura, peso, capacidad auditiva y otros atributos. Se emplearon por primera vez las escalas de estimación y los cuestionarios.

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

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

James McKeen Cattell, un psicólogo norteamericano que fue colaborador de Galton y discípulo de Wundt, siguió los estudios de las diferencias individuales y el uso de los test en los Estados Unidos. En el año de 1890 publicó un artículo llamado *Mental Test and Measurements*, en la revista *Mind* (Cerdá, 1984), donde utilizó por primera vez el término **test mental** para designar a una serie de pruebas psicológicas que se empleaban para conocer las diferencias individuales y la inteligencia. Cattell, consideraba al igual que Galton, que la inteligencia se podía medir a través de la discriminación sensorial y el tiempo de reacción.

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

En 1897 Ebbinghaus elabora pruebas para medir memoria y cálculo aritmético.

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

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

En 1904 se formó una comisión dentro de la Institución Pública francesa para estudiar las formas de educación de los niños subnormales y el diagnóstico de la enfermedad mental.

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

En 1908 apareció la primera revisión de la escala Binet-Simon: se eliminaron los test que no eran satisfactorios de la versión anterior y se agregaron nuevos test. Además se realizó una clasificación de los test de acuerdo a diversos niveles de edad. En esta escala se utilizó por primera vez el término "*Edad mental*", que correspondía a los niños cuya ejecución igualaba la edad cronológica (Anastasi, 1978).

La tercera revisión de esta escala se realizó en 1911: en ésta se le adicionaron varios niveles de edad, contemplando su uso en adultos.

La escala Binet-Simon despertó un gran interés en diversas partes del mundo, se realizaron diversas traducciones y adaptaciones, una de las más importantes fue la que llevó a cabo L.M. Terman en la Universidad de Stanford en 1916, conocida como *Escala Stanford-Binet*, en la cual Stern sugirió el término "*Coficiente Intelectual*", que consideraba la relación entre la edad mental y la edad cronológica (Anastasi, 1978).

Las diversas formas de la escala de Binet, requerían en su aplicación un examinador que interactuará con cada sujeto.

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

Durante esa época Arthur S. Otis había realizado un test colectivo de inteligencia; formado con preguntas de información general, razonamiento y aritmética, que presentó ante al comité, en donde se revisó y adaptó para poderse emplear en la selección de los reclutas. A este test se le conoció como "*Test Alfa*" ("*Army Alpha Test*") (Otis, 1939 a). Se construyó además un test complementario para personas analfabetas y con poco dominio del idioma inglés, que incluía rompecabezas geométricos y análisis de ilustraciones. Estos test fueron cedidos para su uso civil.

Durante la Primera Guerra Mundial, se desarrollaron también test de personalidad, dentro de los cuales destaca el que elaboró Woodsworth para los reclutas y que servía para detectar casos que podrían requerir un mayor estudio clínico. Este test estaba formado por preguntas que se emplearían comúnmente en una entrevista referentes a la adaptación del sujeto en su hogar, en la escuela y con sus amigos.

En Suiza, también durante la Primera Guerra Mundial, el psiquiatra Hermann Rorschach implementaba una nueva forma para evaluar la personalidad. Elaboró una prueba en donde al sujeto se le presentaban 10 láminas con manchas de tinta y se le pedía que las describiera. Esta técnica tenía influencia psicoanalítica y daba relevancia al aspecto simbólico.

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

Anastasi (1978) menciona que el desarrollo de pruebas psicológicas, en un principio se orientó principalmente a medir diversas funciones y a partir de éstas determinar el nivel de inteligencia, estas pruebas medían en su mayoría aptitud verbal y en menor grado aptitudes para manejar relaciones numéricas, abstractas y simbólicas. Posteriormente se descubrió que estas medidas proporcionaban poca información de cada función, por lo que fue necesario crear pruebas que midieran funciones específicas.

Desde antes de la Primera Guerra Mundial los psicólogos comenzaron a realizar pruebas que midieran aptitudes especiales que complementarían a las pruebas de inteligencia, se crearon pruebas para conocer la vocación profesional, para selección y clasificación de personal, entre otras.

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

puntuaciones de una prueba, están presentes diversos factores o rasgos independientes.

Los estudios realizados con el análisis factorial permitieron el desarrollo de baterías de aptitud múltiple, que tenían como finalidad proporcionar una medida de la posición del sujeto en cierto número de rasgos.

En 1941, se publican los Test de Chicago de aptitudes mentales primarias (PMA), siendo este el primer esfuerzo por construir una batería.

Fue después de la Segunda Guerra Mundial, en 1945, cuando los psicólogos militares realizaron investigaciones sobre las pruebas basándose en el análisis factorial dirigidas a la elaboración de aptitudes diferenciales. Durante este período se crearon pruebas especiales para pilotos, bombarderos, operadores de radio dentro de la Army Air Forces 1947-1948.

Por otro lado en el ámbito escolar se experimentaban nuevas técnicas, dentro de las cuales se encontraba la aplicación de pruebas tipificadas de rendimiento, siendo éstas las primeras que consideraban importante que todos los sujetos se sometieran a las mismas circunstancias. En estas pruebas se trataba de abarcar contenidos más amplios y eliminar la posibilidad de favoritismo por parte del examinador durante el proceso de evaluación (Mann, citado en Anastasi, 1978).

Durante esta misma época se establecieron programas de test a escala estatal, regional y nacional, de los cuales el más conocido fue el que realizó el *College Entrance Examination Board* (CEEB). Posteriormente los servicios del CEEB se unieron al *Carnegie Corporation* y los del *American Council on Education* formando el *Educational Testing Service*, quien se encargaron de elaborar pruebas para

universidades, escuelas profesionales, y organismos gubernamentales (Anastasi, 1978).

En 1959 el *American Collage Tesitng Program* elaboraba pruebas para los aspirantes de ingreso en los centros no incluidos en el CEEB.

Dentro del estudio de la personalidad, se crearon pruebas para evaluar los intereses, actitudes y opiniones, dentro de las primeras pruebas que se elaboraron se encuentra la que fue creada en el seminario para graduados celebrado en el Instituto de Tecnología Carnegie 1919-1920 (Fryer, 1931, citado en Anastasi, 1978), esta prueba se denominaba *Strong Vocational Interest Blank (SVIB)*. El grupo del Carnegie creó diversos inventarios de intereses los cuales presentaban preferencias o aversiones de los sujetos para una variedad de actividades específicas, objetos o personas que se encuentran comúnmente en la vida diaria (Anastasi, 1978). Posteriormente se crearon otros inventarios, dentro de los cuales destaca la elaboración del *MMPI (1939)* (Inventario Multifásico de la Personalidad Minnessota) y el *CPI (California Psychological Inventory)*.

Anastasi (1978) señala que para complementar el SVIB, se elaboró un inventario similar de intereses para diversas ocupaciones especializadas y semiespecializadas. Este inventario se llamó *Minnessota Vocational Interest Inventory*, formado por 21 escalas para ocupaciones tales como panaderos, conductores, reparadores de televisión, entre otros.

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

En lo referente a la historia de la medición en psicología es importante señalar que aquí solamente se han citado algunas de las pruebas que marcaron un paso en la historia, y que el desarrollo de pruebas ha seguido hasta la fecha.

1.2. Definición de las pruebas psicológicas

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

Tyler, Leona (1972) señala que una *"Una prueba puede definirse como una situación estándar diseñada para tomar una muestra del comportamiento de un individuo"*.

Pichot (1976) da la siguiente definición *"Un test es una situación experimental estandarizada, sirviendo de estímulo a un comportamiento."*

Anastasi (1978) considera que un test tiene como finalidad medir las diferencias que existen entre diferentes individuos o las reacciones del mismo individuo en diferentes ocasiones. Por lo tanto, para Anastasi *"Un test psicológico constituye una medida objetiva y tipificada de una muestra de conducta"*.

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

"Test es una palabra inglesa que significa prueba y que se deriva del latín testis." (Cerdá, 1984).

Cerdá (1984) menciona la definición que fue aceptada en la Antigua Asociación Internacional de la Psicotécnica de H. Pieron que dice: "*Test es una prueba definida, fijando una tarea a realizar, idéntica para todos los sujetos examinados. Debe disponer de una técnica precisa, que permita discriminar entre soluciones satisfactorias y erróneas o bien de un sistema numérico que permita puntuar el resultado. La tarea puede consistir en poner de manifiesto conocimientos adquiridos (test pedagógicos) o bien funciones sensorio-motoras o mentales (test psicológico).*" Cerdá, señala que está definición sólo incluye test de inteligencia, aptitudes y conocimiento, y no toma en cuenta las pruebas de personalidad.

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

1.3. Características de las pruebas psicológicas

Toda prueba de medición psicológica, debe cumplir con características específicas, para que los datos que ésta proporcione sean objetivos, de utilidad y confiables.

1.3.1. Grado de dificultad y poder discriminativo de los reactivos

Una de estas características es el *grado de dificultad* (Anastasi, 1978) o la *sensibilidad de discriminación* (Pichot, 1976) que debe tener los reactivos de una prueba psicológica. Esta característica permite conocer las diferencias individuales; y se logra mediante un procedimiento que se lleva a cabo en el momento de construir la prueba. Este procedimiento consiste en agregar reactivos que puedan ser contestados por todos los sujetos y reactivos que sólo puedan ser contestados por aquellos más aptos. Generalmente en las pruebas psicológicas que evalúan el rendimiento académico,

inteligencia y aptitudes, los reactivos son ordenados en orden de dificultad, colocando primero los reactivos más fáciles y posteriormente los de mayor dificultad, esto se hace con la finalidad de que los sujetos sientan confianza al empezar la prueba y permitir que los sujetos con menor capacidad, puedan dedicar mayor tiempo a aquellos reactivos que pueden resolver correctamente.

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

1.3.2. Estandarización

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

Dentro de las características más importantes que debe cubrir una prueba psicométrica es que ésta tenga un grado aceptable de confiabilidad y validez. En seguida se presentará una breve definición de ambos términos y posteriormente se revisará detalladamente cada uno de éstos.

La *confiabilidad* de una prueba "se refiere a la consistencia o estabilidad de las puntuaciones obtenidas en un mismo sujeto en diferentes ocasiones o con conjuntos distintos de elementos equivalentes" (Anastasi, 1978).

La *validez* "Es el grado en que un test psicológico mide lo que realmente pretende medir" (Anastasi, 1978), es decir, los elementos que conformen la prueba psicométrica, deben ser representativos de la conducta que desea medir.

1.3.3. Confiabilidad

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

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

Nunnally (1995) menciona que el error de medición es una combinación de procesos sistemáticos y procesos aleatorios. Los procesos sistemáticos son las fallas en el procedimiento, que incluyen desde la elaboración, aplicación y calificación de la prueba, hasta la captura y el análisis de los datos. Los procesos aleatorios son los que

quedan fuera de control del proceso de la medición, como los cambios propios del elemento medido.

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

El error aleatorio tiene mayor importancia en las mediciones psicológicas puesto que los puntajes obtenidos se ven afectados por diversas situaciones tales como la elección adecuada de los reactivos de acuerdo a la muestra de sujetos, es decir, determinada prueba puede arrojar puntuaciones elevadas en un grupo y bajas en otro, probablemente porque uno de estos tuvo mayor acceso a información que le fue útil al contestar la prueba y el otro grupo no; otras situaciones que pueden variar las puntuaciones de un mismo sujeto, es el empleo del azar al contestar los reactivos de la prueba o que no tenga la misma disposición en las diversas mediciones. Este tipo de error es casi imposible de eliminar, por lo que se debe de tratar de reducir al mínimo siempre que sea posible.

Debido a la cantidad de errores que se pueden cometer en las mediciones es importante conocer la confiabilidad de una prueba. Anastasi (1978) define la confiabilidad de una prueba como "la consistencia de las puntuaciones obtenidas por los mismos individuos cuando son examinados con el mismo *test* en diferentes ocasiones, con conjuntos distintos de elementos equivalentes o bajo otras condiciones variables del examen".

Para obtener la confiabilidad de una prueba psicométrica se obtiene un *coeficiente de correlación*, que indica el grado de concordancia que hay entre dos conjuntos de calificaciones. Para estimar el coeficiente de correlación se emplean

algunos métodos estadísticos como la fórmula de correlación de Pearson, Spearman entre otros.

Los valores que puede asumir una correlación van desde -1 hasta $+1$, cuando el valor obtenido de la correlación es positivo se dice que se ha encontrado una correlación positiva, y cuando el valor es negativo la correlación se denomina negativa. Cuando el valor del coeficiente se acerca más a $+1$ se considera que la correlación es significativa. Es importante señalar que los coeficientes de correlación que se encuentran en una prueba psicológica son generalmente positivos. El coeficiente de correlación toma en cuenta la posición del sujeto en el grupo así como la distancia que hay en el puntaje de cada sujeto respecto al promedio de calificaciones obtenidas. Es importante señalar que en las mediciones de atributos psicológicos es imposible llegar a obtener un grado de correlación igual a 1. Este coeficiente también es denominado por algunos autores como *coeficiente de confiabilidad*.

Debido a los diversos factores que pueden alterar la confiabilidad se pueden considerar diferentes tipos de confiabilidad, algunos de éstos contemplan el tiempo que transcurre entre una y otra aplicación como es el caso del *coeficiente de estabilidad*, otros considera la adecuada elección de los reactivos que forman la prueba y proporcionan el *coeficiente de consistencia interna y/o el coeficiente de equivalencia*. Cada uno de éstos presenta una varianza de error distinta, en seguida se describirá brevemente en que consiste cada uno de estos métodos así como el tipo de coeficiente de correlación que aporta y el tipo de varianza de error que se contempla.

MÉTODOS PARA OBTENER LA CONFIABILIDAD

1.3.3.1. Método de confiabilidad test-retest

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

1.3.3.2. Método de confiabilidad con formas equivalentes o paralelas

Consiste en elaborar dos pruebas parecidas que midan el mismo atributo; la construcción de pruebas paralelas consiste en crear dos formas de la prueba que contengan el mismo número de reactivos y el mismo grado de dificultad. Ambas pruebas deben tener formato, instrucciones, límite de tiempo de la administración similares, para posteriormente poder ser comparadas. Una vez construidas ambas se aplica una forma y con un período mínimo de tiempo se aplica la segunda forma al mismo grupo de sujetos. Para estimar la confiabilidad se correlacionan las calificaciones obtenidas en las dos formas y el resultado de ésta se denomina *coeficiente de equivalencia*. La varianza de error muestra las variaciones en la actuación de los sujetos en ambas formas. Este método permite además conocer la magnitud del error de medida debido al muestreo de contenido.

1.3.3.3. Método confiabilidad de división por mitades

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

1.3.3.4. Método confiabilidad de Kurder Richardson

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

1.3.3. Validez

En las pruebas psicológicas no sólo es importante que haya confiabilidad en las mediciones, puesto que esta sólo indica la consistencia en las puntuaciones, un

problema más importante que enfrenta una prueba psicológica es comprobar empíricamente que ésta mide el atributo o rasgo para el que fue construida.

Magnusson (1969) menciona que "la validez de un método es la exactitud con que pueden hacerse medidas significativas y adecuadas con él, en el sentido que midan realmente los rasgos que se pretenden medir". Anastasi (1978) señala que para poder determinar la validez es necesario contar con criterios específicos y fuentes objetivas de información que definan el rasgo que mide la prueba. Una prueba psicológica no tiene una sola validez, sino varias, que están en función del uso para el que ésta se destine y el grupo en el que se utilice.

La validez se estima mediante un coeficiente de correlación, denominado *coeficiente de validez*, que indica la relación existente entre los datos obtenidos de la prueba psicológica y los datos que se utilizan como índices para los puntajes del individuo en la variable de criterio. Magnusson (1969).

Brown (1980) menciona que los diversos procedimientos para conocer la validez de una prueba se clasifican en tres categorías principales según los *Standards for Educational and Psychological Test (1974)*: validez de contenido, validez relacionada con el criterio, y validez de construcción. A continuación se explicará en que consiste cada uno de estos procedimientos.

1.3.4.1. Validez de contenido

Este tipo de validez supone un adecuado muestreo de contenido, es decir, una selección cuidadosa de los reactivos que conforman la prueba psicológica, para el cual se realiza un examen sistemático del contenido de los reactivos que conforman la prueba. En el muestreo de contenido se debe especificar la amplitud que fue

considerada, y esta debe tratar de abarcar todas las áreas posibles. Este análisis permitirá determinar si éstos comprenden una muestra representativa del atributo o rasgo que pretende medir.

Uno de los procedimientos para determinar si una prueba tiene validez de contenido consiste en que un grupo de jueces expertos en el tema decidan si los reactivos de ésta son una muestra representativa del área que se pretende medir, si los jueces deciden que la muestra es representativa, entonces la prueba tiene validez de contenido. Para que este procedimiento pueda ser considerado objetivo es necesario especificar a los jueces el dominio de contenido que fue considerado, así como las categorías y su grado de importancia.

Un procedimiento más para conocer la validez de contenido consiste en administrar la prueba a un grupo de sujetos, someter a este grupo a un curso que contenga información relativa a lo que la prueba pretende medir, y al finalizar el curso se administra nuevamente la prueba. Este procedimiento indicará si la prueba mide la información proporcionada en el curso.

Algunos autores mencionan la *validez aparente o de facie* al describir la validez de contenido. Esta validez, es la menos objetiva, ya que sólo toma en cuenta la apariencia de la prueba, es decir, lo que aparentemente parece medir. Se considera importante que la mayoría de las pruebas tengan este tipo de validez, ya que ésta permite cierto grado de confianza en las personas que han utilizarla. Una prueba que a simple vista no indica que es lo que se pretende medir es probable que genere desconfianza y poca cooperación de los sujetos. Este tipo de validez no debe confundirse con la validez de contenido, puesto que no es una validez en sentido técnico. Es importante señalar que "la validez aparente no se debe substituir por la validez objetivamente determinada" (Anastasi, 1978).

Para obtener una validez aparente es necesario que en el momento en que se elabora se tenga en cuenta la población a la que va dirigida, ya que se requiere de una terminología especial para cada caso, por ejemplo, no se podrían utilizar los mismos términos para evaluar a niños y adultos o a un ingeniero y a un mecánico.

1.3.4.2. Validez empírica o de criterio

Las pruebas psicológicas se emplean en algunas situaciones para predecir conductas en situaciones específicas. En este caso se requiere comparar los resultados de la prueba con un criterio específico. El criterio es una medida directa e independiente de lo que mide la prueba (Anastasi, 1978). Un ejemplo de esto es comparar las puntuaciones obtenidas de un individuo en un test de inteligencia para predecir sus logros académicos en la universidad. Las medidas de criterio para obtener la validez de una prueba se pueden obtener mediante dos formas, al mismo tiempo en que se aplica la prueba o después de un intervalo de tiempo.

La validez de criterio se puede clasificar en validez predictiva y validez concurrente, la primera se emplea como se mencionó anteriormente para predecir la actuación o ejecución de un individuo ante determinada situación, y la medida del criterio se obtiene después de un determinado intervalo de tiempo; este tipo de validez se emplea comúnmente en las pruebas empleadas en la selección y clasificación de personal. La validez concurrente se emplea cuando ya se tienen datos de los individuos (criterio) y éstos se pretenden substituir con los resultados obtenidos en la prueba, esto es con la finalidad de ahorrar tiempo en determinados procesos, esta validez se emplea generalmente en pruebas de diagnóstico.

Debido a que la validez empírica requiere de medidas de criterio, es necesario tener ciertas consideraciones al elegir éstas.

Brown (1980) señala tres principales características para elegir una medida de criterio: relevancia, confiabilidad y desviaciones tendenciosas.

La *relevancia*, considera una evaluación racional de las dimensiones relevantes del criterio conceptual para determinar si están representadas o no en la medida del criterio.

La *confiabilidad* de la medida de criterio, ya que considera que si la ejecución del criterio varía con el tiempo no se puede relacionar de manera consistente con otras medidas.

Y por último es importante que no haya *desviaciones tendenciosas* que afecten las medidas, para evitar esto señala que si han de utilizarse medidas de criterio como juicios, será necesario proporcionar indicaciones específicas y concretas de las características que se van a calificar para que el procedimiento de calificación sea lo más objetivo posible. Brown menciona que una fuente de desviación es la contaminación de criterios que se refiere a la situación en que la calificación de una persona en un criterio se ve afectada por el conocimiento del evaluador de su calificación predictor.

Concluye señalando que la mejor medida de criterio es la que tenga mayor número de ventajas prácticas, es decir, la que sea más fácil de utilizar, que se disponga de ella con facilidad y que no resulte muy costosa.

En seguida se presentan algunos de los métodos que sugieren Brown (1980) y Anastasi (1978) para obtener la validez de criterio.

El método más común para obtener la validez de criterio es el de obtener un *coeficiente de validez*, éste se obtiene mediante la correlación de las calificaciones de la prueba con un determinado criterio. Este procedimiento requiere de cinco etapas que

incluyen la selección de un grupo apropiado; la aplicación de la prueba predictora; la aplicación del tratamiento relevante; la recopilación de los datos del criterio y por último se obtiene la correlación de las calificaciones y el criterio. Es importante señalar que si la correlación de las variables es lineal se puede subestimar la magnitud de la correlación y que esta magnitud estará determinada por el rango de las diferencias individuales.

Otro método es el de *grupos contrastados o diferenciación de grupos* que implica comparar dos grupos, uno que haya sido sometido a tratamiento (ejecución en el criterio) y uno control, si se encuentran diferencias significativas entre ambos se dice que la prueba tiene validez.

La *exactitud en la toma de decisiones o desempeño en la tarea*, es un método más para conocer la validez de criterio y consiste en realizar una serie de registros acerca de la actuación en el criterio determinado y determinar la proporción de éxitos o fracasos en los sujetos que fueron seleccionados con una prueba predictora. La prueba psicológica que tenga mayor validez será entonces aquella que proporcione mayor número de decisiones correctas.

La *utilidad* de la prueba, es un método más para determinar la validez de criterio, éste se logra mediante un análisis de costos y beneficios de diversas acciones. Los resultados de la prueba deben representar algún beneficio y ofrecer un ahorro de costos en la toma de decisiones, para ser considerada válida.

1.3.4.3. Validez de construcción o validez estructural

Este tipo de validez se emplea para comprobar un constructo que se supone mide la prueba. "En la medida en que una variable es abstracta y latente más que

concreta y observable se denomina constructo. Es algo que los científicos construyen y que no existe como dimensión observable de la conducta. Un constructo refleja una hipótesis de que una variedad de conductas se correlacionan entre sí en estudios de diferencias individuales y/o serán afectadas de manera semejante por manipulaciones experimentales" (Nunnally, 1995).

Un ejemplo de un constructo es el estudio de la ansiedad o la personalidad, por la tanto las pruebas psicológicas que tienen validez de construcción miden rasgos para los cuales no hay un criterio externo (Magnusson, 1969) o algún atributo o cualidad que se supone poseen las personas (Brown, 1980).

Nunnally (1995) menciona que "el objetivo del estudio de los constructos es emplear una o más medidas cuyos resultados se generalicen a una clase más amplia de medidas que legítimamente empleen el mismo nombre".

Para obtener la validez de construcción es necesario reunir pruebas suficientes que permitan comprobar que los resultados de la prueba pueden medir un constructo, Brown (1980) señala que no es posible obtener un índice cuantitativo de la validez de construcción.

Los pasos para obtener la validez de construcción son los mismos que utiliza el método científico, que incluyen formulación de hipótesis, observación del fenómeno, recopilación de datos y por último aceptar o rechazar las hipótesis.

Existen diversos métodos para obtener la validez de constructo, en seguida se presenta brevemente en que consiste cada uno de ellos, según la clasificación que presenta Brown (1980).

Método intrapruebas: Consiste en estudiar la estructura interna de la prueba psicológica: su contenido, formas de respuesta, y las relaciones entre las subpruebas o

los reactivos. Este método considera la determinación del universo medido por el constructo y otras variables. La validez de contenido de una prueba representa una fuente de evidencia para la validez de constructo, ya que al especificar el universo conductual se define la naturaleza del constructo que mide la prueba.

Método *interpruebas*: Este método toma en consideración las relaciones recíprocas entre varias pruebas de manera simultánea, indicando cuales son los aspectos que tienen en común éstas, o si miden o no el mismo constructo. Encontrándose en este método la validez congruente que consiste en obtener una correlación entre las calificaciones de una prueba nueva y otras ya existentes; y el análisis factorial.

Nunnally (1995), define el análisis factorial como un "procedimiento estadístico que permite encontrar agrupamientos de variables relacionadas. Cada grupo o factor, consiste en un grupo de variables cuyos miembros se correlacionan de manera más elevada entre ellos mismos de lo que hacen con variables fuera del grupo. Cada factor se toma como un atributo que es medido en mayor o menor grado por instrumentos particulares de acuerdo con su correlación con el factor". Señala además que estas correlaciones son denominadas a veces como "validez factorial" de las medidas, pero es mejor hablar de dichas correlaciones como la "composición factorial".

Método de *estudios relacionados con el criterio*: En este método se toma en cuenta la naturaleza y el tipo de criterio que predice una prueba psicológica, para dar una idea del constructo que la prueba está midiendo. Por lo tanto la validez de criterio puede proporcionar información valiosa para la validez de construcción.

Método de *manipulación experimental*: Consiste en manipular una variable y observar los efectos que esta produce en las calificaciones obtenidas en la prueba psicológica. La variable que se manipulará puede ser el constructo que se pretende

medir. De esta manera se puede observar si los resultados de la prueba comprueban las hipótesis que plantea el constructo.

Método de estudios de capacidad de generalización: Consiste en analizar sistemáticamente la prueba sobre una variedad de condiciones o dimensiones o con diferentes condiciones en la administración.

Este método se basa principalmente en la contribución de Campell y Fiske(1959), (citado en Brown, 1980 y Nunnally, 1995) sobre estudios de la validez de constructo, en el cual se hace mención al método de *matriz de multirrasgo y multimétodo*, el cual sostiene que hay cuatro tipos de correlaciones:

- 1) *Coefficientes de confiabilidad* Describen el grado en que una medida es consistente internamente en el sentido de que todos sus componentes miden lo mismo.
- 2) *Correlación heterorrasgo-monométodo* Denota la correlación entre dos medidas que comparten un método común pero evalúan atributos diferentes.
- 3) *Correlación monorrasgo-heterométodo* Denota una correlación entre dos medidas del mismo rasgo que utilizan métodos diferentes.
- 4) *Correlación heterorrasgo-heterométodo* Es una correlación entre atributos diferentes que usan métodos diferentes.

Es importante señalar que los coeficientes de confiabilidad esperados sean elevados, mientras que las correlaciones heterorrasgo-heterométodo deben ser más bajas puesto que utilizan métodos diferentes y miden rasgos distintos.

1.3.5. Normalización

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

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

Magnusson (1969) señala que la normalización es una transformación de la distribución de los puntajes originales a una distribución normal que permita tener escalas con diferentes niveles de dificultad. De ésta manera las calificaciones de una prueba podrán interpretarse con mayor precisión.

La finalidad de convertir los puntajes a una distribución normal es que ésta proporciona un significado estadístico preciso, en el cual el porcentaje de individuos que se encuentra arriba y abajo de cada puntaje se conoce exactamente cuando se tiene una media y una unidad de medidas conocidas.

Anastasi (1978) señala que los puntajes normalizados tienen una doble finalidad, indicar la situación relativa de un sujeto en relación al grupo normativo que permite valorar su ejecución con relación a otras personas; y proporcionar medidas que hacen posible la comparación directa de la actuación del sujeto en distintas pruebas.

En relación a la determinación de normas, es importante señalar que éstas se deben obtener en una muestra representativa de la población para la que fue construida para que se puedan obtener valores estables.

Por último es importante señalar que sí se han de utilizar normas ya establecidas de una determinada prueba, es necesario considerar las características de la muestra para las que éstas se obtuvieron.

1.4. Clasificación de las pruebas psicológicas

Morales (1975) señala que las pruebas psicológicas, se clasifican de acuerdo a sus características, las cuales contemplan: a) su objetivo, b) tipos de respuesta, c) formas de aplicación, d) ejecución, e) instrucciones, f) material y g) calificación. Brown (1980) hace una clasificación de acuerdo al formato de cada prueba. Se expone en seguida una clasificación considerando a ambos autores.

a) En principio las pruebas psicológicas se clasifican por su *objetivo*, es decir, el objetivo principal para el que fueron construidas y los rasgos o conductas que pretenden medir. Esta clasificación divide a las pruebas en: 1) pruebas de potencia o ejecución máxima y, 2) pruebas de rasgos o ejecución típica, conocidas también como pruebas proyectivas. Una característica de las pruebas de ejecución máxima, es que éstas presentan estímulos y tareas específicas al sujeto, mientras que en las pruebas proyectivas los estímulos y tareas que se le piden al sujeto son ambiguos.

1) Las pruebas de ejecución máxima o de potencia son aquellas en las cuales se evalúa el rendimiento y las aptitudes del individuo, estas pruebas demandan el mayor rendimiento del sujeto en la tarea encomendada. La calificación que obtiene el

individuo está determinada por tres factores principales: habilidades innatas, habilidades aprendidas y la motivación que presente al resolver la prueba.

Dentro de esta clasificación se encuentran las pruebas de inteligencia, rendimiento, aptitudes múltiples y específicas. Las pruebas de rendimiento proporcionan información acerca de los conocimientos, habilidades y capacidades del individuo. Las pruebas de aptitudes permiten hacer una predicción de la conducta futura que tendrá el sujeto en determinadas tareas.

2) Las pruebas de ejecución típica o de rasgos permiten conocer la forma en que se comporta un individuo de manera habitual, la forma en que reacciona, piensa y siente ante diversas situaciones; estas pruebas se pueden clasificar además en estructuradas e inestructuradas.

En esta clasificación se encuentran las pruebas de personalidad, actitudes, intereses, preferencias, valores y técnicas proyectivas.

b) Las pruebas se pueden clasificar por su *forma de respuesta* en: 1) objetivas y 2) subjetivas. En las objetivas se encuentran las pruebas de inteligencia, y en las subjetivas las técnicas de proyectivas.

En esta clasificación se puede considerar además las formas que se presentan al sujeto para elegir su respuesta. Una prueba, puede entonces, presentar respuestas alternativas o respuestas libres, en las primeras el sujeto deberá elegir su respuesta de entre varias opciones, y en las de respuesta libre el sujeto deberá completar frases, hacer historias, realizar un dibujo, etc.

c) La clasificación de acuerdo a la *forma en que se administra* una prueba señala tres formas para aplicar una prueba: 1) individual, 2) colectiva y, 3) autoadministrable. Las pruebas de aplicación individual, requieren la presencia de un

examinador para cada sujeto, en este tipo de aplicación el examinador indicará al sujeto las tareas que habrá de realizar y manipulará los materiales que sean necesarios. Las pruebas de aplicación colectiva son aquellas que se pueden aplicar a más de un sujeto a la vez, generalmente son del tipo lápiz-papel, y no requieren manipulación de materiales, la tarea del examinador será dar las instrucciones, vigilar a los sujetos y atender las dudas que pudieran presentar. En las pruebas autoadministrables, la tarea del examinador será sólo dar las instrucciones de cómo resolver la prueba, pero no será necesario que éste permanezca mientras el sujeto resuelve la prueba.

Dentro de las formas de aplicación cabe señalar, una de las más recientes, a la cual se le denomina *automatización*, esta forma de aplicación requiere de una computadora en la cual se le presentarán los reactivos

d) Las pruebas psicológicas se clasifican de acuerdo a *la manera en que se pueden contestar* en: 1) pruebas de poder y 2) pruebas de velocidad. Las pruebas de poder contienen reactivos de diferentes grados de dificultad, y tienen un tiempo límite suficiente para poder contestar todos los reactivos. Las pruebas de velocidad generalmente contienen reactivos de menor dificultad, y tienen un tiempo límite en el cual el sujeto deberá contestar el mayor número de reactivos que le sea posible.

e) Las *instrucciones* que puede contener una prueba psicológica pueden estar escritas en el protocolo, o pueden requerir ser especificadas por el examinador de manera oral, en algunas ocasiones será necesario que el examinador ejemplifique cómo ha de realizarse la tarea.

f) Las pruebas psicológicas se pueden clasificar de acuerdo al material que empleen en: 1) lápiz-papel, 2) verbal, 3) ejecución, 4) aparatos específicos para realizar alguna tarea y, 5) una combinación de los tres primeros.

g) La *calificación*, de las pruebas se pueden llevar a cabo manualmente o por medio de una computadora, representando este último método un ahorro de tiempo y una mayor objetividad en las puntuaciones de cada sujeto.

h) Por último es importante señalar que las pruebas psicológicas deben tener un límite de aplicación, éste considera la población para la que se elaboró la prueba. De esta manera se clasifican en pruebas para bebés, niños preescolares, niños escolares, adolescentes y adultos.

1.5. Influencia del examinador en cuanto a la aplicación

Una prueba psicológica proporciona información útil para la toma de decisión en diversas actividades, tales como, elaborar un diagnóstico, determinar un tratamiento, elegir la carrera que ha de estudiarse, o aceptar o no a una persona para determinado empleo, y como es de esperarse, estas decisiones pueden influir o determinar de alguna manera en el futuro de las personas, por lo tanto, los resultados que brinda una prueba son información muy delicada, que requiere de un manejo profesional, es decir por personal debidamente capacitado para aplicar, calificar e interpretar los resultados.

En seguida se presentarán algunas de las consideraciones más importantes que señalan diversos autores (Morgan, 1975; Anastasi, 1978; Sattler, 1988) que se deben tener en cuenta tanto para el personal que ha de administrar la prueba (examinador), como para las condiciones que se deben tener en el proceso de aplicación.

El examinador o aplicador de una prueba debe cubrir determinadas características y habilidades para poder desempeñar profesionalmente su labor. En principio debe contar con conocimientos suficientes a cerca del comportamiento humano y de las diferencias individuales que presenta cada sujeto, esto con el

propósito de conocer la importancia del trabajo que ha de realizar, entender a los diferentes sujetos y poder establecer una relación interpersonal adecuada. Las personas que pueden desempeñar esta labor pueden ser psicómetras o psicólogos con preparación psicométrica, es decir todos los conocimientos relacionados con el manejo de pruebas, material, instrucciones y forma en que ha de aplicarse cada prueba.

Morgan (1975) hace algunas sugerencias de las habilidades que requiere el entrevistador completo, que se pueden aplicar al examinador de pruebas, estas habilidades son sociales, de comunicación, analíticas e interpretativas y para la toma de decisiones.

Las *habilidades sociales* le permiten al examinador mantener una relación interpersonal cordial y amable, para de esta manera obtener mayor cooperación y participación de los sujetos, además le permitirán determinar que es correcto o no hacer en las diferentes situaciones que puede afrontar.

Las *habilidades de comunicación* le han de servir para mantener una correcta interacción verbal, que incluye el poder determinar el vocabulario que ha de emplear según la edad, escolaridad y cultura de los sujetos. El aplicador deberá saber modular los tonos de voz que ha de utilizar, y ser claro en sus explicaciones para que no existan problemas en la forma de resolver la prueba.

Las *habilidades analíticas e interpretativas* le serán de utilidad para poder ser objetivo en el momento de calificar e interpretar las puntuaciones de cada sujeto obtenidas en las pruebas psicológicas.

Las *habilidades en la toma de decisiones* le servirán para no dejarse llevar por estereotipos al elaborar los reportes finales con los que se han de tomar decisiones, es decir, que no deberá juzgar a las personas por su sexo, edad, raza o religión.

Las condiciones para poder aplicar adecuadamente una prueba psicológica requieren de un lugar amplio perfectamente iluminado, en el que no haya exceso o falta de luz, con ventilación adecuada y temperatura agradable, deberá contar con una silla cómoda y una mesa para cada sujeto, así como aislado de ruidos y de todo tipo de interrupciones que puedan distraer la atención del sujeto.

En el proceso de aplicación, el examinador necesita establecer un adecuado *rapport*, es decir, deberá realizar un esfuerzo para despertar el interés del sujeto, obtener su cooperación y asegurar que sigue las instrucciones tipificadas de la prueba (Anastasi, 1978).

Para poder establecer el *rapport*, es necesario que haya una empatía con el sujeto, y para lograrla, el examinador deberá empezar el proceso de aplicación saludando al (los) sujeto (s), presentándose, preguntado su nombre al sujeto, en el caso de que la aplicación fuera individual. En esta parte el aplicador se puede ayudar con expresiones faciales y modulaciones de voz apropiadas que le permitirán establecer un ambiente de confianza.

Anastasi (1978), señala que cualquier prueba representa una amenaza implícita al prestigio del individuo, por lo que será necesario ganarse la confianza y cooperación del sujeto, así como darle cierta seguridad al iniciar la prueba. Sattler (1988) menciona que para reducir la angustia que pudiera tener el sujeto, se le debe explicar para que se le está aplicando la prueba y como afectará su futuro.

Por último se presentarán los puntos principales de la lista de sugerencias que propone Sattler (1988) para el examinador, en el proceso de aplicación de pruebas de inteligencia, tomando únicamente aquellos puntos que se pueden extender a la mayoría de las pruebas psicológicas.

Precauciones generales

- A.** Lea, aprenda y *vuelva a leer* las instrucciones.
- B.** Atégase siempre a los procedimientos estandarizados.
 - 1. Use la redacción exacta.
 - 2. Observe con mucho cuidado el tiempo establecido.
 - 3. Presente los materiales en la forma descrita.
 - 4. Siga con todo rigor las instrucciones para calificar.
 - 5. No se conforme con leer las instrucciones impresas, téngalas a la mano para consultarlas.
- C.** Sea objetivo.
 - 1. No haga ninguna insinuación sobre lo correcto o incorrecto de las respuestas del sujeto.
 - 2. No de ninguna pista sobre la respuesta que espera; cuido mucho su tono de voz, recuerde que está en examen, no enseñando.
- D.** Sea natural.
 - 1. Trate de ser impersonal, pero sea amable.
 - 2. Aprenda a usar una dicción estandarizada en forma natural e informal
 - 3. Establezca empatía recíproca y comunicación franca antes de que inicie la prueba; tome una actitud expectante.
- E.** Prepare el ambiente
 - 1. Evite distracciones.
 - a) Visuales: Impida que el sujeto mire hacia puertas y ventanas donde haya mucho movimiento y actividad. Que no vea tampoco hacia grandes espacios abiertos que tengan imágenes, colores, etc. que puedan distraerlo. Evite el ruido.

- b) **Auditivas:** Evítese las áreas ruidosas, otras voces, etc.; aplique la prueba en un lugar aislado.
- c) **Emocionales:** No administre la prueba cuando el sujeto se sienta presionado, perturbado o enfermo.

Administración y calificación

- A. **Eficiencia para la forma en que se aplique la prueba y el manejo de materiales.**
 - 1. **Establezca un orden eficiente y un método apto para manipular los materiales.**
 - a) **Para registrar.**
 - b) **Para consultar el manual *sin que se convierta en obstáculo entre usted y el examinado.***
 - c) **Para guardar y sacar en forma oportuna los materiales.**
 - 2. **Haga una transición suave de una prueba a otra y de un reactivo a otro.**
En cada punto de la prueba debe saberse qué es lo que se va a presentar en seguida.
 - 3. **Conozca el material, las calificaciones lo suficiente como para no alargar la prueba en forma innecesaria. Una prueba demasiado prolongada suele producir cansancio, desinterés o ambos.**
 - a) **Disimule o reoriente los comentarios no importantes.**
 - b) **Reduzca al mínimo los movimientos extraños fomentando el interés, la motivación, la orientación de las tareas.**
 - c) **Prevenga el cansancio y la distracción.**

B. Calificación

- 1. Es esencial que el examinador conozca bien los criterios de calificación.**
Este requisito se refiere en especial a la comprensión del *objetivo* de cada prueba o subprueba.
- 2. Debe recordarse que los criterios de calificación son simplemente lo que dice la "etiqueta". "Criterios" y no respuestas completas y de absoluta veracidad o falsedad.** Suele ser necesario evaluar respuestas equivalentes basándose en las enumeradas en los criterios de calificación, ya que no pueden incluirse todas las respuestas posibles.
- 3. El examinador principiante debe constatar las respuestas con el manual,** para verificar cualquier duda.
- 4. Deben volver a revisarse cada uno de los pasos en el proceso de calificación.**

C. Cuidado del material de la prueba.

- 1. Siempre que cualquier material presentado al niño tenga marcas o esté deteriorado de modo que pudiera influir en su respuesta, habrá que reemplazarlo.**
- 2. Los materiales extravagantes o dañados tendrán que ser substituidos por otros idénticos a los originales.**

De acuerdo al marco teórico que se ha expuesto en este capítulo, en el que se señalan las consideraciones que han de tenerse al elaborar y aplicar una prueba, las características que debe cubrir, confiabilidad, validez, normas, etc., así como al manejo de las calificaciones que ésta proporciona, criterios, normalización, estandarización, es posible concluir que para hacer un uso adecuado y responsable de ellas, es necesario que cubran siempre todas las características mencionadas, y éste se debe limitar sólo a aquellas personas que estén debidamente capacitadas en el manejo de ellas.

CAPÍTULO 2

INTERESES

INTERESES

2.1. Definición de los intereses

Los intereses han sido definidos por diversos autores, en seguida se presentan algunas definiciones de éstos.

"La palabra INTER- ESE es equivalente a "estar entre", o sea, la posibilidad de elección de esa cosa que se encuentra entre otras" (Yoffe, 1988).

Perry (1926, citado en Yoffe, 1988) considera que la noción de interés se halla ligada a todas las nociones afines a instinto, deseo, sentimiento, voluntad y otras análogas.

Pléron (citado en Super, 1967) define: el interés es una tendencia a prestar atención a ciertos objetos y a orientarse a ciertas actividades.

Para Strong (citado en Super, 1967) el interés es una tendencia a ocuparse por ciertos objetos, a amar a ciertas actividades.

Super (1967) basa su definición en la de Fryer y señala, los intereses se pueden considerar como sentimientos agradables o desagradables asociados a objetos y actividades, y como reacciones positivas o negativas que se manifiestan por la atención y por el comportamiento.

Fernández, *et al*, (1970) plantean que los intereses son aquellas áreas en las cuales los sujetos demuestran cierto tipo de gusto o motivación.

Super (1973) señala los valores son objetivos que buscan las personas, los intereses son las actividades y los objetos mediante los cuales se hace la búsqueda.

Ferrater (1980, citado en Yoffe, 1988) menciona "se tiene interés por' algo cuando se orienta hacia ello la apetencia, el deseo o la voluntad":

Angelini (1984) señala "nuestros intereses están representados por aquellas actividades que, causándonos satisfacción, en un sentido más amplio y más profundo, atienden a nuestros sentidos". Considera que éstos representan el aspecto más consciente de los motivos, ya que ciertas actividades pueden o no gustarle al individuo, y esto le permite tener clara conciencia de lo que siente, aunque no pueda aclarar con seguridad el por qué de este sentir. Menciona además que los intereses varían de un individuo a otro, pero también a lo largo de la vida del ser humano, principalmente durante los años de la infancia, señalando como punto importante que a medida que se alcanza la madurez, se van haciendo más estables.

Yoffe (1988) concibe el interés como una expresión de un deseo o voluntad de un objeto a elección, el o la cual se hallan matizados por su valor correspondiente.

Es posible, considerando la opinión de los diversos autores conceptualizar a los intereses como una expresión de gusto o desagrado hacia una actividad o cosa.

2.2. Clasificación de los intereses

Super (1967, 1973) señala que la taxonomía de los intereses es el resultado de las diversas investigaciones, las cuales reportan resultados del análisis estadístico de las respuestas dadas a los inventarios.

La taxonomía de los intereses planteada por Super (1967) contempla las categorías: profesionales, e intereses de los niños.

Los intereses profesionales abarcan las categorías básicas, factoriales y los intereses y los valores.

CATEGORÍAS PROFESIONALES: Contempla diferentes grupos ocupacionales.

La clasificación que realizó Strong planteó baremos (escala convencional de valores que se utiliza como base para valorar o clasificar los elementos de un conjunto, de acuerdo con alguna de sus características) para la calificación de su inventario a través de la intercorrelación con las diferentes profesiones, considerando la mayor parte de las profesiones de nivel intelectual y social elevados. Señaló además en una de sus investigaciones que los intereses son diferentes para cada sexo. Los grupos ocupacionales que plantea son:

Ciencias Biológicas	Bienestar Social
Ciencias Físicas	Detalles Comerciales
Técnicos	Literario y jurídico.

Clark (1961, citado en Super 1967) construyó un inventario cuyos reactivos son más apropiados para los obreros y los artesanos, que permiten diferenciar uno de los otros, construir baremos de calificación para 17 oficios calificados y no calificados. Los 17 oficios se agrupan en 10 oficios de dos categorías principales de 5 oficios cada uno y los 7 restantes no se agrupan en estas categorías, considerando la mayor parte de las profesiones de nivel inferior, una categoría técnica. Las categorías obtenidas son:

Oficios electromecánicos

Oficios de registro y control

Oficios no reagrupables.

Categorías básicas:

G. F. Kuder (1939, citado en Super, 1967,1973) clasificó los intereses de acuerdo a categorías *a priori*, eligió reactivos para evaluar intereses en cada una de estas categorías. Las categorías de Kuder informan de los intereses ligados a actividades profesionales bien definidas y limitadas:

Científico

Cálculo

Mecánico

Persuasivo

Aire Libre

Literario

Servicio Social

Artístico

Oficina

Musical

Clark utilizando (citado en Super, 1967) el mismo método de Kuder construyó baremos de clasificación homogéneos e independientes para las categorías de intereses en su inventario:

Mecánicos	Eléctricas	Limpieza
Biológicas	Carpintería	Aire Libre
Oficina	Gastronómicas	Estético Verbales

Categorías factoriales:

Thurstone(1931, citado en Super, 1967) es el primer psicólogo que emplea un análisis factorial en los intereses a partir de los resultados obtenidos con el inventario de Strong. El primer análisis discriminó cuatro factores:

Científico

Personal

Lingüístico

Comercial

Posteriormente este mismo autor, en otros análisis encontró los factores científico, personal y de acuerdo a este análisis concluyó que los

Guilford, (1954, citado en Super, 1967,1973) basándose en Strong y Kuder, realizó un el análisis factorial más agudo sobre los intereses, con una muestra de 1500 soldados de la aeronáutica de Estados Unidos, identificando diversos factores. Para esta investigación elaboró un inventario compuesto de 1000 reactivos relacionados con las actitudes del sujeto, las actividades preferidas y ciertas descripciones de sí mismo. A través de varios análisis factoriales determino las siguientes categorías:

Científico	Oficina
Mecánico	Comercial
Aire Libre	Expresión Estética
Bienestar Social	Apreciación Estética

Super (1967) señala que los intereses de las mujeres han sido estudiados por Crissy y Daniel (1939) en un instrumento especialmente construido para tal efecto por Strong. De esta investigación se reportó que fue posible agrupar cuatro factores, de los cuales tres son más comunes en los hombres y a las mujeres, los factores de intereses *lingüísticos, científicos y personales*. El cuarto factor fue denominado *interés por la compañía de los hombres*.

Los intereses y los valores:

En el estudio de los intereses se observa que estos son similares a los valores medidos en diversos inventarios de valores. Y de acuerdo a esta similitud Super (1967) señala los diferentes tipos de pasatiempos que se han clasificado.

Pasatiempos suplementarios Son aquellos que se asemejan a las profesiones, particularmente, la profesión de quien se dedica a dichos pasatiempos, permitiéndole así una satisfacción más completa de los intereses que tiene por su profesión.

Pasatiempos complementarios de la profesión, aquellos que no se parecen a la profesión que se práctica, pero tienen alguna similitud con otras profesiones reales, permitiendo satisfacer intereses que no pueden expresarse en la profesión ejercida.

Pasatiempos no profesionales A cerca de éstos no se tiene mucha información ya que no han sido bien estudiado, sin embargo debido a que no tienen ningún parecido con otra profesión, no afectan tanto en la orientación y selección de personal.

CATEGORÍAS DE INTERESES EN LOS NIÑOS: Contemplan las variaciones de los intereses en los niños.

Super (1967) señala que se han realizado diversas investigaciones para conocer los intereses de los niños, encontrando que es más difícil discriminar los intereses específicos de éstos.

La investigación de Larcebeau (citada en Super, 1967) señala que los intereses en los niños se pueden medir fácilmente cuando se trata de intereses intelectuales, técnicos y por el dibujo, pero no las actividades físicas, oficina, artístico y persuasión.

La investigación de Tyler reporta que es posible observar las manifestación de los intereses en los niños, pero esto no es señal de que serán definitivos a lo largo de su vida.

2.3. Factores que determinan los intereses

Super (1967) considera que los intereses están determinados por diversos factores, entre los cuales se encuentran las capacidades propias de cada individuo, su raza y nacionalidad, el nivel socioeconómico, el sexo, la familia y la personalidad.

Las capacidades propias de un individuo se relacionan causalmente con sus aptitudes e intereses, de tal manera que éstas se pueden considerar un factor determinante en los intereses

Raza y nacionalidad estos factores son indudablemente determinantes para los intereses, si se contempla que las tradiciones y actitudes de una cultura determinan el comportamiento y lugar de cada miembro tanto para su rol social como para su participación en la productividad.

Nivel Socioeconómico este factor desempeñará un papel importante en el desarrollo de los intereses, puesto que no tendrán las mismas oportunidades aquellos sujetos que viven en un medio propicio para alentar sus aptitudes, que aquellos que a los cuales se les imponen limitaciones para desarrollarlas. Por lo tanto algunos intereses sólo se podrán encontrar en determinados niveles socioeconómicos.

El sexo Las diferencias físicas, biológicas y psicológicas permiten afirmar que existen intereses distintos para los hombres y para las mujeres, es posible hacer esta afirmación por los roles que determina la cultura a cada sexo.

La familia es tal vez uno de los factores que más influyen en la elección de los intereses del individuo, ya que en está confluyen factores como la identificación herencia y la presión social.

La personalidad debido a que los intereses comienzan a organizarse desde la infancia, de acuerdo a la idea del yo y las actitudes de individuo frente a sí mismo, es la aceptación o el rechazo de un papel lo que orienta la elección y determinación de los intereses.

Por otra parte Angelini (19884) señala que la mayoría de los intereses se adquieren en el ambiente inmediato en que se vive, y, en el mismo caso de los impulsos fundamentales innatos, cuya satisfacción puede determinar un interés por ciertos estímulos.

2.4 Medición de los intereses

Super (1967) plantea que los intereses han sido estudiado en varias investigaciones, por lo que ha sido necesario plantear definiciones operacionales claras para la recolección de datos, en seguida se presentan estas definiciones.

Intereses expresados: Son aquellos donde se le solicita al sujeto exprese el agrado o desagrado por alguna actividad, en forma escrita u oral. Considerando el interés como aquello que se dice encontrar interesante.

Intereses manifiestos: Para conocer este tipo de intereses se realizan observaciones de diversas actividades. El interés, será entonces, lo que se muestra al asistir a un acontecimiento, participar en él, o al actuar sobre los objetos o las personas.

Intereses revelados por medio de test: La fuente de los datos sobre este tipo de intereses la constituyen las pruebas de atención y de memoria. En estas pruebas el sujeto revela sus intereses de manera indirecta, es decir, centrandlo su atención en lo que le interesa y recordando lo que ha visto u oído.

Intereses inventariados El uso de inventarios de intereses permite obtener una calificación objetiva de ellos, ya que estos están formados por preguntas que solicitan la expresión de los intereses y por medio de las respuestas que el sujeto brinda es posible a través de un sistema estadístico de calificación conocer los intereses del sujeto, así como las actitudes frente a estos.

2.5 Instrumentos para evaluar Intereses

Anastasi (1978) considera que la medición de los intereses profesionales no es una tarea sencilla, debido a que la mayoría de las personas no cuentan con información suficiente de los diferentes trabajos y actividades, por lo tanto no pueden juzgar si les gustaría o no determinada actividad, el interés o falta de interés para alguna actividad puede estar determinado por una noción muy limitada de las tareas que ella abarca. Otro factor que influye en los intereses es el predominio de estereotipos de las profesiones, generalmente se tiene una idea errónea de las funciones de cada profesión.

Debido al problema señalado se llegó a la conclusión de que había que encontrar medios más directos y sutiles para determinar los intereses de los individuos. Se presentan en seguida se la lista de algunos de los instrumentos que se han desarrollado para lograr este fin.

En 1908 Parsson (citado en Ackerman, 1964) fundó la primera oficina de orientación profesional, iniciando de esta manera el movimiento de la orientación profesional. Esta orientación estaba fundamentada en bases psicológicas, limitada entonces a exámenes del sentido de la forma, velocidad de lectura y escritura, y el

empleo simultáneo de cuestionarios que contenían preguntas relacionadas con los intereses e inclinaciones profesionales.

En 1907 en Holanda surgieron comités que tenían como función dar información y orientación vocacional (Gamelli, 1968, citado en Valencia, 1972).

Anastasi (1978) menciona que la determinación de los intereses a través de medios directos fue logrado en el Instituto de Tecnología Carnegie durante el año académico 1919-1920, señala además que posteriormente se desarrollaron varios inventarios tipificados por las personas que asistían a ese seminario.

Los inventarios de intereses elaborados por el grupo de Carnegie trataban de las preferencias o aversiones del sujeto para una gran variedad de actividades específicas, objetos o tipos de personas que se encuentran comúnmente en la vida diaria. Las respuestas se pusieron en forma de clave, de forma empírica para las diferentes ocupaciones. Se observó que las personas empleadas en aquellas ocupaciones distintas se caracterizaban por intereses comunes que las diferenciaban de las personas que tenían otras ocupaciones (Anastasi, 1978).

Entre los inventarios del grupo de Carnegie destaca el *Strong Vocational Interest Bank* (SVIB) (hoja de intereses profesionales), que fue elaborada por Strong Jr., el cual posteriormente ha sido investigado, revisado y ampliado.

La forma corriente del SVIB publicada en 1966 consta de 399 elementos agrupados en ocho partes. En las primeras cinco se registra la preferencia del sujeto a través del llenado de óvalos que indican *Gusto*, *Indiferencia*, *Disgusto*, respectivamente. Las tres partes restantes requieren que el sujeto coloque unas actividades dadas por orden de preferencia, compare su interés en parejas de

elementos y valores, sus aptitudes actuales y otras características. Existen claves distintas para hombres y mujeres (Anastasi, 1978).

En 1930 las investigaciones de Kuder lo llevaron a publicar su "Registro de profesiones" que fue diseñado para aplicarse a estudiantes de secundaria, universitarios y adultos (Szekely, 1960).

Los inventarios de Kuder son instrumentos que tienen diversas formas, versiones, y ediciones, son un conjunto de instrumentos que enfocan la medida de los intereses desde diversos ángulos y están designados para finalidades un tanto diferentes (Anastasi, 1978).

El *Kuder Preference Record-Vocational* (hoja de preferencias de profesión, de Kuder) se construyó para indicar el interés relativo en un pequeño número de amplias áreas. Los elementos se formularon y agruparon originalmente por tanteo, basándose en la validez de contenido. Esta prueba proporciona 10 escalas de interés, más una de verificación para registrar el descuido, la equivocación y la elección de respuestas socialmente aceptables. Las escalas de interés incluyen: exterior, mecánico, cálculo, científico, persuasivo, artístico, literario, musical, de servicio social y de oficina. Contiene normas distintas para hombres y mujeres, así como para estudiantes de secundaria, universitarios y adultos. Proporciona perfiles ocupacionales establecidos empíricamente para 41 grupos de ocupaciones homogéneas.

El *Kuder General Interest Survey* (investigación de Kuder sobre el interés general) está destinado a grados de 6 a 12, emplea lenguaje más simple, y un vocabulario más fácil que la forma original y sólo requiere un nivel de lectura de sexto grado. Este inventario se elaboró a través de procedimientos clave de criterios similares a los seguidos por el SVIB. La puntuación del sujeto en cada escala de ocupaciones se expresa como una correlación entre su esquema de intereses y el

modelo de intereses del grupo ocupacional dado. Generalmente da puntuaciones para 79 ocupaciones 29 carreras universitarias para hombres y 56 ocupaciones y 25 carreras para mujeres.

Akermann (1964) señala que la Primera Guerra Mundial fue un gran obstáculo en el desarrollo de la orientación profesional.

En 1954 Baumgarten empleó un método para evaluar los intereses de los adolescentes, este método consistió en hacer preguntas relacionadas con el agrado o desagrado de su profesión, el porqué de su elección, etc. a 4500 sujetos, método tal vez muy subjetivo, pero fue el primer acercamiento del estudio de la relación posible entre intereses y profesiones. (Super, 1967).

El inventario de Thurstone fue traducido y adaptado en Brasil en 1953 por Angelini. Este inventario contiene los siguientes grupos de profesiones: Ciencias físicas, ciencias biológicas, cálculo, negocios, profesiones ejecutivas, profesiones lingüísticas, profesiones humanitarias, profesiones artísticas y profesiones musicales.

"El Inventario de Intereses Profesionales de Hereford (IIPH), fue desarrollada por el Dr. Carl Hereford durante su estancia en México como profesor visitante, dando un seminario en el entonces colegio de Psicología de la U.N.A.M." (Valencia, 1972).

Yoffe (1984) señala el IIPH es una prueba "HECHA EN MÉXICO".

El IIPH es una prueba de tipo vocacional cuya intención central es la de explorar áreas de interés, producto del gusto, deseo, motivaciones; y con base a la información permite elaborar ciertas predicciones acerca de la actividad futura de la persona que haya contestado el inventario. De esta manera el sujeto puede detectar aquellas ocupaciones con las que parece tener mayor afinidad (Fernández, *et al*, 1970).

Fernández, G., *et al* (1970), señalan que en su versión original contenía 100 reactivos, y la versión que se empleaba desde 1970 se redujo a 90 reactivos debido a la eliminación del área verbal.

Posteriormente, en una investigación inédita (Varela, R., Villegas, E., Cordero, E., 1995), llevada a cabo con 1356 sujetos, el área verbal se incluyó nuevamente, siendo esta versión la que se utilizó en esta investigación. Cada reactivo está formado por una frase referente a alguna clase de actividad relacionada con el área que pretende medir. El sujeto debe calificar estas frases con una escala de 1 a 5, en la cual 1 equivale a "Me disgusta mucho", 2 "Me disgusta", 3 a "Me es indiferente", 4 a "Me gusta" y 5 a "Me gusta mucho".

Los 100 reactivos comprenden 10 áreas de interés vocacional, cada una compuesta de 10 reactivos. La calificación máxima por área es de 50 y la mínima de 10 (Ver Anexo 1).

Las áreas por las que está formada el inventario son:

Área I:	Cálculo	Área VI:	Verbal
Área II:	Científico Físico	Área VII:	Literario
Área III:	Científico Biológico	Área VIII:	Ejecutivo-Persuasivo
Área IV:	Mecánico	Área IX:	Artístico Plástico
Área V:	Servicio Social	Área X:	Artístico Musical

Para concluir este capítulo, se señala que se presentaron los principales aspectos que se consideran en la medición de los intereses, así como algunos de los instrumentos disponibles para tal efecto.

Señalando como factor importante que los intereses son una característica relativamente estable en el individuo, permitiendo esto incrementar su validez predictiva (Angelini, 1984). Y por tal motivo se eligió como tema para estudiar la confiabilidad y validez en una prueba automatizada en esta investigación.

CAPÍTULO 3
AUTOMATIZACIÓN
DE
PRUEBAS PSICOLÓGICAS

AUTOMATIZACIÓN

3.1. Definición de la automatización de pruebas psicológicas

El proceso de automatización de pruebas psicológicas consiste en substituir la forma tradicional (lápiz-papel) de aplicar un instrumento estandarizado con el uso de una computadora, a través de la cual los reactivos que forman la prueba se presentan por medio de la pantalla del monitor de ésta y las respuestas se obtienen a través de los diferentes dispositivos de entrada de la computadora como lo son el teclado, el *mouse* y pluma óptica.

La automatización de pruebas, permite almacenar los puntajes de cada uno de los reactivos para proporcionar una puntuación total bruta. El puntaje global de la prueba se obtiene en puntuaciones en crudo y es posible convertirlo posteriormente a puntuaciones percentiles que proporcionan una calificación final de la prueba, estos resultados pueden aparecer en la pantalla del monitor de la computadora y/o imprimirse; la automatización en algunos casos ofrece la interpretación de esta calificación. En inventarios automatizados es posible realizar una gráfica del perfil del sujeto. La información que se obtiene de cada sujeto en la aplicación automatizada, puede ser almacenada fácilmente y estar disponible para una consulta posterior.

Stout (1981) señala que el término prueba por computadora ha sido utilizado para describir el "proceso de recolección de datos para estimar el puntaje de un sujeto sobre una o más dimensiones medidas a través de un instrumento estándar".

Murphy (1994) sostiene que el uso de la computadora para la aplicación o interpretación de una prueba o selección de opciones representa una potencialidad, un cambio fundamental en la práctica de las pruebas.

El uso de pruebas psicológicas automatizadas, requiere ofrecer resultados confiables y válidos como los que ofrece una prueba psicológica tradicional, sin embargo no existen estudios realizados en esta área (Fleishman, 1988; Jäger, 1991; Murphy, 1994). Esta investigación pretende conocer si una prueba automatizada brinda resultados confiables y válidos.

3.2. Características de una computadora

Aiken (1995) plantea que una computadora común PC moderna está constituida esencialmente por tres unidades: unidad de entrada (*input*), unidad central de procesamiento y unidad de salida (*output*). En seguida se presenta una breve descripción de cada una de ellas:

Unidad input o dispositivos de entrada: Esta unidad es la encargada de llevar la información a la unidad central de procesamiento a través de sus diferentes periféricos como: el teclado, *mouse*, discos flexibles, *scanner* óptico y la pluma óptica.

Unidad central de procesamiento (CPU): Esta unidad es considerada como la más importante, ya que es el "cerebro de la computadora", en ella se interpreta y procesa la información. El CPU efectúa los cálculos, hace comparaciones y manipula el control de las demás unidades; a su vez el CPU está constituido por 3 unidades: a)unidad de memoria para instrucciones de programas y datos, b)unidad lógico-aritmética (encargada de hacer cálculos y tomar decisiones) y c)unidad de control (encargada de seguir instrucciones dadas por los programas). La unidad de control dirige el flujo de la información recibida por los dispositivos de entrada, utiliza datos para que la unidad lógico-aritmética tome

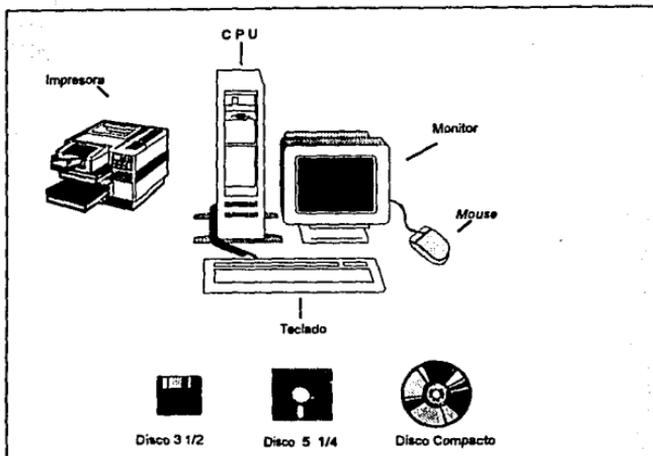
decisiones y presente las necesidades de la computadora a través de los dispositivos de salida.

Unidad output, o dispositivos de salida: Esta unidad es la encargada de llevar la información procesada por el CPU a los dispositivos de salida como la pantalla del monitor, cinta o discos magnéticos o la impresora.

Una microcomputadora contiene un microprocesador formado por una red de miles de circuitos microscópicos para combinar las funciones de la unidad lógico-aritmética.

La computadora tiene dos formas básicas de almacenamiento y estas son la memoria ROM (memoria de sólo lectura) que almacena tanto las instrucciones de la computadora como la información que recibe a través de los dispositivos de entrada y la memoria RAM (memoria de acceso aleatorio) encargada de almacenar temporalmente la información que recibe de las o cintas o discos de almacenamiento permanente, entre los cuales se encuentran discos flexibles de 5 1/4 ó 3 1/2 , discos compactos y disco duro, este último se localiza en el interior de la computadora.

La computadora adquiere mayor utilidad con el uso de programas (*software*), éstos son una serie de instrucciones que permiten dar entrada a los datos y salida a la información, un programa procesa los datos para después mandar la información necesaria al usuario. Los programas generalmente se escriben en lenguajes de computación y se adaptan para poder ser interpretados por la computadora.



3.3. Estudios realizados

La revolución de las computadoras ha traído un cambio rápido y significativo en el mundo actual, en psicología se ha transformado el carácter de los experimentos psicológicos y otros campos de la investigación psicológica han empezado a hacer uso del potencial de la nueva tecnología disponible (Beaumont, 1981).

En el campo de la psicometría el uso de las computadoras hasta 1980, se reducía únicamente al almacenamiento de respuestas y computo de calificaciones de las pruebas, posteriormente las computadoras han sido utilizadas para diversos propósitos en las pruebas, tales como la administración, calificación e interpretación, al grado de que en algunos países la computadora empieza a reemplazar el método tradicional de la aplicación lápiz-papel (Murphy, 1994).

Beaumont en 1981 afirma "hasta estas fechas los trabajos relacionados con la automatización de pruebas han sido pocos debido al costo tan elevado que presentaban las computadoras y a la dificultad que se presentaba para tener acceso a ellas". Menciona que entre los estudios de los que se tiene reporte se encuentra la aplicación de un sistema para interpretar la batería neuropsicológica Halstead-Reitan, elaborado por Russell, Neuringer y Goldstein en 1970; señala además que Gilberstadt, Lushene y Bueguel (1976), citados en Vicent (1980) han escrito acerca de otros sistemas para interpretación de baterías de pruebas, incluyendo aquellas destinadas a medir funciones cognitivas.

Beaumont (1981) considera que una de las mayores aportaciones a este campo es el trabajo realizado por Jonhson, Williams y sus colegas en la Unidad de Evaluación Psiquiátrica, quienes desarrollaron un sistema que en un principio utilizaba un sistema de computadora a gran escala y empleaba una terminal diseñada especialmente, pero en 1978 Jonhson y Williams hicieron modificaciones para que operara en PDP-11/V03 con terminales CRT estándar, los estudios de confiabilidad y validez de este sistema reportaban resultados alentadores. También reporta otras aplicaciones que están bajo el control de una computadora, entre las cuales se encuentra el *Halstead Category Test* (Beaumont 1975), que utiliza una PDP LINC-8, y el Peabody Picture Vocabulary Test que emplea una terminal especial conectada a una IBM 1800 (Kinge y Rodziewicz, 1976). Señala que ambas aplicaciones se veían limitadas por el costo y las dificultades prácticas relacionadas con el *hardware* de la computadora, y no tanto por la aplicación automatizada. Sin embargo, concluye que el panorama ha ido cambiando radicalmente gracias a que las computadoras han adquirido un precio más accesible y a la abundancia del buen *software* disponible, y que pronto se publicarían un gran número de proyectos de menor escala.

Este mismo autor considera que las normas para una pruebas automatizada, pueden ser las mismas de la prueba tradicional, siempre y cuando ésta sea una forma paralela de la prueba original.

Entre los desarrollos de automatización de pruebas se encuentra la evaluación adaptativa por medio de la computadora (Vale, 1981). Las pruebas adaptativas fueron aplicadas por primera vez en 1970 para conocer diferencias individuales, a través de la medición de habilidades, y se denominaron pruebas IQ de Binet (Weiss, D y Vale, 1987). En este tipo de evaluación una prueba es adaptada de acuerdo a las habilidades del sujeto; estas pruebas cuentan con un banco de reactivos muy extenso, al iniciar la prueba se presenta un grupo de reactivos, y dependiendo de las respuestas que el sujeto de, se aplica el siguiente grupo de reactivos y así sucesivamente hasta determinar el nivel de habilidad del sujeto, en este momento la prueba termina.

El proceso de adaptación se hace mientras el sujeto realiza la prueba, los reactivos que se aplican dependen tanto de la habilidad de cada sujeto como de la habilidad que se esté evaluando, al iniciar la prueba se aplica un reactivo fácil y dependiendo de la respuesta que éste de, se le aplica el siguiente reactivo incrementando el grado de dificultad, y así sucesivamente hasta llegar a los reactivos que el sujeto no pueda contestar, de esta manera se determina el nivel de habilidad del sujeto.

En la aplicación automatizada de una prueba adaptativa el sujeto se coloca frente a una terminal interactiva en la cual se le presentan los reactivos en forma gráfica o de texto. El examinado responde a los reactivos presionando una o más teclas de un teclado que puede variar de acuerdo a su complejidad, desde un teclado simple de dos botones (cierto-falso) hasta un panel parecido al teclado de una máquina de escribir. Posteriormente la respuesta que da el sujeto al reactivo es evaluada por la

computadora, para presentar el siguiente reactivo. Finalmente la computadora califica la prueba. La automatización permite dar una retroalimentación durante el proceso de aplicación o al final de la prueba, y ésta puede o no ser presentada al sujeto examinado.

Vale (1981) señala que la selección simultánea de reactivos cuenta con una gran variedad de estrategias de evaluación que permiten acercarse a presentar la alternativa más óptima.

El uso de las computadoras no se ha limitado sólo al campo de la aplicación y calificación de pruebas psicológicas, sino que además ha trascendido en un tema de mucha controversia, y es el de la generación de reportes psicológicos a través de la computadora.

Los reportes psicológicos se clasifican en dos procesos diferentes la *evaluación psicológica* y la *interpretación de pruebas psicológicas* (Tallent, 1987), la evaluación psicológica "es un proceso variable dependiendo de la pregunta de la persona involucrada, del tiempo establecido, y de otros factores innumerables que conducen a soluciones de problemas específicos" (Malloney y Ward, 1976, citados en Tallent, 1987); el segundo se refiere a la interpretación de las calificaciones que proporciona una prueba.

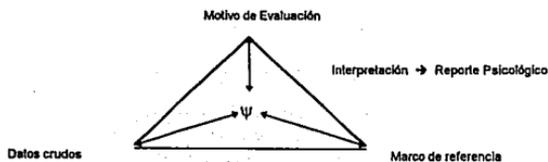
Tallent (1987) considera que las computadoras pueden elaborar una buena evaluación psicológica ya que señala: "ellas tienen una memoria muy eficaz y una gran capacidad; son rápidas, eficientes y económicas, eliminan la dificultad de escribir los reportes psicológicos y pueden ser aplicadas en muchas evaluaciones psicológicas (clínicas, industriales, educacionales, etc.). Pueden procesar volúmenes de datos y al mismo tiempo trasladarlos, en términos de correlaciones empíricas más o menos

aceptables dentro de los estatutos conductuales que pueden constituir los reportes psicológicos y en una sintaxis aceptable”.

Newman (1966, citado en Tallent, 1987) menciona que la necesidad de un desarrollo de guías para el uso de la tecnología de las computadoras en la evaluación psicológica fue reconocido oficialmente por la *American Psychological Association (APA)* a mediados de los sesentas y la primera publicación de guías fue después de veinte años en 1986. Estas guías fueron elaborados en conjunto por el *Committee on Professional Standars* y el *Committee on Psychological Test and Assesment*, y contemplaban los principios y estándares técnicos, profesionales y aplicaciones éticas.

Tallent (1987) señala que la evaluación psicológica es una interacción múltiple y compleja que demanda significativamente un compromiso personal y responsabilidad por parte del psicólogo, ante esta demanda formuló un esquema (Véase Figura 1) en 1983, donde sugiere que el psicólogo es el centro de todo este proceso complejo. En el contexto general de interacción, el motivo de evaluación, los datos de origen (resultados de la prueba) y el marco de referencia del psicólogo (o la teoría) son los elementos que contribuyen a la interpretación y al reporte psicológico.

Figura 1



Tallent (1987) menciona que la importancia de este esquema radica en que el psicólogo juega la parte más importante al igual que la razón de evaluación siendo los datos de la prueba los que tienen menor valor dentro del rol del psicólogo. Y concluye su aportación señalando que la computadora es una máquina eficiente, aunque existen áreas importantes dentro de la psicometría, en las cuales la computadora no puede intervenir.

En 1987, Weiss y Vale retoman el tema de las pruebas adaptativas para medir habilidades y otras variables psicológicas por medio de la computadora al cual denominan *Computerized Adaptive Test (CAT)*.

Las CAT utilizan una computadora interactiva que permiten remplazar al administrador de la prueba de Binet, la computadora utiliza diversas funciones para presentar los reactivos a los examinados, recibe y califica las respuestas de cada reactivo, posteriormente elige el siguiente grupo de reactivos que será administrado y termina la prueba cuando es necesario. Weiss y Vale (1987) plantean que un sistema CAT adecuado, es capaz de escribir un reporte de interpretación que resuma la administración de la prueba y reporte los puntajes adecuados.

Señalan además que las CAT no son simplemente una prueba asistida por la computadora, ya que además de la administración de la prueba, calificación e interpretación proporciona algunas ventajas psicométricas sobre las pruebas tradicionales que se resumen a dos palabras: *eficiencia* y *precisión*, estas ventajas serán tratadas en una sección posterior.

Por otra parte, se tiene una opinión europea acerca del uso de la computadora en las aplicaciones prácticas de la evaluación. Jäger (1991) señala "los avances de la computadora se ha implementado internacionalmente y estos son capaces de realizar las tareas del diagnóstico". Demuestra que no existe una teoría general para elaborar

un diagnóstico por computadora, pero sin embargo, señala que un *programa* apropiado justifica el uso de la computadora, problema que no se puede dejar de lado, ya que el uso de ésta es un proceso vigente.

Jäger (1991) menciona que los desarrollos en el campo de la evaluación han tomado dos caminos, por una parte están aquellos que cuando generan desarrollos pierden el enfoque de lo que es un ser humano, predominando en ellos un pensamiento que sólo considera términos cuantitativos y esto genera serias implicaciones en los profesionistas que trabajan en el campo de la psicología aplicada (Schötzau-Fürwentsches y Grubitzsch, 1991).

Por otra parte es importante que las computadoras brinden un trato más amigable, considerando los aspectos cognoscitivos es posible que se pueda tener un mayor control en el diagnóstico de la práctica profesional (Jäger, 1991).

Jäger (1991) señala que se han descrito diversos estudios relacionados con el diagnóstico por computadora, que van desde la presentación de los reactivos en la pantalla de la computadora (Calvert y Waterfall, 1982; Evan y Miller, 1969; Hoffman y Lunderg, 1976; Honacker, 1988; Skinner y Allen, 1989) hasta abarcar el uso simultáneo de otros medios como el video (Bayard-White y Hoffos, 1988; Fricke, 1991; Hornke, 1982; Koblitz y Wenzel, 1989) y la implementación de sistemas expertos para diferentes propósitos (Beaumont, 1991; Hayes-Roth, Waterman y Lenat, 1983; Huba, 1987; Langner, 1988; Maurer, Lögler, Kühne, Iig y Biehl, 1990; Tinger, 1990, Vale y Keller, 1987, Warzecha, 1991a, b). Considera que el término diagnóstico asistido por computadora subraya los factores en los cuales la computadora puede ser usada para varias tareas del proceso del diagnóstico.

Este mismo autor menciona que el uso de la computadora en las pruebas psicológicas se enfrenta con las algunas limitaciones ya que los programas disponibles

sólo permiten ser utilizados en las pruebas que se presentan tradicionalmente en formato lápiz-papel, donde los reactivos de la prueba son presentados en la pantalla de la computadora (Calvert y Waterfall, 1982; Hodgkinson, 1986; Moser, Seelig y Rebstock, 1990; Dann, Irvine y Collins, 1991; Tanzer, 1987), o para convertir calificaciones crudas a calificaciones normalizadas (Bartram, 1989; Bartram y Bayliss, 1984; Reinert, 1982; Krug, 1986 y Böttcher, 1991).

Sin embargo, Jáger (1991) menciona que la equivalencia de las versiones automatizadas y las pruebas de lápiz-papel, como lo señala la APA (1986) y el Testkuratorium (1986) no ha sido comprobada.

Jäger (1991) presenta una lista de las pruebas que están disponibles en versión automatizada:

- ψ Test Anxiety (Homke, 1981);
- ψ Vocational Aptitude (Moreno, Wetzel, McBride y Weiss, 1984);
- ψ Depressive disorders (Bayer *et al.*, 1984);
- ψ Intelligence (Gilberstadt, Lushene y Buegel, 1986; Kubinger, 1986, 1989; Kubinger y Wild, 1989);
- ψ Concentration (Fay y Stumpf, 1980; Fay y Meyer, 1990);
- ψ Various areas of educational psychology (Weiss y Kingsbury, 1984; Wilson, 1991; Mills, 1991; Corte y Verschaffel, 1991);
- ψ Aptitude assessment (Eckardt, 1977; Harrell, Honacker, Hetu y Oberwager, 1987; Bakassa, Kisser y Wenninger, 1990; Steege, 1988; Wildgrube, 1990);

- ☒ ψ Occupational stress and resources (Krieger, in press; Krieger, Becker, Schoerer y Kamm, 1989);
- ☒ ψ Clinical Assessment (Beaumont y French, 1987);
- ☒ ψ Marketing (Meiner, 1989), entre otros.

Jäger (1991) plantea que el propósito de este tipo de programas es meramente reemplazar la versión lápiz-papel, y substituir la calificaciones automatizadas por el procedimiento tradicional. Pero es importante que las características de los reactivos en cuanto a confiabilidad y validez no varíen en relación a la versión original.

Jäger (1991) considera que es importante tener en cuenta los factores que han planteado diferentes autores, (unos en contra y otros a favor), en relación a la equivalencia de ambas forma de aplicación, entre los cuales destacan la ansiedad, la oposición a la nueva tecnología, la poca familiaridad con las computadoras y la cultura.

Menciona que no hay investigaciones realizadas que han tenido como centro de atención temas como la presentación de pruebas automatizadas contra las pruebas convencionales, (Sacher y Fletcher, 1978), actitudes y resultados de la aceptación del uso de computadoras (Koch y Patience, 1978), y problemas especiales en la metodología (Reckase, 1974 Kubinger, 1989).

Murphy (1994) habla del uso de las computadoras en pruebas psicológicas, su aplicación e interpretación, plantea que las computadoras pueden ser utilizadas para aplicar la misma prueba de formato lápiz-papel o para administrar en forma breve una prueba de acuerdo a las características del examinado, además indica, las computadoras permiten elaborar interpretaciones de las pruebas psicológicas. Sostiene que el uso de las computadoras en las pruebas psicológicas representa sólo un aspecto del incremento del uso de las computadoras en todo tipo de mediciones,

siendo éste un caso especial, ya que el uso de la computadora puede cambiar el significado de las puntuaciones de las pruebas psicológicas, especialmente en aquellas que son adaptadas a las necesidades del sujeto.

Ante el problema del uso de la computadora en pruebas psicológicas y la variación en las calificaciones, Murphy (1994) menciona que se han realizado diversas investigaciones que reportan resultados pequeños, simples e inconclusos. Sin embargo, señala que los estudios de Elwood (1969), Johnson y Mihal (1973) y Katz y Dalby (1981) sugieren importantes conclusiones: 1) las calificaciones que se obtienen en una prueba presentada en computadora se correlacionan altamente con las calificaciones de un prueba lápiz-papel, 2) hay evidencia de que las pruebas aplicadas por computadora son posiblemente equivalentes a las pruebas de lápiz-papel, y 3) existen evidencias de que las diferencias culturales pueden ser menores en las pruebas por computadora que en las pruebas de lápiz-papel. Considera que si esta apreciación es posible, entonces las computadoras pueden ser utilizadas para presentar muchas formas de la prueba con una pérdida menor de su confiabilidad.

Murphy (1994) plantea que hay evidencias que sugieren que la presentación de los reactivos de la prueba contra los reactivos presentados en una hoja de papel, no afecta substancialmente las características de la prueba, aunque no es posible hacer una generalización a todo tipo de pruebas, especialmente aquellas destinadas a medir ansiedad. Sugiere que para lograr equivalencia entre ambas formas de aplicación es necesario tener en cuenta ciertos factores que han sido señalados por diversos autores:

El estudio de Mazzeo y Harvey's (1988) señalaron que las diferencias que pueden existir entre los dos tipos de pruebas dependen de factores como la velocidad en que es presentada la prueba, el uso de gráficos, el tamaño del texto en que es

presentado el reactivo (sí ocupa todo el despliegue de pantalla, dificulta su lectura) y la posibilidad de regresarse u omitir algunos reactivos.

Finalmente el estudio de Fleishman (1988) concluye que comúnmente se asume la equivalencia de ambas formas, pero que en realidad no hay investigaciones que demuestran dicha equivalencia, además de que las reglas para desarrollar pruebas por computadora son considerablemente pocas y menos definidas, a las reglas que existen para desarrollar pruebas de lápiz-papel.

En seguida se presentan los trabajos mencionados por Murphy (1994) relacionados con el uso de la interpretación por computadora de las pruebas psicológicas (CBTI):

El primero uso extenso de los sistemas CBTI fue cuando se desarrollaron en 1960 para ayudar a la interpretación clínica de las calificaciones del Inventario Multifásico de la Personalidad Minnesota (MMPI), (Moreland, 1990).

Posteriormente se desarrollaron sistemas CBTI para la interpretación de pruebas de inteligencia y habilidades (Kramer, 1988), sistemas para intervenciones clínicas estructurales (Spitzer, Williams, Gibbon, y Furst, 1990), para diagnosticar problemas neuropsicológicos (Adams y Brown, 1986). En la actualidad existen más de 400 servicios de interpretación que están comercialmente disponibles.

3.4. Recomendaciones para elaborar una prueba automatizada

El desarrollo de pruebas psicológicas es algo complejo, y como Jäger (1991) señala, no existe una teoría general específica para elaborar este tipo de desarrollos,

no hay normas, no hay reglas, pero el trabajo de los investigadores ha sugerido algunas consideraciones, que serán presentadas en esta sección.

Beaumont (1981) menciona dos consideraciones importantes que se debe tener en cuenta en el desarrollo de la automatización, el sistema (*hardware*) y el programa (*software*) en el cual se va trabajar.

El sistema debe ser eficiente en términos de rapidez y de requerimientos de memoria, debe brindar facilidades periféricas y gráficas de alta resolución que permitan que la evaluación automatizada sea apropiada, de manera que el investigador tenga la posibilidad de automatizar cualquier tipo de pruebas y podrá brindar al sujeto examinado diversas posibilidades en la elección de sus respuestas.

El programa debe ser de fácil operación, es decir, debe contemplar tanto las necesidades del sujeto que va a ser examinado como las del aplicador (quien puede no estar familiarizado con las computadoras). El manejo del programa debe ser tan sencillo, como operar una grabadora, en este punto será necesario programar un lenguaje en el cual sea posible escribir, debe ser diseñado para que el aplicador sólo requiera de encender la computadora, insertar un disco y responder al dialogo con la computadora para seleccionar la prueba y establecer los parámetros particulares de la prueba. El programa debe contar con un sistema de protección contra errores en el manejo de la información, que permita al aplicador dar una solución ante los problemas que se pudieran presentar.

Beaumont (1981) recomienda además, que el programa tenga una presentación amigable, tratando de mantener un equilibrio entre lo complejo e impersonal y lo falsamente humano, de manera que éste logre inspirar confianza en el sujeto en cuanto a su interacción con el sistema. La forma en que la computadora demanda las

respuestas, debe por lo tanto, permitirle al sujeto sentir que tiene un cierto grado de control sobre el proceso de la prueba.

Vale (1981) sugiere algunas especificaciones del lenguaje de programación que se debe utilizar en el desarrollo de pruebas por computadora, señala que éste debe ser diseñado para permitir la flexibilidad que se requiere para cambiar la estrategia de los árboles de decisiones de un simple movimiento hacia arriba o hacia abajo para elegir el tipo de respuesta (cierto-falso), además debe ser capaz de proporcionar un mínimo de respuesta entre la presentación de un reactivo y otro.

Vale (1981) plantea tres tipos de módulos para administrar una prueba: 1) estructura lineal, en el cual se incluye una lista de reactivos, los cuales deben ser administrados de principio a fin hasta terminarlos, en este programa no es posible volver a entrar, 2) rama, en este módulo se especifica un grupo de reactivos que se interrelacionan con diversas ramificaciones, una ramificación puede ser de reactivo a reactivo o de reactivo a subtest y ambos pueden utilizar las mismas especificaciones y 3) estructura de búsqueda en la cual se especifica una lista de reactivos de entre los cuales es posible seleccionar el que se ajuste mejor al sujeto.

Jäger (1991) señala que el primer prerrequisito para utilizar programas de evaluación psicométrica es contar con una computadora y sistema de *hardware* y *software* adecuado. Una PC típica permite que el teclado y el *mouse* sean utilizados para operar el programa de evaluación, almacenar, imprimir e ilustrar los resultados

Jäger (1991) considera, que la configuración del *hardware* y *software* debe estar diseñada para facilitar las tareas del diagnóstico. Para la presentación de los reactivos de las pruebas convencionales de lápiz-papel en la pantalla, es necesario tener memoria disponible para almacenar todos los reactivos que serán desplegados. Para las pruebas adaptativas es necesario contar con algoritmos que permitan tener un

control en la selección de reactivos, en la presentación de éstos la pantalla y para almacenar las respuesta del reactivo.

3.5. Ventajas

La automatización de pruebas psicológicas presenta un número importante de ventajas con relación a las pruebas tradicionales, en esta sección se presenta la opinión de algunos autores.

Beaumont (1981) considera que si las pruebas tradicionales se convirtieran a formas paralelas automatizadas, esto representaría un gran ahorro de tiempo y recursos, ya que las pruebas automatizadas se podrían emplear en el período en el cual otras pruebas podrían ser desarrolladas. Una de las ventajas que representa automatizar pruebas tradicionales estandarizadas, es que éstas cuentan ya con normas.

Stout (1981) sugiere que las principales ventajas del uso de la computadora frente al formato de lápiz-papel es que éstas pueden registrar una amplia gama de información auxiliar a través de la respuesta de los examinados. Las latencias de respuesta y la fuerza con que ha sido presionada una tecla pueden ser medidas por medio del movimiento por detección ultrasónica, considerada una técnica no intrusiva o que afecte al sujeto. El reporte de estas medidas permite en combinación con los chequeos tradicionales de consistencia y validez facilitar la detección de respuestas inválidas.

Tallent (1987) plantea entre las ventajas de la computadora, que ésta cuenta con una memoria muy eficaz, son rápidas eficientes y económicas, permiten procesar y trasladar grandes volúmenes de datos, así como prestar conclusiones de probabilidad

de ocurrencia de conductas. Sin embargo, considera que aunque la computadora puede tener un valor incuestionable, existen áreas dentro de la psicometría dentro de las cuales la computadora no puede intervenir, ya que carece de un "pensamiento psicológico".

Weiss y Vale (1987) mencionan que las ventajas que proporciona la computadora se pueden resumir en dos palabras, eficiencia y control:

El control de eficiencia en las CAT permite obtener una calidad de medición equivalente al de las pruebas tradicional aplicando un número menor de reactivos (o mejor calidad que la que proporciona una prueba tradicional con el mismo número de reactivos). En las CAT es posible reducir el tamaño de la prueba hasta un 50% sin comprometer la calidad de la medición (Brow y Weiss, 1977; Kiely, Zara, y Weiss, 1983; McBride y Martin, 1983; Moreno, Wetzel, McBride y Weiss, 1985, citados Weiss y Vale, 1987). De manera que el tiempo que se ahorra en estas mediciones puede ser empleado para realizar mediciones de otras habilidades del individuo e incrementar la validez predictiva de una batería de pruebas.

El control sobre la precisión de la medición se refiere a que cada sujeto puede ser medido con el mismo grado de exactitud, es decir el mismo error estándar, ya que en las pruebas tradicionales en las cuales se aplica un número fijo de reactivos a cada sujeto, no es posible hacer una medición exacta sin reducir su precisión global, puesto que en estas todos tienen el mismo grupo de reactivos y la precisión se maximiza igualando la dificultad de la prueba a la habilidad promedio examinada; y sólo las personas que se aproximen al promedio serán las que estén siendo correctamente evaluadas, mientras que las que se encuentren en las colas de la distribución serán evaluadas de manera equivocada o aproximada, en cambio debido a que las CAT se adaptan a las habilidades del sujeto y sólo presentan los reactivos apropiados para

cada sujeto, hacen posible que la pruebas se construyan a la medida de éste, permitiendo además terminar la prueba cuando la habilidad del sujeto ha sido estimada con un grado predeterminado de precisión.

Murphy (1994) considera que la eficiencia de las computadoras permite tener el control de latencias, o tiempo que se requiere para responder a cada reactivo, señala como ventaja única de la computadora el hecho de poder medir las habilidades de espacio y percepción que eran imposibles de medir en una prueba con formato lápiz-papel. Menciona la necesidad de utilizar una computadora cuando se presente una tarea que requiera tiempos precisos, control de estímulos, presentación de gráficas, o una limitada exposición de material en cada reactivo de la prueba.

Entre las ventajas que proporcionan las pruebas adaptativas Murphy (1994) señala que éstas son altamente eficientes en el sentido de que este campo es demasiado responsable y preciso en sus puntuaciones utilizando un menor número de reactivos. Este tipo de pruebas permiten una mayor seguridad en su distribución que las pruebas tradicionales pues estos pueden ser fácilmente copiados y distribuidos en personas que desean obtener un puntaje adecuado y memorizan respuestas para lograr ser aceptados en trabajos o escuelas, en cambio, en una prueba adaptativa por computadora una persona que haya intentado memorizar los reactivos, sólo será expuesto a un número pequeño de éstos.

Aiken (1995) considera que la computadora proporciona una ventaja en el momento de la recolección de datos ya que algunos sujetos tienen una mayor disposición para brindar información personal, particularmente de naturaleza emocional a alguien impersonal, que no juzga, como lo haría un entrevistador. Otra ventaja es que las computadoras son extremadamente eficientes, se adaptan a las diferencias

individuales tanto mentales como físicas y proporcionan calificaciones e interpretación de resultados.

3.6. Desventajas

Tallent (1987) plantea que una de las principales desventajas que se presentan en la generación de reportes es que estos pueden ser utilizados por personas que no están adecuadamente entrenadas o en aquellas que no hacen un uso responsable de éstos, las cuales representan un serio peligro, porque no sólo distorsionarían la información sino que además podrían conducir a tomar decisiones erróneas, eliminado por lo tanto el trabajo que proporcionan los servicios computarizados y de los profesionistas que proporcionan éste tipo de servicios.

Murphy (1994) señala que las pruebas por computadora representan una desventaja para aquellas personas que no requieren de hacer aplicaciones a un gran número de personas, puesto que el costo es relativamente alto, además de que requiere una inversión necesaria en *hardware* y su mantenimiento.

Otra desventaja señalada por Murphy (1994) se relaciona con el uso de reportes de interpretación generados por la computadora, ya que considera que ésta se basa únicamente en una calificación tomada en un punto del tiempo, mientras que el juicio clínico considera muchos factores que no incluyen estos reportes, considera además que éstos no proporcionan una interpretación completa y confiable de las calificaciones de las pruebas, ya que pueden ser defectuosos.

Aiken (1995) considera como desventajas: el costo de los sistemas, el uso de *software* inadecuado, el hecho de que la computadora tiene como limitación el no ser apropiada a cierto grupo de personas (niños, retrasados mentales, personas en crisis,

psicóticos entre otros), el hecho de no contar con investigaciones que reporten confiabilidad en la recolección y almacenamiento de información personal, además señala que un serio problema se presenta cuando la información es guardada en la misma computadora y cualquiera puede acceder a ella, por lo tanto se viola el derecho de mantener la información personal en forma confidencial.

Finalmente Aiken (1995) señala que muchos de los servicios que se ofrecen para hacer interpretaciones no han sido adecuadamente validados y utilizan normas inadecuadas. Menciona que otra desventaja es que la venta de programas de autoevaluación al público en general puede causar serios disturbios y muchos de ellos no tienen una validación adecuada.

En este capítulo se ha presentado una definición de la automatización de pruebas psicológicas, así como algunos de los estudios realizados en este tema, se señaló además las recomendaciones que han de tenerse en cuenta en su desarrollo y finalmente se indicaron las ventajas y desventajas que representan al utilizar pruebas automatizadas. Consideraciones que fueron tomadas dentro de este estudio.

CAPÍTULO 4

MÉTODO

MÉTODO

4.1 Pregunta de investigación

¿La forma de aplicación automatizada (por computadora) del Inventario de Intereses Profesionales de Hereford ofrece resultados confiables y validos?

4.2. Hipótesis conceptual

Si una prueba psicométrica tiene confiabilidad y validez en las puntuaciones obtenidas (Inventario de Intereses Profesionales de Hereford) éstas se conservarán independientemente de su forma de aplicación.

4.3. Hipótesis de trabajo

Los resultados obtenidos de la aplicación en forma automatizada del Inventario de Intereses Profesionales de Hereford son confiables y validos.

4.4. Hipótesis nula

Los resultados obtenidos de la aplicación en forma automatizada del Inventario de Intereses Profesionales de Hereford, no son confiables ni validos.

4.5. Definición de variables

Variable Dependiente: Las respuestas al Inventario de Intereses Profesionales de Hereford en cada forma de aplicación (automatizada y lápiz-papel).

Variable Independiente: Forma de aplicación del Inventario de Intereses Profesionales de Hereford: automatizada y tradicional.

4.6. Población

La población de alumnos del sistema escolarizado de la Facultad de Psicología es aproximadamente de 3,000 inscritos en la licenciatura y 400 en el posgrado.

4.7. Sujetos

El estudio se llevó a cabo dentro de las instalaciones de la Facultad de Psicología. Participaron en un inicio 120 alumnos de 18 a 39 años de edad, que se encontraban inscritos en los diferentes semestres de la licenciatura y maestría, pero es importante señalar que a la segunda aplicación del IIPH sólo se presentaron 101 sujetos, siendo éste el número total de la muestra con que se trabajó.

La muestra de sujetos se dividió aleatoriamente en dos grupos para efectos de control, donde a un grupo se le aplicó en la primera fase del experimento (test) la prueba automatizada y en la segunda fase (re-test) la prueba tradicional. Al segundo grupo se le aplicó en la primera fase la prueba tradicional y en la segunda fase la prueba automatizada.

4.8. Muestreo

Se empleó un muestreo probabilístico al azar contemplando a los diversos horarios que existen en la Facultad de Psicología, para el procedimiento de selección se utilizaron los horarios y las listas de alumnos inscritos al sistema escolarizado.

La muestra quedó conformada por 120 alumnos.

Mediante un procedimiento aleatorio se eligieron 5 grupos de licenciatura y 1 grupo de posgrado.

4.9. Diseño

Se empleó un diseño de investigación no-experimental ex-post-facto, debido a que en éste no se manipulan directamente las variables, Este tipo de investigación observa los fenómenos tal y como se dan en su contexto natural, para después analizarlos.

Se empleó un diseño de una sola muestra, en el cual un grupo de sujetos es medido en dos los tiempos o momentos para evaluar el cambio entre las variables, este diseño permite conocer los cambios grupales así como individuales (Hernández, *et al.* 1991).

GRUPO 1	GRUPO 2
TEST	RE-TEST
Aplicación automatizada	Aplicación lápiz-papel
Aplicación lápiz-papel	Aplicación automatizada

4.10 Instrumento y materiales

El Inventario de Intereses Profesionales de Hereford (IIPH), es una prueba psicológica "de tipo vocacional que pretende descubrir áreas de interés en las cuales los sujetos demuestran cierto tipo de gusto o motivación y qué, en base a ésta información, permite hacer ciertas predicciones acerca de la actividad futura para la cual tales sujetos parecen ser más afines" (Fernández, G., Lara, L., Hereford, C. 1970).

En esta investigación se empleó la versión revisada por Varela, R., Villegas, E., Cordero, E., 1995 (en investigación inédita), (Ver capítulo Intereses).

CONFIABILIDAD Y VALIDEZ DEL INSTRUMENTO

En investigación inédita (Varela,R., Villegas,E., Cordero,E., 1995), con una muestra de 1356 sujetos se obtuvieron los datos referentes a la confiabilidad y validez del instrumento.

El análisis de confiabilidad por área del instrumento se obtuvo mediante el procedimiento Alpha de Cronbach. Los resultados de este análisis señalan confiabilidades aceptables, ya que todas las correlaciones son significativas y positivas: Cálculo $\alpha=.93$, Científico físico $\alpha=.84$, Científico biológico $\alpha=.85$, Mecánico $\alpha=.92$, Servicio Social $\alpha=.87$, Verbal $\alpha=.79$, Literario $\alpha=.88$, Ejecutivo-persuasivo $\alpha=.83$, Artístico plástico $\alpha=.89$ y Artístico musical .82.

La validez de constructo del instrumento se obtuvo mediante un análisis factorial con rotación varimax. Los resultados indican que las 10 áreas definidas en su construcción se definen claramente como factores independientes.

La prueba original tiene una forma de aplicación de lápiz-papel, para efectos de la presente investigación esta versión fue automatizada, es decir, se diseñó un programa para computadora en el que se presentó al sujeto el inventario, donde cada uno de los 100 reactivos, así como su opción de respuesta fueron presentados en pantalla, y por medio del teclado y/o mouse contestaron el inventario.

Se utilizó el Inventario de Intereses Profesionales de Hereford en sus dos formas de aplicación, lápiz-papel y automatizada:

Para la aplicación lápiz-papel se utilizó el inventario en su forma original formado por 3 hojas con los 100 reactivos y sus 5 opciones de respuesta escritos y un lápiz por cada sujeto.

Para la aplicación automatizada se empleó una computadora por cada sujeto y un mouse, donde se encontraba instalado el programa de automatización del inventario.

4.11 Procedimiento

Al cambiar el tipo de aplicación, se pretendía que los resultados obtenidos en el IIPH en la versión automatizada arrojarán resultados similares a la versión de lápiz-papel.

La aplicación se realizó en forma contrabalanceada para que el orden de aplicación de la prueba (lápiz-papel y automatizada) no afectará los resultados obtenidos.

FASE PRELIMINAR

Se diseñó un programa por medio de la computadora que presentará el IIPH y permitiera al sujeto contestar la prueba. La automatización se llevó a cabo mediante el programa "Authorware", este programa se utiliza en ambiente Windows, la programación está orientada a objetos, y además permite la compilación de las aplicaciones desarrolladas en él para una fácil distribución.

FASE 1

En esta fase se realizó la primera aplicación del procedimiento test-retest, de manera que a la mitad de los sujetos de la muestra se le aplicó el IIPH en la versión

automatizada, y a la otra mitad de la muestra se le aplicó el IIPH en su versión lápiz-papel.

FASE 2

Después de un mes, se realizó la aplicación re-test del IIPH en forma contrabalanceada a la misma muestra de sujetos, aplicándose la versión lápiz-papel a la mitad de la muestra que se le había aplicado en el test la versión automatizada, y la versión automatizada a los sujetos que se les aplicó la versión lápiz-papel en el test.

FASE 3

Una vez realizadas ambas aplicaciones test-retest se procedió a realizar la captura de los datos de la aplicación lápiz-papel para el análisis de los datos. Los datos de la aplicación automatizada eran capturados directamente por el programa mientras el sujeto contestaba el inventario.

CAPÍTULO 5

RESULTADOS

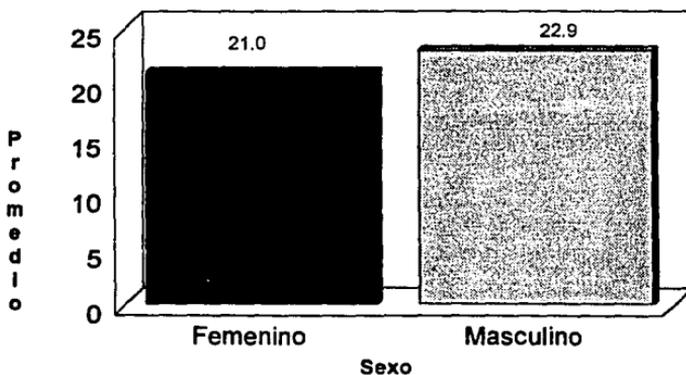
RESULTADOS

PROCEDIMIENTO ESTADÍSTICO.

Descripción de la muestra

La investigación se llevó a cabo con 101 sujetos, 83 mujeres y 18 hombres, el promedio de edades por sexo se muestra en la Gráfica 1.

Promedio de Edad por Sexo de la Muestra



Gráfica 1

Para obtener la confiabilidad y validez del proceso de automatización se realizaron diversos análisis estadísticos, mediante el programa estadístico para las ciencias sociales SPSS para Windows versión 6.1 (1994).

CONFIABILIDAD

El análisis de confiabilidad se realizó para cada área del instrumento y para cada una de las formas de aplicación. Para tal fin se empleó el procedimiento Alpha de Cronbach.

En la tabla 1 se muestran los resultados de este análisis, encontrando que las confiabilidades son aceptables, y sus valores son mayores a 0.74 para ambos tipos de aplicación.

Tabla 1
Resultados del Alpha de Cronbach para cada
Área del Inventario de Intereses Profesionales de Hereford

Área	Aplicación Lápiz-Papel	Aplicación Automatizada
Cálculo	.9298	.9278
Científico Físico	.8389	.8279
Científico Biológico	.8206	.7895
Mecánico	.9202	.8987
Servicio Social	.8642	.8490
Verbal	.7444	.7309
Literario	.8984	.8912
Ejecutivo-Persuasivo	.7741	.7659
Artístico Plástico	.8784	.8534
Artístico Musical	.8630	.8381

VALIDEZ

Análisis por área

Para conocer si existía validez de criterio en el procedimiento de automatización de esta prueba psicométrica, se realizó una correlación producto-momento de Pearson, con las puntuaciones obtenidas en cada área del IIPH de cada sujeto en ambas aplicaciones, con el propósito de conocer la equivalencia de las formas de aplicación. Con este análisis se obtuvieron correlaciones significativas, altas y positivas.

En la tabla 2, se muestran los resultados de este análisis, encontrando que todas las correlaciones son significativas y mayores a 0.76. La correlación más alta se encuentra en el área Artístico Musical ($r = 0.90$), y las más bajas en las áreas: Verbal ($r = 0.76$) y Ejecutivo Persuasivo ($r = 0.78$). Estos resultados indican un alto grado de validez para la aplicación automatizada.

Tabla 2

Correlaciones por Escala entre la Aplicación Automatizada y la Aplicación Lápiz-Papel

Área	r
Cálculo	.89
Científico Físico	.81
Científico Biológico	.86
Mecánico	.81
Servicio Social	.84
Verbal	.76
Literario	.86
Ejecutivo- Persuasivo	.78
Artístico Plástico	.87
Artístico Musical	.90

$p < 0.001$

Análisis por reactivos

Para tener una idea más específica de ambas aplicaciones, se repitió el análisis de correlación entre formas de aplicación, sólo que a nivel de cada reactivo.

En la tabla 3 se muestran los resultados de las correlaciones de cada reactivo de las áreas: Cálculo, Científico Físico, Científico Biológico, Mecánico y Servicio Social, y en la tabla 4 se encuentran las correlaciones de las áreas: Verbal, Literario, Ejecutivo-Persuasivo, Artístico Plástico y Artístico Musical.

De este análisis se resalta el hecho de que las correlaciones de todos los reactivos de la prueba, son significativas y positivas, ($p < 0.001$). La correlación más alta se encuentra en el reactivo 91 ($r = 0.85$) del área Cálculo, y la más baja en el reactivo 55 ($r = 0.24$) del área de Servicio Social. En seguida se presenta la correlación más alta y más baja de cada área.

En el área Cálculo, la correlación más alta se localiza en el reactivo 91 ($r = 0.85$) y la correlación más baja en el reactivo 21 ($r = 0.45$); el promedio de correlación de esta área fue de 0.69.

En el área Científico Físico, la correlación más alta se localiza en el reactivo 62 ($r = 0.71$) y la correlación más baja en el reactivo 22 ($r = 0.54$), el promedio de correlación de esta área fue de 0.62.

En el área Científico Biológico, la correlación más alta se localiza en el reactivo 33 ($r = 0.78$) y la correlación más baja en el reactivo 13 ($r = 0.52$), el promedio de correlación de esta área fue de 0.65.

En el área Mecánico, la correlación más alta se localiza en el reactivo 44 ($r = 0.74$) y la correlación más baja en el reactivo 24 ($r = 0.46$), el promedio de correlación de esta área fue de 0.62.

En el área Servicio Social, la correlación más alta se localiza en el reactivo 75 ($r = 0.77$) y la correlación más baja en el reactivo 55 ($r = 0.24$), el promedio de correlación de esta área fue de 0.56. Es importante señalar que la correlación del reactivo 55 es significativa y positiva.

En el área Verbal, la correlación más alta se localiza en el reactivo 76 ($r = 0.69$) y la correlación más baja en el reactivo 36 ($r = 0.36$), el promedio de correlación de esta área fue de 0.57.

En el área Literario, la correlación más alta se localiza en el reactivo 77 ($r = 0.74$) y la correlación más baja en el reactivo 47 ($r = 0.56$), el promedio de correlación de esta área fue de 0.68.

En el área Ejecutivo-Persuasivo, la correlación más alta se localiza en el reactivo 68 ($r = 0.70$) y la correlación más baja en el reactivo 58 ($r = 0.50$), el promedio de correlación de esta área es igual a 0.62.

En el área Artístico Plástico, la correlación más alta se localiza en el reactivo 89 ($r = 0.75$) y la correlación más baja en el reactivo 29 ($r = 0.53$), el promedio de correlación de esta área fue de 0.66.

En el Área Artístico Musical, la correlación más alta se localiza en el reactivo 80 ($r = 0.82$) y la correlación más baja en el reactivo 30 ($r = 0.65$), el promedio de correlación de esta área fue de 0.72.

Tabla 3
Correlaciones por Reactivo entre la Aplicación Automatizada y la Aplicación
Lápiz-Papel
(ÁREA 1 A 5)

Reactivo	Cálculo	Reactivo	Científico Físico	Reactivo	Científico Biológico	Reactivo	Mecánico	Reactivo	Servicio Social
R1	.58	R2	.57	R3	.60	R4	.59	R5	.64
R11	.75	R12	.65	R13	.52	R14	.56	R15	.67
R21	.43	R22	.54	R23	.57	R24	.46	R25	.50
R31	.73	R32	.64	R33	.78	R34	.68	R35	.69
R41	.66	R42	.66	R43	.73	R44	.74	R45	.64
R51	.69	R52	.66	R53	.70	R54	.54	R55	.24
R61	.74	R62	.71	R63	.62	R64	.73	R65	.56
R71	.84	R72	.57	R73	.74	R74	.66	R75	.77
R81	.68	R82	.63	R83	.57	R84	.66	R85	.44
R91	.85	R92	.62	R93	.69	R94	.57	R95	.48
\bar{X}		\bar{X}	.62	\bar{X}	.65	\bar{X}	.62	\bar{X}	\bar{X}

p < 0.001

Tabla 4
Correlaciones por Reactivo entre la Aplicación Automatizada y la Aplicación
Lápiz-Papel
(ÁREA 6 A 10)

Reactivo	Verbal	Reactivo	Literario	Reactivo	Ejecutivo-Peruasivo	Reactivo	Artístico Plástico	Reactivo	Artístico Musical
R6	.60	R7	.70	R8	.68	R9	.66	R10	.71
R16	.67	R17	.73	R18	.59	R19	.67	R20	.66
R26	.51	R27	.61	R28	.69	R29	.53	R30	.65
R36	.36	R37	.67	R38	.63	R39	.65	R40	.76
R46	.68	R47	.56	R48	.55	R49	.58	R50	.72
R56	.44	R57	.71	R58	.50	R59	.72	R60	.73
R66	.56	R67	.70	R68	.70	R69	.67	R70	.72
R76	.69	R77	.74	R78	.64	R79	.71	R80	.62
R86	.61	R87	.70	R88	.69	R89	.75	R90	.76
R96	.59	R97	.69	R98	.59	R99	.70	R100	.72
\bar{X}	.57	\bar{X}	.68	\bar{X}	.62	\bar{X}	.66	\bar{X}	.72

p < 0.001

VALIDEZ DEL INVENTARIO DE INTERESES PROFESIONALES DE HEREFORD (IIPH)

En las tablas 5a, 5b, 5c, 6a, 6b y 6c se presentan los resultados del análisis factorial con rotación varimax que se aplicó a la matriz de correlación de los reactivos que forman el IIPH en sus dos versiones.

El análisis tuvo como finalidad conocer si la versión automatizada del IIPH conserva la validez de constructo, que se definió para la versión lápiz-papel. Por lo que se esperó que se definieran claramente las diez áreas, como factores independientes en el análisis factorial, con los reactivos específicos de cada área

Los resultados de las tablas 5a, 5b, y 5c muestran que en la versión automatizada sólo se distinguen claramente las áreas Literario, Cálculo, Servicio Social y Mecánico, en menor grado las áreas Artístico Musical, Científico Físico y Artístico Plástico, y las áreas Científico Biológico, Verbal, y Ejecutivo-Persuasivo, presentan mayor dificultad para discriminar el área. Se observa además que los siguientes reactivos presentan carga en más de un factor:

9	16	27	38	47	62	69	87
12	17	28	39	48	63	70	92
13	18	29	41	49	65	76	93
14	19	30	42	50	66	78	96
15	24	37	46	52	69	82	98

Los resultados de las tablas 6a, 6b, y 6c se observa que en la versión lápiz-papel se distinguen claramente las áreas Literario, Servicio Social, Cálculo, Mecánico, Científico Físico, Artístico Musical y Artístico Plástico, en menor grado el área Verbal, y las áreas Científico Biológico y Ejecutivo-Persuasivo, presentan mayor dificultad para

discriminar el área. Se observa además que los siguientes reactivos presentan carga en más de un factor:

6	22	29	40	52	68	82
10	23	32	41	53	72	83
12	24	33	46	56	74	88
13	26	37	47	58	76	95
16	27	38	48	63	78	98
18	28	39	49	66	79	100

Se presenta en seguida una lista de los reactivos que presentan carga en más de un factor en ambas versiones:

12	18	28	38	46	49	66	82
13	24	29	39	47	52	76	98
16	27	37	41	48	63	78	

Es importante mencionar que estos resultados se deben considerar preliminares, ya que la muestra con que se trabajó, es muy pequeña para este análisis (sólo 101 sujetos), cuando el instrumento consta de 100 reactivos.

Sin embargo, resulta interesante observar que aún con esta limitación varias de las áreas se distinguen claramente.

TESIS SIN PAGINACION

COMPLETA LA INFORMACION

Tabla 5a
Resultados del Análisis Factorial para la Validez del Inventario
(Versión Automatizada)

	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4	FACTOR 5	FACTOR 6	FACTOR 7	FACTOR 8	FACTOR 9	FACTOR 10
R1	.65308									
R2										
R3		.33541		.34683		.50319				
R4				.52426						
R5			.63828							
R6	.35424		.55579							
R7	.67452									
R8	.62895									
R9						.37022	.41973			
R10						.70491				
R11		.82697								
R12										
R13	.32655			.33475		.56277		.44044		
R14				.69200						
R15			.52656			.30658				
R16	.52845		.44926							
R17	.73083					.33715				
R18	.30585					.38441				
R19						.40144	.35848			
R20						.73321				
R21		.61648								
R22						.33896				
R23										
R24		.31019		.3913						
R25			.67601							.42844
R26										
R27	.37317					.43368				
R28	.46509									
R29	.34185		.39876							
R30						.50214	.47851			
R31		.72809				.56528				
R32									.63531	
R33									.68964	
R34										
	Literario	Cálculo	S. Social	Mecánico	A. Musical	C. Físico	A. Plástico.	C. Biológico.	Verbal	Elec. Pers.

Tabla 6a

Resultados del Análisis Factorial para la Validez del Inventario
(Versión Lápiz-Papel)

	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4	FACTOR 5	FACTOR 6	FACTOR 7	FACTOR 8	FACTOR 9	FACTOR 10
R1			.47638							
R2				.34882						
R3					.72528					.59625
R4		.65887								
R5		.54351								
R6	.41974									
R7	.74668									
R8	.60818									
R9						.68965				
R10			.85960		.73671	.34518				
R11										
R12				.44055	.43337					
R13				.43165	.40243					
R14			.82460							
R15	.65384									
R16	.46532									
R17	.72664									
R18		.33768								
R19						.63228				
R20										
R21		.63816								
R22		.38621		.42757						
R23			.43632							
R24			.54219					.34214		.33778
R25		.63797								
R26		.37716								
R27	.50938					.33537				
R28	.30969	.50970				.34628				
R29										
R30						.48627				
R31						.72135				
R32		.63173								
R33		.30407								
R34				.81250						
	Literario	S. Social	Cálculo	Mecánico	C. Físico	A. Musical	A. Plástico	Verbal	C. Biología	Ejec.-Pers.

Tabla 6b

Resultados del Análisis Factorial para la Validez del Inventario
(Versión Lápiz-Papel)

	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4	FACTOR 5	FACTOR 6	FACTOR 7	FACTOR 8	FACTOR 9	FACTOR 10
R35		.78780								
R36		.47708								
R37	.69272				.32671					
R38	.48855				.40782	.31896			-.38306	.36359
R39	.41861				.48857					
R40		.74710								
R41				.30003						
R42				.60604						
R43				.71248						
R44					.75101					
R45	.71339									
R46	.32671						.47723			
R47	.30807						.40014			
R48	.30057						.33882			
R49		.31200				.51472				
R60		.85702			.44686					
R61		.50510		.48223				.56728		
R62			.39660							
R63			.78849							
R64										
R65	.32435									
R66	.74158									
R67	.34229						.74453			
R68					.61632					
R69		.80544								
R70				.54459						
R71				.45571						
R72										
R73			.78849					.48548		
R74		.73744								
R75		.32387								
R76	.76073									.41602
R77	.37718									
R78										.57696
	Literario	S. Social	Cálculo	Mecánico	C. Físico	A. Musical	A. Plástico	Verbal	C. Biolog.	Ejec.-Pers.

CAPÍTULO 6

DISCUSIÓN Y CONCLUSIONES

DISCUSIÓN Y CONCLUSIONES

El objetivo principal de esta investigación era conocer si los resultados que ofrece la aplicación automatizada del Inventario de Intereses Profesionales de Hereford (IIPH) eran confiables y válidos, en seguida se describe la manera en que este objetivo fue cubierto.

CONFIABILIDAD Y VALIDEZ

El análisis estadístico de esta investigación demuestra que los resultados de la aplicación automatizada del IIPH presentan una alta confiabilidad, en su consistencia interna.

Es posible afirmar que existe consistencia interna ya que los valores Alpha obtenidos para cada una de las 10 áreas del IIPH en la versión automatizada son mayores a 0.73 y para la versión lápiz-papel son todos mayores a 0.74.

Los resultados obtenidos para la validez de esta investigación son satisfactorios y muy alentadores, tanto para la forma de aplicación como para los resultados que ofrece la versión automatizada del IIPH.

La automatización del IIPH presenta validez de criterio, ya que la correlación monorango-heterométrico que se realizó entre ambas formas de aplicación reporta coeficientes mayores a 0.76 en todas las áreas del inventario, con una probabilidad menor de 0.001. El criterio utilizado para obtener la validez fue la misma prueba en su versión original de lápiz-papel.

Además las correlaciones que se obtuvieron para cada uno de los reactivos reportan una media mayor a 0.57 para todas las áreas del inventario lo cual muestra que es posible afirmar la validez de criterio de la aplicación automatizada.

El análisis que se realizó para probar la validez de constructo no fue tan alentador, ya que la discriminación por áreas sólo se logró claramente para algunas de ellas y no para todas, en ambas formas de aplicación. En este punto es importante señalar que el tamaño de la muestra es demasiado pequeño para los requisitos que exige un análisis factorial (idealmente 1000, mínimo 500 sujetos), por lo que estos resultados deben considerarse como preliminares y señalan la necesidad de hacer este análisis con una muestra mayor.

Estos resultados permiten aceptar la hipótesis de trabajo que se planteó para esta investigación la cual sostiene que los resultados de una aplicación automatizada son confiables y válidos.

Debido a que no existen investigaciones realizadas en el campo de la equivalencia entre ambas formas de aplicación, (Fleishman, 1988; Jäger, 1991; Murphy, 1994), el procedimiento aplicado para esta investigación puede ser considerado como una forma de demostrar dicha equivalencia, este procedimiento permite conocer la confiabilidad y validez de los resultados obtenidos en una aplicación automatizada.

Las ventajas de contar con una versión automatizada de esta naturaleza son, evidentemente, la rapidez de calificación de la prueba misma, manejo simultáneo de grandes volúmenes de información (Tallent, 1987), mayor control en la precisión de la medición (Weiss y Vale, 1987), por lo que los costos de tiempo y recursos se disminuyen de manera importante, si se cuenta con la infraestructura necesaria.

Además, se cuenta con la posibilidad de generar bases de datos que se capturarían de manera automática por el mismo programa, para de esta manera obtener normas adecuadas para las poblaciones de interés. De hecho, la programación misma puede realizarse de manera que las normas de calificación de la prueba puedan ajustarse conforme se van evaluando diferentes grupos de sujetos, lo que redundaría en generar perfiles más precisos.

Asimismo, estas versiones automatizadas de los instrumentos permiten aprovechar las bondades de las computadoras para obtener mediciones adicionales del sujeto como lo son el tiempo de reacción ante cada respuesta (Stout, 1981).

Por supuesto, la finalidad de esta investigación no pretende plantear que el uso de computadora sustituya las funciones del psicólogo como tal, sino más bien que el evaluador aproveche los recursos de la computadora como una herramienta que le permite obtener mediciones rápidas y precisas (también adicionales si se desea), para poder abocarse más a la integración e interpretación de baterías que incluyan la entrevista misma realizada al sujeto.

Finalmente se presentan algunas consideraciones que han de tenerse en investigaciones posteriores.

Las recomendaciones que han de tenerse en cuenta en el desarrollo de la automatización de pruebas son la adecuada elección de los colores, tipo de letra, duración de mensajes informativos, instrucciones, la presentación de reactivos y control en el acceso a los datos.

Los colores que utilizará la prueba deben ser elegidos tomando en cuenta su brillo y luminosidad, ya que algunos colores pueden resultar molestos y cansados a la vista del sujeto examinado y dificultar el proceso de contestar la prueba.

El tipo de letra que desplegará los reactivos deberá tener un tamaño adecuado para poder ser leído por cualquier persona; el utilizar letra demasiado pequeña dificultaría la lectura en algunas personas, de la misma manera el tipo de letra debe ser claro, no se recomienda utilizar tipos de letra que tengan adornos que dificulten su lectura.

La duración de los mensajes informativos que aparecen a lo largo de programa deberá tener un tiempo suficiente como para que una persona sin problemas de lectura lo pueda leer por lo menos dos veces, o en el mejor de los casos se recomienda que el mensaje desaparezca cuando el sujeto lo decida.

Las instrucciones de la prueba deberán estar presentes durante todo el proceso de aplicación, para que el sujeto las pueda consultar en cualquier momento como lo haría en una prueba tradicional de lápiz-papel.

Se recomienda que la presentación de un nuevo reactivo esté determinada por el sujeto, esto es con la finalidad de que el sujeto tome el tiempo necesario para decidir su respuesta (si la prueba lo permite) en cada reactivo.

Una de las recomendaciones más importantes se refiere al control de acceso a la información que proporciona la prueba automatizada, en este punto es necesario destacar que la ventaja que brinda una automatización de calificar la prueba y proporcionar resultados finales debe tener un acceso restringido a psicólogos instruidos en la interpretación de dichos resultados.

ANEXO

TESIS SIN PAGINACION

COMPLETA LA INFORMACION

INVENTARIO DE INTERESES PROFESIONALES DE HEREFORD

Nombre _____ Edad _____ Sexo _____

Estado Civil _____ Escolaridad _____ Grupo _____ Fecha _____

Instrucciones: A continuación encontrará una serie de afirmaciones que representan cien actividades que es posible que le agrade realizar. Por favor, indique cuál es la actividad que le agrada o disgusta colocando el número en el paréntesis que corresponde según su elección de acuerdo a la siguiente escala:

- 1.- ME DISGUSTA MUCHO
- 2.- ME DISGUSTA
- 3.- ME ES INDIFERENTE
- 4.- ME GUSTA
- 5.- ME GUSTA MUCHO

- | | | |
|----|--|-----|
| 1 | Resolver rompecabezas numéricos | () |
| 2 | Tomar fotografías de las fases de un eclipse | () |
| 3 | Hacer análisis de sangre | () |
| 4 | Reparar una licuadora | () |
| 5 | Visitar orfanatos | () |
| 6 | Funcionar como un guía, y explicar cosas de interés a los extraños | () |
| 7 | Escribir cuentos para una revista | () |
| 8 | Participar en debates y discursos | () |
| 9 | Pintar paisajes | () |
| 10 | Aprender a leer música | () |
| 11 | Realizar operaciones aritméticas | () |
| 12 | Recibir un telescopio como regalo | () |
| 13 | Hacer colecciones de plantas | () |
| 14 | Manejar un tomo o taladro eléctrico | () |
| 15 | Colaborar con otros para bien de ellos y de sí mismos | () |
| 16 | Dar la bienvenida públicamente a un visitante | () |
| 17 | Ser escritor de novelas | () |
| 18 | Ayudar a los candidatos políticos | () |
| 19 | Recibir un juego de pinturas de óleo como regalo | () |

- 1.- ME DISGUSTA MUCHO
 2.- ME DISGUSTA
 3.- ME ES INDIFFERENTE
 4.- ME GUSTA
 5.- ME GUSTA MUCHO

- 20 Saber distinguir y apreciar la música ()
 21 Convertir radianes a grados ()
 22 Informarse sobre la energía en clase ()
 23 Aprender a practicar primeros auxilios ()
 24 Hacer dibujos de máquinas ()
 25 Ayudar a encontrar empleo a personas de escasos recursos ()
 26 Hacer elogio de un personaje ()
 27 Leer a los clásicos ()
 28 Hacer campañas estudiantiles ()
 29 Asistir a exposiciones de pintura ()
 30 Tener discos de música clásica ()
 31 Calcular el área de un cuarto para alfombrarlo ()
 32 Hacer una colección en rocas ()
 33 Cuidar un pequeño acuario ()
 34 Instalar un contacto eléctrico ()
 35 Servir como consejero en un club de niños ()
 36 Contar sus experiencias en un grupo ()
 37 Saber distinguir y aprender la literatura ()
 38 Convencer a otros para que hagan lo que tú crees que deben hacer ()
 39 Leer libros sobre arte ()
 40 Escuchar los conciertos en las plazas públicas ()
 41 Usar una regla de cálculo ()
 42 Observar el movimiento aparente de las estrellas ()
 43 Observar las costumbres de las abejas ()
 44 Observar cómo el técnico repara la televisión ()
 45 Leer cuentos a los ciegos ()
 46 Expresarse con facilidad en clase ()
 47 Asistir a la biblioteca en una tarde libre ()
 48 Ser promotor de artículos nuevos ()

- 1.- ME DISGUSTA MUCHO
2.- ME DISGUSTA
3.- ME ES INDIFFERENTE
4.- ME GUSTA
5.- ME GUSTA MUCHO

- 49 Diseñar escenarios para las representaciones teatrales ()
50 Obtener el autógrafo de un músico famoso ()
51 Calcular porcentajes ()
52 Estudiar el espectro luminoso de la luz ()
53 Asistir a una operación médica ()
54 Soldar alambre o partes metálicas ()
55 Escuchar a otros con paciencia y comprender su punto de vista ()
56 Conversar con otros en las reuniones y fiestas ()
57 Participar en concursos literarios ()
58 Defender un punto de vista de alguna persona ()
59 Hacer diseños para tapices ()
60 Asistir a conciertos ()
61 Consultar tablas de logaritmos y raíces ()
62 Visitar una exposición científica ()
63 Leer libros sobre el funcionamiento de los organismos vivos ()
64 Observar cómo los mecánicos hacen reparaciones de coches ()
65 Visitar casas humildes para determinar qué es lo que necesitan ()
66 Entrevistar a los solicitantes de empleo ()
67 Corregir composiciones o artículos periodísticos ()
68 Ser líder de un grupo ()
69 Dibujar o delinear personas o cosas ()
70 Escuchar la música de diferentes países como la India, el Japón, etc. ()
71 Ayudar a otras personas en problemas matemáticos ()
72 Estudiar los cambios de clima y sus causas ()
73 Hacer colecciones de insectos ()
74 Desarmar y armar un reloj ()
75 Cuidar a mis hermanos menores ()
76 Hablar con extraños en el camión o tren ()

- 1.- ME DISGUSTA MUCHO
- 2.- ME DISGUSTA
- 3.- ME ES INDIFFERENTE
- 4.- ME GUSTA
- 5.- ME GUSTA MUCHO

77	Escribir reseñas críticas de un libro	()
78	Organizar y dirigir festivales, excursiones o campañas sociales	()
79	Saber distinguir y apreciar la pintura	()
80	Tomar parte en un conjunto coral	()
81	Ilustrar problemas geométricos con la ayuda de las escuadras, la regla T y el compás.	()
82	Visitar un observatorio astronómico	()
83	Cultivar plantas exóticas	()
84	Amar o componer muebles comunes	()
85	Participar en campañas contra la delincuencia juvenil	()
86	Jugar juegos de palabras con mis amigos	()
87	Escribir cartas narrativas a mis amigos o parientes	()
88	Dirigir un grupo o equipo en situaciones difíciles	()
89	Hacer un proyecto de decoración de interiores	()
90	Tocar un instrumento musical	()
91	Resolver problemas matemáticos.	()
92	Experimentar con la necesidad de oxígeno para la combustión	()
93	Observar a menudo como transportan las hormigas su carga	()
94	Observar las máquinas cuando son montadas	()
95	Impartir conocimientos a aquellas personas que no los tienen	()
96	Estudiar una lengua extranjera.	()
97	Escribir artículos en el periódico	()
98	Mostrar un nuevo producto al público	()
99	Hacer mosaicos artísticos para decoración	()
100	Ser compositor de música	()

REFERENCIAS BIBLIOGRÁFICAS

- Acevedo M., Zarabozo D. (1993). Un programa para la calificación automatizada del Inventario Multifásico de la Personalidad de Minnesota (MMPI). Revista Mexicana de Psicología. Vol. 10. págs. 195- 202.
- Aiken, L.R. (1995) Psychological Testing and Assessment. 8ª Ed. Allynd and Babon U.S.A.
- Akermann, A. (1964). Psicología aplicada. Ed. Morata México.
- Anastasi A. (1978). Test Psicológicos. Ed. Aguilar. España.
- Angelini, A: (1984). Inventario de intereses. Ed. Trillas México
- Beaumont, J. G. (1981). Microcomputer aided assesment using standard pshometric procedures. Behavior Research Methods & Instrumentation. Vol.13 (4). pág. 430.433.
- Brown F.G. (1980). Principios de la medición en psicología y educación. Ed. Manual Moderno.
- Cerdá, E. (1984). Psicometría General. Ed. Herder. Barcelona.
- Fernández, G. (1970). El inventario de intereses personales Hereford. Revista Interamericana de Psicología. Vol.4 (1).págs.
- Guevara M.A. y col. (1993). Un sistema para la adquisición y el análisis automatizado de potenciales evocados en humanos. Revista Mexicana de Psicología. Vol. 10. págs. 189-194.
- Hernández R., Fernández C. Baptista P. (1991). Metodología de la Investigación. Ed. Mc.Graw Hill, México.

- Jäger, S. (1991). Computer Diagnostics -A Survey Practical Applications of computerized Assessment Teoretical Principles and Perspectives-. Revue Européenne de Psychologie Appliquée. Vol. 41 (4). pág.247-268.
- Kerlinger F.(1988) Investigación del Comportamiento. Ed. McGrawHill, México.
- Magnusson, D. (1969). Teoría de los test. Ed. Trillas. México.
- Marsá, F. (1990). Diccionario Planeta de la lengua española usual. Ed.Planeta Barcelona.
- Morales, M.L. (1975). Psicometría aplicada .Ed. Trillas. México
- Morgan, H. y Cogger J.W. (1975). El manual del entrevistador. Ed. Manual Moderno. México.
- Murphy K. R. y Shofer D. (1994). Psychological Testing: Principles and Applications. 3ª Ed. Prentice-Hall. New Jersey. USA.
- Nunnally, J.C. (1970). Introducción a la medición en psicología. Ed. Paidos. Buenos Aires.
- Nunnally, J.C. (1995). Teoría Psicométrica. Ed. McGraw Hill. México.
- Pichot, P. (1976). Los test mentales. Ed. Paidos. Buenos Aires.
- Pick de Weiss S. (1986). Como investigar en Ciencias Sociales. Ed. Trillas. México.
- Sattler, J. (1984). Evaluación de la inteligencia infantil y habilidades especiales. Ed. Manual Moderno. México.
- Super, D (1967). Psicología de los intereses y las vocaciones. Ed. Kapelusz. Buenos Aires.
- Super, D. (1973). Psicología ocupacional. Ed. Cía. Editorial Centinal Sn. México.

- Stout, R. L. (1981). New approaches to the design of computerized interviewing and testing systems. Behavior Research Methods & Instrumentation. Vol. 13 (4). pág. 436-442.
- Szekely, B. (1960) Los test: Manual de técnicas de exploración psicológica. 4ª Ed. Ed. Kapelusz. Buenos Aires.
- Talent, N. (1987) Computer-Generated Psychological Reports: A Look at the Modern Psychometric Machine. Journal of Personality Assessment. Vol 51(1). Pág. 95-108.
- Tyler L. (1972). Pruebas y medición en Psicología. Ed. Prentice/Hall Internacional. Colombia.
- Vale, D. (1981). Design and implementation of a microcomputer-based adaptive testing system. Behavior Research Methods & Instrumentation. Vol. 13 (4). pág. 399-406.
- Valencia, R. (1972). Estudio psicométrico de dos pruebas vocacionales. Tesis de Licenciatura. Facultad de Psicología. UNAM. México.
- Weiss, D. y Vale, D. Computerized Adaptive Testing for Measuring Abilities and Other Psychological Variables. En Butcher, J. (1987). Computerized Psychological Assessment. Ed. Basic Books, Publishers. New York. USA. Cap. 16.
- Yoffe, H. (1988). Relación y correlación de valores e intereses en adolescentes. Tesis de Licenciatura. Facultad de Psicología. UNAM. México.