

7
2e)

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE PSICOLOGIA



DETERMINACION DE LA VALIDEZ DE LA PRUEBA DE RAVEN EN UNA MUESTRA DE NIÑOS QUE ASISTEN A ESCUELAS SECUNDARIAS PARTICULARES DE LA CIUDAD DE MEXICO

T E S I S
QUE PARA OBTENER EL TITULO DE
LICENCIADO EN PSICOLOGIA
P R E S E N T A N
ELVIRA ANGELICA ALFARO SANCHEZ
SARA RODRIGUEZ MATA

DIRECTORA: MTRA. CRISTINA HEREDIA

MEXICO, D. F.

1994

TESIS CON
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.

"Qué suerte he tenido de nacer,
para estrechar las manos de un amigo
y poder asistir como testigo
al milagro de cada amanecer.

Qué suerte he tenido de nacer,
para tener la opción de la balanza,
sopesar la derrota y la esperanza
con la gloria y el miedo a caer.

Qué suerte he tenido de nacer,
para entender que el honesto y el perverso,
son dueños por igual del universo
aunque tengan distinto parecer.

Qué suerte he tenido de nacer,
para callar cuando habla el que más sabe,
aprender a escuchar, esa es la clave,
si se tiene intenciones de saber.

Qué suerte he tenido de nacer,
y lo digo sin falsos triunfalismos
la victoria total, la de uno mismo,
se concreta en el ser y en el no ser.

Qué suerte he tenido de nacer,
para contarlo a la gente y a la rosa
y al perro y al amor y a cualquier cosa
que pueda el sentimiento recoger."

A. Cortez / C. Gentile

A la maestra Ma. Cristina Heredia S.
Directora de Tesis

Nuestro agradecimiento por el apoyo e interés
que siempre mostró a lo largo de esta investigación.

E. Angélica y Sara

A la Lic. Lourdes Monroy J.
Profesora de la Unidad de Cómputo, Informática
e Instrumentación de la Facultad de Psicología

En reconocimiento por la ayuda brindada para
elaboración de este trabajo.

E. Angélica y Sara

A Roberto y Avelino

*Por que gracias a ustedes somos capaces
de recorrer nuevos caminos.*

E. Angélica y Sara

A Ismael

*Por toda su paciencia y apoyo incondicional
al compartir sus conocimientos
en la concepción de este trabajo.*

E. Angélica y Sara

Para Angela

*Toda nuestra gratitud por que con su valioso
apoyo hizo posible la realización de este trabajo.*

E. Angélica y Sara

*En tu amistad he encontrado
pan en la austeridad
compañía en la soledad
aliento en mi empeño
consuelo a mi tristeza
regocijo en mi alegría
comprensión a mis errores y
respuesta a mis preguntas
por esto y más
cuentas conmigo.*

A mis hijas

*Por permitirme conocer el milagro de la creación
y así poder tocar el cielo con mis manos.*

*Por ser el aliento y razón de mi vida,
todo mi amor para Iris y Vero.*

E. Angélica

A David

*Por que tu amor ha permitido
que haya espacio en nuestra unión
por la confianza que me has brindado
y el regocijo ante mis logros
todo mi cariño para ti.*

E. Angélica

A mis padres Manuel y Elvira

*Por que con su ejemplo me han enseñado
a luchar por mis anhelos.*

*Por el cariño, estímulo y confianza,
gracias por siempre.*

E. Angélica

*A mis hermanos
Manuel, Lusi, Mario, Elvira, Arturo y Guillermo*

*Por compartir conmigo sus alegrías
por siempre estar a mi lado y
enorgullecerse de mis triunfos.*

E. Angélica

Para mis hermanos
Alberto, Tina, Claudia y Alma

Por que sin su apoyo nunca hubiera encontrado
la tan ansiada libertad
aquella que va más allá de lo esperado y permitido
aquella por la cual pueda ser como soy y lo que soy.

Dara

Para mi papá

Porque traer en mi sangre tu sangre
me ha llevado a recorrer los caminos
en busca de la ilusión.

Dara

En especial para ti

*Por que sin importar de donde y cuándo llegue
siempre has mantenido
encendida la luz del sendero
que me ha permitido
encontrar el camino de regreso
Gracias mamá*

Sara

*A todos aquellos que de una u otra forma
hicieron posible, con su apoyo y estímulo,
la realización de este trabajo.*

Gracias

Angélica y Sara

INDICE

INTRODUCCIÓN	1
CAPÍTULO I FUNDAMENTOS TEÓRICOS DE LA MEDICIÓN	5
I.1	Medición.....	5
I.2	Problemas generales de medida.....	6
I.3	Escalas de medición y valor escalar.....	7
I.4	Confiabilidad.....	11
I.5	Validez.....	16
I.6	Medición en Psicología.....	22
I.7	Estandarización.....	28
CAPÍTULO II MEDICIÓN DE LA INTELIGENCIA	37
II.1	Desarrollo del Constructo de Inteligencia.....	37
II.2	Desarrollo de las Pruebas de Inteligencia.....	42
II.3	Teorías sobre la estructura intelectual.....	47
CAPÍTULO III PRUEBAS DE INTELIGENCIA	59
III.1	Escala de Inteligencia Revisada para el Nivel Escolar - WISC-R.....	59
III.2	Escala de Inteligencia Revisada para el Nivel Escolar-WISC-RM.....	67
III.3	Estudios realizados en México con las escalas de Wechsler.....	70
III.4	Prueba de Matrices Progresivas de Raven.....	79
III.5	Estudios realizados en México con la prueba de Raven.....	84
CAPÍTULO IV METODOLOGÍA	91
IV.1	Planteamiento del problema.....	92
IV.2	Hipótesis.....	93
IV.3	Diseño.....	96
IV.4	Variables.....	96
IV.5	Descripción de la muestra.....	97
IV.6	Escenario.....	98
IV.7	Instrumentos.....	98
IV.8	Procedimiento.....	100
CAPÍTULO V RESULTADOS Y DISCUSIÓN	105
CAPÍTULO VI CONCLUSIONES	131
CAPÍTULO VII LIMITACIONES Y SUGERENCIAS	137
	Bibliografía.....	139

INTRODUCCIÓN

El papel del psicólogo en el ámbito educativo, es de suma importancia, tiene la responsabilidad -entre muchas otras- de determinar el perfil psicológico del estudiante, ésto a través de la aplicación de diversas pruebas psicométricas. Labor que se ve obstaculizada en la práctica por la gran demanda del servicio.

Generalmente hay un solo psicólogo por cada escuela, con una población aproximada de entre 100 y 150 alumnos, como mínimo, debido a que el presupuesto destinado para esta labor en comparación con otras actividades es mínimo.

En este contexto, el psicólogo hecha mano de diversas herramientas que le permitan atender a la mayor población en el menor tiempo posible utilizando en la mayoría de los casos pruebas psicológicas de aplicación accesible, sin tomar en cuenta si el instrumento utilizado, esta estandarizado para la población en estudio.

Un ejemplo de ésto es la prueba de Matrices Progresivas de Raven, la cual es ampliamente utilizada, tanto en el ámbito educativo como laboral. Situación por demás equívoca, si tomamos en cuenta "... que los resultados de una prueba son únicamente válidos para los grupos que representan la población muestreada" (Morales 1991).

No obstante el uso frecuente en nuestro país, es de llamar la atención las pocas investigaciones que se han llevado a cabo con el mismo, entre ellos se encuentra un solo estudio que tiene como finalidad la estandarización de dicha prueba en una población mínima de adolescentes infractores (Fuentes y López 1988).

La prueba de Raven se fundamenta en la teoría bifactorial de Spearman, quien considera que la inteligencia está formada por múltiples habilidades y que todas las habilidades del hombre tienen un factor común o general a todas ellas, FACTOR G, y un factor específico, FACTOR E, a cada una de ellas.

En algunas habilidades el factor G es el principal y en otras lo es el factor E, también existen factores de grupo que son comunes a muchas de las habilidades.

La prueba de Raven es una prueba no verbal diseñada para valorar la capacidad intelectual del sujeto al establecer relaciones y correlatos mediante figuras y diseños geométricos, puede aplicarse en forma individual o en forma colectiva y tiene la ventaja de calificarse en forma muy rápida debido a que la respuesta que se le pide al sujeto examinado debe ser seleccionada de una serie de opciones.

Desafortunadamente al tratar de utilizar esta prueba en ámbitos escolares, no tenemos la certeza de que sea válida para nuestra población, a pesar de estar estandarizada en varios países sudamericanos.

García (1987) señala que existen evidencias empíricas de que resulta injusto para los

niños mexicanos ser evaluados con instrumentos no estandarizados ya que pueden ubicarse en una categoría descriptiva inferior.

En un intento de superar esta problemática se estandarizó en México para la población de primaria y secundaria oficiales del Distrito Federal el WISC-RM. Sin embargo, ésta resulta una prueba poco práctica ya que al ser de tipo individual requiere de mayor tiempo tanto para su aplicación como para su calificación, lo que la hace poco atractiva a los psicólogos que se desenvuelven en ámbitos escolares con gran cantidad de alumnos a los que hay que evaluar.

De tal manera, cuando intentamos evaluar alumnos de escuelas particulares nos enfrentamos ante la problemática de no disponer de un instrumento que sea válido y que nos proporcione las características que se requieren de acuerdo al tiempo que se dispone para dicha evaluación.

En virtud de lo anterior y tomando en cuenta los fundamentos teóricos de Spearman quien señala que el factor G es un factor cuantitativo de la inteligencia, el común y fundamental a todas las funciones cognitivas del mismo individuo (Raven 1988), se plantea la siguiente pregunta de investigación al aplicarse las pruebas WISC-R, WISC-RM y Raven, a una muestra de estudiantes de dos escuelas secundarias particulares.

¿Cuál es el nivel de validez concurrente entre la escala de Matrices Progresivas de Raven y las escalas de Wechsler WISC-R y WISC-RM?

Para responder a esta pregunta se llevó a cabo una investigación con alumnos de escuelas secundarias particulares, a quienes se aplicaron las pruebas señaladas.

La muestra fue dividida en dos grupos, aplicándole a ambos la prueba de Matrices Progresivas y posteriormente a cada uno de los grupos una de las escalas de Wechsler, el WISC-R o el WISC-RM.

En la investigación realizada se utilizó tratamiento estadístico de correlación lineal para conocer la validez concurrente en donde la variable de criterio estuvo representada por las escalas de inteligencia de Wechsler, WISC-R y WISC-RM, y la prueba de Matrices Progresivas de Raven fue la prueba a la cual se determinó la validez para la población señalada.

El interés de este estudio se originó por la necesidad de contar con una prueba de inteligencia válida, que pudiera ser aplicada en forma colectiva en el ámbito educativo privado. Anastasi (1978) enumera las ventajas de utilizar este tipo de pruebas, entre las que destacan:

- Se pueden aplicar simultáneamente a tantas personas como se puedan acomodar confortablemente en el espacio disponible.
- Debido a que la misión del examinador se reduce al mínimo, las pruebas

colectivas pueden proporcionar condiciones más uniformes que las individuales.

- La puntuación es característicamente más objetiva ya que proporciona reglas mejor establecidas y se emplean elementos de elección múltiple.
- Debido a que en estas pruebas las contestaciones se escriben en las hojas de respuestas, los folletos se pueden utilizar indefinidamente, lo que supone una economía considerable.

CAPITULO I

FUNDAMENTOS TEÓRICOS DE LA MEDICIÓN.

I.1 MEDICIÓN.

En la vida, la palabra medición tiene un significado claro y conciso. para medir en situaciones prácticas, generalmente disponemos de instrumentos que nos dan resultados fijos y precisos en forma de puntajes, tal es el caso de todos los instrumentos con los que medimos longitud, peso, espacio, tiempo, etc. donde los resultados se dan en centímetros, gramos, decilitros o segundos. medidas de clase, que se obtienen con instrumentos físicos, casi nunca presentan problemas prácticos en el momento de medir, ni al interpretar los resultados.

La definición de medida más común según Magnusson (1990): es asignar números a las cantidades de las propiedades de los objetos de acuerdo con reglas dadas cuya validez puede probarse empíricamente. Dicho en forma más simple, medir es dar la magnitud de cierta propiedad de uno o más objetos con ayuda del sistema numérico.

Wilfor (1954) dice que la medición es la descripción de datos en términos de números; Stevens (1951) define a la medición como la asignación de números a objetos o eventos, de acuerdo a reglas específicas, ambos autores son citados en Brown (1980).

La tarea más ardua e interesante de la medición es la determinación de la regla, que es una guía, un método o una ordenanza que nos señala lo que hemos de hacer.

Kerlinger (1985) menciona que el primer y más elemental paso en cualquier procedimiento de medición consiste en definir los objetos que integran el universo de discurso.

La forma más sencilla de medición sería clasificar o categorizar a todos los objetos en base a la posesión o ausencia de alguna característica.

La gente pasa gran parte de su vida clasificando cosas, hechos y personas; la vida no sería posible sin tal clasificación, pero se antoja difícil vincular este proceso con la medición. Una vez terminada la clasificación del universo de los subconjuntos designados, se cuentan los miembros de cada conjunto.

Para poder determinar correctamente las reglas que deberán tomarse en cuenta en la asignación de numerales a un objeto o evento, es importante tener en cuenta los postulados básicos de la medición; empecemos por conocer la definición de postulado:

Postulado es una suposición que constituye un requisito esencial para realizar alguna operación o línea de pensamiento. En este caso se refiere a las relaciones que existen

entre los objetos que están siendo medidos. Aunque suele suponerse la veracidad de los postulados, en la medición hay que someterlos a prueba siempre que sea posible:

Postulados de medición.- Hay cuatro niveles generales de medición: nominal, ordinal, de intervalo y de razón. Todos ellos conducen a cuatro clases de escalas.

Primer postulado: A es igual o diferente a B pero no ambas cosas; este postulado es indispensable para la clasificación, debemos tener la capacidad de afirmar que un objeto es igual a otro en lo que respecta a una característica o bien que no lo es.

En la medición el adjetivo igual no denota necesariamente identidad absoluta, puede significar igualdad suficiente como para ser clasificado con los miembros del mismo conjunto; para poder afirmar que los dos son iguales, necesitamos un criterio o un conjunto de criterios; el criterio tienen que ser lo bastante definido para ser posible la clasificación, es decir, para cumplir con la condición que establece el postulado.

Segundo postulado: Si A es igual a B y si B es igual a C, A será igual a C (de transitividad). Este postulado permite al investigador establecer la igualdad de los miembros de un conjunto en base a una característica mediante la comparación de los objetos. Y lo que es más importante aun, si el postulado se satisface, los objetos que normalmente no son susceptibles de observación podrán asignarse a los subconjuntos de un universo.

Tercer postulado: Si A es mayor que B y si B es mayor que C, A será mayor que C (de transitividad), en dicho postulado se fundan la mayoría de las mediciones psicológicas. Hemos de estar en condiciones de hacer enunciados de tipo ordinal o de ordenación por rangos, tales como A posee cierta propiedad en mayor grado que B; y B posee esa propiedad en mayor grado que C; por tanto A la posee en grado mayor que C.

1.2 PROBLEMAS GENERALES DE MEDIDA:

La calidad de los procedimientos de medición empleados se reflejará en la calidad de los resultados de la investigación que nos formulemos, de ahí que sea de suma importancia la formulación adecuada de las cuestiones a medir o investigar y las definiciones claras de los conceptos implicados.

Selltiz, dice que debe conocerse en primer lugar que es lo que se quiere medir, el concepto a que apunta la medida se denomina concepto - objetivo.

Al respecto Kerlinger (1985), menciona que "Se ha dicho que a veces la medición carece de sentido, para evitar caer en este error es preciso que no se deslinden de la realidad la definición de los conjuntos de objetos en cuestión y la definición de los conjuntos numéricos a partir de los cuales asignamos numerales a los objetos ni las reglas de asignación o correspondencia. Se dice que el método de medición y el sistema numérico son isomórficos con respecto a la realidad lo que significa que hay una identidad o semejanza de forma. La interrogante fundamental que debe plantearse

todo método es si existe isomorfismo entre el procedimiento y la realidad; El único problema es que rara vez descubrimos con tanta facilidad el grado de correspondencia que guardan con la realidad nuestras mediciones.

Se afirma que medimos objetos, pero eso no es del todo cierto; lo que medimos son los indicadores de las propiedades, por lo tanto al decir que medimos objetos lo que realmente queremos dar a entender es que medimos los indicadores de sus propiedades.

Es imposible observar de manera directa la hostilidad, la moral, la angustia, la inteligencia, la creatividad y el talento; siempre hemos de inferir estas propiedades o características a partir de la observación de sus supuestos indicadores."

¿Qué se pretende medir? es un requisito previo a la decisión de como va a ser medido, es decir, al establecimiento de procedimientos de medida.

Un procedimiento de medida consiste en una técnica de recogida de datos más una serie de normas para la utilización de estos datos.

El objetivo de las diversas técnicas de recogida de datos es el de obtener evidencia fidedigna que sea significativa para las preguntas de investigación que van a ser formuladas.

Los datos pueden ser recogidos de muy diversas formas: por observación de conducta, por cuestionarios o entrevistas, por técnicas psicométricas, por el examen de registros existentes, etc. Las reglas para la utilización de tales datos como medios de elaboración de afirmaciones acerca del fenómeno en el que se está interesado pueden ser elaboradas dentro de la misma técnica de recogida de datos, o puede ser llevada a cabo como un complemento de la misma.

1.3 ESCALAS DE MEDICIÓN Y VALOR ESCALAR.

Brown (1985) dice que para poder entender la precisión de las mediciones, es necesario tomar en consideración la naturaleza de la escala utilizada. Dependiendo de las suposiciones matemáticas y lógicas que se hagan, resultan posibles varios tipos de escalas.

Esos niveles de escala son jerárquicos ya que las de nivel superior satisfacen todos los requisitos de las de nivel inferior más otros adicionales que caracterizan a ese nivel dado.

Del orden inferior al superior, de lo más simple a lo más complejo, las escalas se llaman, nominal, ordinales, de intervalos y de razón de proporción.

Sattler (1988) menciona que hay varios métodos para ordenar datos, los cuales pueden ser agrupados en cuatro tipos de sistemas (escalas) o reglas de asignación de valores

a características o rasgos mensurables y que a partir de estos se definen la clase de escalas y el nivel de medición al que se quiere llegar.

Las escalas las clasifica al igual que Brown.

Magnusson (1990) nada más diferencia tres escalas de medición o niveles de medida, las cuales difieren en la cantidad de información llevada por los números que representan las magnitudes de las cualidades.

Los números pueden dar dichas magnitudes sobre una escala ordinal, una escala de intervalo o una escala de proporción.

Escalas nominales:

En una escala nominal, se les asigna un número o nombre a cada cosa que se está midiendo, estos números o nombres por lo general representan categorías que se excluyen mutuamente y no pueden organizarse en forma ordenada.

Brown (1985) indica que la operación fundamental es la de determinar si hay dos personas que sean miembros de la misma categoría o clase; o sea si poseen características comunes.

Las cifras se utilizan sólo con fines de identificación, para denotar la pertenencia a cierta clase; no implican magnitud.

Este tipo de medición en una escala nominal tiene una utilidad limitada, puesto que sólo permite efectuar una clasificación, mientras que, en la mayoría de los casos, nos interesa tener alguna estimación de la magnitud.

Kerlinger (1985) al respecto dice que el nivel de medición más bajo es el nominal donde las cifras asignadas son de índole numérica pero sin el significado de número; no se las puede ordenar ni sumar; son denominaciones muy semejantes a las letras que sirven para designar conjuntos, los símbolos dados a los objetos, o más bien a grupos de objetos, constituyen escalas nominales.

Puesto que se satisface el primer postulado de medición (A igual o diferente de B) y es posible comparar y contar los miembros de los conjuntos nombrados, entonces cabe decir que los procedimientos nominales son mediciones.

Escala ordinal:

Al nivel de la escala ordinal los números proporcionan solamente el orden de los objetos con respecto al rasgo que se mide.

Cuando conocemos las posiciones de los objetos en una escala ordinal podemos indicar sus relaciones entre sí por medio de los signos mayor que, igual a, menor que.

Las escalas ordinales clasifican a la persona en alguna dimensión. Por ejemplo, podemos clasificar a los colegiales según su altura, del más alto al más bajo; sin embargo, no podremos saber hasta que punto es más alto uno de los niños que el otro; sólo será posible comparar sus alturas relativas.

La medición ordinal exige que los objetos de un conjunto sean ordenados por rangos según una característica o propiedad definida en términos operacionales. Si se justifica el postulado de transitividad ordinal, será posible esta medición.

El método puede generalizarse de tres maneras: la primera es que cualquier número de objetos de toda clase puede medirse ordinariamente con la sola extensión A, B, C, ... N; la segunda generalización consiste en valerse de propiedades o criterios combinados, en vez de una sola propiedad se usan dos o más; la tercera extensión se logra al aplicar otros criterios que no sean el de mayor que, o el criterio menor que, pueden emplearse criterios como precede, esta arriba o es superior a.

Los números ordinales indican el orden por rangos y nada más, no indican cantidades absolutas, ni que los intervalos entre los números sean iguales.

Por ejemplo, una escala ordinal dirá, quien fue primero, segundo o tercero, pero no se podrá saber la distancia entre el primero y el último.

Magnusson (1990) dice que cuando medimos variables psicológicas, generalmente no llegamos a un nivel superior al de la escala ordinal sin que hagamos algunas suposiciones. La medición en este nivel, sin embargo, da suficiente información en muchas situaciones de importancia práctica para la psicología ya que existe una metodología estadística para tratar datos ordinales.

Escala de intervalos :

En una escala de intervalos, los números también dan información acerca del tamaño de las diferencias entre los objetos con respecto a la magnitud del rasgo medido, o sea para poder medir al nivel de una escala de intervalo es necesario tener unidades iguales en la escala, Magnusson (1990).

Permite conocer no sólo la dirección del continuo, sino también tenemos unidades iguales en la región del continuo donde los individuos toman sus posiciones y donde queremos hacer las mediciones. Por lo tanto podemos determinar las diferencias entre las posiciones de los objetos y compararlas entre sí.

Brown (1985) dice que en una escala de intervalos, una diferencia de cierta magnitud significa lo mismo en todos los puntos de la escala. O sea las unidades de clasificación se pueden demostrar que son equivalentes en todos los puntos de la escala. Así, una escala de intervalos implica clasificación, magnitud y unidades de tamaño iguales.

Sattler (1988) dice que se obtiene una unidad específica de medición, que es de tal naturaleza, que la distancia o diferencia entre cualesquier dos números adyacentes es idéntica a la de cualesquier otros dos números.

Esta técnica de medición tiene que especificar un procedimiento para determinar los intervalos idénticos, pero no ofrece ninguna operación que permita la definición de cero.

Las escalas intervalares o de intervalos iguales poseen las características de las escalas nominales y de las ordinales, en particular, la que se refiere al orden por rangos. Además, en ellas las distancias numéricamente iguales representan distancias iguales con respecto a la propiedad que se está midiendo. Otro modo de expresar la idea de intervalos iguales consiste en afirmar que éstos pueden ser sumados y restados.

Magnusson (1990) al respecto dice que el significado de las expresiones "unidades iguales" e "intervalos iguales" no es muy clara, para demostrar esto menciona el siguiente ejemplo: "si los individuos A, B, C y D, saltadores de longitud, saltan 4.25, 4.50, 8.00 y 8.25 metros, respectivamente, podemos decir de inmediato que en sentido puramente métrico la diferencia entre los saltos de A y B es igual a la diferencia entre los saltos C y D. sin embargo, difícilmente podríamos concluir que las diferencias de los saltos son la misma en cualquier otro sentido que no sea el métrico.

Kerlinger (1985) dice que a los psicólogos les interesa obtener escalas de intervalos por dos razones:

1. Las calificaciones se pueden transformar en cualquier tipo de calificación lineal, esto significa que se puede sumar o restar una constante, o bien, multiplicar o dividir las calificaciones por una constante, sin destruir las relaciones existentes entre resultados. En consecuencia, las calificaciones de una escala se pueden convertir a las de otra que utilice unidades diferentes.
2. Una escala de intervalos es de las estadísticas que más se utilizan.

Pueden interesar las diferencias entre los individuos en el mismo continuo, los cuales reciben el nombre de diferencias interindividuales o bien las diferencias entre las posiciones de un solo individuo en diferentes continuos, llamadas diferencias intraindividuales.

Escala de proporción o de razón:

En el nivel de la escala de proporción los números dan información, no sólo del orden del rango de los objetos y del tamaño relativo de las diferencias, sino también de la relación entre las proporciones, Magnusson (1990).

La medición al nivel de la escala de proporción esta basada en la suposición que se conoce el punto cero, y que se tienen unidades iguales a partir de este punto a todo lo largo de la escala.

Este tipo de escalas (medición de razón) constituyen el nivel más alto de medición que es el ideal científico.

Una escala de medición de razón, además de poseer las características de las escalas anteriores, tiene un cero absoluto o natural provisto de significado empírico. Si en ella una medición es cero, existen razones para decir que un objeto carece totalmente de la propiedad en cuestión. Hay un cero absoluto o natural, y por ello son posibles todas las operaciones aritméticas tales como la multiplicación y la división. Los números escalares indican los grados completos de la propiedad que se está midiendo.

Brown (1985) dice que puesto que es virtualmente imposible definir un punto cero absoluto significativo para la mayoría de las características psicológicas y puesto que las escalas de intervalo son suficientes para la mayoría de los fines, no se ha estado interesado en desarrollar pruebas que midan sobre escalas de razón.

Con el fin de poder entender la precisión de las mediciones psicológicas es necesario tomar en consideración la naturaleza de la escala utilizada.

Luego de asignar numerales a objetos o eventos según ciertas reglas, es importante conocer que tan confiable es el instrumento de medición.

1.4 CONFIABILIDAD.

Las técnicas de recogida de datos y las normas para su utilización, para ser útiles, deben ofrecer una información que no solamente sea significativa, sino correcta.

Dos aspectos decisivos de la corrección son la fiabilidad o confiabilidad (es decir, el grado en que las medidas proporcionan resultados consistentes) y la validez (es decir el grado en que éstos corresponden a la posición verdadera de la persona u objeto en la característica que se esta midiendo).

Definición de confiabilidad:

Kerlinger (1985) dice que para que los datos obtenidos con diferentes tipos de instrumentos de medición puedan usarse en situaciones prácticas, éstos deben satisfacer ciertas condiciones:

Primero, el instrumento de medida que se usa en un caso y con un propósito dados debe realmente medir el rasgo que se intenta medir.

Segundo, el instrumento debe dar medidas confiables, de manera que se obtengan los mismos resultados al volver a medir el rasgo, bajo condiciones similares del objeto o

individuo en cuestión.

Por confiabilidad Sattler (1988) mencionando a Anastasi (1976) se refiere a la consistencia de los puntajes obtenidos por la misma persona cuando se le examina en diferentes ocasiones con la misma prueba, u otros conjuntos de reactivos equivalentes, o en diversas situaciones de prueba.

Los datos deben ser confiables desde dos puntos de vista: deben ser significativos y reproducibles Magnusson (1990).

Los resultados obtenidos con el instrumento en una determinada ocasión, bajo ciertas condiciones, deben ser reproducibles, es decir, los resultados deben ser los mismos si volvemos a medir el mismo rasgo en condiciones idénticas. Este aspecto de la exactitud de un instrumento de medida es su confiabilidad en la acepción técnica de este instrumento, aquí la confiabilidad es la exactitud de la medición, independientemente de que uno este realmente midiendo lo que ha querido medir.

Sattler (1988) dice que la confiabilidad se expresa por el coeficiente de confiabilidad, o error estándar de medida derivado del mismo.

Una prueba cuya confiabilidad es muy baja, simplemente no genera confianza.

La confiabilidad se ocupa de la pregunta ¿Hasta que punto será consistente la ejecución de una persona si toma otra forma de prueba psicológica (equivalencia) o si toma la misma forma en un momento diferente (estabilidad).

La mayoría de los índices de confiabilidad se expresan como coeficientes de correlación y, por ende, se denominan coeficientes de confiabilidad. Estos coeficientes no son mas que la correlación entre dos conjuntos de calificaciones, obtenidos de una misma muestra de personas y que se utiliza como índice de consistencia de la medición.

Según Magnusson (1990) el grado de acuerdo entre medidas hechas en diferentes ocasiones puede computarse por medio de los métodos de correlación.

El coeficiente de correlación para el acuerdo entre medidas repetidas bajo condiciones similares, constituye el valor numérico de la confiabilidad de los datos que pueden obtenerse con un instrumento dado.

Este coeficiente de correlación es llamado coeficiente de confiabilidad, y puede tomar valores entre cero y uno pero no puede ser negativo.

Si el instrumento es insensible a los factores debidos al azar, los puntajes del individuo en medidas sucesivas serán idénticos, y sus posiciones en las distribuciones que podemos construir en cada ocasión de medida serán las mismas, por lo que la correlación entre las distribuciones será 1.0. las medidas que pueden hacerse con tal

instrumento son completamente confiables.

Correlación y predicción:

Magnusson (1990) la describe como la relación que hay entre variables, esta relación es reportada en grados y se expresa por un coeficiente de correlación.

El procedimiento en forma simplificada es asignar un puntaje para cada una de las dos variables y estudiar su correlación.

El coeficiente de correlación indica el grado que los puntajes en una de las variables mantiene una relación lineal sistemática con los puntajes en la otra.

La predicción a través del conocimiento de la relación entre dos variables es posible por ejemplo, si podemos verificar que la relación obtenida en un grupo se mantiene también en el otro grupo.

Cuando es expresado el grado de la relación entre dos variables como coeficiente de correlación y se conoce el valor para un cierto objeto para una de las variables, podemos predecir el valor más probable de la otra.

La exactitud de la predicción que puede hacerse de una variable a partir de la otra, aumenta con el grado de relación entre los puntajes; cuanto mayor es el valor absoluto del coeficiente de correlación, tanto más cierta será la predicción.

La incertidumbre de la predicción aumenta cuando la correlación entre las variables disminuye.

El coeficiente de correlación puede tener valores entre 1.0 y -1.0. El valor 1.0 significa que el acuerdo entre los dos conjuntos de puntajes para los que se calcula la correlación es perfecto y positivo. El valor -1.0 significa una relación perfecta pero negativa; en este caso los puntajes estándar de los individuos en las dos distribuciones son los mismos, pero tienen signos contrarios.

Si no hay una relación lineal sistemática entre los puntajes de las dos distribuciones, el coeficiente de correlación es cero.

Formas de estimaciones de confiabilidad

Se supone que muchas características psicológicas son relativamente estables en el tiempo, los resultados de la medición de esas características no varían mucho a través del tiempo. por consiguiente, una medida posible de confiabilidad es la correlación entre medidas repetidas (o sea entre una prueba y una reaplicación), esta estimación se denomina coeficiente de estabilidad.

El procedimiento ordinario para conocer este coeficiente es someter al mismo grupo a

la misma prueba en dos ocasiones diferentes.

Esta definición de confiabilidad está planteada en términos de estabilidad, certeza y predecibilidad.

La correlación obtenida, representa el grado de consistencia de la prueba en el transcurso del tiempo.

Una segunda forma de confiabilidad se refiere a la exactitud, en ésta se contempla que las mediciones obtenidas por medio de una prueba corresponden a las mediciones verdaderas. Es reconocida como confiabilidad de consistencia interna o mitades iguales. Sattler (1988) dice que se obtiene al administrar la prueba a un grupo de individuos y luego dividirla en dos mitades. Esta división crea dos formas alternas de la prueba. El método más común de dividirlo es asignar los reactivos impares a una forma y los pares a otra. En este procedimiento se asume que todos los reactivos miden el mismo rasgo.

Una tercer forma de estimación de confiabilidad se refiere a la consistencia sobre las formas de una prueba, se denomina coeficiente de equivalencia, en ésta se espera que las calificaciones sean consistentes de una muestra de reactivos a otra - o sea, de una forma de prueba a otra.

También es conocida como confiabilidad de formas alternas, Sattler (1988) o paralelas, se obtiene cuando se administran dos pruebas equivalentes al mismo grupo de examinados. Si ambas son equivalentes tienen las mismas medidas y variancias y un elevado coeficiente de confiabilidad. Si no hay errores de medición, es de esperar que el sujeto obtenga el mismo resultado en ambas.

La combinación de estos procedimientos mediante la administración de formas equivalentes dando un intervalo en el tiempo, es llamada coeficiente de equivalencia y estabilidad.

Otra forma de enfocar la confiabilidad, que no sólo ayuda a formular y resolver problemas teóricos y prácticos se refiere al error de medición que puede tener un instrumento, este error se debe a la varianza aleatoria, los errores de medición son entonces aleatorios, y están formados por la suma o producto de un número de causas, tales como, fatiga temporal o momentánea, circunstancias fortuitas que afectan temporalmente al objeto o al instrumento de medición, fluctuaciones de la memoria o del estado de ánimo etc.

Dicho de otro modo, la confiabilidad se puede definir como la carencia relativa de error de medición de un instrumento.

La confiabilidad de una prueba es importante para su interpretación, si no se tiene seguridad en los resultados de la medición de las variables, es importante precisar con confianza las relaciones de las variables.

La confiabilidad es en extremo importante, si bien no constituye el aspecto esencial de la medición. Un alto grado de confiabilidad no garantiza que se obtengan buenos resultados científicos, pero no puede haberlos sin ella.

En síntesis, la confiabilidad es una condición necesaria, pero no suficiente, del valor de los resultados de la investigación y de su interpretación.

Los factores que afectan a la confiabilidad de una prueba con más frecuencia son, según Sattler (1988):

- El tamaño de la muestra: a mayor número de reactivos y homogeneidad, suele ser mayor la confiabilidad.
- Intervalo entre pruebas (test - retest): a mayor intervalo entre pruebas, mayor probabilidad de cambio, y por tanto más probable que sea baja la confiabilidad. Esto debido a que en un lapso de tiempo mayor pueden ocurrir cambios que afecten las habilidades de los niños o sus aptitudes para tomar la prueba.
- Variabilidad de puntaje: la varianza de una prueba suele estar en relación directa con la confiabilidad. Los cambios pequeños en el rendimiento influyen más en la confiabilidad de una prueba cuando el rango o dispersión de los puntajes es estrecho, y menos cuando es grande. Por consiguiente, las muestras con varianza pequeña suelen producir cálculos de confiabilidad más bajos que muestras parecidas con una varianza grande.
- Adivinación: entre más se adivine, es decir, se responda al azar a los diversos reactivos, suele ser menor la confiabilidad. Incluso la adivinación acertada causa errores en los reactivos finales.
- Variación dentro de la situación de prueba: las variaciones en la situación de prueba (por ejemplo, instrucciones erróneas o mal entendidas, equivocarse de lugar en las hojas de respuesta, enfermedad o ensoñaciones, etc.), suelen estar en relación inversa con la confiabilidad. Estos factores introducen una cantidad indeterminada de error en la administración de la prueba.
- Otros factores: entre mayores sean los efectos de la memoria y práctica, afectará más la confiabilidad.

El error estándar de la medida:

Debido al error de medición vinculado con la falta de confiabilidad, siempre hay algo de incertidumbre respecto a los resultados finales verdaderos.

El error estándar de medida proporciona información acerca de la certeza o confianza con que puede interpretarse el puntaje obtenido de las pruebas, y proporciona también

los límites dentro de los cuales quizá se encontrará el verdadero puntaje de un individuo.

Entre más grande sea el error estándar de medida, mayor será la incertidumbre relacionada con el puntaje del sujeto.

El error estándar de medida (S_n) es un cálculo aproximado del error que suele acompañar al puntaje obtenido de un examinado. Es llamado también error estándar de calificación, se refiere a la desviación estándar de la distribución de puntajes de error, y puede computarse del coeficiente de confiabilidad de la prueba mediante la siguiente fórmula:

$$S_m = DE \sqrt{1 - r_{xy}}$$

Donde la desviación estándar (DE) de la prueba se multiplica por la raíz cuadrada de uno, menos el coeficiente de confiabilidad r_{xy} de la prueba.

Intervalos de confianza:

Para poder hablar de la probabilidad de que el puntaje de un examinado refleje su verdadera calificación se obtiene un intervalo o banda de confianza - un rango de puntajes alrededor de lo que obtuvo el examinado.

El error estándar de medida es la base para formar este intervalo de confianza. El intervalo puede ser grande o pequeño, según el grado de confianza deseado.

Suele escogerse criterios que representan niveles de 68, 95 o 99% de confianza, aunque también se usan los niveles de 85 y 90.

Si se construye un intervalo de confianza de 95% entonces la probabilidad es sólo de 5 en 100, que el verdadero puntaje de error de una persona esté fuera del intervalo de confianza.

Los intervalos de confianza representan una banda o rango de puntajes en la que hay una gran probabilidad de que se encuentre el verdadero puntaje de un examinado.

No puede construirse ningún intervalo de confianza que prediga siempre, con absoluta certeza, el sitio en que habrá de encontrarse el verdadero puntaje de una persona.

1.5 VALIDEZ:

En psicología diferencial muchos instrumentos fueron contruidos para medir ciertos rasgos y empleados con este propósito; se desacreditaron más tarde cuando un examen cuidadoso reveló que otras variables eran las que determinaban los resultados individuales. Es necesario, por consiguiente, probar empíricamente que el instrumento mide la variable que se intenta medir en cada caso específico. A esto se le llama investigar la validez del instrumento y es una fase importante en el trabajo de

construcción de nuevos instrumentos en psicología diferencial. El conocimiento del grado de validez del instrumento es necesario para que los datos obtenidos con él puedan usarse significativamente, Magnusson (1990).

Significa también lo apropiado de las inferencias, según los resultados. Estos se usan para fines diferentes, como la colocación en programas educativos especiales o de capacitación, determinan la aptitud para algún empleo y evalúan la personalidad para algún empleo.

A no ser que la prueba sea válida para su fin particular, los resultados no podrán usarse con ningún grado de confianza.

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.

La validez del instrumento rara vez es un problema cuando tratamos con medidas físicas tales como longitud, peso, etc. obviamente el peso se mide con un balanza. Sin embargo, con los métodos usados para medir variables psicológicas es necesario probar empíricamente que el instrumento es válido en todos los casos.

La aplicación de un instrumento psicológico no debe basarse en la confianza subjetiva de que el instrumento funciona bien en situaciones prácticas. Los estudios empíricos muestran en la práctica clínica que las impresiones de este tipo son bases muy poco dignas de confianza para juzgar la seguridad de un método dado.

Debe mantenerse el requerimiento de que la validez de cualquier método tiene que probarse empíricamente en las diferentes situaciones donde va a usarse.

Las tres principales variedades de validez son, según Sattler (1988), **de contenido, criterio y constructo:**

Validez de contenido:

Se refiere a un examen sistemático del contenido de la prueba para saber si los reactivos son o no representativos del campo que se pretende medir, Brown (1985).

Para poder hacer esto es necesario preguntarse: ¿Las preguntas son apropiadas para la prueba y esta mide el objeto que interesa? ¿Contiene la prueba la información necesaria para abarcar bien lo que se supone que tiene que medir? ¿A qué nivel de dominio se evalúa el contenido?.

La respuesta satisfactoria de estas preguntas representan la validez de contenido. Es un requisito muy importante en el proceso de validación de cualquier prueba educacional o psicológica.

Este tipo de validez no se debe confundir con la validez aparente de un instrumento. Para que se pueda hablar de que una prueba tiene validez de contenido es necesario probarlo teóricamente.

Brown (1985) menciona que este tipo de validez es aplicable cuando estimamos el grado en que una prueba escolar, por ejemplo, abarca algún campo de estudio. Los reactivos de la prueba pueden considerarse como una muestra de una población que representa el contenido y las metas del curso.

La validez de contenido se determina entonces, por el grado en que la muestra de reactivos de la prueba es representativa a la población total.

Antes de que pueda estimarse la validez de contenido, es necesario definir explícitamente las metas del curso, el material que los alumnos deberán aprender, la importancia relativa de las diferentes partes del curso, etc. A diferencia de la validez predictiva o concurrente, la validez de contenido no puede expresarse por un coeficiente de validez.

Validez de criterio:

Se refiere a la relación entre los resultados de una prueba y algún tipo de criterio y resultado, por ejemplo, calificaciones, clasificaciones, u otro puntaje o resultado.

Entre el criterio y la prueba debe de haber una relación complementaria, ya que es a través de ésta que se puede determinar si la prueba mide, en realidad, el rasgo o característica para cuya medición fue estructurada.

Existen dos tipos de validez relacionadas con el criterio, éstas son, la validez concurrente y la validez predictiva.

En el concepto de validez relacionada con el criterio va implícita la idea de que las pruebas psicológicas se utilizan como parte de un proceso de toma de decisiones, Brown (1985).

Uno de los usos mas frecuentes de las pruebas de psicología es el de predecir la ejecución futura de un individuo en alguna variable significativa (el criterio). Un índice de esta precisión de predicción es una medida de la validez de la prueba.

La validez concurrente: Sattler (1988) menciona que se refiere a si los resultados de una prueba están o no relacionados con alguna medida de criterio disponible en la actualidad.

Si una prueba está muy correlacionada con un procedimiento de uso actual y además su funcionamiento a sido comprobado (validado), esto puede tomarse como señal de que la prueba en cuestión puede medir el mismo rasgo.

Al respecto, Kerlinger (1985) dice que en este caso se dispone de la medida de la variable de criterio en el momento en que se obtienen los resultados de la prueba. El diagnóstico clínico de lesión cerebral, sobre el cual estuvieron de acuerdo varios neurólogos, puede usarse, por ejemplo, como criterio para la validez de una prueba de lesión cerebral. La razón para construir una prueba que mida una variable de la cual ya se tienen datos es que en la mayoría de los casos, ésta última ahorra tiempo y esfuerzo y da el mismo resultado que la medida de criterio. Como la validez predictiva, la validez concurrente se expresa por un coeficiente de validez.

La validez concurrente se usa para pruebas en situaciones de diagnóstico. Cuando evaluamos los coeficientes de validez concurrente, debemos recordar que los criterios usados, por ejemplo, en situaciones clínicas, varían grandemente en calidad.

La validez predictiva: Sattler (1988) menciona que la validez predictiva se refiere a la correlación entre puntajes de prueba y rendimiento conforme a un criterio importante, con un intervalo de tiempo entre la administración de la prueba y el rendimiento conforme al criterio. Responde a la pregunta: ¿Predice el puntaje obtenido en una prueba, el rendimiento posterior en el criterio seleccionado?, la precisión con que una prueba de aptitudes o preparación indica el éxito futuro escolar depende de su validez predictiva. Por consiguiente, ésta es muy importante en muchos contextos psicoeducativos.

Cuando computamos la validez predictiva, deseamos usar la prueba para predecir las posiciones de los individuos sobre una distribución de la que sólo podemos disponer más tarde. La prueba predice cierto resultado después de un tiempo dado. Los datos de criterio consisten en alguna medida del resultado; por ejemplo, las calificaciones obtenidas al completar los estudios, o las estimaciones después de un cierto periodo de empleo. La validez predictiva, computada como un coeficiente de validez, necesita estimarse por las pruebas usadas en la orientación vocacional y en la selección y clasificación de los individuos para propósitos de adiestramiento o trabajo.

En la validez predictiva, la cuestión es la exactitud con que la prueba predice lo que se intenta predecir, problema que es especialmente importante en la orientación vocacional, en la selección y la clasificación de personal. Hay que señalar aquí que una buena capacidad predictiva presupone una buena capacidad de diagnóstico. Si algo no puede medirse exactamente, no puede usarse como base de predicciones válidas.

Está claro que no hay diferencias en principio entre los métodos para computar la validez predictiva y la concurrente. Muy frecuentemente, ambos se determinan computando la correlación entre los puntajes de la prueba y las medidas de la variable de criterio, y la validez es expresada en ambos casos por un coeficiente de correlación.

El coeficiente de validez da la validez de la prueba con respecto a la variable definida por la medida de criterio.

En la validez concurrente, la cuestión es si la prueba mide lo que se intenta medir, y el coeficiente de validez indica que tan adecuados son los datos de la prueba como base para el diagnóstico, en el sentido más amplio de la palabra.

Cuando computamos los coeficientes de validez, es necesario que los datos de la prueba y los del criterio se determinen en forma independiente.

Validez de construcción:

El concepto de validez de construcción, fue creada en 1954 por el Comité Técnico designado por la Asociación Americana de Psicología, el propósito de este comité fue definir la terminología y formular reglas para la estandarización de pruebas.

La validez de construcción, a diferencia de la validez predictiva y concurrente, no se expresa por un solo coeficiente que represente la correlación entre la prueba y las medidas de criterio.

El concepto de validez de construcción es especialmente útil en relación a las pruebas que miden rasgos para los cuales no hay un criterio externo.

La validez se determina mostrando las consecuencias que pueden predicarse con base a la teoría. Con respecto a los datos de la prueba pueden, en lo fundamental, confirmarse por una serie de pruebas.

La validez de construcción puede comprobarse de varias maneras, ejemplo de éstas son:

1. El estudio de las diferencias entre grupos que deberían diferir de acuerdo con la teoría de la variable.
2. El estudio de como los resultados de la prueba son influidos por los cambios en los individuos o en el medio que, de acuerdo a la teoría, deberían respectivamente influir o dejar de influir en las posiciones de los individuos sobre el continuo.
3. La correlación entre pruebas diferentes que se supone que miden la misma variable. Aquí debe procurarse que las correlaciones entre las medidas no surjan como resultado de semejanzas en el método. Esto podría suceder si, por ejemplo, las respuestas a las pruebas requiriesen de alguna aptitud especial diferente de la que está bajo consideración. Un posible acuerdo entre las medidas podría ser entonces puramente efecto de las diferencias individuales con respecto a esta aptitud especial.
4. La correlación entre reactivos aislados o diferentes partes de la prueba. Las partes de la prueba deben tener una alta intercorrelación para que se considere que miden una variable unitaria.

Sattler (1988) la define como el grado en que una prueba mide un constructo o rasgo psicológico. Se evalúa por varios procedimientos, todos ellos diseñados para determinar como se relacionan los reactivos con los constructos teóricos que pretende medir la prueba.

La validez de constructo de una prueba de inteligencia puede evaluarse al examinar la forma en que sus diversos reactivos se relacionan con una teoría de inteligencia.

El análisis factorial también permite examinar la validez de constructo, por ejemplo, en las baterías Wechsler, el análisis factorial suele demostrar que tienen componentes verbales y de ejecución, y por consiguiente avalan el uso de puntajes verbales y de ejecución separados.

Factores que afectan a la validez:

Como los coeficientes de validez son de correlación, también los afecta factores como el rango de talento que se esta midiendo y la duración de los intervalos al administrar ambas medidas.

La validez predictiva de C.I. (y de otros resultados) puede ser afectada por diversas razones Sattler citando a Deutsch, Fishman, Kogan, North y Whiteman, 1964:

- Factores relacionados con la misma prueba, por ejemplo: habilidades para ejecutar pruebas, ansiedad, motivación, rapidez, comprensión de las instrucciones, grado de novedad de los reactivos o del formato, simpatía entre examinador y examinado, impedimentos físicos, bilingüismo, deficiencias educativas, desconocimiento del material de la prueba, desviaciones de cualquier tipo del grupo normativo de estandarización.
- Factores relacionados con el criterio: las calificaciones escolares, criterio usado con mucha frecuencia, son afectados por motivación, conducta en el salón de clase, apariencia personal y hábitos de estudio. Si los examinados tienen problemas en cualquiera de estas áreas, suele disminuir la validez predictiva de las pruebas de inteligencia.
- Sucesos inesperados tales como, estados o perturbaciones emocionales, etc., los estados agudos de perturbación, con frecuencia destruyen la eficiencia intelectual, por consiguiente originan resultados de pruebas poco representativas.
- Cuando una intervención terapéutica -por ejemplo, fármacos, psicoterapia, colocación en hogar adoptivo, o manipulación del ambiente- pueden mejorar el rendimiento del examinado, habrá que dudar de la validez o representatividad de los resultados iniciales de la prueba.

1.6 MEDICIÓN EN PSICOLOGÍA.

Retomando la definición general de medición, la medición en psicología, es entonces, el proceso de asignar valores numéricos a la ejecución de una persona, de acuerdo a reglas específicas.

En el caso de las dimensiones físicas, el proceso de medición es relativamente directo. En general hay operaciones aceptadas comúnmente y reglas específicas que seguir, así como también escalas bien definidas que sirven para expresar los resultados.

Es muy distinto medir la cooperación, dependencia y creatividad de un niño que su talla, peso o desarrollo del hueso de la muñeca. Aunque el proceso fundamental es el mismo, resulta mucho más difícil la prescripción de las reglas de medición. Más aún, las observaciones de las variables psicológicas difieren muchísimo más de las cualidades reales, que las observaciones de las propiedades físicas. Esto representa quizá, el mayor problema de la medición psicológica; es necesario especificar un conjunto de procedimientos para medir el rasgo específico, en este tipo de mediciones es preciso controlar más variables para poder obtener datos significativos.

El problema de medición en psicología es considerablemente más complicado que en los campos donde se emplean instrumentos físicos comunes de medición.

Una variable psicológica se define como una prioridad o característica que poseen diferentes individuos en cantidades distintas. Al medir variables como independencia, neurotismo, capacidad de pensamiento lógico o aptitudes de aprendizaje, nos encontramos con problemas de escalamiento muy complejo.

Cuando se desean medir variables psicológicas tales como memoria, agresividad o capacidad aritmética, es generalmente imposible ir más allá del nivel de la escala ordinal sin hacer ciertas suposiciones. Nos pueden interesar las diferencias entre los individuos en el mismo continuo, las cuales reciben el nombre de diferencias interindividuales o bien las diferencias intraindividuales. En tales casos, es necesario medir la magnitud de los rasgos por medio de escalas de intervalo.

Naturaleza de la medición psicológica.

Desde tiempos remotos el hombre, como producto de una curiosidad natural relativa a nuestros congéneres, ha tratado de estudiar las diferencias individuales, tanto en sus características físicas como en sus características de personalidad.

Las personas no sólo difieren en aspecto y habilidades físicas sino que varían también en sus características de personalidad y en sus habilidades psicológicas.

Debido a estas variaciones, el estudio de las diferencias individuales ha sido un punto continuo de enfoque del interés de la psicología.

Brown (1980) dice que las mediciones psicológicas son relativas, ya que se ocupan primordialmente de las diferencias entre personas debido a que la mayoría de los procedimientos de las pruebas comparan la ejecución de una persona con otra, en lugar de hacerlo con una escala absoluta.

Para el psicólogo, uno de los mayores problemas es el de saber como descubrir la naturaleza y la amplitud de esas diferencias individuales en forma precisa y exacta.

Debido a gran número de formas para describir la conducta y las características de los individuos, el psicólogo ha tratado de reducir ésto, para facilitar su manejo. El método primario que se ha utilizado es el de la teoría de los rasgos (Anastasi, 1948), la cual define un rasgo como un conjunto de conductas interrelacionadas o intercorrelacionadas, ésto es, un rasgo es un término que describe un grupo de conductas que tienden a producirse juntos.

Con el fin de que resulten útiles como construcciones descriptivas, los rasgos tienen que ser relativamente estables en el tiempo y en diferentes situaciones. Si la conducta de una persona no es constante, las predicciones de su conducta y sus reacciones de una situación a otra y de un momento a otro serán inexactas y tendrán muy poca utilidad.

Debido a la estabilidad relativa de la mayoría de las características, muchas personas creen que los rasgos se determinan genéticamente. Aun cuando las diferencias entre personas pueden tener bases genéticas firmes, no es necesario considerar que los rasgos tienen una base genética.

Un punto de vista alternativo (Ferguson 1954, citado por Brown, 1970) concibe los rasgos como hábitos sobreaprendidos como patrones de conducta y reacción que se han aprendido tan bien que la persona se conduce de manera similar en muchas situaciones.

Así, en opinión de Ferguson, la consistencia en la conducta se debe a la dependencia de modos habituales de reacción y no a patrones de reacción determinados genéticamente.

Otro punto importante es el de que los rasgos se infieren de los patrones conductuales. No son características personales tangibles y directamente mensurables. En otras palabras, las mediciones psicológicas son siempre indirectas.

En consecuencia nos enfrentamos siempre a la pregunta de si medimos en realidad lo que intentamos medir.

La adopción de un método de rasgos tiene otra implicación importante, las mediciones psicológicas son más bien relativas que absolutas.

Procesos mediante los cuales se miden las características individuales:

Aun cuando se pueden utilizar muchos métodos, tales como las observaciones, la escala de estimación, las muestras de trabajo y las medidas de ejecución, el método más desarrollado hasta ahora para medir las características psicológicas son las pruebas psicológicas.

Las pruebas psicológicas o escalas constituyen el método de observación y captación de datos, que más se utiliza en las ciencias del comportamiento.

Definición de pruebas y escalas:

Las pruebas se consideran como métodos para la observación sistemática de la conducta de una persona.

Brown (1980) define a la prueba como un proceso sistemático para medir una muestra de conducta.

Para Kerlinger (1985), prueba psicológica es un procedimiento sistemático, mediante el cual a un individuo se le presenta una serie de estímulos artificiales ante los que reacciona; sus respuestas permiten al examinador asignarle un numeral con conjunto de numerales con los que se hacen inferencias sobre el grado en que poseen las cualidades que mide la prueba.

La escala consiste en una serie de símbolos o números contruidos de tal modo que puedan ser asignados, según determinadas reglas, a los individuos (a su comportamiento) a quienes se aplica; la asignación depende de que el individuo posea aquello que mide la escala.

Al igual que la prueba psicológica, la escala es un instrumento de medición, si no fuera por el significado tan basto asociado al término prueba, podríamos decir que tanto la prueba como la escala se definen en forma análoga. En rigor, la escala tiene dos sentidos: designa al instrumento de medición y a los numerales de dicho instrumento.

Las pruebas son escalas pero no todas las escalas son pruebas.

Procedimiento sistemático indica que una prueba se construye, se administra y califica, según reglas preestablecidas.

El uso de reglas predeterminadas para la evaluación (calificación) de las respuestas, asegura un acuerdo entre diferentes personas que tengan que calificar la prueba, permite que las calificaciones dadas a distintas personas se puedan comparar directamente.

La utilización de procedimientos estándar sirve para reducir al mínimo la influencia posible de variables personales y situacionales irrelevantes, sobre las calificaciones de

la prueba.

Brown (1985) dice que una de las confusiones más comunes en las pruebas psicológicas es la de no establecer una distinción clara entre medición y evaluación. La medición responde a la pregunta ¿Cuánto? Es decir, que proporciona una descripción de la ejecución de una persona, no nos dice nada sobre el valor de dicha ejecución.

Sin embargo, cuando interpretamos ésta, le atribuimos cierto valor, tratamos de responder a la pregunta ¿Hasta que punto es bueno?

En resumen, la descripción objetiva de la ejecución (la calificación en la prueba) es una medición; el juicio subjetivo de su calidad (el grado) es una evaluación.

Uso de las pruebas.

Las pruebas como instrumentos de medición, miden la naturaleza y la amplitud de las diferencias individuales. Si se aplica una prueba que mide un rasgo dado a un grupo de personas, se podrá determinar como se clasifica cada persona en esas características. Si un individuo se somete a varias pruebas, podremos describir su desempeño relativo en varias dimensiones. Brown (1985) diferencia el uso de las pruebas en:

Usos teóricos - aplicados.

Con frecuencia, se utilizan pruebas para proporcionar datos que nos ayuden a tomar decisiones prácticas; también pueden ayudar a desarrollar teorías psicológicas.

Las teorías en psicología, al igual que en otras disciplinas, incluyen tanto leyes como construcciones. Uno de los medios de definir estas últimas es por medio de pruebas. En estos casos las pruebas se utilizan para definir una construcción teórica.

Usos predictivos - descriptivos.

Las pruebas se utilizan para proporcionar descripciones de un individuo.

En general, los datos puramente descriptivos tienen una utilidad limitada. En la mayoría de las situaciones, se desea saber que resultados se pueden predecir a partir de las calificaciones obtenidas en las pruebas.

Ejecución máxima - típica.

En las pruebas de ejecución máxima, quienes se someten a ella tratan de obtener la máxima calificación posible; la meta es medir los límites de su capacidad, todas las pruebas de rendimiento, aptitudes y habilidades son medidas de ejecución máxima. En contraste, las medidas de ejecución típica intentan evaluar las reacciones y conductas

habituales o usuales en las personas. En este caso interesa conocer su conducta o reacción habituales.

La prueba como ayuda en la toma de decisiones.

Hay varias clases de situaciones en las que se utilizan las pruebas como ayuda para la toma de decisiones, la más común es para la selección - ya sea en ambientes académicos, laborales, etc. El papel de la prueba, en este caso, es el de identificar a los sujetos más prometedores.

La clasificación es otra que tiene como fin hacer coincidir a los individuos con las alternativas de manera óptima.

El diagnóstico implica la comparación de la ejecución de un individuo en varios campos, para determinar sus puntos fuertes y debilidades relativas.

También se utilizan con frecuencia en las investigaciones psicológicas, para la prueba de hipótesis. También pueden utilizarse para construir hipótesis.

Otro uso de las pruebas es en la evaluación, una prueba de ejecución no sólo da una indicación de diferencias entre individuos, puede señalar también áreas específicas conflictivas o no.

Todos estos usos implican alguna decisión. En la selección, la decisión consiste en aceptar o rechazar a un solicitante; en la clasificación, el curso alternativo de acción que se debe instigar; en el diagnóstico, el tratamiento de corrección; en la comprobación de hipótesis, la exactitud de la formulación teórica; en la construcción de hipótesis, las pruebas a las informaciones adicionales que se requieren y en la evaluación, la calificación que se le dará a un sujeto o el punto hasta que el procedimiento será o no eficiente.

Clasificación de las pruebas psicológicas según Brown (1980):

Las diversas pruebas y escalas que existen se pueden clasificar en las clases siguientes:

Pruebas de personalidad:

La medición de rasgos de la personalidad es el problema más intrincado de la psicología cuantitativa ya que la personalidad humana es la organización de rasgos del individuo.

El problema básico en la medición de la personalidad es la validez. Para medir los rasgos legítimamente se requiere que de antemano se conozca la que son, la manera en que interactúan y cambian, y el modo en que se relacionan entre sí.

Hay dos enfoques generales en lo que se refiere a la elaboración y validación de medidas de personalidad: el método a priori y el método teórico o de constructor. En el primero se construyen reactivos que reflejen la dimensión que se desea medir.

El método teórico o de constructor hipotético, son que se elaboran las medidas de personalidad subraya las relaciones que la variable medida tiene con otros; es decir, las relaciones sugeridas por la teoría en que se funda la investigación.

Escalas de actitudes:

La actitud es considerada como la predisposición organizada para pensar, sentir, percibir y comportarse en cierta forma ante un referente o un objeto cognoscitivo.

Es una estructura estable de creencias que predispone al individuo a comportarse selectivamente ante referentes actitudinales.

Existen tres tipos principales de escalas:

Escalas de clasificaciones sumadas: son un conjunto de preguntas sobre actitud que según se supone tienen de manera aproximada el mismo valor actitudinal y a las que los sujetos responden en grados diversos de acuerdo o desacuerdo.

Escalas de intervalo constante: a diferencia de la anterior a cada reactivo se le adscribe un valor escalar y éste indica la fuerza de una actitud o es considerado como un conjunto ordenado; es decir, difieren en su valor escalar.

Escalas acumulativas: constan de un número relativamente pequeño de reactivos homogéneos que son (supuestamente) unidimensionales. La escala unidimensional mide una variable y solamente una.

Escalas de valores:

Los valores son, en parte fruto del ambiente cultural y denotan preferencias por cosas, ideas, personas, instituciones y conductas, expresan preferencias por determinados modos de conductas y estados finales de existencia, los valores expresan lo bueno, lo malo, los imperativos de la conducta humana.

Este tipo de pruebas somete ideas, cosas y conductas a un continuo de aprobación - desaprobarción, suponen una elección entre diversas cursos de acción y de pensamiento.

Pruebas de inteligencia y aptitudes:

Estas pruebas miden la capacidad para realizar ciertos tipos de tareas (Brown, 1980) las calificaciones de una prueba de inteligencia son un índice del nivel de desarrollo de ciertas capacidades, así como la capacidad potencial para el aprovechamiento.

Para medir la variable inteligencia se pueden administrar pruebas cortas llamadas medidas generales, una medición general es aquella que en un sólo instrumento incluye reactivos de distintas clases: verbales, numéricos, espaciales y de otra índole. En su mayor parte son de carácter verbal y se correlacionan principalmente con el aprovechamiento escolar.

Al medir las aptitudes, se adopta por lo común una visión de rasgos y factores de las capacidades humanas. Esto quiere decir que estas últimas se conceptualizan en el sentido de que están organizadas y que se expresan mediante combinaciones de respuesta y conductas intercorrelacionadas.

I.7 ESTANDARIZACIÓN:

En la medición al administrar una prueba, la meta es obtener una estimación tan precisa como sea posible de la ejecución de los sujetos, Brown (1985).

La estimación precisa en las pruebas psicológicas, igual que en otros procedimientos científicos, depende del control de los errores - o sea, de la minimización de la influencia de factores irrelevantes para los fines de la medición. Esto se logra, haciendo que las situaciones de prueba sea lo más semejante posible para todos los individuos.

El proceso de desarrollo de esos controles se denomina estandarización. Este término tiene distintos significados pero en general la estandarización implica que a cada sujeto le sean presentados los mismos reactivos (o equivalentes) y que existan reglas específicas para la administración y la calificación de la prueba. Otros autores añaden el requisito de que se disponga de normas de ejecución, requisitos necesarios para interpretar los datos.

Elementos indispensables de la estandarización de una prueba son:

- a) **CONTENIDO.**- El primer elemento esencial es el conjunto de reactivos comunes que se aplican a todos los sujetos que se someten a la prueba.
- b) **APLICACIÓN:** Aun cuando se aplicaran los mismos reactivos a todos los sujetos, las calificaciones no serán comparables a menos que se aplicaran en las mismas condiciones, es posible eliminar gran parte de la variabilidad extraña, utilizando instrucciones prescritas para la aplicación de las pruebas, límites de tiempo estándar y métodos objetivos de calificación.
- c) **CALIFICACIONES:** El tercer elemento de la estandarización es el de las calificaciones objetivas. La objetividad implica un acuerdo entre dos o más calificadores competentes.

Los requerimientos de calificación objetiva se pueden clasificar en tres etapas básicas:

la primera es la del registro inmediato y carente de ambigüedades de la respuesta. Este registro permanente evita las distorsiones posibles debidas a pérdida de la memoria y proporciona las bases para la clasificación de las respuestas.

Es segundo requisito es una lista de respuestas estándar o correctas o sea, una clave de calificación. Si respuestas diferentes reciben pesos diferentes, estos pesos deberán indicarse en la clave.

El tercer requisito es un procedimiento para comparar las respuestas dadas por el sujeto con las que figuran en las claves, éste es, un procedimiento para clasificar objetivamente las respuestas.

La aplicación de los procedimientos analizados antes, sólo asegurará una cosa: que la calificación dada a un individuo refleje sus capacidades o sus características de personalidad, y no las peculiaridades o las vicisitudes de la situación de prueba.

Brown (1980) al respecto de los datos normativos dice que los resultados tienen por sí mismo poco significado. Sólo alcanzan un significado cuando se compara una calificación individual con los resultados obtenidos por otras personas en la misma prueba.

Al comparar las calificaciones de un individuo con las de otra persona en una población específica (que se llama grupo normativo), obtenemos una indicación de su desempeño relativo en comparación al de otro de la misma población.

Antes de ocuparnos de los diversos tipos de calificaciones, será conveniente tomar en consideración que información se deriva de las calificaciones de las pruebas. Una calificación individual en cualquier prueba será siempre una función de tres condiciones: su composición genética, su aprendizaje y sus experiencias antes de las pruebas y las condiciones de la situación de prueba propiamente dicha. Los tres aspectos desempeñarán un papel importante en la determinación de su ejecución. Aun cuando es difícil separar los componentes genéticos y los de la experiencia, las condiciones estandarizadas de un examen aseguran que la ejecución de una persona refleje sus capacidades y características y no la situación particular del examen.

Tipos de calificaciones:

Calificaciones brutas, originales o directas:

Después de administrar una prueba, las respuestas de un individuo se comparan con una clave para obtener su calificación en la prueba.

Cualquier calificación obtenida directamente a partir de la prueba, se denomina calificación bruta u original.

Brown (1980) dice que las calificaciones brutas de una persona serán el número de

reactivos a los que responda en la dirección indicada.

Calificaciones discretas o continuas:

En una escala discreta los resultados pueden caer sólo en distintos puntos a lo largo de ella, por ejemplo, el número de personas en una habitación y el número de puntos obtenidos en un partido de fútbol.

En las escalas continuas, los resultados pueden caer en cualquier punto de la escala, con las limitaciones de la precisión que permitan las técnicas de medición y la que se desee que tenga la medición. Así, las dimensiones de una habitación se pueden describir en metros, en metros y centímetros, etc.

Aun cuando las calificaciones de una prueba sean discretas, se tratan por lo común como si fueran continuas, bajo la suposición de que si fuera necesario, se podrían hacer distinciones más precisas.

Es raro que las calificaciones brutas sean significativas por sí mismas. Sólo cuando una prueba cubre un universo definido explícitamente tendrá sentido una calificación bruta. De otro modo, tendrá que compararse con las calificaciones obtenidas por individuos comparables o con algún estándar definido, o bien, expresando términos de alguna consecuencia o criterio.

Calificaciones transformadas:

Los resultados brutos se transforman con frecuencia a otras escalas, con el fin de facilitar el análisis y la interpretación.

Las calificaciones resultantes de estas transformaciones se llaman calificaciones derivadas o transformadas.

Brown (1980) dice que las transformaciones más comunes incluyen la suma o la resta de una constante, la multiplicación o la división por una constante o ambas. A este tipo de transformaciones se le conoce como lineal, debido a que los resultados brutos y los transformados tendrán una relación lineal y la relación entre pares comparables de resultados será la misma en todas y cada una de las escalas.

Formas posibles de clasificación de las calificaciones derivadas:

Calificaciones relacionadas con el contenido:

Aunque las mediciones psicológicas y educativas suelen estar relacionadas con las normas, es frecuente que este método no proporcione información más específica, se refiere a una descripción mucho más específica de la ejecución del sujeto y no en función de su ejecución en relación a un grupo.

Este método es más frecuente en el ámbito educativo, sin embargo, el desarrollo de

calificaciones relacionadas con el contenido se ha llevado a cabo con lentitud. Una de las razones para ello es que este método requiere una especificación precisa del universo de contenido; esta tarea es difícil, excepto en áreas muy circunscritas. El segundo problema es el de definir un nivel aceptable de destreza.

Calificaciones relacionadas con las consecuencias:

Una tercera posibilidad es la de expresar la ejecución en función de una conducta o una calificación pronosticada; tienen una ventaja sobre otros tipos de calificaciones ya que incluyen datos de validez en interpretación de la prueba. O sea que la interpretación es esencialmente una predicción de la conducta futura. Para poder hacer esta predicción necesitamos datos de validez que indiquen que conductas se pueden predecir a partir de las calificaciones de las pruebas, suelen ser datos de validez relacionadas con el criterio.

Este método requiere más trabajo, puesto que se deben de realizar estudios de validez e incluir sus resultados en el procedimiento de interpretación de las calificaciones, da casi siempre interpretaciones más significativas.

Calificaciones relacionadas con la norma:

Lo mas frecuente es que la ejecución de un individuo se compare con las calificaciones obtenidas por otras personas en un grupo de referencia pertinente, el grupo normativo. Este último se compone de personas que comparten ciertas características con el individuo.

El desarrollo de las calificaciones relacionadas con las normas implica:

Identificar un grupo pertinente de comparación; obtener las calificaciones en la prueba de los miembros de este grupo y convertir las calificaciones brutas a una escala que exprese la ejecución como clasificación relativa dentro de ese grupo normativo.

Las calificaciones en las pruebas psicológicas se expresan por lo común en escalas que no son de calificaciones brutas, por esta razón el constructor de pruebas debe desarrollar también escalas apropiadas para expresar las calificaciones.

El uso de calificaciones relacionadas con las normas hace hincapié en que las mediciones psicológicas son más bien relativas, que absolutas.

En las mediciones psicológicas es raro que se disponga de normas o escalas absolutas, así en la mayoría de situaciones, las diferencias entre individuos son más importantes o al menos más interesantes que las similitudes. Estos dos factores argumentan la expresión de la ejecución en términos comparativos; o sea, en escalas relacionadas con normas.

Normas de grupo:

En la mayoría de las pruebas, se interpretan las calificaciones comparando la ejecución de un individuo con las de otros del grupo normativo. El grupo normativo proporciona una base de comparación que muestra las calificaciones de un grupo estándar, definido, de referencia. Potencialmente, hay cierto número de posibles grupos normativos para cualquier prueba. Puesto que la calificación relativa de una persona puede variar mucho, dependiendo del grupo normativo que se utilice para la comparación, la composición del grupo normativo es un factor crucial para la interpretación de las calificaciones relacionadas con las normas.

Los grupos normativos se deben elegir como representantes de poblaciones, si la prueba está diseñada para evaluar las aptitudes de los alumnos de la preparatoria, el grupo normativo deberá consistir en alumnos de preparatoria.

Requisitos de las normas:

- La composición de las normas deberá definirse con claridad, tomando en cuenta la finalidad y los usos de la prueba.
- El grupo normativo deberá ser una muestra representativa de la población designada, el no obtener una muestra representativa hará que se desvíen los datos normativos, lo que complicará la interpretación de las calificaciones.
- El procedimiento de muestreo se debe describir con claridad, al igual que en el caso de la descripción de la población, cuanto más completa y precisa sea la descripción, tanto mejor.
- Los grupos normativos deberán de formarse con una muestra de tamaño adecuado, ésto debido a que la cantidad de error de muestreo varía en proporción inversa al tamaño de la muestra, esta última será mejor cuanto mayor sea su tamaño.
- Las normas se deben actualizar periódicamente.

Tipos de calificaciones relacionadas con la norma:

Percentiles. Los rasgos percentiles de una calificación se definen como el porcentaje de personas del grupo normativo que obtienen las calificaciones más bajas. Así como un rango percentil de 78 indica que 78% de las personas en el grupo normativo obtuvieron calificaciones más bajas; un rango percentil de 5 indica que sólo el 5% del grupo normativo obtuvieron calificaciones más bajas. En otras palabras un rango percentil indica la clasificación relativa de la persona en porcentaje.

La ventaja principal de los percentiles es su facilidad de interpretación, de conocer la clasificación relativa de una persona en un grupo pertinente de comparación es, para la

mayoría de los individuos, un índice de ejecución simple, significativo y fácil de entender.

Brown (1980) menciona que tienen dos limitaciones importantes: siendo una escala ordinal no se puede sumar, restar o dividir. Esto no constituye una limitación importante al interpretar las calificaciones; pero constituye una dificultad grave en los análisis estadísticos. Una segunda limitación es que los rangos percentiles tienen una distribución rectangular, en el que las distribuciones de las calificaciones de las pruebas se aproximan en general a la curva normal, como consecuencia de ello, las pequeñas diferencias de las calificaciones brutas cerca del centro de la distribución dan como resultado grandes diferencias en percentiles. A la inversa, las grandes diferencias de las calificaciones brutas en los extremos de la distribución producen solo pequeñas diferencias en percentiles. A menos que se recuerden estas relaciones, los rangos percentiles se pueden malinterpretar con facilidad; en particular, las diferencias aparentemente grandes alrededor del centro de la distribución tienden a interpretarse con exageración.

Para desarrollar rangos percentiles se requiere determinar la proporción de personas del grupo normativo con calificaciones inferiores a una calificación específica.

La primera etapa consiste en preparar una distribución de frecuencias después, se determina el número de personas que obtuvieron calificaciones inferiores a la calificación particular en cuestión (frecuencia acumulativa).

La división de este número por el número total de calificaciones que hay en la muestra de la proporción de casos que caen debajo de las calificaciones (proporciones acumuladas), al multiplicar esta proporción por 100 las calificaciones se transforman en rangos percentiles.

Calificación estándar:

Los rangos percentiles se miden en una escala ordinal. Por varias razones, sobre todo cuando se realizan análisis estadísticos con las calificaciones de las pruebas, sería conveniente expresar las calificaciones en una escala de intervalos - o sea, tener una escala con unidades de tamaño iguales. Las calificaciones estándar son expresadas en una escala de intervalos.

Una calificación estándar (z) es la desviación que tiene una calificación bruta de la media, en unidades de desviación estándar.

$$z = \frac{X - \bar{X}}{S}$$

Puesto que la unidad básica de la escala es la desviación estándar, se dice que esas calificaciones son estándar.

Brown (1980) enumera las propiedades de las calificaciones estándar:

- Se expresan en una escala que tiene una medida de cero y una desviación estándar de uno.
- El valor absoluto de una calificación z indica la distancia a la que se encuentra la calificación bruta de la media de la distribución. El signo de las calificaciones z indican si la calificación cae por encima o por debajo de la media.
- Son expresadas en una escala de intervalos, por lo que se pueden someter a operaciones algebraicas comunes.
- La transformación de las calificaciones brutas a estándar es lineal.

Calificaciones estándar normalizadas:

Se le llama a la transformación de calificaciones brutas en una distribución normal, estos se pueden concertar directamente en rangos percentiles.

Esta transformación es posible, debido a que en una distribución normal, hay una relación especificable entre las calificaciones estándar y las áreas situadas a lo largo de la curva, o sea, la proporción de casos que caen entre dos puntos.

Incluso cuando las calificaciones brutas o directas no estén distribuidas normalmente, se puede realizar una transformación de área y forzar las calificaciones a entrar en una distribución normal.

Brown (1980) dice que las calificaciones estándar normalizadas se calculan sólo cuando una distribución obtenida se acerca a la normalidad pero, debido a los errores de muestreo, es ligeramente diferente. Esta situación se presenta con frecuencia en la normalización de pruebas en muestras grandes y heterogéneas.

Otras calificaciones estándar:

Calificación T y Estanina:

Son otros dos tipos de calificaciones normalizadas, la primera se caracteriza por que las calificaciones se dan en una escala con una media de 50 y una desviación estándar de 10.

La escala estanina es una escala de calificaciones estándar de nueve intervalos, con una media de 5 y una desviación estándar de 2 puntos.

Todas las calificaciones estándar indican la posición relativa de un individuo, en función de desviaciones estándar de la media.

Puesto que las calificaciones estándar expresan las calificaciones de las pruebas en

una escala de intervalos, son valiosos para cuando se necesitan análisis estadísticos posteriores, además, las calificaciones estándar normalizadas permiten referirse a una distribución estándar (la curva normal) a una conversión directa a percentiles, a modo de simplificar la interpretación.

Permite además la comparación directa de las calificaciones en dos o más pruebas o escalas. La transformación de las calificaciones en calificaciones estándar corrige las diferencias en medias y de las diversas distribuciones, de modo que las calificaciones de las diversas pruebas se encuentran en la misma escala.

Escalas de desarrollo:

En estas escalas una calificación individual indica el nivel de desarrollo que caracteriza su ejecución. Como podría esperarse, las escalas de desarrollo señalan generalmente la ejecución como equivalente a la edad cronológica o al grado escolar.

Las ventajas de las escalas de desarrollo son.- Dan las calificaciones en unidades que se pueden leer con facilidad, en términos de equivalentes de grados o edades; proporcionan una comparación directa con la ejecución de los coetáneos del estudiante y brindan una base para la comparación entre individuos y para el estudio del crecimiento a través del tiempo.

Sus principales desventajas son.- Que sólo son apropiadas cuando la característica medida cambia sistemáticamente con la edad, por lo que se limita primordialmente al uso con los niños pequeños; se ven afectados por la composición, o sea, las experiencias y la educación anterior del grupo normativo; las unidades de la escala a diversas edades o distintos grados no son iguales y las extrapolaciones de un nivel de grado o edad a otros son arriesgadas.

Escala de índices y cocientes:

Se le llama al proceso por el cual se utiliza la razón entre dos calificaciones, el más conocido de todos, el cociente intelectual, es definido como la razón entre la edad mental y la edad cronológica.

Debido al hecho de que el crecimiento intelectual no aumenta linealmente con los incrementos de la edad, las razones de C.I. de razón o proporción no se utilizan ya en las principales pruebas de inteligencia.

En lugar de ello, se emplean en la actualidad calificaciones estándar normalizada, basadas en una muestra representativa de la población a cada nivel, llamadas C.I. de desviación.

CAPITULO II

MEDICION DE LA INTELIGENCIA.

II.1 DESARROLLO DEL CONSTRUCTO DE INTELIGENCIA.

Desde tiempos muy remotos, se ha utilizado el concepto de inteligencia, aún cuando las palabras para nombrarla varíen de acuerdo al tiempo y al lugar.

La literatura más antigua que hace uso de algún concepto de inteligencia se encontraba en los escritos de los antiguos filósofos griegos como Platón (427-347 A.C.) y Aristóteles (384-322 A.C.), quienes planteaban la necesidad de clasificar a los individuos a partir de sus "facultades mentales".

Se cree que el momento en que se inicia el estudio e investigación exhaustiva de las capacidades del ser humano y que ha conducido a lo que en la actualidad se ha considerado como fundamental en la comprensión y estudio tanto de la teoría acerca de la inteligencia, como del manejo de la instrumentación para su evaluación, ocurrió durante la segunda mitad del siglo pasado. (Morales, 1990)

Se atribuye el mérito de introducir el término inteligencia, en la psicología a Herbert Spencer (1855), quien antes subrayó su papel en la biología. Luego de definir la vida como "el ajuste continuo de las relaciones externas". Spencer creía que el ajuste se lograba gracias a la inteligencia en el hombre y gracias a los instintos en los animales inferiores. Definió la inteligencia como "el poder de combinar muchas impresiones separadas", vinculando el concepto con la doctrina de la evolución.

El término inteligencia ha sido utilizado con cierto recelo debido a que tiene diversas connotaciones, lo empleamos sólo con el propósito de describir un aspecto conductual que a pesar de ser infinitamente complejo, nos permita expresar "algo" que pueda ser comprendido con facilidad por aquellos que estudian el campo de la aplicación de la medición en psicología.

No podemos separar la investigación acerca de la naturaleza de la inteligencia del desarrollo simultáneo de instrumentos de medición que puedan demostrar y comprobar las hipótesis planteadas.

Históricamente hablando existen tres aproximaciones fundamentales para el estudio de este aspecto psicológico:

- a) La psicología experimental,
- b) las investigaciones genéticas, y
- c) los métodos multivariados.

- a) Los psicólogos experimentales han realizado su investigación en los laboratorios, tratando de encontrar una solución a los problemas psicológicos básicos, haciendo estudios acerca de la percepción, cognición, aprendizaje y memoria destinados a la comprensión de la inteligencia.

Los psicólogos experimentales empezaron a explorar las capacidades mentales utilizando técnicas cuantitativas, como las psicofísicas de Ernest Weber (1795-1878) y Gustav Fechner (1801-1887), para medir la agudeza visual, la auditiva y la táctil. También encontramos los trabajos de Herman Von Helmholtz (1821-1894), sobre la rapidez de reacción ante los estímulos y los estudios precursores de Herman Ebbinghaus (1850-1909), sobre el aprendizaje de memoria y la retención. Estos primeros psicólogos se interesaron más por las funciones perceptuales o intelectuales en los seres humanos promedio o típicos, que por las diferencias individuales.

- b) Desde el punto de vista genético, podremos considerar sus cimientos en los naturalistas, quienes se vieron influenciados por Ch. Darwin (1809-1882). Con Darwin se inicia un período de gran actividad en la investigación científica; si se planteaba la "supervivencia del más apto entre las plantas y los animales inferiores" debería existir también la "supervivencia del más apto entre los seres humanos".

Los naturalistas se interesaron por la evolución de las capacidades en las especies animales, desde los tropismos reflejos o instintos rígidos de las especies inferiores, hasta la adaptabilidad de los seres humanos. Su interés se vio complementado por estudios anatómicos y fisiológicos del sistema nervioso central, los cuales mostraron que conforme las interconexiones neurológicas, sobre todo las de los centros corticales del cerebro se hacen más complejas, ocurre lo mismo con la conducta. Así, la inteligencia se concibió como un atributo o propiedad innata que diferenciaba principalmente a los seres humanos de las especies subhumanas (Stemberg, 1988).

Jean Piaget nació en Suiza en 1896. Realizó estudios de ciencias naturales y luego de psicología. Estudió simultáneamente los fundamentos de la lógica y la formación de la inteligencia en el niño; su cultura científica le permitió abordar la psicología genética de una manera absolutamente nueva. Piaget al referirse a la inteligencia distingue tres aspectos de ella: el contenido, la función y la estructura. Así, explica que el contenido es la expresión observable de la conducta, la función se refiere a los principios de actividad intelectual, los cuales están genéticamente relacionados con la edad o con el estado de desarrollo del individuo; de este modo, Piaget afirma que la función está constituida por conceptos y leyes que los científicos infieren del contenido observado. La estructura se refiere fundamentalmente al conocimiento, el cual varía con la edad y la experiencia y se desarrolla a través de la actividad. La estructura de la inteligencia es el principal enfoque de las investigaciones de Piaget, debido a su interés por el estudio de la genética (Morales, 1990).

Piaget demuestra que la adquisición de conocimientos se efectúa según dos procesos complementarios: la acomodación (mediante el cual el sujeto se ajusta a las condiciones externas) y la asimilación (incorporación de los datos de la experiencia en las estructuras innatas del sujeto).

- c) La aproximación multivariada considera la existencia de un conjunto de variables que conforman el comportamiento inteligente. Esta aproximación maneja los llamados factores a fin de explicarla a partir de habilidades intelectuales básicas y diferenciadas.

Los métodos multivariados más apropiados son las interrelaciones y el análisis factorial.

Desde esta aproximación, se considera que más que una inteligencia general, existen habilidades específicas primarias y básicas, además de que existen tanto diferencias interindividuales como intraindividuales, dependiendo no sólo de las que potencialmente tiene el individuo al nacer, sino de la facilitación u obstaculización que el ambiente le proporcione para el desarrollo de todas y cada una de tales habilidades. Así, los métodos multivariados permiten describir a cada persona conforme a un perfil de funcionamiento intelectual y por la otra, comparar sus expresiones conductuales e intelectuales con las del grupo de individuos que sea necesario según la finalidad de la investigación.

De acuerdo a las tres grandes aproximaciones arriba mencionadas, podemos darnos cuenta que no hay una cosa concreta que se llame inteligencia, las definiciones de ésta parten de diferentes puntos de vista o consideraciones teóricas.

Por otra parte, la inteligencia realmente es algo abstracto que no podemos medir ni evaluar directamente, sino sólo a través de las diversas manifestaciones de la conducta, y la conducta a su vez está influida por una diversidad de factores tanto biológicos, cognitivos y de personalidad, los cuales están conjuntamente, interaccionando unos con otros de tal manera que se confunden y no puede establecerse donde termina uno y comienza el otro. Así también, el propio ambiente en donde se desenvuelve el individuo facilitará o dificultará su desarrollo.

Con las diferencias individuales se parte de una aproximación que ha permitido realizar un avance en el estudio de la inteligencia. Considerando que en cada individuo difieren los factores tanto biológicos de personalidad y sociales y por lo tanto, la interacción de cada uno de los mismos, las manifestaciones conductuales e intelectuales diferirán. Algunos individuos se manifiestan más competentes cuando se trata de comprender lo que se ha leído, de encontrar la solución de un problema, de responder apropiadamente a intercambios verbales, etc. Debido a estas variaciones, el estudio de las diferencias individuales ha sido un punto de enfoque del interés de la psicología.

El origen de la psicología individual y de las pruebas mentales se atribuye a Galton

(1822-1911) en Inglaterra, con la publicación en 1883 de su libro *Inquiries into Human Faculty and its Development*.

Las investigaciones psicológicas de Galton estaban arraigadas en su principal preocupación con el problema de la evolución humana y son, por tanto, un ejemplo excelente de la influencia de la teoría de Darwin (Boring, 1978).

Galton suponía que todo conocimiento necesariamente debería pasar a través de los sentidos. La mayoría de las pruebas construidas por Galton se relacionan con la discriminación sensorial. En 1884 fundó el primer laboratorio antropométrico, en donde realizaba mediciones a través de una serie de pruebas sensoriales y físicas. Para poder analizar los datos obtenidos, Galton utilizó por primera vez métodos estadísticos. Su principal interés consistió en establecer las semejanzas entre padres e hijos, para lo cual consideró que era necesaria una medición de asociación, dando así los primeros pasos para el desarrollo del análisis correlacional (Morales, 1990).

Galton creía que las pruebas de discriminación sensorial podían servir para medir el intelecto, idea que le fue sugerida por las teorías de Locke. Así Galton (en Anastasi 1978) escribió: "La única información que llega hasta nosotros sobre los acontecimientos exteriores pasa, al parecer, por la avenida de nuestros sentidos, y cuanto más perciban la diferencia nuestros sentidos, mayor será el campo sobre el que puede actuar nuestro juicio e inteligencia".

Mientras Galton trabajaba en Inglaterra, Cattell, en los Estados Unidos desarrolló instrumentos de medición psicológica. Cattell siguió la escuela de Wundt y aprendió de él que es menester ejercer un riguroso control al llevar a cabo las observaciones de la conducta; sin embargo, rechaza el concepto de que la meta fundamental en psicología es la generalización de los hallazgos a cerca de los fenómenos psicológicos ocurrientes en los seres humanos, y considera, sobre todo, las diferencias individuales en la ejecución de la conducta como variable principal en los estudios científicos.

Años después, Cattell en contacto con Galton, estableció laboratorios de psicología en las Universidades de Pennsylvania y Columbia, en donde se analizó el desarrollo y depuración de los instrumentos de medición psicológica, a los cuales, Cattell designó con el nombre de pruebas mentales (1890). Estos instrumentos habían sido creados sólo para la medición de capacidades sensoriales como tiempo de reacción, capacidades auditivas y visuales, sensibilidad al dolor, gusto, etc., (Morales, 1990).

Pearson (1857-1936), entusiasmado con el éxito que había obtenido Galton, desarrolló los métodos estadísticos para el estudio de las diferencias individuales; desarrolló el coeficiente de correlación producto momento (P.M), la correlación múltiple y la correlación parcial; asimismo, estableció las bases del análisis factorial y otros métodos de análisis multivariado.

En 1901, junto con Galton y Weldon fundaron la revista *Biométrica* para publicar investigaciones matemáticas en Biología y Psicología (Boring, 1978).

Por la misma época en que Pearson desarrollaba los métodos estadísticos, Alfred Binet (1857-1911) se interesó en investigar también las diferencias individuales. su interés surgió como consecuencia de sus trabajos con hipnosis en los que supuso que existían diferencias individuales en las personas con posibilidades de ser hipnotizadas. Descubrió que la psiquiatría y la psicología, por su propia naturaleza, estaban relacionadas con las diferencias individuales.

Binet, siguiendo en cierto sentido las teorías de Galton, trabajó principalmente con simples variables de discriminación sensorial y con atributos físicos.

A principios de este siglo fue comisionado por el gobierno francés para estudiar el problema de los niños de bajo rendimiento intelectual. Trabajando en colaboración con Theodore Simon, Binet construyó un instrumento que pudiera medir la inteligencia general, considerando que la inteligencia no se expresa en segmentos de conducta, sino más bien como una operación mental combinada, en la cual todo proceso funciona como total unificado.

Definió la inteligencia como "la tendencia a tomar y mantener una dirección, la capacidad de adaptación destinada a obtener la o las metas deseadas y el poder de autocrítica".

A Binet se le atribuye el mérito de haber descubierto una forma apropiada de evaluar la inteligencia o capacidad mental, su trabajo dio origen al concepto de C.I. Aunque el método mismo que utilizaba fue abandonado por los investigadores modernos (Matarazzo, 1976).

Thorndike (en Morales, 1990) fue el primero en desarrollar la idea de que la medición de la inteligencia consistía en la evaluación cuantitativa de los productos mentales en función del número, la eficiencia y la velocidad con que se realizaran. Afirmaba que las habilidades son producciones mentales ordenadas en diferentes clases o tipos de operaciones. Sugirió que la inteligencia debía agruparse en tres categorías:

- a) inteligencia abstracta o verbal, mediante el manejo fundamental de símbolos.
- b) inteligencia práctica, por medio de la manipulación de objetos, y
- c) inteligencia social, facilitadora de las confrontaciones con los seres humanos.

Charles E. Spearman (en Boring, 1978) publicó su famoso trabajo sobre General Intelligence, objectively determined and measured, en 1904, en el que establecía la teoría de los dos factores. Spearman estuvo mucho más interesado en el entendimiento de las habilidades humanas que justamente en medirlas. El, como Binet, se interesó en las habilidades humanas, pero con una aproximación basada en la naturaleza de las relaciones existentes entre las teorías psicológicas y apoyado en modelos matemáticos para el estudio de tales habilidades.

Los esfuerzos de Binet para medir la inteligencia se fundaban en la hipótesis de que existe solo un factor de inteligencia. La teoría de Spearman plantea como hipótesis sujeta a comprobación la presencia de un "factor general". Concluye que las diferencias individuales en todas las pruebas de habilidad pueden considerarse a partir de dos factores: un factor general (G) y un factor específico, el factor G fue interpretado como la inteligencia.

Tomando en cuenta las agrupaciones que de la inteligencia sugirió Thorndike y apoyándose en los estudios de Spearman, David Wechsler desarrolló su primer instrumento de medición de la inteligencia, el cual fue publicado en 1939.

Wechsler le denominó a la inteligencia agregado o capacidad global porque consideraba que está compuesta de elementos o habilidades que, aún cuando no son totalmente independientes, pueden ser cualitativamente diferenciables, por tanto, a través de la medición de estas habilidades evaluamos, en última instancia, la inteligencia, entendiendo que la inteligencia no es la mera suma de habilidades.

Wechsler considera tres importantes razones para fundamentar lo anterior:

- a) que los productos finales de la inteligencia no están sólo en función del número de habilidades, sino también de la forma en que se combinan para operar, es decir, lo que él llama su "configuración",
- b) que existen en cierto sentido, otros factores independientes de las habilidades intelectuales, que influyen en la conducta intelectual, como son los incentivos, los impulsos y las motivaciones, etc., y
- c) el hecho de que en ciertas tareas se exija una mayor "cantidad" de habilidad intelectual específica no implica un "aumento" en la efectividad o eficiencia de la conducta intelectual como un total.

Wechsler inició su tarea intentando definir el concepto de inteligencia de la siguiente manera: la inteligencia es el agregado o capacidad global del individuo para actuar propositivamente, para pensar racionalmente y para conducirse adecuada y eficientemente en su ambiente. (Morales, 1990).

II.2 DESARROLLO DE LAS PRUEBAS DE INTELIGENCIA.

Las definiciones de inteligencia, así como sus sistemas de evaluación, han sido tema de constante polémica entre los estudiosos de la conducta, especialmente de los psicólogos. Ma. Vicker (en Sattler, 1988) refiere que al tratar de conocer la naturaleza de las capacidades intelectuales, los estudiosos han cuestionado los sistemas de medición, la variabilidad de una prueba y las probables consecuencias de las decisiones tomadas en esta materia para la educación y el mejoramiento de la raza humana.

Como se ha señalado anteriormente, los psicólogos experimentales empezaron a explorar las capacidades mentales utilizando técnicas cuantitativas, como las psicofísicas de Weber. Estos primeros psicólogos se interesaron más por las funciones perceptuales e intelectuales en los seres humanos promedio o típicos, que por las diferencias individuales.

Posteriormente, con Francis Galton surge una "nueva" psicología de la Gran Bretaña. Galton se dedicaba principal, aunque no totalmente, a aclarar el problema de las diferencias individuales. En 1883 publicó su famoso libro *Inquiries into Human Faculty and its Development*, el cual ha sido considerado algunas veces como el comienzo de la psicología individual y de las pruebas mentales.

Binet y Henri, en un artículo publicado en Francia en 1895, criticaron la mayoría de las series de pruebas existentes, por ser en gran medida sensoriales y concentrarse, indebidamente, en aptitudes especializadas sencillas. Argumentaban, además, que en la medida de funciones más complejas no es necesaria una gran precisión, puesto que las diferencias individuales son mayores que ellas.

Binet y sus colaboradores dedicaron muchos años a una investigación ingeniosa y activa sobre las maneras de medir la inteligencia. Se probaron numerosos métodos que incluían hasta la medida de rasgos físicos, el análisis grafológico y la quiromancia. Sin embargo los resultados condujeron a la creciente convicción de que la medida directa, aunque burda de las funciones intelectuales complejas, era la mejor solución.

Binet como psicólogo experimental, se interesó en el estudio de aquellos individuos que presentaban desviaciones conductuales, particularmente los niños. Después de minuciosos estudios en retrasados, normales y avanzados Binet y Victor Henri crearon una serie de instrumentos destinados a medir "procesos mentales muy complejos". Se propuso una extensa y variada lista de pruebas que abarcaban funciones tales como la memoria, la imaginación, la atención, la comprensión, la sugestibilidad, la apreciación estética y muchas otras.

Entonces se presentó una situación particular que llevó los esfuerzos de Binet a un resultado práctico inmediato. En 1904, el Ministro de Instrucción Pública de Francia nombró una comisión para estudiar los procedimientos para la educación de los niños subnormales que asistían a la escuela de París. Con el fin de satisfacer esta demanda práctica, Binet en colaboración con Simon, preparó la primera escala Binet-Simon. Esta escala constaba de 30 pruebas que estaban destinadas a cubrir una amplia variedad de funciones, con especial preferencia para el juicio, la comprensión y el razonamiento, que Binet consideraba componentes esenciales de la inteligencia.

Esta investigación de Binet fue más prometedora en cuanto a los datos que con ella podían obtenerse, al grado de que en la actualidad aún se sigue utilizando, naturalmente después de haber pasado por múltiples revisiones y estudios que la han clasificado como un instrumento útil, predictor del comportamiento intelectual, y sobre

todo válido.

Debido a este instrumento a Binet se le atribuye el mérito de haber descubierto una forma apropiada de evaluar la inteligencia o capacidad mental, su trabajo dio origen al concepto de C.I.; aunque el método mismo que el utilizada fue abandonado por los investigadores modernos (Matarazzo, 1976).

La escala Binet-Simon proporciona una calificación de Edad Mental para cada niño, es decir, relaciona el rendimiento de una persona en una serie determinada de pruebas con el promedio de edad en que éstas son realizadas con éxito por niños de diferentes edades. El hecho de que sólo se obtenga una calificación presupone que la inteligencia es general o que únicamente está considerándose un factor en los elementos de la prueba (Morales 1990).

Esta escala sufrió modificaciones y revisiones en muchos idiomas. En América, la más famosa es la llevada a cabo, en 1916, bajo la dirección de L.M. Terman en la Universidad de Stanford, conocida como la revisión Stanford-Binet. En esta prueba se utilizó por primera vez el coeficiente de Inteligencia (C.I.) o razón entre la edad mental y la edad cronológica o real.

En 1937 se hizo una segunda revisión de la escala, formando dos series equivalentes, L y M, con 129 reactivos. En esta revisión se eliminaron aquellos reactivos que no resultaron adecuados en la de 1916.

La última revisión hecha en 1960, incluye una importante innovación en la prueba Stanford-Binet; esta es la sustitución del C.I. de proporción por el C.I. de desviación, empleado en las formas anteriores. Estos C.I. de desviación son puntuaciones típicas con una media de 100 y una desviación estándar de 16.

La ventaja principal en este tipo de C.I. es que proporciona puntuaciones comparables en todos los niveles de edad.

En relación con la validez de contenido de la prueba de Stanford-Binet, las tareas que se eligieron desarrollan claramente funciones como la exactitud de observación, el juicio práctico, la memoria para diversas clases de material, la aptitud para seguir direcciones, la capacidad de ver en el espacio y el razonamiento y manejo de conceptos abstractos.

Las aptitudes verbales predominan claramente en las edades mentales superiores. Los conocimientos que se adquieren en la escuela, como la lectura y la aritmética, son indispensables para la actuación satisfactoria en los niveles de edad superiores. En la medida en que todas estas funciones son pertinentes a lo que comúnmente se considera como inteligencia, cabe decir que la escala tiene validez de contenido.

Las pruebas de Binet, así como todas sus revisiones, son individuales, en el sentido de que únicamente pueden aplicarse a una persona a la vez.

Las pruebas colectivas fueron creadas para satisfacer una urgente necesidad práctica, cuando en 1917 los E.U. decidieron participar en la primera guerra mundial y requirieron clasificar rápidamente a millón y medio de reclutas respecto a su nivel intelectual grupal. Tal información sirvió de ayuda en muchas decisiones administrativas. Con este fin, los psicólogos del ejército recurrieron a todos los materiales de pruebas disponibles, y especialmente a un grupo de pruebas de inteligencia no publicadas, preparadas por Arthur S. Otis y que este entregó al ejército. Estas pruebas se han llegado a conocer como el Alfa y el Beta del Ejército.

El Alfa del Ejército estaba destinado a su aplicación grupal; y el Beta del Ejército era una escala no verbal, empleada con analfabetos o con reclutas nacidos en países extranjeros que no dominaban el inglés. Ambos eran apropiados para su aplicación en grupo.

Poco después del fin de la primera guerra mundial, las pruebas del ejército fueron cedidas para su uso civil. Estas pruebas pasaron por muchas revisiones y sirvieron como modelo para la mayoría de las pruebas de inteligencia colectivas. El desarrollo de las pruebas experimentó un gran avance. Pronto se idearon pruebas de inteligencia colectivas para todas las edades y tipos de personas. Como las pruebas colectivas eran instrumentos para la prueba de masas, no solamente permitían el examen simultáneo de grandes grupos, sino que simplificaban también las instrucciones y procedimientos de aplicación, de manera que requerían un mínimo de formación por parte del examinador.

Otro instrumento, de aplicación colectiva, es la Prueba de Matrices Progresivas de Raven, la cual está construida sobre la base de la teoría bifactorial de Spearman. La teoría bifactorial de Spearman establece el concepto de inteligencia, diciendo que ésta consiste en la educación de relaciones y la educación de correlatos, es decir, la capacidad de cada individuo de establecer relaciones, desde las más simples hasta las más complejas.

El Raven es una prueba no verbal, diseñada para valorar la capacidad del sujeto, a fin de establecer relaciones y correlatos mediante diseños y figuras geométricas.

El análisis factorial ha demostrado que las Pruebas de Matrices Progresivas miden con suficiente extensión el factor g (factor general) con poca aplicación del factor de percepción espacial.

Sólo para ejemplificar otra prueba, construida sobre las mismas bases, mencionaremos la prueba de Anstey (de dominó). El análisis factorial ha demostrado que el grado de correlación existente entre los factores que miden ambas pruebas no es suficientemente alto para considerar equivalentes a dichas pruebas.

Un punto de vista que ha provocado gran división entre los teóricos de este campo es el de si los aspectos cognoscitivos y no cognoscitivos de las diferencias individuales

podrían separarse. Históricamente se sabe que existe una distinción entre dos características de la naturaleza humana: el razonamiento y la emoción, la habilidad y la motivación, etc. Muchos psicólogos han estado de acuerdo en que tales divisiones no son más que artificios llevados a la práctica y que quienes evalúan la inteligencia tienen que admitir plenamente que la motivación y la personalidad son aspectos no eliminables de la medición mental.

David Wechsler ha insistido en que la inteligencia y otros tipos de características como el temperamento no pueden estar separadas.

Wechsler, trabajando en el hospital psiquiátrico de Bellevue, Nueva York, desarrolla su primer instrumento de medición intelectual, el cual difería en muchos aspectos de los elaborados hasta entonces, particularmente de el de Binet.

Este instrumento denominado escala Wechsler Bellevue, cuyos fundamentos lógicos están sustentados en el concepto de "inteligencia general", con una validez racional más que empírica, fue publicado en 1939.

En general, las escalas de Wechsler están basadas en la teoría bifactorial de Spearman, en el sentido de que para Wechsler existe un factor fundamental en las funciones intelectuales; no obstante, al igual que Binet, Wechsler analiza la idea de la "inteligencia general".

Empero, debe aclararse que el factor g de inteligencia, que para Spearman era un factor primario, para Wechsler es un factor g secundario, o sea, hace la función de unificador para la expresión unitaria de una conducta que requiere la participación de varias operaciones e, incluso de una serie de habilidades específicas que al combinarse dan un producto intelectual.

Las escalas de Wechsler son escalas compuestas, verbales y de ejecución, constituidas por una serie de pruebas específicas que, al pasar a formar parte de una escala, se denominan "subpruebas". Son, por una parte, escalas heterogéneas interpruebas, debido a que miden diferentes funciones y, por tanto, distintos factores en cada una de ellas, y, por otra parte, son escalas homogéneas intraprueba, porque cada subprueba mide un solo factor en toda su amplitud.

Las subpruebas están estructuradas de la siguiente forma:

Los reactivos: se clasifican en tres grupos; el primero corresponde a aquéllos que toda persona puede contestar, tenga o no instrucción escolar, el segundo abarca reactivos que se espera sean contestados por sujetos que han adquirido instrucción escolar media, y el tercero se refiere a aquéllos reactivos que sólo pueden ser contestados por sujetos con cierta especialización.

En cuanto a las subpruebas, estas se clasifican en varios grupos, atendiendo a los factores específicos que evalúan cada una de las subpruebas. Así, tenemos que: El

primer grupo es el denominado estrictamente verbal. El segundo grupo se denomina de atención y concentración.

En la escala de ejecución también encontramos dos grupos: Los visuales y los visomotrices.

Desde el punto de vista de la validez de las escalas de Wechsler, la escala ha sido determinada sobre un amplio rango de edades y sujeta a una serie de análisis a fin de establecerla. Se han efectuado interrelaciones entre subpruebas, para determinar la presencia o ausencia del factor g.

Un criterio significativo de validez concurrente ha sido el establecido con la prueba de Stanford-Binet; ambas pruebas están midiendo lo mismo en un alto grado.

Respecto a la confiabilidad, las correlaciones "test-retest", han dado resultados satisfactorios.

En 1955 aparece una forma revisada de la escala de Wechsler, conocida como WAIS (escala de inteligencia Wechsler para adultos), la cual no tiene cambios significativos en cuanto a su contenido, construcción, organización y calificación. Lo significativo fue su aplicación en cuanto a las muestras que originalmente se habían utilizado para su estandarización, algunas modificaciones prácticas en cuanto a las instrucciones y calificaciones, una revisión minuciosa del contenido para situaciones especiales (por ejemplo, en la evaluación de sujetos con bajo nivel de eficiencia mental).

En 1949, Wechsler había planteado la necesidad de elaborar una escala designada especialmente a la evaluación de la inteligencia en los niños de cinco a quince años de edad. Sobre la misma base, construye la Escala de Inteligencia Wechsler para niños WISC y posteriormente en 1965, la Escala de Inteligencia Wechsler para Preescolares.

II.3 TEORÍAS SOBRE LA ESTRUCTURA INTELECTUAL

II.3.1 Teoría Bifactorial

Existió un remarcado paralelo durante los primeros años de este siglo entre el trabajo de Binet en Francia y el de Charles Spearman en Inglaterra. El trabajo de Spearman está resumido en su libro "The Abilities of Man" de 1927.

Como Binet, Spearman estuvo interesado en la naturaleza de las habilidades humanas, pero su interés fue una aproximación mucho más académica basada sobre las teorías psicológicas relacionadas a la naturaleza de las habilidades y apoyada por modelos matemáticos para el estudio de tales habilidades (Morales, 1990).

El problema de la organización de los rasgos se estableció por primera vez sobre bases empíricas en 1904, con un artículo de Spearman en el cual presentaba un

método de investigación nuevo, el cual se basa en ciertas relaciones entre las intercorrelaciones de las puntuaciones de la prueba. El análisis se estableció después sobre bases más sistemáticas con el desarrollo de la ecuación tetrádica, denominada así porque reúne pruebas en series de cuatro, sin embargo, esta técnica resulta ineficaz y de difícil manejo cuando el número de pruebas es grande.

No obstante, el desarrollo de la ecuación tetrádica de Spearman, abrió el camino a la investigación estadística de las relaciones entre datos (Brody y Brody, 1976).

Las investigaciones de Spearman lo condujeron a la llamada Teoría de los dos factores o Bifactorial, la cual presentó en su famoso trabajo "General Intelligence, objectively determined and measured", publicado en 1904.

Spearman se propuso interpretar la correlación entre dos variables como indicación de la existencia de un factor común y un factor específico en cada variable. Las mediciones de habilidades mentales aparentemente diferentes presentan muchas veces correlación. Spearman concluyó que la existencia de correlaciones positivas se debe a la presencia de una habilidad general común a todo tipo de ejecuciones. Llamó a este factor general, factor G, y fue interpretado como la inteligencia. Parecía bastante lógico analizar dos habilidades en tres factores (lo que es común a ambas y lo que es específico a cada una). (Boring, 1978)

En 1912, Spearman y Hart desarrollaron una técnica de una matriz jerárquica de coeficientes de correlación que separaba gran variedad de ejecuciones en el factor general, y varios factores específicos E; E1, E2, etc. En 1916, Godfrey H. Thomson señaló que cuando se tiene más de dos ejecuciones puede haber otras solapadas además de G. Por ejemplo, tres ejecuciones podrían tener lo que es común a todos (G), lo que es común para cada par (R1, R2, R3) y lo que es específico a cada una de las tres (E1, E2, E3) (Boring, 1978).

La correlación positiva entre dos funciones cualesquiera, se atribuía pues al factor general. Cuanto más saturadas estuvieran dos funciones con el factor G, más alta sería la correlación entre ellas. Por otra parte, la presencia de los factores específicos tendía a hacer descender la correlación entre las funciones. Aunque en esta teoría se afirman dos tipos de factores, el general y el específico, sólo el factor G explica la correlación.

De la teoría de los dos factores se sigue que el objetivo de la aplicación de las pruebas psicológicas debería ser el de medir la totalidad del factor G de cada individuo. Si este factor aparece a través de todas las aptitudes nos proporciona la única base para la predicción de la ejecución del sujeto al pasar de una situación a otra. De acuerdo con ellos Spearman propuso substituir, por una sola prueba, muy saturado de G, la heterogénea colección de elementos que se encuentran en las pruebas de inteligencia, sugirió que las pruebas que se ocupan

de relaciones abstractas son probablemente, las mejores medidas de G, y que se podrían usar con este fin (Anastasi, 1980).

Algunas pruebas tienen más G y menos E que otras y las pruebas que no tienen G fueron consideradas como externas al dominio cognoscitivo, ellas pueden ser pruebas de habilidades puramente sensoriales o motoras o de rasgos de personalidad, por lo que no se podrían llamar cognitivas o intelectuales (Jensen, 1980).

Spearman advirtió desde un principio que la teoría de los dos factores necesitaba ciertas modificaciones. Cuando las actividades que se comparan son muy similares, puede resultar cierto grado de correlación además del atribuible al factor G. Así, en adición del factor G, y a los específicos, debe existir otra clase intermedia de factores que no sean tan universales como G ni tan específicos como E. Estos factores comunes a un grupo de actividades, pero no a todas, han sido designados como factores de grupo (Anastasi, 1978).

Se suscitaron controversias porque la teoría de los dos factores y la teoría del grupo de factores parecían incompatibles, hasta que J.C. Maxwell Garnet se propuso demostrar que la teoría de los dos factores no es más que la forma más simple de factorización de grupo (Boring, 1978).

Hacia 1945, época en que murió Spearman, había moderado sus puntos de vista e incluso aceptado los hallazgos de algunos de sus discípulos, de la existencia de factores de grupo, como los de aptitudes aritméticas, mecánicas y lingüísticas. (Matarazzo, 1976).

Spearman originalmente trató de comprender la naturaleza psicológica de G al analizar factorialmente más de un ciento de pruebas, cada una completamente unitaria u homogénea en contenido y entonces comparó sus cargas G. Su análisis estuvo basado en el método clásico sobre acuerdo y diferencia. Él investigó en que forma todas las combinaciones de pares posibles de las varias pruebas: 1) son similares, y en que forma ellas, 2) son diferentes cuando a) ambas pruebas tienen cargas grandes G, b) ambas pruebas tienen cargas pequeñas G, y c) una prueba tiene una carga G grande y la otra una pequeña. Al tratar de contestar estas preguntas para cada uno de los pares de pruebas Spearman llegó a la conclusión de que "G esta esencialmente caracterizada por la combinación de neogénesis con abstracción".

Por "neogénesis" Spearman quiere decir: la "edución de relaciones y correlatos", ésto es percibir relaciones, inducir de lo general a lo particular y deducir de lo particular a lo general. La neogénesis es inductiva o inventiva si es comparada con la conducta reproductiva o de aplicar reglas.

"Abstracción" se refiere a las ideas, relaciones y conceptos en contraste a propiedades que pueden ser directamente percibidas por los sentidos.

Spearman encontró que la mayoría de las pruebas altamente cargadas de G eran aquellas que incluían tanto neogénesis como abstracción. La carga sustancial de G depende de la combinación de estas dos características. En un análisis realizado por Spearman encontró que la prueba más cargada de G fue la de Matrices Progresivas de Raven, la cual depende casi enteramente de percibir características claves y relaciones y descubrir las reglas abstractas que gobiernan las diferencias entre los elementos de la matriz (Jensen, 1980).

Aunque en un sentido, las teorías y descubrimientos de Spearman tendían a apoyar el tipo de prueba desarrollada por Binet, Spearman propuso un método más refinado para medir G. Primero sería necesario determinar que tareas miden realmente G más bien que estar dominadas por factores específicos; segundo, para medir G requeriría un peso apropiado de cada tarea intelectual. Spearman estuvo alarmado por la práctica de Binet de reunir una mezcla de problemas sin primero examinar la presencia de un factor general o sin pesar adecuadamente los problemas en términos de sus cargas sobre el factor general. (Nunnally, 1970).

Hacia el final de su vida, Spearman reconoció que su teoría no encajaba con todos los hechos observados. Así Spearman y Wynn Jones dicen que la teoría Bifactorial sólo indica el grado inicial de análisis pero ciertamente no el definitivo. Consideran las pruebas acumuladas sobre factores de grupo, es decir, aptitudes que cubren muchas actividades aunque no todas y su conclusión es que tales factores han sido encontrados pero que son pocos o raros.

A pesar de todo, Spearman ha estado, probablemente, más cerca de la verdad de lo que mucha gente en el estado actual de la opinión psicológica está dispuesta a admitir y su trabajo ha sido importante por dos razones principales.

Primero: mientras desarrollaba los métodos matemáticos para estudiar G, puso las bases para el análisis factorial. Una vez que mostró que las correlaciones entre las pruebas de habilidad podían ser estudiadas con respecto a un factor general, alentó a otros a extender la lógica a los métodos para la investigación de cualquier número de factores entre prueba.

La segunda razón de importancia es que, en contraste a la tradición del trabajo aplicado que siguió de las pruebas de Binet, estableció una tradición académica en la investigación de las habilidades humanas (Butcher, 1979).

II.3.2 Teoría factorial jerárquica.

Después de Spearman, la mayor influencia sobre las ideas de los psicólogos británicos con respecto a la inteligencia ha sido ejercida por Sir Cyril Burt. La innovación más importante de Burt fue el rechazar la postura que Spearman defendió decididamente, de que sólo se necesitaban dos clases de factores para

explicar las correlaciones entre los factores G y E.

En las décadas de los años 20 y 30 de este siglo se reunieron suficientes pruebas por parte de Kelly, Stephenson y El Koussy, entre otros, que demostraban definitivamente la aparición de factores de grupo, por ejemplo, para problemas verbales o problemas espaciales, y que en cambio no están representados en otras clases de tareas (Butcher, 1979).

Burt observó que una ligera rotación de dos o tres factores podía ser emprendida para llegar a un factor general y uno o dos factores de grupo principales. Como un resultado de la investigación de Burt, una estructura jerárquica de las habilidades humanas fue formulada.

Burt concibió su modelo de tipo jerárquico que puede aplicarse a la totalidad de la mente humana, con la mayor dicotomía entre las características intelectuales. Este investigador identifica varios niveles de bifurcación: en el primero y más alto nivel existen relaciones; en el segundo asociaciones, en el tercero percepciones y en el cuarto sensaciones.

Sin embargo, en su modelo, Burt tuvo que separar todas aquellas categorías que contenían más de dos factores. A nivel de asociaciones, reconoció que había divisiones en lo que se ha denominado memoria con un factor de retentividad general, donde fundamentalmente existen factores visuales, auditivos, cenestésicos, de memoria verbal, etc., y asociación productiva con un factor general de "inventiva", donde quedan suspendidos factores de fluidez y originalidad.

Otra asociación general puede considerar factores de habilidad verbal, habilidad de lenguaje y habilidad aritmética, bajo cada una de las cuales se encuentran agrupados dos o tres factores. (Morales, 1990).

La teoría jerárquica sugerida en primer lugar por Burt es un adelanto, porque reconoce la dependencia que existe entre el patrón factorial y la selección de tareas y sujetos, y explícitamente concede que un factor a un nivel dado puede una vez aparecer como general e importante y en otra ocasión como relativamente restringido y específico cuando se analiza un conjunto mucho mayor y variado de resultados. Uno de los méritos de Burt ha sido darse cuenta agudamente de la relatividad de las estructuras factoriales y de su dependencia respecto a la muestra de variables que se encuentran en la batería.

Otro investigador que ha propuesto una teoría jerárquica de la inteligencia fue Vernon. El modelo de Vernon está basado sobre g, el factor intelectual general, más factores de grupo principales y de menores o secundarios. Esto es: jerárquicamente, bajo el factor g se fundamentan dos grandes factores, el educacional-verbal y el K-M, que implica aspectos pragmáticos, como el modelo de Burt. Estos factores de grupo principales se subdividen en tres tipos: habilidad

espacial, habilidad manual e información mecánica (Morales, 1990).

Vernon admite que no existe una estructura final, ya que mucho depende de la población examinada, su heterogeneidad y fondo educacional, las pruebas particulares elegidas y las técnicas de factorización y rotación empleadas. Ha seguido el uso británico de nombrar a los factores con letras pequeñas para diferenciarlos de los correspondientes primarios americanos de los cuales el elemento g no ha sido eliminado (Vernon, 1965).

La teoría jerárquica está comprometida con el factor g. Se cree que este factor refleja un rasgo predominantemente heredado. Cuando a niños de varios años de edad, que tuvieron el beneficio de la preparación académica les son administrados una batería de pruebas cognitivas, la matriz factorial en adición a g generalmente produce un factor de grupo principal verbal-educacional y otro factor de grupo principal espacial-mecánico. El factor verbal-educacional puede ser de nuevo reducido a factores de grupo menores como verbal y numérico. Similarmente el factor de grupo espacial-mecánico puede ser reducido a factores menos académicos pero aún cognitivos. Estos factores de grupos principales y menores son reconocidos como las habilidades aprendidas. Cuando una batería de pruebas es administrada a niños pequeños o a una población generalmente no escolarizada los factores de grupo principales parecen ser menos definidos. La estructura jerárquica de las habilidades humanas nos da un mejor entendimiento de los problemas asociados con la medición de la inteligencia.

Existe un compromiso con g, a causa de que generalmente explica más del 50% de la varianza factorial común. Esto es, conocer acerca de g es tener una fuente principal de información acerca de un individuo. Las pruebas de factores de grupo mayores y menores proporcionan información adicional. (Lemke y Wiersma, 1976)

La teoría jerárquica tiene más ventaja que todos los demás modelos de las aptitudes humanas. Esta teoría explica la proliferación de los hallazgos aparentemente conflictivos y la multiplicidad de los factores de aptitud que han sido descritos y etiquetados con un nombre.

Una segunda importante ventaja que tiene este modelo es que posibilita reconocer que las diferentes clases de factores, y las pruebas basadas en ellos, sirven a diferentes propósitos de valoración y predicción. En igualdad de circunstancias, la medida de un factor general explicará una mayor parte de la varianza en los resultados y proporcionará una predicción sobre una serie más amplia de tareas que la medida de los factores de grupo mayores o menores. Esta es naturalmente la razón fundamental para continuar utilizando medidas de inteligencia general. Pero para familias particulares de destreza y para lograr una mayor precisión en una área más limitada puede ser muy necesario usar pruebas que midan factores de grupo mayores o menores (Butcher, 1979).

II.3.3 Teoría Multifactorial.

Tanto las teorías de Spearman en Inglaterra como las de Binet en Francia fueron rápidamente aplicadas en los Estados Unidos; sin embargo, debe hacerse notar que los psicólogos norteamericanos mostraron mayor interés por el análisis factorial de Spearman. De 1920 a 1930, estos psicólogos estudiaron exhaustivamente los métodos matemáticos del análisis factorial. Destacaron en esta actividad personalidades como T.L. Kelly y K.J. Kalzinger. (Morales, 1990).

Sin embargo, la teoría de Spearman también contó con opositores, se encuentran entre otros el norteamericano Edward Lee Thorndike (1874-1949). Su investigación junto con la de sus colaboradores (1927) les llevó a postular la existencia de un gran número de aptitudes específicas independientes, cada una de ellas con un sustrato neuronal, y cada una mostrada en diferentes combinaciones de las distintas pruebas mentales. Thorndike destacó la independencia relativa de cada función específica y la concepción de que la inteligencia no era homogénea a su juicio, la inteligencia debería ser considerada como la suma total de muchas funciones distintas.

E.L. Thorndike fue el primero en desarrollar claramente la teoría de que la medida de la inteligencia consiste esencialmente en una evaluación cuantitativa de los productos mentales en términos de números y la eficacia y velocidad con que son realizados. Afirmaba que las habilidades son producciones mentales ordenadas en diferentes clases o tipo de operaciones. Así, por ejemplo, sugirió que la inteligencia debía agruparse en tres categorías: a) inteligencia abstracta o verbal, mediante el manejo fundamental de símbolos; b) inteligencia práctica, por medio de la manipulación de objetos, y c) inteligencia social, facilitadora de las confrontaciones con los seres humanos. (Morales, 1990).

Posteriormente, L.L. Thurstone (1887-1955) se concentró en el desarrollo matemático del análisis factorial; construyó técnicas matemáticas para lo que él llamó análisis factorial múltiple, a fin de explorar las habilidades humanas. Esta investigación llevó a Thurstone en 1941 a desarrollar un punto de vista teórico que diera alguna alternativa al problema de la naturaleza de la inteligencia. dicha teoría fue llamada más tarde "Teoría de los factores múltiples" o "multifactorial". (Butcher, 1979).

En 1930, aún se discutía si Spearman estaba en lo cierto por lo que respecta a su enfoque acerca de la estructura de la inteligencia. Thurstone rechazó la afirmación de Spearman en este sentido e intentó contestar a la pregunta: ¿Cuántos y qué tipos de factores presentan correlación con las pruebas de inteligencia? Para dar respuesta satisfactoria a esta pregunta llevó a cabo minuciosos estudios, a fin de mejorar los métodos de análisis factorial existentes. Para estos fines específicos empleó más de sesenta pruebas psicológicas desarrolladas. Cuando realizó sus investigaciones no tuvo a su disposición un equipo suficiente para el procesamiento de datos, es decir, para computar el

material obtenido. Esta fue una de las razones principales por las que utilizó el método centroide y de rotación de factores. Así, una vez que se aplicaron los instrumentos elaborados para la investigación, administrados adecuadamente, tabulados y llevados a los procedimientos estadísticos específicos, se obtuvo una tabla de interrelaciones factorizada, en la cual aparecieron siete factores mediante el método centroide. Al revisar la matriz factorial rotada aparecieron los siguientes factores: a) comprensión verbal, b) fluidez verbal, c) factor numérico, d) factor espacial, e) velocidad perceptual, f) memoria (memoria de asociación y memoria de significado de relaciones) y g) factores de razonamiento general (deducción o diseño de conclusiones y educación de relaciones) (Morales, 1990).

En los tempranos días del desarrollo del análisis factorial, los teóricos habían encabezado argumentos con respecto a si los factores debían ser rotados, y si lo eran, exactamente cómo deberían ser rotados. Antes de proseguir con nuestra exposición sobre la teoría multifactorial, brevemente resumiremos las cuestiones sobre la rotación de factores.

La matriz factorial "sin rotar" significa que los componentes principales justamente son dados conforme ellos emergen del análisis matemático, cada uno explicando el componente lineal más grande posible de varianza que es independiente de la varianza explicada por todos los componentes anteriores. Cuando rotados a otra posición, ellos ya no son componentes principales, sino factores rotados. El primer factor después de la rotación ya no es un factor general en el sentido de que explique la máxima cantidad de varianza en todas las pruebas. La varianza total permanece, claro inalterada. La rotación meramente cambia los ejes de referencia. La rotación frecuentemente es hecha porque por lo general clasifica y simplifica la identificación, interpretación y nombramiento de los factores de grupo; la rotación es totalmente análoga a tomar una fotografía del mismo objeto desde un ángulo distinto.

Con respecto a la posición en que deben ser rotados los factores, no existe alguna regla a seguir. La idea principal es que al rotar los factores a cualquier posición se tenga una descripción más clara de la estructura factorial de todas las pruebas. Thurstone en 1947 propuso un criterio para la rotación factorial que él nombró "estructura simple".

Thurstone desarrolló este principio de estructura simple a partir de la creencia intuitiva de que en un conjunto amplio y representativo de tareas mentales estarán implicadas aptitudes que facilitarán grandemente algunas de las tareas, pero en cambio no afectarán a otras. Por ejemplo, no se puede esperar que la aptitud numérica influya en los resultados de una tarea de fluidez verbal (Butcher, 1979).

Una matriz factorial idealizada con estructura simple perfecta daría como resultado que cada prueba representara un único factor. Tales pruebas serían llamadas pruebas de factor puro, porque ellas medirían sólo un factor, no contaminado por los otros. Fue el sueño de Thurstone inventar tales pruebas de

factor puro para medir las siete "Habilidades mentales primarias" representadas por los siete factores que él tuvo éxito en extraer confiablemente de una multitud diversa de pruebas altamente cognitivas (Jensen, 1980).

El criterio de la estructura simple de Thurstone ciertamente no admite la posibilidad de descubrir un factor general. El factor general conspiró contra el sueño de Thurstone. Penetraba en todas las pruebas y así hacía imposible hacer nada más que una aproximación a la estructura simple; las pruebas siempre tenían cargas sustanciales sobre más de un factor porque la rotación despliega el factor general sobre los varios factores rotados, así que la estructura simple, aunque puede ser más o menos aproximada no puede completamente ser alcanzada mientras exista un factor general. Para abordar este problema Thurstone adoptó el método de rotación oblicua de factores.

Lo que es notable acerca de la investigación de Thurstone es que nos regresa a la cuestión presumiblemente planteada por Spearman en 1904, es decir, ¿es la inteligencia una o muchas cosas, y pueden los puntajes sobre pruebas de inteligencia apropiadamente ser definidos en términos de un único número?

Al principio, Thurstone estuvo inclinado a argumentar que sus resultados estaban en completa contradicción con los de Spearman.

Sin embargo, al encontrar Thurstone necesario usar factores oblicuos para obtener soluciones que encajaran con el criterio de la estructura simple, estaba utilizando el llamado análisis factorial de segundo orden. Cattell, en 1940 señaló que el análisis factorial de segundo orden resultaría en un acercamiento de los puntos de vista de Spearman y Thurstone.

Desde este punto de vista g emergía como un factor de segundo orden. Thurstone estuvo en acuerdo sustancial con este punto de vista. Sin embargo, él señaló que aunque el análisis factorial de segundo orden producía un factor similar a la g de Spearman, no era el único factor de segundo orden que emergía.

Por otra parte, Spearman podría señalar que los resultados de la investigación analítica factorial de Thurstone de la habilidad realmente no contradecía la existencia de g .

En resumen, podría ser argumentado que los estudios de Thurstone no habían mostrado que g no existía sino que g podría ser distribuida y dividida en componentes que estuvieran entre sí positivamente relacionados. Ya que Thurstone reconoció la existencia de g y Spearman reconoció la existencia de los factores de grupo representando habilidades especiales, podemos decir que es aparente la diferencia entre Spearman y Thurstone.

J.P. Guilford.

La teoría de la inteligencia de Guilford representa una separación radical de la tradición Spearman-Thurstone. Guilford había rechazado los ampliamente aceptados conceptos de factor g y s y en su lugar propuso un modelo tridimensional de la estructura del intelecto (Matarazzo, 1976).

Guilford ha enfatizado las relaciones que existen entre los factores de habilidad humana y los procesos. El descubrimiento de los componentes de la inteligencia se ha llevado a cabo a través de experimentos con el método del análisis factorial. Nunnally (en Morales, 1990) indica que cada uno de los componentes o factores intelectuales constituye una habilidad única y necesaria para "hacer algo bien", cuando un sujeto tiene que ejecutar alguna tarea (o prueba).

Guilford, utilizando ampliamente el análisis factorial desarrolló su teoría de estos problemas a la que denominó "estructura del intelecto". Así, aun cuando Guilford considera que cada factor es perfectamente distinto de otro para ser apreciado mediante el análisis factorial, en los últimos años se ha considerado necesario clasificarlos, dado que son similares en algunos aspectos. El criterio de clasificación empleado por Guilford ha sido en función del tipo de proceso o de operación ejecutada. Así, este investigador obtuvo cinco grupos de habilidades intelectuales: factores de cognición, de memoria, de pensamiento convergente, de pensamiento divergente y de factores de evaluación.

Desde otro punto de vista, a los factores intelectuales los clasifica Guilford de acuerdo con la clase de material o contenido que utilizan. Así, hasta ahora se han encontrado tres clases de materiales: figuras, símbolos y un contenido de tipo semántico.

Estas tres clases de materiales pueden ser representados por un modelo sólido (en forma de cubo) que simboliza el "intelecto humano" o "estructura del intelecto" en donde se pueden observar las dimensiones, y a través de cada una de ellas operan diversos elementos; en la primera, distintos tipos de operaciones; en la segunda, productos de tipo variado, y en la tercera diversas clases de contenido. Además se ha agregado una cuarta categoría, a la cual se le ha denominado "contenido conductual" para representar en cierto modo la llamada "inteligencia social".

La cognición es básica para cualquier tipo de operaciones intelectuales, por lo que, en orden de clasificación de dimensiones, aparece primero. Así sin cognición no hay memoria, y sin ésta no hay producción. Asimismo, sin cognición ni producción, no hay evaluación.

Cada celdilla del modelo corresponde a cierta clase de habilidad, que se puede describir mediante operaciones, contenidos y productos; es decir, cada celdilla es la intersección de una combinación de tipos de operaciones, contenidos y

productos. (Morales, 1990).

Raymond B Cattell.

La amplia difusión de la aplicación del análisis factorial múltiple a la investigación de las habilidades parece haber llevado al análisis factorial más allá de sus limitaciones descriptivas y conceptuales como una herramienta de investigación, resultando en un número incrementado de factores de la "mente". Resulta paradójico que aunque el modelo factorial múltiple desarrollado por Guilford (estructura del intelecto), tratara de salvaguardar a la investigación sobre las habilidades, de este caos empírico, de hecho puede haber contribuido a un adicional "empiricismo" al esforzarse en llenar las celdas vacías del modelo cúbico de 150 factores.

Fue hasta con Raymond B Cattell (un primer asistente de Spearman y de Cyril Burt) que se proporcionó una nueva reorientación hacia dimensiones más amplias de ejecuciones intelectuales y no tan específicas como los factores de Guilford.

La teoría de Cattell sobre la inteligencia fluida y cristalizada -la teoría gf-gc- es una de las teorías sobre amplios factores que más ha atraído la atención en los últimos 15 años.

La teoría de Cattell sobre la inteligencia es una síntesis contemporánea de las tradiciones de Spearman y Thurstone. Como Spearman, él es un defensor de la importancia de g. Y como Thurstone, creyó en el uso de rotaciones oblicuas y derivar a g como un factor de segundo orden. Los análisis de Cattell de la inteligencia toman como su punto de inicio un análisis de las habilidades primarias bastante análogo a los procedimientos seguidos por Thurstone, esto es, existe un intento para muestrear un amplio rango de pruebas y usar rotaciones oblicuas para satisfacer el criterio de la estructura simple (Brody y Brody, 1976).

Si analizamos factorialmente un número grande de pruebas mentales altamente diversas y rotamos los ejes factoriales tanto como nos sea permitido para que estén oblicuos (es decir correlacionados) para aproximarse al criterio de la estructura simple tan estrechamente como sea posible y entonces analizamos factorialmente las correlaciones entre los factores de segundo orden, los cuales también rotamos a la estructura simple, eventualmente revelaremos ya sea un factor o dos factores grandes, esto es, ya sea g o gf y gc (Jensen, 1980).

La teoría gf-gc primero fue presentada en 1941, estuvo basada en parte sobre las observaciones en el laboratorio de Spearman hechas en los años 1930, las cuales indicaron que las pruebas, incluyendo clasificaciones perceptuales y analogías, estaban altamente intercorrelacionadas y parecieron ser medidas particularmente claras de g. Por otro lado, las pruebas que parecieron estar más relacionadas con el rendimiento escolar y reflejar el conocimiento adquirido en una situación escolar tendieron a cargar algo más bajo en g y estar algo

separadas de las pruebas de clasificación perceptual y analogías. Las pruebas que eran medidas relativamente puras de *g* fueron similares a aquellas que venían a ser usadas en pruebas "libres de cultura" o de "cultura reducida" (Brody y Brody 1976).

La distinción esencial entre inteligencia general fluida y cristalizada o *gf* y *gc* en resumen, puede ser rebuscada, en parte, a partir de notar las clases de pruebas que cargan muy pesadamente en uno o en otro factor. Las pruebas cargadas principalmente sobre *gf* son aquellas que tienen poco contenido informacional pero demandan la habilidad para ver relaciones, frecuentemente relaciones complejas, entre elementos relativamente simples: series de números, series de letras, clasificación de figuras, analogías de figuras, visualización espacial, pruebas de cierre, figuras que deben ser encajadas diseños de bloques y matrices. Las pruebas cargadas principalmente sobre *gc* son aquellas que tienen contenido informacional y representan los conocimientos y habilidades ya adquiridas del sujeto: información general, vocabulario, aritmética, información mecánica e identificación de herramientas, silogismos verbales y otros problemas de razonamiento lógico formal y analogías verbales difíciles de entender.

La distinción *gc-gf* no es la misma cosa que la distinción verbal-no verbal. Por ejemplo, las analogías verbales basadas sobre palabras altamente familiares pero demandando un alto nivel de educación de relaciones están cargadas sobre *gf*, mientras que las analogías basadas sobre palabras difíciles o especializadas en términos raramente encontrados fuera del contexto de la educación formal están cargados sobre *gc* (Jensen, 1980).

De acuerdo a Cattell, tanto la inteligencia fluida como la cristalizada representan "inteligencia" como comúnmente es conceptualizada, es decir: procesos de percibir relaciones, descubrir correlatos, abstraer en tareas aceleradas y no aceleradas, pero *gf* es considerada a representar influencias grandemente constitucionales y fisiológicas, mientras que *gc* está más relacionada a influencias educacional-experiencias, a técnicas y estrategias adquiridas, y así representa la sabiduría acumulada de una cultura (Undheim, 1987b).

CAPITULO III

PRUEBAS DE INTELIGENCIA

III.1 WISC-R.- ESPAÑOL. ESCALA DE INTELIGENCIA REVISADA PARA EL NIVEL ESCOLAR.

Como indicamos anteriormente Wechsler elaboró, en 1939, su primera escala de inteligencia, Wechsler-Bellevue, la cual era aplicada en un amplio rango de edades que iban desde los diez hasta los sesenta años. Esta escala quedó conformada por seis subpruebas verbales y cinco de ejecución. Dichas subpruebas, en el orden siguiente, se denominan:

Escala Verbal

- a) Subprueba de información
- b) Subprueba de comprensión general
- c) Subprueba de aritmética
- d) Subprueba de semejanzas
- e) Subprueba de retención de dígitos
- f) Subprueba de vocabulario.

Escala de Ejecución

- a) Subprueba de completamiento de figuras
- b) Subprueba de diseño con cubos
- c) Subprueba de ordenamiento de figuras
- d) Subprueba de ensamble de objetos
- e) Subprueba de claves

Esta primera escala sufrió, en 1949 una revisión en la cual se dividió en dos escalas, la de niños y la de adultos.

La escala de Wechsler para niños fue construida bajo los mismos principios que la de adultos, es decir, quedó conformada por una escala verbal y una de ejecución. Asimismo, la unidad de calificación fue el C.I. total, el cual se obtiene de los C.I.s verbal y de ejecución.

La diferencia que existe entre la escala de Wechsler para niños (WISC), respecto de las demás escalas, es que ésta consta de doce subpruebas, de las cuales dos son opcionales o suplementarias (retención de dígitos y laberintos).

Asimismo, contiene una subprueba que sustituye al de Claves para niños menores de ocho años y que consiste en una prueba de las llamadas de código (correspondencia de figuras geométricas con números).

Los datos de confiabilidad son buenos, aunque no óptimos, lo que puede deberse al

error de medición propiciado por el desarrollo psicológico y normal del niño.

Morales (1990) también reporta que los estudios de validez demuestran que tanto las subpruebas como la escala completa miden factores comunes. Mientras cada subprueba tiene poco en común con las demás, las subpruebas verbales combinadas tienen más en común con cada Subprueba verbal individual. Lo mismo ocurre con las subpruebas de ejecución: en ellas, la validez concurrente alcanza altos niveles de correlación tanto en la escala de Standford-Binet como en la de Raven, sin embargo, Morales no aporta más datos acerca de estos estudios, los cuales resultarían de suma importancia, dado que en la presente investigación nos interesa conocer la correlación de las escalas de Wechsler para niños con el Raven.

El WISC se ha caracterizado como una herramienta práctica en la clínica y en el diagnóstico, cuyo uso se ha incrementado en las áreas de evaluación educativa, de apreciación del aprendizaje y de algunas incapacidades. Sin embargo, en base a los comentarios críticos que emitieron psicólogos con amplia experiencia en la aplicación y evaluación del WISC, en 1974 se llevó a cabo la revisión del mismo.

La revisión del WISC representa la conjunción de dos metas opuestas; primero, la conservación hasta donde fue posible, del WISC 1949, tanto por su uso extendido, como por su amplia aceptación y segundo, la modificación o eliminación de reactivos ambiguos, obsoletos o injustos para determinados grupos de niños.

El WISC-R, Escala de Inteligencia Revisada para el Nivel Escolar, quedó construida por las mismas doce subpruebas del WISC publicado en 1949 (seis pertenecientes a la escala verbal y seis a la escala de ejecución). Las doce subpruebas, al igual que la subescala denominada como "Visual Motor Memory" se aplicaron a la muestra completa de estandarización, esta última subescala fue eliminada más tarde del WISC-R porque representaba dificultades para aplicarse y por las estrictas normas estadísticas que se establecieron.

Solamente diez de las subpruebas del WISC-R se consideran como básicas. El C.I. se calcula a partir de cinco subpruebas verbales y cinco de ejecución, que se mencionan a continuación (el número asignado a cada una de ellas corresponde al orden de aplicación).

Escala Verbal	Escala de Ejecución.
1 Información	2. Figuras incompletas
3 Semejanzas	4 Ordenación de dibujos
5 Aritmética	6 Diseño de cubos
7 Vocabulario	8 Composición de objetos
9 Comprensión	10 Claves
(Retención de dígitos)	(Laberintos)

Tanto retención de dígitos en la escala verbal, como laberintos en la de ejecución no se utilizan para establecer los cuadros de los C.I.s. Se conservaron como subpruebas complementarias que tienen por objeto equilibrar la calificación global del sujeto y se utilizan como sustitutos cuando existe un impedimento para aplicar alguna de las subpruebas básicas o cuando una de ellas queda invalidada. Sin embargo, una sustitución no puede hacerse simplemente porque el niño puntúa bajo en una Subprueba particular, verbal o de ejecución.

Siempre es permisible aplicar las doce subpruebas; aún más, en situaciones clínicas, el incluir las subpruebas complementarias es muy recomendable porque proveen información cualitativa y diagnóstica. No obstante, si el examinador aplica Retención de Dígitos o Laberintos adicionalmente a los básicos, no puede incluir éstos en el cálculo del C.I. del niño.

En cuanto a la dicotomía en escala verbal y de ejecución, es en primer lugar una manera de identificar los dos principales modos por medio de los cuales se expresan las habilidades del humano. Los estudios factoriales se han incrementado a través de los años y han confirmado la validez de esta amplia dicotomía.

Es importante considerar que las normas del WISC-R se derivan de grupos representativos de la población de niños en E.U.A. En la estandarización se adoptó un plan de muestreo estratificado para asegurar que en la muestra normativa se incluirían proporciones representativas de distintas clases de niños. Desde el principio de la estratificación se seleccionaron variables de acuerdo a los reportes del Censo de E.U.A. del año 1970.

Las variables que se utilizaron fueron: edad, sexo, raza (blancos y no blancos), región geográfica, ocupación del jefe de familia y residencia (urbana y rural).

De acuerdo con la teoría de medición del autor, que hace énfasis en la comparación de un niño con la edad cronológica de compañeros de su misma edad, los C.I. del WISC-R están basados en puntuaciones normalizadas derivadas separadamente para cada grupo de edad.

Al construirse los cuadros de C.I. para el WISC-R se obtuvieron tres sumas de puntuaciones normalizadas (de la escala verbal, de ejecución y total), de las cuales se obtienen el C.I. verbal, C.I. de ejecución y C.I. total. Para cada uno de los tres C.I. de escalas, la media y la desviación estándar de la suma correspondiente a la puntuación normalizada se igualó a 100 y 15 respectivamente.

Un C.I. de 100 en cualquiera de las escalas define la ejecución del niño promedio, de una edad determinada, en esa escala. Los C.I. de 85 y 115 corresponden a una desviación estándar de la media, mientras que los C.I. de 70 y 130 se encuentran a 2 desviaciones estándar de la media.

En consonancia con la teoría que especifica los niveles de inteligencia por medio de la definición estadística, uno inevitablemente vincula las categorías cualitativas y diagnósticas históricamente utilizadas, a los valores numéricos de C.I. El siguiente cuadro proporciona una clasificación de los equivalentes de C.I. para términos diagnósticos de uso común.

Clasificación de la Inteligencia

C.I.	Clasificación
130 y por encima	Muy Superior
120-129	Superior
110-119	Arriba del normal (Brillante)
90-109	Normal
80-89	Abajo del normal (torpe)
70-79	Límitrofe
69 y hacia abajo	Deficiencia mental

El examinador debe intentar comprender la capacidad del niño basándose en los C.I. de las escalas del WISC-R y en las subpruebas individuales. "Pero de ninguna manera, el examinador debe estimar o inferir los C.I. de la escala total a partir de la puntuación normalizada de una sola subprueba, o a partir del promedio de las puntuaciones normalizadas de una sola subprueba, o a partir del promedio de las puntuaciones normalizadas de un número limitado de subpruebas". Los C.I. deben calcularse en la forma prescrita en el capítulo correspondiente del manual.

Medidas de g.

De acuerdo a Kaufman (1982), hay varios métodos para determinar el grado en que una subprueba mide g, o inteligencia general; entre éstos se encuentran el uso de análisis de factores de segundo orden, el análisis jerárquico de factores y el primer

factor sin rotar en el análisis de factores principal. Todas estas técnicas parecen producir resultados similares y la elección de algún procedimiento particular es arbitraria.

En general, las doce subpruebas del WISC-R, de acuerdo al análisis de factores de Kaufman, quedan agrupados en tres categorías dependiendo de su carga promedio para todas las edades consideradas:

Buenas medidas de g	Medidas regulares de g	Medidas deficientes de g
<ul style="list-style-type: none">- Vocabulario.- Información.- Semejanzas.- Diseño de Cubos.- Comprensión	<ul style="list-style-type: none">- Aritmética.- Completamiento de figuras- Composición de objetos.- Ordenamiento de figuras.	<ul style="list-style-type: none">- Retención de dígitos.- Laberintos- Claves

Descripción de las Subpruebas del WISC-R, de acuerdo a Morales (1990)

Escala verbal

Información:

El principal soporte del grado de información de una persona, considerado como una buena medida de inteligencia, fue obtenido del Army Alfa (el alfa del ejército). La subprueba de información consta de 30 reactivos, el orden en el que se presentan es el que indicó como necesaria la muestra sobre la cual se hizo la estandarización.

Las primeras funciones que contiene la subprueba de información se presentan a partir de la composición factorial de cada una de ellas, en este caso, en primer lugar una carga considerable del factor de memoria, operaciones de asociación y organización de la experiencia. Es una subprueba sujeta a la influencia de las oportunidades culturales, del medio ambiente y de los intereses.

Semejanzas.

La carga factorial de esta subprueba se ubica en el factor g, fundamentalmente a través de todos los niveles de edad. No obstante, es de vital importancia la carga del factor verbal, lo que hace que esta subprueba conduzca, primeramente, a la medición de habilidades de generalización y de abstracción.

La subprueba consta de diez y siete reactivos apareados, es decir, diez y siete pares de palabras en las cuales se le pide al sujeto que dé una respuesta en relación con la razón del porqué están o deben estar juntas.

Aritmética.

Esta subprueba consta de una carga factorial muy alta de factor memoria, aunque, como cabe esperarse, presenta una gran saturación del factor g. De acuerdo con algunos análisis factoriales como el de Davis parecen encontrarse cargas significativas no sólo en el razonamiento general sino también en factores de fluidez numérica y de conocimiento e información.

Consta de dieciocho problemas que exigen concentración y atención específica. En este tipo de subpruebas se requieren operaciones de razonamiento con abstracciones, formación de conceptos y retención de procesos aritméticos. Además, dicha subprueba se ve grandemente influida por la atención inmediata y automática, así como por la oportunidad de adquirir los procesos aritméticos fundamentales.

Vocabulario.

Este tipo de subpruebas no sólo señala el índice de la escolaridad de una persona, sino que constituye una buena y confiable medida de inteligencia, debido a que el número de palabras que una persona conoce, constituye una medida tanto de su habilidad para aprender, como de su información verbal y de la amplitud general de sus ideas. Este tipo de actividad está influida por las oportunidades culturales y educacionales de la persona.

Esta subprueba consta de treinta y dos palabras seleccionadas de un diccionario sobre una base de "grado de dificultad" y elegidas al azar, de las cuales, el sujeto debe dar una definición. La definición es de gran importancia, porque muestra el nivel cultural y educacional de las personas.

Comprensión.

Esta subprueba depende, en individuos normales, del factor de comprensión verbal y del factor g. Contiene diez y siete situaciones-problema, en las que el sujeto debe comprender cuál es el problema contenido en la situación y debe dar la o las respuestas adecuadas, dentro de lo que se espera en un grupo "normal" de comportamiento. El éxito de la prueba depende de las funciones u operaciones de razonamiento abstracto, organización del conocimiento y capacidad de utilizar en la práctica las experiencias y aprendizajes obtenidos.

Retención de dígitos.

Desde el punto de vista operacional, esta subprueba consiste en la repetición de números como única habilidad o habilidad específica para su ejecución. Además, se encuentra fuertemente influido por el factor memoria y con baja correlación con el factor g. Parece ser que la habilidad para atender y concentrarse es fundamental para realizar correctamente esta tarea.

Wechsler afirma que la habilidad o inhabilidad para repetir o no los dígitos depende de lo que él ha llamado "Factores no intelectivos de la inteligencia". Tal vez la posibilidad o libertad de poder distraerse y no responder correctamente constituye uno de esos factores.

Las principales funciones que se advierten en esta prueba son: el recuerdo inmediato, la imaginación auditiva y la imaginación visual simultánea; y como factores que influyen notablemente en la ejecución se encuentran la atención inmediata y la atención dirigida.

Esta prueba consta de series de dígitos que varían en longitud, desde tres hasta nueve dígitos (en orden directo) y desde dos hasta ocho dígitos (en orden inverso).

Escalas de Ejecución

Figuras incompletas.

Esta subprueba consta de veintiséis reactivos consistentes en tarjetas que muestran una figura a la cual le falta una parte importante, que el sujeto debe encontrar y dar su nombre.

Factorialmente, en esta subprueba aparece una situación especial: presenta una carga sobre factores no verbales en la mayoría de los análisis efectuados, y muestra poca influencia del factor g, la más baja de todas las subpruebas de ejecución. Sin embargo surge un único y nuevo factor que parece importante en el comportamiento intelectual y que no ha podido ser identificado con precisión. Intervienen en él factores como la percepción visual (análisis) y la imaginación visual.

Dadas las operaciones a ejecutar en esta tarea, puede, obviamente, considerarse que influye de manera notable la experiencia ambiental y la agudeza visual. Asimismo, mide ostensiblemente habilidades perceptuales y conceptuales implicadas en el reconocimiento e identificación de objetos y formas conocidas.

Ordenación de dibujos.

Esta subprueba consta de doce series de figuras. Cada serie se le presenta al sujeto desordenadamente, para que la ordene en una secuencia correcta y forme una historia. De este modo, se intenta medir la comprensión y la evaluación de una situación por parte del sujeto, sin la intervención del lenguaje hablado.

Factorialmente, presenta una carga de factores no verbales, aunque menos significativamente que la subprueba de completamiento de figuras o ensamble de objetos.

Las principales funciones implicadas en la solución de los problemas son: percepción visual de relaciones, y síntesis de material no verbal, las cuales están grandemente

influidas por la agudeza visual y por un mínimo de oportunidades culturales, así como la función de planeación y anticipación. La anticipación que se requiere, es una habilidad para prever las consecuencias de los actos o situaciones iniciales e interpretar situaciones sociales.

Diseño de cubos.

Esta subprueba consta de once diseños diferentes, los cuales, al igual que las demás subpruebas de esta escala, son progresivas en grado de dificultad. Cada diseño está constituido por cubos de colores (rojos y blancos); el sujeto debe emplear sólo cuatro de ellos para construir o reproducir los ejemplos que le son presentados en la primera parte de la prueba. Cuando el examinador lo indique, debe emplear los nueve cubos.

Por otra parte, la carga factorial indica un factor de organización no verbal, y tiene gran influencia del factor g. También se ha encontrado que esta subprueba muestra una carga mayor no sólo en visualización, sino, también en factores de conocimiento e información mecánica.

Las principales funciones en esta subprueba son: percepción de forma (análisis) e integración visomotora, las cuales son factores que influyen en la ejecución motora y en un mínimo de capacidad discriminativa del color.

Composición de objetos.

Esta subprueba está constituida por cuatro figuras, las cuales le son presentadas al sujeto en forma de rompecabezas, es decir, cortados de tal manera que puedan ser acomodadas para obtener la forma que represente un objeto familiar al examinado.

Se ha encontrado que dicha subprueba correlaciona poco con el resto de las subpruebas de la escala y produce poca diferenciación entre los individuos, pero ayuda al examinador a comprender los estilos de percepción, el grado de confiabilidad que se puede tener en los métodos de ensayo y error, y la manera en que se reacciona a los errores. Asimismo, esta subprueba expresa un factor analítico: con excepción del factor g, casi toda la varianza está en función de factores de organización no verbales: percepción visual (síntesis), integración visual y precisión de la actividad motora.

Claves.

Esta subprueba consta de un rectángulo dividido en nueve cuadros o celdillas dobles. En las celdillas de la parte superior se encuentran los números del uno al nueve y en el de la parte inferior (las cuales corresponden a cada uno de los números) se halla un símbolo. Enseguida, se encuentran cuatro rectángulos divididos en cien celdillas dobles, en las cuales están distribuidos al azar los números en la parte superior.

En esta subprueba, la tarea consiste en asociar los números con los símbolos que aparecen en la parte superior. En realidad, lo que se le pide al sujeto es que asocie

símbolos con símbolos en un tiempo dado, con lo cual determinará la medida de habilidad intelectual en esta área.

Las funciones fundamentales que implica esta prueba son: el recuerdo rutinario inmediato (memoria a corto plazo), la integración visomotora y la imaginación visual.

Factorialmente, a partir de g, esta subprueba muestra una fuerte carga factorial en memoria.

Claves implica un proceso de codificación verbal. Este proceso se refiere al hecho de que muchos de los símbolos tienen etiquetas distintas a la vista. Por ejemplo, el símbolo + se puede clasificar como "signo de suma" o "cruz" y el v como la letra "v". La ejecución mejora cuando los símbolos se recodifican en términos de clasificaciones verbales. Por tanto, también se puede decir que la subprueba de claves mide la habilidad para aprender combinaciones de símbolos, figuras y hacer asociaciones rápidas y correctas. (Sattler, 1988).

Laberintos

Esta subprueba consta de nueve laberintos con grado creciente de dificultad, de los cuales el examinado debe encontrar la salida siguiendo los caminos disponibles desde el centro del mismo, procurando no entrar a callejones sin salida.

Esta subprueba proporciona información acerca de la capacidad de planeación y de la capacidad de seguir un patrón visual.

III.2 WISC-RM ESCALA DE INTELIGENCIA REVISADA PARA EL NIVEL ESCOLAR

El psicólogo clínico mexicano que intenta hacer evaluaciones de la inteligencia por medio de algún instrumento psicométrico, se ve confrontado con dos grandes problemas. Por un lado, no hay instrumentos desarrollados específicamente para medir la inteligencia de los niños mexicanos. Por el otro, los instrumentos de que dispone no están estandarizados, lo cual hace que la confianza que pueda tener en sus resultados sea mas bien tentativa.

García (1987) señala que autores tales como Ahumada y Díaz Guerrero (1964), Reyes-Lagunes (1966), Reyes-Lagunes y Díaz Guerrero (1967), han emprendido esfuerzos para modificar lo antes señalado, pero por alguna razón, los trabajos de dichos investigadores no han recibido la difusión apropiada y, por lo tanto, no han tenido la aceptación que ameritarían.

Asimismo, señala García, que hay evidencias empíricas de que resulta injusto para los niños mexicanos el ser evaluados con instrumentos no estandarizados; tales evidencias indican que, al evaluar a un niño mexicano con el WISC-R, utilizando normas desarrolladas para niños norteamericanos, se comete un error de hasta 10 puntos en su coeficiente intelectual total, lo cual implica ubicar al niño erróneamente en

la categoría descriptiva inferior.

Por lo anterior, el Departamento de Investigación, dependiente de la Dirección Técnica, de la Dirección General de Educación Especial, llevó a cabo la estandarización del WISC-RM, publicándose la misma en el año 1983, con el propósito de contar un instrumento que nos proporcionara normas de evaluación más acordes o apropiadas a la población mexicana.

La estandarización del WISC-RM se llevó a cabo con una población de niños y adolescentes mexicanos, en México, D.F., los cuales se encontraban inscritos en Escuelas Primarias y Secundarias oficiales de la Secretaría de Educación Pública, durante el ciclo escolar 80-81, en turnos matutinos y vespertinos.

La muestra incluyó a 100 estudiantes en cada uno de los 11 grupos de edades, desde los 6 1/2 hasta los 16 1/2 años; la muestra total comprendió 1100 casos.

La muestra se seleccionó al azar, partiendo de la población definida anteriormente, asimismo, la selección de estudiantes estuvo determinada, en cuanto a la cantidad de sujetos a investigar, por los requerimientos mínimos necesarios para procesar los datos estadísticamente.

Para la derivación de los puntajes normalizados, la construcción de los cuadros de C.I. y la interpretación de los C.I. se siguió el mismo procedimiento utilizado en el WISC-R.

Asimismo, las consideraciones para la aplicación son iguales, ya que las subpruebas están conformadas de la misma forma que el WISC-R.

Naturaleza General de los Cambios.

La revisión del WISC-RM representa la conjunción (al igual que la revisión del WISC-R en 1974) de dos metas opuestas; primero, la conservación hasta donde fue posible, del WISC 1949 (y del WISC-R 1974) tanto por su uso extendido, como por su amplia aceptación, y segundo, la modificación o eliminación de reactivos ambiguos, obsoletos o injustos para determinados grupos de niños.

A continuación se describen los cambios que sufrieron las subpruebas, de acuerdo al Manual de Aplicación del WISC-RM

Cambios en las subpruebas verbales.

- **Información**

En el WISC-RM se eliminaron nueve reactivos del WISC-R sobre la base de la correlación reactivo-escala, los efectos de los reactivos individuales sobre la confiabilidad de la escala, así como el porcentaje de la muestra que paso o fracasó en el reactivo. De estos nueve reactivos eliminados, siete fueron sustituidos, quedando la subprueba constituida por veintiocho reactivos. El orden

de los mismos varió, con respecto al WISC-R.

Aplicación y calificación: En el WISC-R los niños entre 8 y 10 años empiezan con el 7 y los de 14-16 años con el 11; en el WISC-RM, los niños entre 11 y 12 años empiezan con el reactivo 5 y los de 13-16 años con el 11.

- **Semejanzas**

Únicamente se cambió el orden, quedando las subpruebas constituidas por los mismos diez y siete reactivos del WISC-R.

- **Aritmética**

El único cambio en el WISC-RM fue en el orden de los reactivos 13 y 14.

- **Vocabulario.**

En el WISC-RM se eliminaron siete reactivos sobre las bases de la correlación reactivo-escala, los efectos de los reactivos individuales sobre la confiabilidad de la escala y el porcentaje de la muestra que pasó o fracasó el reactivo. De los reactivos eliminados, seis fueron sustituidos. El orden también fue substancialmente cambiado.

- **Comprensión**

En el WISC-RM se eliminaron 5 reactivos del WISC-R y fueron substituidos por cinco nuevos reactivos sobre la base de la correlación reactivo-subprueba, los efectos de los reactivos individuales sobre la confiabilidad de subprueba y el porcentaje de la muestra que pasó o fracasó en el reactivo. El orden también fue cambiado.

- **Retención de dígitos**

Esta subprueba queda en el WISC-RM sin ninguna modificación.

Cambios en las Subprueba de Ejecución

- **Figuras Incompletas.**

En el WISC-RM se eliminaron 3 reactivos , para mejorar la confiabilidad, dejando 23 reactivos. También se cambió el orden de los reactivos.

- **Ordenación de Dibujos.**

En el WISC-RM el único cambio fue respecto al inicio de aplicación; los niños entre 6 y 8 años empiezan con el reactivo 1, y los de 9-16 con el 2.

- **Diseños con Cubos.**

En esta subprueba también el único cambio fue respecto al inicio de aplicación; en el WISC-RM, los niños entre 6 y 9 años empiezan con el reactivo 1 y los de 10-16 con el 2.

Las siguientes subpruebas quedan en el WISC-RM sin ninguna modificación

- Composición de objetos.
- Claves A
- Claves B
- Laberintos.

CUADRO DE COMPARACION DEL WISC-R Y LOS CAMBIOS EFECTUADOS EN EL WISC-RM

ESCALAS *	Número de reactivos en el WISC-RM		
	Total	Del WISC-R 1974	Substancialmente Modificados o nuevos
Escala Verbal			
Información (30) +	28	21	7
Semejanzas (17) +	17	17	0
Aritmética (18) +	18	18	0
Vocabulario (32) +	31	25	6
Comprensión (17) +	17	12	5
Retención dígitos (14)	14	14	0
Escala de Ejecución			
Figuras Incompletas (26) +	23	23	0
Ordenación de dibujos (12)	12	12	0
Diseño con cubos (11)	11	11	0
Composición de objetos (04)	4	4	0
Claves A (45)	45	45	0
Claves B (93)	93	93	0
Laberintos (09)	9	9	0

* El número de reactivos del WISC-R 1974 está entre paréntesis, después del nombre de cada subprueba.

+ Se incluyen cambios en el orden de reactivos.

III.3 ESTUDIOS REALIZADOS EN MÉXICO CON LAS ESCALAS DE WECHSLER.

Como se ha señalado con anterioridad, las escalas de Wechsler son instrumentos destinados a medir la capacidad intelectual. El WISC, Escala de Inteligencia para niños, apareció en 1949, pero debido a las observaciones de los psicólogos que la utilizaban en la práctica profesional, para hacer evaluaciones, David Wechsler realizó la revisión de la escala, con el propósito de modificar o eliminar aquellos reactivos

ambiguos, obsoletos e injustos para determinados grupos de niños. Esta revisión se realizó utilizando una muestra representativa de la población de niños de E.U.A. y apareció en el año 1974.

Ambas versiones, el WISC y el WISC-R son ampliamente utilizados en las diferentes áreas de la psicología con fines de evaluación, selección, diagnóstico o en la investigación, sin embargo, como señala García (1989), el trabajo del psicólogo mexicano ha sido objeto de severas críticas entre las cuales destacan el hecho de utilizar pruebas desarrolladas en otros países, con normas apropiadas para niños de otras culturas e inapropiadas para la evaluación de niños mexicanos, lo que resta eficacia y precisión a la evaluación.

Con el propósito de subsanar estas dificultades, se llevó a cabo la estandarización del WISC-R, utilizando una muestra de niños mexicanos; dicha estandarización se publicó en el año de 1984.

De tal manera, actualmente se cuenta con tres versiones de las escalas de Wechsler para niños (WISC, WISC-R y WISC-RM) además del WPPSI, Escala de Inteligencia Wechsler para niveles preescolar y primario, los cuales pueden ser utilizadas indistintamente por los psicólogos mexicanos.

Por lo anterior, se presenta a continuación una serie de investigaciones realizadas en México con dichas pruebas, con el propósito de contar con un panorama referente al resultado de la utilización de las mismas.

En el año de 1990 Castelar y Martínez llevaron a cabo una investigación con la Escala de Inteligencia Wechsler para niveles preescolar y primario (WPPSI), utilizando una muestra obtenida de la población en educación preescolar de escuelas particulares, escuelas oficiales y CENDIS.

El planteamiento de las hipótesis fue el siguiente: Determinar si existen o no diferencias estadísticamente significativas entre las puntuaciones obtenidas en la aplicación de la escala, dependiendo del sexo del sujeto, de la edad y del subsistema educativo en el que se encontrara cursando el nivel preescolar (escuela oficial, particular o CENDI).

Asimismo, otro objetivo era determinar si existían diferencias estadísticamente significativas entre las normas originales de la Escala y las obtenidas en la investigación por ellas realizada.

La investigación se llevó a cabo aplicándola a una muestra de 270 sujetos, los cuales se valoraron de acuerdo a los criterios establecidos por Wechsler para obtener los puntajes.

A partir de la investigación mencionada se llegó a las siguientes conclusiones:

La traducción de la escala de inteligencia Wechsler para niveles preescolar y primario (WPPSI) es confiable y válida para aplicarla en niños mexicanos, ya que las diferencias encontradas entre las tabulaciones realizadas por Wechsler y las resultantes de la investigación son similares.

A partir de la investigación se comprobó que el C.I. de desviación propuesto por Wechsler es vigente, es decir, se mantiene una desviación estándar de 3 con una media de 10 para cada subprueba y una desviación estándar de 15 con una media de 100 para puntajes totales.

Los criterios establecidos por Wechsler para la aplicación de las escalas siguen siendo confiables, ya que son perfectamente comprensibles las instrucciones para el sujeto y el orden de presentación de las escalas facilita el correcto desempeño del mismo.

La investigación indica que la población mexicana de los tres subsistemas educativos, presenta características heterogéneas, ya que se encuentran puntajes más altos en las escuelas particulares que en las escuelas oficiales o CENDIS. Uno de los criterios que se comprueba es que la habilidad intelectual aumenta conforme aumenta la edad.

En cuanto al sexo, se encontraron diferencias significativas en la mayoría de las subpruebas, siendo preponderante el sexo masculino.

Sin embargo, las autoras sugieren que es necesario que se realice la estandarización de la Escala de Inteligencia Wechsler para los niveles preescolar y primario (WPPSI), ya que es necesario ajustar algunos de los reactivos a las necesidades mexicanas (moneda actual, mexicanismos, etc.), además de que algunos de los reactivos de la escala original son anacrónicos y por tanto, incomprensibles para el sujeto actual.

Chávez y Lara (1982) señalan que en la consulta psicológica un alto porcentaje (60%) de los niños que requieren atención es debido a problemas de aprendizaje y de que una vez detectados estos problemas en los servicios asistenciales, estos niños son derivados a las escuelas especiales, en donde en varias ocasiones, por la demanda existente, el tiempo de espera puede llegar a ser de varios meses o incluso años, porque el número de escuelas es insuficiente.

Las autoras mencionan además que es importante notar que la mayoría de esos niños tiene un C.I. normal y aún superior, no obstante, por presentar algunas deficiencias su capacidad potencial de aprendizaje no se logra manifestar adecuadamente y en numerosas ocasiones los maestros suelen etiquetarlos como "retrasados mentales", "burros", "flojos", etc., obstaculizando así no sólo su rendimiento, sino creando problemas emocionales.

Tomando en cuenta lo anterior, las autoras consideran que es importante detectar futuras dificultades, a edad temprana, así como reducir el tiempo de atención paciente porque la demanda es considerable, por lo que es poco costeable aplicar una batería completa de pruebas.

De tal manera, se cuestionan si es posible utilizar la prueba de WISC para detectar además del C.I., los problemas de percepción que darían como consecuencia problemas de aprendizaje, dado que la prueba del WISC requiere para su aplicación y calificación menos tiempo que una batería completa.

La hipótesis planteada consideraba que era posible encontrar una correlación entre las subpruebas de las escalas verbal y ejecutiva de WISC con las cinco áreas de la Prueba de Percepción Visual de Frostig.

Para aceptar o desechar la hipótesis trabajaron con una muestra de 32 niños, tomada de una escuela oficial. El análisis estadístico consistió en una correlación utilizando el índice de correlación de Pearson producto-momento.

De acuerdo a los resultados, las autoras concluyeron que no es posible sustituir la prueba de Frostig por la prueba de WISC, ya que aunque existe correlación entre algunas subpruebas del WISC, como diseño con cubos, comprensión, aritmética y ordenamiento de dibujos con algunas de las áreas de Frostig, el número de correlaciones obtenidas no fue significativo para poder aceptar la hipótesis de trabajo.

Soberanis (1984) llevó a cabo un estudio comparativo entre las escalas de Wechsler (WISC y WPPSI) y el Terman-Merril. Considerando que se trata de escalas ampliamente utilizadas en el Servicio de Psicología del Instituto de Salud Mental DIF debido a que a este acuden gran cantidad de menores que requieren ser evaluados en cuanto a su capacidad intelectual y a que se utiliza la escala correspondiente "según el caso lo amerite", Soberanis se interesó por determinar si existen diferencias significativas en relación al sexo y al C.I. proporcionado por estas tres escalas, en una muestra de niños de 5 años a 6 años y 7 meses.

Sin embargo, aunque Soberanis menciona que se selecciona la escala de acuerdo al criterio del personal especializado del instituto, no especifica a que se refiere cuando señala que se manejan las escalas "según el caso lo amerite". Esta investigación se llevó a cabo utilizando un diseño factorial 2 X 2, donde las Variables Independientes fueron el sexo y la edad y la Variable Dependiente fue el Cociente Intelectual obtenido en las tres escalas.

La muestra se obtuvo de una población de 67 niños y quedó constituida por un total de 30 niños; de los cuales 15 eran del sexo masculino y 15 del sexo femenino.

Los datos arrojados por las tres escalas se procesaron a través de un análisis descriptivo, el cual reveló que el Terman fue la escala que obtuvo la mayor frecuencia de C.I.s altos y el WPPSI, el mayor número de C.I.s bajos en ambos sexos. En relación a las categorías de diagnóstico sólo los niños presentaron el nivel "Muy superior" y "Limitrofe", a diferencia de las niñas cuya tendencia fue el nivel promedio.

Para el análisis estadístico se empleó el análisis de varianza, que demostró que no

existen diferencias significativas, por lo que se concluyó que cualquiera de estas pruebas pueden incluirse en un estudio psicológico.

Sin embargo, Soberanis concluye que el Terman es una escala más práctica y funcional en niños de 2 a 5 años y niños con diferentes tipos de problemas como retardo mental, síndrome de Down y sin escolaridad y que de las escalas de Wechsler, el WISC representa una de las mayores contribuciones al campo de los exámenes de la inteligencia destinados a niños escolares.

Por otra parte, otro atributo del WISC es que permite hacer observaciones clínicas respecto a la conducta del niño.

En cuanto a las sugerencias de Soberanis, una de ellas es que se elaboren formas abreviadas de las escalas de Wechsler, y que puedan aplicarse en forma grupal.

Cuevas (1993) señala que las pruebas psicológicas son un pilar fundamental en la formación y desempeño profesional del psicólogo, dado que estas son ampliamente utilizadas en las diferentes áreas de la psicología.

Menciona, sin embargo, que uno de los principales errores en la evaluación a través de las pruebas es el de adaptación y estandarización, ya que la mayoría de las que se utilizan en México son únicamente traducidas y los resultados que estas pruebas arrojan carecen, evidentemente de confiabilidad y validez.

Por otra parte, menciona que aunque algunas pruebas han sido adaptadas y estandarizadas en México, esto ha ocurrido en muestras pequeñas y muy específicas, por lo que su uso no puede ser generalizado, este es el caso del WISC-RM, el cual surgió considerando únicamente a la población que asistía a las escuelas oficiales, por lo que la muestra quedó sesgada a un tipo de población.

Tomando en cuenta lo anterior, y considerando que la escala de inteligencia para niños de David Wechsler es una de las pruebas frecuentemente utilizadas y en forma indistinta por los psicólogos mexicanos, en sus tres versiones, realiza un estudio exploratorio con el que se pretende detectar cual de estas tres versiones (WISC, WISC-R Y WISC-RM) proporciona resultados más confiables y válidos para una muestra de población mexicana.

Para lograr ese objetivo, Cuevas realizó un análisis interno de la estructura de las tres versiones, en grupos de diferente edad y sexo, tomando dicha muestra tanto de escuelas oficiales como privadas. Con su estudio pretendió dar respuesta, entre otras, a las siguientes preguntas:

¿En cuál de las tres versiones estarán mejor graduados los reactivos en relación a su dificultad?

¿Cuál de las tres versiones poseerá una mayor consistencia interna?

¿Las puntuaciones crudas de verbal, ejecución y total de cada una de las tres versiones del WISC serán semejantes o diferirán significativamente?

De acuerdo a los resultados observados, y en cuanto a las preguntas anteriores, Cuevas reportó lo siguiente:

"En las tres versiones, WISC, WISC-R y WISC-RM se encontró en relación a los grados de dificultad, que el orden que presenta el manual es diferente al orden establecido al examinar a esta muestra, sobre todo en las subpruebas de la parte verbal, que es donde se registran la mayor cantidad de cambios en los grados de dificultad, no encontrando tanta variabilidad en el orden de presentación de las preguntas en la parte de ejecución.

Las tres versiones del WISC poseen coeficientes de consistencia interna adecuada, variando entre .71 a .98

La versión RM es la que resulta más confiable, alcanzando coeficientes de confiabilidad de .80 a .98 cuando se utiliza el método de reactivos noes y pares. Con la técnica de Cronbach los coeficientes Alpha varían de .80 a .92, exceptuando la subprueba de comprensión en donde se obtiene un coeficiente con valor de .74

Los coeficientes de consistencia interna obtenidos a través del método de división por mitades tienden a incrementarse ligeramente con respecto a los valores obtenidos con el Alpha de Cronbach.

Las subpruebas más consistentes dentro de la Escala Verbal son: Información, Vocabulario y Semejanzas; en la escala de Ejecución, el más confiable es Diseño con cubos. Los coeficientes de confiabilidad para la Escala Verbal resultan más altos que los de Ejecución y los de la total, fluctúan entre .88 a .98

Para las escalas Verbal y Total, se observaron discrepancias entre las tres versiones, en donde WISC-RM obtuvo la media aritmética más baja; estas diferencias quizá sean debidas a que en la escala verbal es donde se reportan mayor cantidad de cambios y modificaciones en la presentación de los reactivos en cuanto a su contenido, indicándonos la influencia del medio en donde se desarrolla el sujeto.

En la escala de Ejecución las discrepancias se observaron entre las versiones WISC con WISC-R y WISC con WISC-RM; no existiendo diferencias entre las versiones WISC-R y WISC-RM".

De acuerdo a los resultados anteriores, Cuevas concluye que la estructura interna de las tres versiones (WISC, WISC-R y WISC-RM) es adecuada, sin embargo, la versión que resulta más confiable, debido a que obtuvo los más altos coeficientes de consistencia interna es el WISC-RM.

Menciona, no obstante, que en la experiencia profesional se ha observado que al aplicar estas tres versiones se obtienen coeficientes intelectuales diferentes, razón por la que se cuestiona a que se debe esta discrepancia, si la consistencia en las tres versiones es satisfactoria.

Cuevas plantea la hipótesis de que las discrepancias de los coeficientes intelectuales obtenidos en las tres versiones son debidos a los puntajes escalares, por lo que resulta de imperiosa necesidad realizar una nueva estandarización de este instrumento, adaptándolo a las características socioculturales de la población mexicana, lo que permitirá que los psicólogos que utilizan dicho instrumento realicen sus evaluaciones de manera más objetiva y sus diagnósticos sean confiables y válidos.

Heredia (1993) señala que el psicólogo necesita valerse de una serie de técnicas que le permitan dar con mayor precisión y seguridad sus psicodiagnósticos, valiéndose de pruebas psicológicas para ello, con lo que, cuando éstas cumplen con los criterios de estandarización y se está familiarizado con su forma de aplicación y calificación, resultan de gran utilidad, ya que permiten objetivar a través de la medición algunas características de las personas y tomar en base a dichos resultados, una serie de decisiones.

Asimismo, señala que dada la trascendencia que tienen los resultados que se obtienen de las pruebas psicológicas, es necesario emplear aquellas que tienen una mayor confiabilidad y validez y que sus normas estén adaptadas a la población con la que se trabaja.

Dado que las escalas de Wechsler para niños es una de las pruebas más empleadas en el psicodiagnóstico porque mide la inteligencia con un alto grado de certeza y validez, además de que en México se utilizan indistintamente las diferentes versiones de la misma (WISC, WISC-R y WISC-RM) Heredia se plantea la pregunta de si estas diferentes versiones serán equivalentes.

Por lo anterior lleva a cabo una investigación con el propósito de conocer el comportamiento de los puntajes escalares de las diferentes versiones que existen en el mercado, así como determinar si los coeficientes intelectuales que arrojan éstas a través de sus diferentes procesos de estandarización son equivalentes.

En dicha investigación se contó con una muestra de 110 niños que contaban con una edad de 6.0 a 6.11 años y de 10.0 a 10.11 años.

El criterio para establecer la edad estuvo determinado por los límites de los bloques de edad que señalan los manuales para iniciar la aplicación en cada una de las subpruebas.

De acuerdo a los resultados obtenidos Heredia reporta lo siguiente:

"En la Escala Verbal, las medias de los puntajes crudos arrojadas por cada una de las

subpruebas en las diferentes versiones, varían en uno o dos puntos, excepto en la subprueba de vocabulario donde se presentan marcadas diferencias del WISC-RM con las otras dos versiones. En éste la media fue de 16.90 mientras que en el WISC y el WISC-R es de 25.62 y 23.63 respectivamente. Aquí hay que considerar que la cantidad de reactivos varía de una versión a otra. En el RM existen 31 reactivos, mientras que en el WISC son 40 y en el WISC-R 32.

Por lo que respecta a los puntajes escalares, éstos son de 10.04 para el WISC, 10.29 en el WISC-R y 11.52 para el WISC-RM.

Dentro de la Escala de Ejecución la subprueba de Figuras Incompletas del WISC tiene una media en el puntaje crudo de 5 puntos inferior a la del WISC-R que es la que alcanza mayor puntaje; esto se puede deber a que en el WISC son 20 reactivos, en tanto que en el WISC-R y el RM son 26 y 23 respectivamente. Las medias de los puntajes crudos obtenidas son de 10.14, 15.53 y 13.69 respectivamente. Por lo tanto la diferencia en las medias se podría considerar como mínima, si se toma en cuenta la proporción de los reactivos que contiene cada una de las versiones. Las medias de los puntajes escalares son de 11.70 para el WISC, en el WISC-R de 10.73 y en el WISC-RM de 12.31. Aunque el puntaje crudo del WISC-R es el mayor, en el escalar es el menor.

Otra diferencia en la media de los puntajes crudos de 3 puntos entre una versión y otra, se encuentra en la subprueba de Diseño con Cubos. El menor puntaje se obtiene en el WISC (19.80), debido quizás a que tiene únicamente 10 reactivos, en tanto que el WISC-R (23.66) y el RM (23.36) tienen 11 reactivos. Los puntajes escalares son de 12.42, 11.44 y 13.28 respectivamente. Es importante hacer notar que aunque en el WISC-R y el WISC-RM los puntajes crudos son similares, en el puntaje escalar hay dos puntos de diferencia; y que el WISC, que es el de menor puntaje crudo, obtiene un puntaje escalar intermedio entre las tres versiones.

En la subprueba de laberintos, también las discrepancias en las medias de los puntajes crudos son de 4 puntos, ya que el WISC es de 14.98, en el R de 18.29 y en el RM de 18.53. Esto puede deberse a que en el WISC se presentan 8 reactivos, mientras que en el WISC-R y en el WISC-RM son 9 y los criterios de calificación son más flexibles y los rangos de puntuación son mayores. Los puntajes escalares son de 12.15, 10.58 y 11.87, respectivamente.

Nuevamente entre el WISC-R y el WISC-RM las medias de los puntajes crudos son prácticamente iguales, pero el puntaje escalar difiere significativamente favoreciendo al WISC-RM. En el WISC el puntaje crudo es el más bajo en tanto que en el escalar es el más elevado.

Una de las preguntas planteadas al realizar la investigación, fue con respecto a si las medias de los puntajes normalizados son semejantes o discrepan entre las diferentes versiones.

Al respecto, los resultados obtenidos muestran que existen diferencias significativas entre los puntajes escalares de las diferentes versiones. La escala que más diferencias muestra con respecto a las otras dos es el WISC-RM, debido a que casi todas las subpruebas presentan discrepancias con el WISC, o con el WISC-R, dado que los puntajes escalares son significativamente más elevados que en las otras dos versiones. Únicamente en las subpruebas de Aritmética, Semejanzas y Laberintos no se encuentran los puntajes escalares más elevados, con respecto a las otras dos Escalas.

El WISC-RM, en las únicas subpruebas que no muestra diferencias significativas en sus promedios de puntajes escalares, es en Comprensión, Aritmética y Claves. En las restantes discrepan en alguna de las versiones o en las dos. En las subpruebas que los puntajes son similares con el WISC son: Comprensión, Semejanzas, Claves y Laberintos. Con el WISC-R presentan puntajes similares únicamente con la subprueba de Aritmética. Esto se debe a que en el WISC-R los puntajes escalares tienden a ser los más bajos de las tres versiones, en tanto que el WISC-RM presenta los puntajes más altos.

Las diferencias más marcadas de los puntajes escalares del WISC-RM en relación con las otras dos versiones, se encuentran en las subpruebas de Información y Retención de Dígitos.

Si consideramos las medias obtenidas en números enteros, podemos ver que en Información el promedio de los puntajes escalares es de 9 para el WISC y el WISC-R, en tanto que en el WISC-RM es de 12. Llama la atención de que existen tres puntos de diferencia entre la media del WISC-RM y los otros dos; en tanto que en los puntajes crudos sólo hay un punto de diferencia con el WISC-R y .50 con el WISC.

En Retención de Dígitos las medias de los puntajes escalares son de 10 para el WISC y el WISC-R, en cambio en el WISC-RM es de 14. En los puntajes crudos estos promedios son de 8, 9 y 10, respectivamente.

En cuanto a lo que se refiere al tipo de escuela, el cual puede estar señalando la diferencia del medio sociocultural en el que se desenvuelven los niños, Heredia reporta lo siguiente:

"Al analizar el comportamiento de las subpruebas en las diferentes versiones, se observa que existen algunas diferencias significativas entre las escuelas públicas y privadas. Las únicas subpruebas que no muestran diferencias significativas entre las escuelas públicas y privadas en las tres versiones son: Comprensión y Aritmética. Y las que sí difieren significativamente son: Semejanzas, Ordenamiento de Dibujos, Diseño de Cubos y Claves. De éstos últimos hay que considerar que las tres primeras se encuentran entre las subpruebas que miden mejor el factor "G".

Las diferencias significativas entre las escuelas públicas y privadas indican que el medio socio-cultural, influye en las funciones de análisis y síntesis, formación de

conceptos, planeación, anticipación y coordinación visomotora."

Tomando en cuenta todo lo anterior Heredia concluye que al medir la capacidad intelectual de los niños mexicanos con el WISC-RM, se sobrevalora la misma y por lo tanto las acciones que se tomen a partir de ellas pueden resultar falseadas.

Por otro lado menciona que aunque esta escala arroja los Coeficientes Intelectuales más elevados, las correlaciones entre las subpruebas muestran la consistencia más elevada, por lo que se considera que al conservar la prueba una estructura interna adecuada, únicamente es conveniente ajustar las normas de calificación a nuestra población.

III.4 PRUEBA DE MATRICES PROGRESIVAS DE RAVEN.

La Prueba de Matrices Progresivas de Raven es un instrumento destinado a medir la Capacidad Intelectual, para comparar formas y razonar por analogías, con independencia de los conocimientos adquiridos.

J.C. Raven, psicólogo inglés, Director de Investigaciones Psicológicas de "The Crixton Royal", con la colaboración inicial de Penrose, y con fines de investigación, imprimió por vez primera sus "Progressive Matrices" en el año 1936. tras algunas revisiones, dos años después su prueba había tomado forma definitiva y estaba ya tipificado. Se le identificaba como Progressive Matrices 1938, Sets A, B, C, D y E, fecha que sirve para distinguirlas de las otras formas para edades y capacidades específicas y de una revisión de la misma escala que Raven construyó posteriormente: una escala para niños de 3 a 10 años y adultos deficientes (Progressive Matrices 1947, Sets, A, AB, B), que se presenta en forma de tablero o de cuaderno.

Raven construyó su instrumento tomando como fundamento teórico la teoría de los factores de Spearman, la que, en resumen, considera lo siguiente:

Todas las habilidades del hombre tienen un factor común, un factor general a todas ellas (factor g), y un factor específico a cada una de ellas (factor e). En cada habilidad se dan los dos factores g y e, pero dichos factores no desempeñan el mismo papel en todas las habilidades: en tanto en algunas de ellas es g el factor principal, en otras lo es e. La investigación ulterior demostró la existencia de otros factores, los llamados factores de grupo, que se encuentran en gran parte en un conjunto de habilidades afines. Por consiguiente, pese a su denominación de bifactorial, la teoría de Spearman ha llegado a distinguir, en rigor, tres factores de inteligencia: factor g, factor e y factores de grupo.

Considerando que g es un factor cuantitativo de la inteligencia, sólo un factor, pero el común fundamental de todas las funciones cognoscitivas del mismo individuo, Spearman propuso construir pruebas homogéneas de alta saturación de g. Se concluyó, tras ciertas investigaciones, que las pruebas homogéneas más saturadas de g son las de razonamiento matemático y gramatical, las de percepción de relaciones

complejas con material visual y, en especial, las de material no verbal, perceptivos, a base de problemas de educación de relaciones.

Dentro de esta línea de investigación psicométrica que busca la máxima saturación posible para g, compuso Raven su prueba de Matrices Progresivas. La prueba de Raven busca, en efecto, medir el factor general. Suministra información directa de la magnitud individual de ciertas funciones cognitivas (observación y razonamiento) y una cierta evaluación de casi todas ellas: la medida de capacidad intelectual general -denominador común de la totalidad de las operaciones de la inteligencia. Ello ha dado lugar a la errónea identificación de la prueba como una prueba de inteligencia general, y a la consiguiente aclaración de Raven orientada a despejar el equívoco. Dado que g es casi todo en algunas habilidades y algo en la totalidad de ellas, establecer la magnitud de g en un individuo dado, tal como se hace en el Raven, significa establecer su habilidad general, su capacidad general: casi todo con respecto a algunas habilidades y algo de casi todas ellas.

Pero debe advertirse que el factor g "sólo mide un factor común a todas las operaciones, y no la operación entera"; o sea, que la determinación de g no implica una medición completa de ninguna operación intelectual dada, en cada caso faltará medir el valor del factor e.

En suma: para un conocimiento completo de las habilidades de un individuo debe complementarse la averiguación de g (Matrices Progresivas) con la averiguación de e mediante las pruebas factoriales correspondientes.

VALIDEZ.

Entre los análisis que se han efectuado para establecer la composición factorial de Raven, es especialmente importante el que realizó Vernon, (en Bernstein, 1988), sobre los resultados obtenidos en el ejército británico. Halló que su saturación de g es de 0.79 y su saturación en factor específico espacial es de 0.15, saturación en factor e relativamente alta, que se debe particularmente a los primeros 30 problemas. En cambio, no se ha podido determinar la naturaleza de los factores que componen el pequeño residuo de 0.6. En general, los expertos coinciden en la opinión de que las "Matrices Progresivas" constituyen en el presente un instrumento de trabajo altamente satisfactorio.

CONFIABILIDAD

En cuanto a la confiabilidad de la prueba de Matrices Progresivas de Raven, Bernstein (1988), reporta una investigación realizada en La Plata, con 187 sujetos examinados en una primera ocasión en forma colectiva. En una segunda aplicación, en forma individual, sólo seis sujetos rebajaron su percentil, pero en forma tan ostensible, que no cupo otra explicación satisfactoria, de que en la forma colectiva copiaron al compañero. El 97% de los casos mantuvieron su percentil, y sólo el 3% degradaron por copia. Los resultados obtenidos permiten confiar en que las notas percentilares de sujetos

examinados en forma colectiva, con adecuada vigilancia, son altamente confiables.

APLICACIONES

La Prueba de Raven es utilizable como instrumento de investigación básica y aplicada. En esta última, sirve como instrumento de clasificación educacional, militar e industrial y como prueba clínica. Se le emplea:

- En los establecimientos de enseñanza primaria y secundaria para ayudar a la formación de clases homogéneas mediante la agrupación de los alumnos de capacidad intelectual semejante. Para discriminar el tipo predominantemente perceptivo o lógico de los alumnos, particularmente en aquellos que se destacan por una alta o baja capacidad o rendimiento.
- En los gabinetes de orientación vocacional y selección profesional: para la evaluación de los estudiantes y la calificación de los aspirantes a ingresar en oficinas y establecimientos comerciales e industriales. Es útil para un diagnóstico rápido del nivel de la capacidad intelectual y la eventual determinación del método y fallas del razonamiento del sujeto.
- En el ejército: para una ágil y económica medición y selección de cuadros. Con este objetivo fue empleado en gran escala, en la última contienda, por la armada británica.
- En los centros de investigación psicológica y sociológica: para estudios diferenciales y sociales de capacidad intelectual, según edad, sexo, medio, status económico-social y profesional.

Descripción de la Prueba de Matrices Progresivas de Raven.

La Prueba de Matrices Progresivas de Raven consta de Manual de Aplicación, Cuaderno de Matrices, Protocolo de Prueba, Rejilla de Calificación y Tabla de Baremos.

Se presenta al examinado el Cuaderno de Matrices y el Protocolo de Prueba en donde el sujeto deberá anotar sus respuestas.

El Cuaderno de Matrices consta de 60 láminas de figuras geométricas abstractas lacunarias (ordenadas en 5 series denominadas A, B, C, D y E, de 12 reactivos cada una) que plantean problemas de completamiento de sistemas de relaciones (matrices), en complejidad creciente, y para cuya resolución el examinado debe educir relaciones en las primeras 24 y correlaciones en las últimas 36, es decir, el examinado ha de descubrir la matriz de pensamiento implicado en cada figura y para ello debe realizar una tarea de observación, comparación y razonamiento analógico.

Raven, (en Bernstein, 1988), considera que la totalidad de sus reactivos entrañan

problemas analíticos de educación de relaciones y correlaciones. Sin embargo, debe distinguir, a grandes trazos, dos grandes grupos de problemas. Las primeras 24 matrices (series A y B) son de tipo fuertemente gestáltico: plantean problemas de percepción de totalidades y el sujeto debe integrar o "cerrar" una figura inconclusa y por tanto, ha de ser capaz de percibir las semejanzas, diferencias, simetría y continuidad de las partes en relación con la estructura o forma del todo. Las 36 matrices últimas (series C, D y E) que son definitivamente sistemas de relaciones, plantean problemas de razonamiento y exigen operaciones analíticas de pensar y discernimiento. Algunos ejemplos son: problemas de serie, del tipo progresiones de adición cuantitativa o espacial (C1), o de movimiento (C7), o de adición y movimiento (C5), de progresión numérica (adición y sustracción) C11, de alternancia y simetría (D2); problemas de analogías, simples (D4) o complejas (D11), y en fin, combinaciones de principios varios (E10).

En cada caso, el sujeto demostrará haber descubierto la solución, si logra reconocerla entre varias inadecuadas que se dan al pie de la lámina respectiva.

En general, se trata de una prueba:

- Sencilla: El aprendizaje de su administración y evaluación es singularmente breve.
- De amplio margen: se aplica desde los 12 a los 65 años de edad.
- Económica: En personal, tiempo y material. Se administra o autoadministra, en forma individual o colectiva, a grupos hasta de cien sujetos, en tiempos que oscilan entre 30 y 60 minutos y el cómputo de los resultados, fácil y ágil, se efectúa en breves minutos. El material de prueba sirve para numerosas aplicaciones, pues sólo se inutilizan los protocolos de registro.
- No verbal: Se aplica a todo sujeto, cualquiera que sea su idioma, educación y capacidad verbal.
- No manual: se aplica a todo sujeto, cualquiera que sea su estado o capacidad motora. Interesante para el examinado: mantiene vivo interés del sujeto en todo el transcurso de la prueba.
- Agradable para el examinador cuya tarea se desenvuelve en su totalidad de un modo simple y grato.

ADMINISTRACION

La prueba consiste en presentar a la inspección del sujeto (o sujetos) uno a uno 60 problemas de completamiento ordenados por dificultad en aumento, cada uno con la respuesta correcta mezclada entre varias otras erróneas (5, en las series A y B y 7 en las C, D y E), y en pedirle que ante cada problema señale cuál es, en su opinión la

acertada entre las 6 u 8 soluciones brindadas a su elección. A medida que el examinado va produciendo sus respuestas, él mismo, o el examinador, las anota en el protocolo especial para el registro de las soluciones propuestas a cada problema. Cuando la prueba se aplica en forma colectiva debe ejercerse cierta vigilancia para asegurarse que los examinados no salten ningún problema, anoten sus soluciones en las casillas pertinentes y muy especialmente que no se copien entre sí.

EVALUACION.

1. En primer término debe calificarse el acierto o el error de cada una de las soluciones propuestas por el examinado, superponiendo la clave matriz de corrección.
2. Cada respuesta positiva se computa como un punto a favor: el total de puntos acreditados nos da el puntaje obtenido por el sujeto (puntaje natural o bruto). Por consiguiente, la escala de puntajes brutos posibles oscila entre 0 y 60.
3. Raven brinda un procedimiento para establecer en que grado un puntaje bruto cualquiera puede estimarse expresión fiel de la capacidad que se procura medir. Ha calculado los puntajes medios que en cada una de las cinco series deben componer cada puntaje total. Si en la prueba colectiva el sujeto obtuvo v. gr., 36 puntos, debe esperarse que este puntaje resulte de la obtención de 11 puntos en la serie A, de 8 en la B y de 8 en la C., 7 en la D y 2 en la E. En base a estos puntajes parciales esperados por serie para cada puntaje total, ha compuesto dos tablas de "Composición de puntaje normal": una para medir la consistencia de pruebas individuales, y la otra para los exámenes colectivos. La consistencia del puntaje se verifica obteniendo, por sustracción, la discrepancia entre los puntajes obtenidos por el sujeto en cada una de las cinco series (puntajes parciales reales) y los puntajes que de él se esperaban para cada serie en razón de su puntaje total (puntajes parciales esperados). El puntaje total será consistente, cuando los puntajes parciales reales y esperados sean iguales (discrepancia: 0), o cuando cada diferencia sea igual o menor que +2 ó -2. Por tanto, en cada serie sólo debe acusarse una de estas discrepancias 0, ó +1, ó -1, ó +2, ó -2: cuando en una serie se diese una discrepancia mayor que 2 (+3, -3), el puntaje total obtenido por el sujeto no podrá considerarse como consistente.
4. El puntaje obtenido en la prueba de Raven se estima según una escala en percentiles (P), procedimiento que tiene la ventaja de brindar resultados de clara significación: el rango que por su capacidad intelectual un sujeto ocupa dentro de un grupo de cien sujetos de su edad. Para transformar el puntaje del sujeto en el percentil que corresponda se consultan las Tablas de Normas Percentilares. La tarea consiste en convertir el puntaje bruto del sujeto, en la norma equivalente para los sujetos de su edad consultando el Baremo correspondiente.
5. Estableciendo el percentil que corresponde a un puntaje, se tiene ya la estimación del nivel de capacidad intelectual del examinado en uno de estos cinco rangos

principales:

TABLA DE DIAGNOSTICO DE CAPACIDAD INTELECTUAL

Puntaje	Norma	Perc.	Rango	Diagnóstico de capacidad
Igual o superior a	P95	95	I	Superior Superior al término medio
	P90	90	II +	
	P75	75	II	
Superior a	P50	50	III +	Término medio
Igual a	P50	50	III	
Inferior a	P50	50	III -	
Igual o menor a	P25	25	IV +	Inferior al término medio
	P10	10	IV	
	P5	5	V	Deficiente

III.5 ESTUDIOS REALIZADOS EN MÉXICO CON LA PRUEBA DE MATRICES PROGRESIVAS DE RAVEN.

Con la presente investigación se inició una revisión bibliográfica que se pretendía fuera de los estudios realizados de cinco años a la fecha. Suponíamos que dado que se trata de una prueba que, de acuerdo a la experiencia personal y a referencias de colegas, es ampliamente utilizada en diferentes áreas, con el propósito de conocer la capacidad intelectual de los sujetos evaluados, se encontrarían investigaciones recientes con la misma. Sin embargo, no ocurrió así, las investigaciones encontradas son de años atrás; y en éstas se encontró, en general dos observaciones:

1. Que, la Prueba de Matrices Progresivas de Raven es ampliamente utilizada en diferentes áreas de la psicología, no sólo por su validez científica, sino también porque, dadas sus características que permiten su aplicación en forma grupal, sencillez para la calificación, etc., resulta sumamente atractiva para los psicólogos responsables de las evaluaciones.
2. De acuerdo a los resultados de dichas investigaciones, se señala que es necesario estandarizarla a nuestra población debido a que se trata de una prueba elaborada en una cultura muy diferente a la nuestra y a que los diferentes Baremos con que cuenta, también fueron elaborados en base a muestras de otros países.

A continuación se presenta un reporte de las investigaciones realizadas en México con la Prueba de Matrices Progresivas de Raven, que están relacionadas con el objetivo de esta investigación.

Zavala (1958) destaca que la Prueba de Matrices Progresivas de Raven es una de las más utilizadas en nuestra cultura occidental para la captación del rendimiento intelectual, no sólo por su validez científica, sino por su gran sencillez y facilidad de aplicación en poblaciones de diferentes culturas, razas, idiomas, edades, sexos, etc.

Por tal razón, convencido de su gran eficacia, incluye dicha prueba en la batería que se utiliza en el estudio de la personalidad total del estudiante del Instituto Politécnico Nacional.

A principios del año 1957 se inició en dicha institución la aplicación de las Matrices Progresivas de Raven en las diferentes escuelas tecnológicas, con el fin de averiguar el grado de capacidad intelectual de esa población que aún no llegaba a determinar su futura profesión. Zavala refiere que fue muy satisfactorio encontrar índices muy altos de inteligencia, con respecto a Baremos realizados en otros países; menciona, sin embargo, que estos puntajes no eran reales, hecho que los obligó a pensar en sus propios Baremos.

En el año 1958 se prolongó dicha investigación, que se había iniciado en una población heterogénea, a sujetos que ya se hallaban ubicados a seguir una determinada profesión. Fue así como se realizó la aplicación del Raven a 1480 alumnos de las diferentes escuelas vocacionales del I.P.N.

Esto último, considerando que los estudiantes politécnicos se "realizan en un medio cultural diferente, educacional y con un concepto de vida distinto", por lo que los resultados variarían con relación a los obtenidos en poblaciones muy disímiles.

Por tal motivo se consideró obligatorio que el Departamento de Orientación penetrara en el problema y encontrara los baremos adecuados para cada escuela vocacional, encontrando los valores que revelaran el verdadero rendimiento del estudiante politécnico. De acuerdo al análisis estadístico realizado, se observó heterogeneidad en la población de una vocacional a otra, por lo que hubo necesidad de elaborar baremos independientes entre éstas.

En las escuelas tecnológicas únicamente se hizo un baremo general, ya que el análisis de varianza no dio muestras de heterogeneidad.

Zavala resume que los resultados de la investigación dieron nuevas luces acerca de la potencialidad intelectual de los estudiantes politécnicos y se pudo apreciar que se contaba con sujetos capaces de observar en forma productiva los programas de la institución técnica de México. Sin embargo, refiere que se pudo apreciar, con otras pruebas, cierta problemática social y emocional que perturba el buen rendimiento intelectual de los alumnos.

Asimismo, recalca Zavala que los altos índices de puntaje en las Matrices Progresivas de Raven, no quieren decir que se trate de una población superdotada intelectualmente, sino que esto se debe a que es ya de hecho, un grupo

preseleccionado, pues para llegar a esos cursos hubo de pasar cierta selección natural.

Jiménez (1965) menciona que, dado que es indiscutible el valor que las pruebas presentan a la clínica, seleccionó la Prueba de J.C. Raven, la que por sus características ofrece una forma objetiva y rápida de seleccionar e investigar las funciones intelectuales del individuo en comparación a su "población".

Jiménez considera de suma importancia contar con baremos adecuados para la aplicación de la prueba a las poblaciones de los Hospitales Psiquiátricos de México, debido a que las condiciones imperantes, determinan que frecuentemente los diferentes "tipos de población" sean medidos con baremos estandarizados en poblaciones extranjeras, violando de esta manera una de las características más importantes de las pruebas mentales, lo que da como resultado que se cuestione la validez de los resultados.

Por lo anterior Jiménez consideró de suma importancia teórico-práctico el contar con escalas adaptadas a las condiciones particulares del "universo" que se ha propuesto para estudio.

Su investigación tuvo como objetivo además de elaborar los baremos para la población estudiada, el dar respuesta, entre otras, a las siguientes hipótesis.

1. Determinar si en los sujetos que se han implantado una sintomatología psicótica, los puntajes totales de la prueba de Raven son inferiores a los de la población normal.
2. Existe diferencia significativa entre los puntajes totales de los sujetos hospitalizados con diagnóstico clínico de psicosis y los de organicidad.

La investigación se llevó a cabo con una muestra de 100 sujetos seleccionados al azar y de donde 50 eran del sexo masculino y 50 del sexo femenino. Esta muestra se obtuvo del Pabellón Piloto del Manicomio General y del Hospital Granja "José Sáyago". Dichos sujetos fueron seleccionados al azar para representar a la población de esos departamentos e investigar sobre su rendimiento.

Se realizó la tabla percentilar, la que indica el grado de afectación en el factor general de inteligencia de los cuadros tipificados para fines del estudio, como orgánicos y psicóticos. También se obtuvieron los percentilares para la población masculina y femenina internada en esos nosocomios.

Por otro lado, el manejo estadístico arrojó las siguientes conclusiones: En los sujetos donde se ha implantado una sintomatología psicótica u orgánica, sus puntajes son inferiores a los esperados para su edad cronológica.

Los rendimientos, tanto en los puntajes parciales, como en los puntajes totales,

presentan un grado significativo de homogeneidad, lo que impide intentar un diagnóstico diferencial.

Los tiempos empleados en los sujetos donde existe un proceso de deterioro orgánico son muy pequeños, en contraposición con aquellos cuadros donde el deterioro parece ser menos severo.

Jiménez señala que se juzga necesario la ampliación de los baremos, mediante una población mayor que sea representativa de la población de los hospitales psiquiátricos de México.

Franco (1969) menciona que es sabido que toda prueba mental debe cubrir los requisitos de validez, confiabilidad y sensibilidad, sin embargo, es posible que aunque una prueba cubra estos requisitos y funcione adecuadamente en la sociocultura para la cual fue creada, ésto no será así para otras socioculturas, por tal motivo, ella se planteo la interrogante de que pasaría al aplicar dos pruebas de inteligencia (Raven y Barsit) realizadas cada cual en socioculturas distintas (Inglaterra y Venezuela) al aplicarlas en una tercera (México). Franco deseaba saber que correlación alcanzarían al aplicarlas en un grupo reducido de esta tercera sociocultura.

Para dar respuesta a esta incógnita, Franco llevó a cabo una investigación en un grupo de 413 personas que aspiraban a ingresar a un determinado puesto en los Ferrocarriles Nacionales de México.

La muestra quedó dividida en cinco grupos de edades, todos ellos hombres y mujeres que poseían un nivel de escolaridad de sexto grado.

En sus conclusiones Franco reporta los siguiente:

En cuanto a las medidas obtenidas en los cinco grupos, no hay diferencias significativas ni en el Raven, ni en el Barsit. Por otro lado tanto las medidas aritméticas totales del Raven (33.89) como las del Barsit (33.44) son bajas comparadas con otros grupos.

Franco supone que ésto se debió a que durante el tiempo en que se efectuó la investigación, predominaron los aspirantes al puesto "ayudantes auxiliares", es decir, un puesto catalogado como para obreros no calificados y para el cual no se requiere un elevado nivel intelectual.

La desviación estandard tanto para el Raven como para el Barsit va de 9 a 12 en los 5 grupos.

Franco supone que esta desviación estandard se debió a la heterogeneidad de la muestra, ya que en esa se encontraban aspirantes a 43 grupos diferentes de trabajo. En cuanto a las correlaciones obtenidas entre ambas pruebas reporta lo siguiente:

La correlación para los grupos 1, 2, 3 y 4 fue significativa pero moderada (relación substancial). Pero el grupo 5 obtuvo una correlación baja.

Franco señala que hipotéticamente, puede decir que los sujetos del grupo 5 son lo que habían conocido con anterioridad "las respuestas correctas del Raven". Menciona asimismo, que esto es común en los primeros ingresos debido a que el Raven es una prueba que está muy difundida en el medio laboral, en cambio, el Barsit es una prueba que no está muy difundida, ya que en el Departamento de Selección, de Ferrocarriles Nacionales, sólo se había utilizado por espacio de dos años.

Finalmente concluye que la correlación total entre el Raven y el Barsit no fue muy alta (0.54) "lo cual era de esperarse" ya que se trata de dos pruebas elaboradas para medir distintos aspectos de la inteligencia: El Raven mide el factor general y el Barsit mide factor específico (aspectos de la inteligencia verbal, razonamiento numérico, elementos cognoscitivos, lógico-verbal y de información general)

Fuentes y López (1988) plantean que en México existe frecuentemente el problema de no contar con pruebas adaptadas a poblaciones específicas para realizar con mayor exactitud los estudios. En el Consejo Tutelar, es necesario contar con un diagnóstico de inteligencia con el propósito de canalizar a los menores a un lugar adecuado para el desarrollo de sus potencialidades en el proceso de readaptación, sin embargo, señala que se ha observado que el diagnóstico de inteligencia reportado en base a las pruebas es inferior a la impresión diagnóstica del psicólogo, lo que representa problemas al decidir la canalización.

Por tal motivo ella se plantea ¿cuál será el baremo más adecuado para un diagnóstico de inteligencia en la población del Consejo Tutelar?

La pregunta o problema de investigación queda así planteada ¿cuál baremo de la prueba de Raven sería el adecuado para un diagnóstico más preciso de la inteligencia en la población del Consejo Tutelar?

Con dicha pregunta se entiende que se refiere a los diferentes baremos que se incluyen en la prueba, sin embargo, la hipótesis conceptual plantea que "dada la homogeneidad de la población del Consejo Tutelar, se puede establecer el baremo más adecuado para la calificación e interpretación de la prueba de Raven" y la hipótesis nula establece que "no hay diferencia significativa entre el baremo adaptado a la población de Córdoba y el baremo adaptado a la población del Consejo Tutelar".

Dada esta descripción de las hipótesis y de la pregunta de investigación, no queda claro si sólo se elegirá entre los baremos que ya tiene la prueba o se adaptará otro baremo.

Sin embargo, lo que primero se adapta es la prueba, al ordenar los reactivos en orden creciente de dificultad en cada serie, de acuerdo a los resultados obtenidos por la muestra estudiada. De esta forma conforman la prueba adaptada a la población del

Consejo Tutelar. Se obtuvieron las normas de calificación a través del rango percentilar.

La confiabilidad de la prueba se obtuvo a través de la Prueba de Equivalencia Racional de Kuder Richardson y el método de Pruebas divididas. Finalmente se comprobó la significancia entre los baremos.

Como conclusiones, las autoras sugieren realizar una estandarización de la prueba para la población de todos los Consejos Tutelares, ya que las características de la población la hacen única y por lo tanto diferente a la de los países que elaboran las pruebas que aquí se utilizan.

CAPITULO IV

METODOLOGIA

En la actualidad las pruebas de inteligencia tienen una aplicación muy importante en el área laboral, educativa y en la clínica.

En el área educativa son utilizados con diferentes propósitos:

- a) Para formar grupos homogéneos que permitan un avance académico paralelo de los estudiantes.
- b) Como un instrumento de diagnóstico útil para determinar las causas del fracaso escolar y así poder tomar las medidas necesarias.

Sin embargo, a nivel nacional, en México nos encontramos con varias dificultades al tratar de evaluar la capacidad intelectual de los estudiantes, debido a que la mayoría de las pruebas de que disponemos han sido elaboradas y estandarizadas en otros países. De acuerdo con Morales (1991) "la definición estadística de inteligencia obtenida sobre cualquier muestra particular o específica, es válida sólo para los grupos que representan la población muestreada" (pág. 105)

García (1987) señala que existen evidencias empíricas de que resulta injusto para los niños mexicanos ser evaluados con instrumentos no estandarizados, ya que pueden ubicarse en una categoría descriptiva inferior, por lo que con el propósito de contar con un instrumento que nos proporcionara normas de evaluación más acordes o apropiadas a la población mexicana, el Departamento de Investigación, dependiente de la Dirección Técnica, de la Dirección General de Educación Especial, llevó a cabo la estandarización del WISC-RM.

El WISC-RM, por estar estandarizado para una población mexicana, ha tenido gran difusión en los diferentes ámbitos en los que se requiere evaluar la capacidad intelectual de los infantes, pero a pesar de ello, su precedente el WISC-R también sigue siendo ampliamente utilizado por los psicólogos mexicanos.

Sin embargo, tanto el WISC-R como el WISC-RM son pruebas de tipo individual, por lo que requieren un tiempo importante para su aplicación, como para su calificación, lo que las hace poco prácticas a los psicólogos que se desenvuelven en ámbitos escolares con gran cantidad de alumnos a los que hay que evaluar.

De acuerdo a los requerimientos del área educativa, se necesita de una prueba que nos proporcione las ventajas de la Prueba de Matrices Progresivas de J.C. Raven:

- Que sea de aplicación grupal con el propósito de evaluar en el menor tiempo posible a un grupo mayor de alumnos.

- Que no distraiga mucho tiempo al estudiante debido a que en la mayoría de los casos esa evaluación debe realizarse dentro del horario de clases.
- Que sea de rápida calificación.

Desafortunadamente al tratar de utilizar la prueba de Raven en ámbitos escolares, no tenemos la certeza de que sea válida para nuestra población, a pesar de estar estandarizada en varios países sudamericanos.

De tal manera, cuando intentamos evaluar alumnos de escuelas secundarias particulares nos enfrentamos ante la problemática de no disponer de un instrumento que sea válido y que nos proporcione las características que se requieren de acuerdo al tiempo que se dispone para dicha evaluación.

IV.1 PLANTEAMIENTO DEL PROBLEMA.

En virtud de lo anterior y tomando los fundamentos teóricos de Spearman quien señala que el factor G es un factor cuantitativo de la inteligencia, el común y fundamental a todas las funciones cognitivas del mismo individuo, se plantea la siguiente pregunta de investigación:

¿Cuál es el nivel de validez concurrente entre la escala de Matrices Progresivas de Raven y las escalas de Wechsler, (WISC-R y WISC-RM).

La inquietud de responder a esta pregunta nace de la necesidad de contar con una prueba de inteligencia, válida y confiable que pueda ser aplicada en forma colectiva en el ámbito educativo privado. Anastasi (1978) enumera las ventajas de utilizar este tipo de pruebas, entre las que destacan las siguientes:

- Se puede aplicar simultáneamente a tantas personas como se puedan acomodar confortablemente en el espacio disponible.
- Debido a que la misión del examinador se reduce al mínimo, las pruebas colectivas pueden proporcionar condiciones más uniformes que las individuales.
- La puntuación es característicamente más objetiva ya que proporciona normas mejor establecidas al emplear elementos de elección múltiple.
- Debido a que en estas pruebas las contestaciones se escriben en las hojas de respuestas, los folletos se pueden utilizar indefinidamente, lo que supone una economía considerable.

Rimoldi (En Bernstein, 1988) al hablar de la estandarización del Raven, en Mendoza, Argentina, también reporta lo siguiente:

"Creemos que estas pruebas (matrices) deben ser utilizadas extensivamente por su

fácil manejo, por su valor discriminativo... creemos también que la influencia de los factores ambientales en esta prueba es relativamente escasa".

Por otra parte, el que la prueba de Raven cuente con un procedimiento para establecer la consistencia del puntaje bruto obtenido por los sujetos da la posibilidad de saber que los resultados son expresión fiel de la capacidad del examinado y de no ser así, descartarlos y no considerar una puntuación errónea.

IV.2 HIPÓTESIS.

Considerando que el objetivo de la investigación es determinar el grado de validez concurrente de la prueba de J.C. Raven utilizando como criterios de comparación las pruebas WISC-R y WISC-RM dado que se plantean las siguientes hipótesis:

- Ho** No existe correlación estadísticamente significativa entre las puntuaciones crudas obtenidas con la prueba de Raven y las puntuaciones crudas obtenidas con la escala verbal de WISC (R / RM).
- H1** Existe correlación estadísticamente significativa entre las puntuaciones crudas obtenidas con la prueba de Raven y las puntuaciones crudas obtenidas con la escala verbal de WISC (R / RM).
- Ho** No existe correlación estadísticamente significativa entre las puntuaciones crudas obtenidas con la prueba de Raven y las puntuaciones crudas obtenidas con la escala de ejecución de WISC (R / RM).
- H1** Existe correlación estadísticamente significativa entre las puntuaciones crudas obtenidas con la prueba de Raven y las puntuaciones crudas obtenidas con la escala de ejecución de WISC (R / RM).
- Ho** No existe correlación estadísticamente significativa entre las puntuaciones crudas obtenidas con la prueba de Raven y las puntuaciones crudas obtenidas con la escala total de WISC (R / RM).
- H1** Existe correlación estadísticamente significativa entre las puntuaciones crudas obtenidas con la prueba de Raven y las puntuaciones crudas obtenidas con la escala total de WISC (R / RM).
- Ho** No existe correlación estadísticamente significativa entre las puntuaciones crudas obtenidas con la prueba de Raven y las puntuaciones crudas obtenidas con la escala verbal de WISC - R.
- H1** Existe correlación estadísticamente significativa entre las puntuaciones crudas obtenidas con la prueba de Raven y las puntuaciones crudas obtenidas con la escala verbal de WISC - R.

- Ho No existe correlación estadísticamente significativa entre las puntuaciones crudas obtenidas con la prueba de Raven y las puntuaciones crudas obtenidas con la escala de ejecución de WISC - R.
- H1 Existe correlación estadísticamente significativa entre las puntuaciones crudas obtenidas con la prueba de Raven y las puntuaciones crudas obtenidas con la escala de ejecución de WISC - R.
- Ho No existe correlación estadísticamente significativa entre las puntuaciones crudas obtenidas con la prueba de Raven y las puntuaciones crudas obtenidas con la escala total de WISC - R.
- H1 Existe correlación estadísticamente significativa entre las puntuaciones crudas obtenidas con la prueba de Raven y las puntuaciones crudas obtenidas con la escala total de WISC - R.
- Ho No existe correlación estadísticamente significativa entre las puntuaciones crudas obtenidas con la prueba de Raven y las puntuaciones crudas obtenidas con la escala verbal de WISC - RM.
- H1 Existe correlación estadísticamente significativa entre las puntuaciones crudas obtenidas con la prueba de Raven y las puntuaciones crudas obtenidas con la escala verbal de WISC - RM.
- Ho No existe correlación estadísticamente significativa entre las puntuaciones crudas obtenidas con la prueba de Raven y las puntuaciones crudas obtenidas con la escala de ejecución de WISC - RM.
- H1 Existe correlación estadísticamente significativa entre las puntuaciones crudas obtenidas con la prueba de Raven y las puntuaciones crudas obtenidas con la escala de ejecución de WISC - RM.
- Ho No existe correlación estadísticamente significativa entre las puntuaciones crudas obtenidas con la prueba de Raven y las puntuaciones crudas obtenidas con la escala total de WISC - RM.
- H1 Existe correlación estadísticamente significativa entre las puntuaciones crudas obtenidas con la prueba de Raven y las puntuaciones crudas obtenidas con la escala total de WISC - RM.
- Ho No existe correlación estadísticamente significativa entre las puntuaciones normalizadas obtenidas con la prueba de Raven y las puntuaciones normalizadas obtenidas con la escala verbal de WISC (R / RM).
- H1 Existe correlación estadísticamente significativa entre las puntuaciones normalizadas obtenidas con la prueba de Raven y las puntuaciones normalizadas

obtenidas con la escala verbal de WISC (R / RM).

- Ho No existe correlación estadísticamente significativa entre las puntuaciones normalizadas obtenidas con la prueba de Raven y las puntuaciones normalizadas obtenidas con la escala de ejecución de WISC (R / RM).
- H1 Existe correlación estadísticamente significativa entre las puntuaciones normalizadas obtenidas con la prueba de Raven y las puntuaciones normalizadas obtenidas con la escala de ejecución de WISC (R / RM).
- Ho No existe correlación estadísticamente significativa entre las puntuaciones normalizadas obtenidas con la prueba de Raven y las puntuaciones normalizadas obtenidas con la escala total de WISC (R / RM).
- H1 Existe correlación estadísticamente significativa entre las puntuaciones normalizadas obtenidas con la prueba de Raven y las puntuaciones normalizadas obtenidas con la escala total de WISC (R / RM).
- Ho No existe correlación estadísticamente significativa entre las puntuaciones normalizadas obtenidas con la prueba de Raven y las puntuaciones normalizadas obtenidas con la escala verbal de WISC - R.
- H1 Existe correlación estadísticamente significativa entre las puntuaciones normalizadas obtenidas con la prueba de Raven y las puntuaciones normalizadas obtenidas con la escala verbal de WISC - R.
- Ho No existe correlación estadísticamente significativa entre las puntuaciones normalizadas obtenidas con la prueba de Raven y las puntuaciones normalizadas obtenidas con la escala de ejecución de WISC - R.
- H1 Existe correlación estadísticamente significativa entre las puntuaciones normalizadas obtenidas con la prueba de Raven y las puntuaciones normalizadas obtenidas con la escala de ejecución de WISC - R.
- Ho No existe correlación estadísticamente significativa entre las puntuaciones normalizadas obtenidas con la prueba de Raven y las puntuaciones normalizadas obtenidas con la escala total de WISC - R.
- H1 Existe correlación estadísticamente significativa entre las puntuaciones normalizadas obtenidas con la prueba de Raven y las puntuaciones normalizadas obtenidas con la escala total de WISC - R.
- Ho No existe correlación estadísticamente significativa entre las puntuaciones normalizadas obtenidas con la prueba de Raven y las puntuaciones normalizadas obtenidas con la escala verbal de WISC - RM.

- H1 Existe correlación estadísticamente significativa entre las puntuaciones normalizadas obtenidas con la prueba de Raven y las puntuaciones normalizadas obtenidas con la escala verbal de WISC - RM.
- Ho No existe correlación estadísticamente significativa entre las puntuaciones normalizadas obtenidas con la prueba de Raven y las puntuaciones normalizadas obtenidas con la escala de ejecución de WISC - RM.
- H1 Existe correlación estadísticamente significativa entre las puntuaciones normalizadas obtenidas con la prueba de Raven y las puntuaciones normalizadas obtenidas con la escala de ejecución de WISC - RM.
- Ho No existe correlación estadísticamente significativa entre las puntuaciones normalizadas obtenidas con la prueba de Raven y las puntuaciones normalizadas obtenidas con la escala total de WISC - RM.
- H1 Existe correlación estadísticamente significativa entre las puntuaciones normalizadas obtenidas con la prueba de Raven y las puntuaciones normalizadas obtenidas con la escala total de WISC - RM.

IV.3 DISEÑO.

El diseño de la investigación, según la clasificación de Kerlinger (1985) es *expost-facto*. En este diseño no se tiene control directo sobre las variables independientes porque sus manifestaciones ya han ocurrido; únicamente se hizo una medición de las características de la muestra, tratando de controlar variables extrañas, que pudieran interferir en los resultados con el propósito de inferir las características del instrumento.

Según la clasificación de McGuigan (1973), en la investigación se utilizó el método no experimental, debido a que las variables evaluadas no podían ser manejadas deliberadamente, por lo tanto puede decirse que se utilizó el diseño cuasi experimental.

IV.4 VARIABLES.

Variables dependientes:

- a) Puntaje crudo de cada uno de los instrumentos.
- b) Capacidad Intelectual.

Variables independientes

- a) Prueba de Matrices Progresivas de Raven.
- b) WISC-R
- c) WISC-RM

Variables controladas:

- a) Sistema educativo: tradicional
- b) Sexo: misma proporción entre hombres y mujeres
- c) Edad: 13 a 15 años
- d) Grado escolar: nivel secundaria
- e) Nivel socio económico: Medio
- f) No repetición de grado escolar
- g) Aplicación de pruebas
 - * Prueba de Matrices Progresivas de Raven: grupal.
 - * WISC-R y WISC-RM: individual

Variables extrañas:

- a) Experiencia previa con pruebas utilizadas
- b) Constelación familiar.
- c) Rendimiento escolar.

IV.5 DESCRIPCIÓN DE LA MUESTRA.

Población

La población de donde se obtuvo la muestra estuvo constituida por alumnos de dos escuelas secundarias particulares de la Cd. de México, mixtas y de turno matutino. La población de ambas escuelas hacía un total de 270 alumnos, los cuales fluctuaban entre los 12 y los 15 años de edad.

El nivel escolar, medio básico, era homogéneo en ambas escuelas tanto en el método de enseñanza, como en las materias impartidas.

El nivel socioeconómico es medio.

Muestra

Para la selección de la muestra se utilizó el método no probabilístico intencional que según Kerlinger (1985) se caracteriza por el empleo de criterios y de un esfuerzo deliberado para obtener muestras representativas mediante la inclusión de parejas típicas o grupos supuestamente típicos en la muestra.

El procedimiento para la selección de la muestra fue el siguiente:

1. Se solicitó a los directores de las escuelas las listas de los alumnos inscritos; en dichas listas se proporcionó la edad de cada alumno, excluyendo a aquellos que se encontraban recursando el grado escolar.
2. Se descartaron aquellos alumnos que tenían 12 años de edad, por tratarse de una minoría.
3. Se seleccionaron al azar los sujetos, por edad y sexo.

Los sujetos quedaron distribuidos de acuerdo a la siguiente tabla:

Edad	Hombres	Mujeres
13	8	8
14	8	8
15	8	8
TOTAL	24	24

IV.6 ESCENARIO.

La investigación se realizó en las instalaciones de las dos escuelas particulares.

Para la aplicación del Raven se utilizó el salón de Música, el cual contaba con espacio suficiente entre un alumno y otro para evitar la contaminación. La aplicación del WISC-R y WISC-RM se realizó en el salón de Orientación Vocacional.

IV.7 INSTRUMENTOS.

Para aceptar o rechazar las hipótesis planteadas se utilizaron los siguientes instrumentos de medición:

- Prueba de Matrices Progresivas de Raven, de la cual se determinaría la validez.

Como criterio para determinar la validez concurrente se utilizaron las escalas de Wechsler:

- Escala de Inteligencia revisada para el nivel escolar, WISC-R

- Escala de Inteligencia revisada para el nivel escolar, WISC-RM.

Prueba de Matrices Progresivas de Raven.

La prueba de Raven es no verbal tanto por la índole del material como por la respuesta que se pide al examinado. Es un instrumento destinado a "medir la capacidad intelectual para comparar formas y razonar por analogía independientemente de los conocimientos adquiridos".

Como material de prueba utiliza una serie de figuras abstractas (geométricas) e incompletas. Cada figura geométrica implica un patrón de pensamiento (incompleto), una matriz. Cada figura representa la madre o fuente de un sistema de pensamiento. La escala incompleta integra una escala de matrices en orden de complejidad creciente, construida de tal manera que permite revisar en la forma más completa posible los sistemas de pensamiento del desarrollo intelectual.

El examinado ha de descubrir la matriz de cada pensamiento implicada en cada figura y para ello debe realizar una tarea de observación, comprensión y razonamiento analógico.

Es una prueba de selección múltiple; para que el sujeto brinde su solución se le ofrece para cada una de las figuras varios trozos (6 u 8) entre los cuales se encuentra el único adecuado, a fin de que indique cual de ellos es el que considera apropiado.

Es una prueba autoadministrable y por lo tanto puede aplicarse indistintamente en forma individual o colectiva.

Escalas Wechsler (WISC-R Y WISC-RM).

La escala WISC-R se diseñó y organizó como un instrumento para evaluar la inteligencia general.

Este instrumento surgió de la revisión del WISC, la cual se llevó a cabo utilizando una muestra representativa de los niños de E.U.

El propósito de este instrumento es evaluar la ejecución del niño bajo un conjunto de condiciones establecidas; se desarrolló para usarse principalmente con niños y adolescentes de 6 a 16 años de edad.

El WISC-RM surgió de la necesidad de estandarizar los instrumentos de medición diseñados en otros países, con el propósito de contar con una prueba adecuada a las condiciones de la población mexicana.

La escala se basó en un muestreo de estudiantes de escuelas primarias y secundarias oficiales incorporadas a la Secretaría de Educación Pública, de ambos sexos y con

valores de edad de 6 1/2 a 16 1/2 años, residentes en México, D.F.

El WISC-R y el WISC-RM están constituidos por las mismas doce subpruebas del WISC publicado en 1949, (seis pertenecientes a la escala verbal y seis a la escala de ejecución), mencionadas a continuación:

Verbal	Ejecución
Información	Figuras incompletas
Semejanzas	Ordenación de dibujos
Aritmética	Diseño con cubos
Vocabulario	Composición de objetos
Comprensión	Claves
Retención de Dígitos	Laberintos

Tanto Retención de Dígitos como Laberintos no se utilizan para obtener el C.I., pero se aplican cuando existe algún impedimento para aplicar alguna de las subpruebas básicas. Tener subpruebas suplementarias tiene por objeto equilibrar la calificación global del sujeto cuando por alguna razón no se puede aplicar alguna de las otras subpruebas (Morales, 1990).

Siempre es permisible aplicar las doce subpruebas; aún más en situaciones clínicas, ya que el hacerlo proporciona información cualitativa y diagnóstica.

Las subescalas verbal y de ejecución se alternan para hacer más interesante y variada la sesión de aplicación.

Las instrucciones para la calificación de las subpruebas del WISC-R y del WISC-RM son objetivas. Cada una de las subpruebas cuenta con criterios fijos establecidos para la calificación, asimismo, en el manual se presentan ejemplos de posibles respuestas y su calificación correspondiente, con el objeto de auxiliar al examinador en esa tarea. Sin embargo, para las subpruebas de Vocabulario y Comprensión deben de considerarse los criterios del examinador; es tarea del examinador evaluar la respuesta del niño de acuerdo a las reglas de calificación para cada reactivo.

IV.8 PROCEDIMIENTO.

Se estableció contacto con los directores de las dos escuelas secundarias particulares, a fin de obtener su autorización para llevar a cabo la investigación. Posteriormente se les solicitó las listas de los alumnos con el propósito de obtener la muestra, de acuerdo a lo especificado en el apartado correspondiente.

Una vez seleccionada la muestra:

1. Se dividió en dos grupos iguales, con el propósito de determinar a quienes se les

aplicaría el WISC-R y a quienes el WISC-RM, cuidando que en ambos grupos quedaran sujetos de las dos escuelas.

Los sujetos quedaron asignados a los grupos de la siguiente manera:

Grupo 1

Aplicación
RAVEN / WISC-R

Edad	Hombres	Mujeres
13	4	4
14	4	4
15	4	4
Subtotal	12	12

Grupo 2

Aplicación
RAVEN / WISC-RM

Edad	Hombres	Mujeres
13	4	4
14	4	4
15	4	4
Subtotal	12	12

2. Se aplicó la Prueba de Matrices Progresivas, en forma grupal, en el salón de música a la totalidad de la muestra de cada escuela.

Antes de iniciar se explicó a los alumnos el propósito de la aplicación, se solicitó su colaboración y se les informó la forma en que habían sido seleccionados, que se les aplicarían dos instrumentos diferentes, así como que los resultados de los mismos no influirían en sus calificaciones escolares; todo lo anterior con el propósito de establecer un adecuado rapport para la ejecución de las pruebas.

Las instrucciones para el Raven fueron dadas de acuerdo a lo establecido en el manual correspondiente. Durante la ejecución se vigiló a los sujetos a fin de verificar que no se copiaran entre sí.

La calificación del Raven se obtuvo de acuerdo a lo establecido en el manual del mismo. Primero se obtuvieron los puntajes naturales o crudos de cada sujeto; posteriormente se verificó la consistencia de los mismos de acuerdo a lo señalado por Raven, con lo que nos percatamos que seis pruebas quedaban invalidadas debido a que la discrepancia entre los puntajes parciales reales y los puntajes parciales esperados eran mayores de +2 ó -2. Por tal motivo no se podía considerar la calificación obtenida por esos sujetos, como expresión fiel de su capacidad intelectual y hubo la necesidad de substituirlos, seleccionando a otros sujetos de la población.

Esta segunda selección se llevó a cabo de la siguiente forma:

De las listas de los alumnos con que se contaba, se seleccionaron al azar dos sujetos de la misma edad y sexo de aquellos cuyas pruebas habían sido invalidadas. Para esta segunda muestra se extrajeron dos sujetos por cada uno de aquellos cuyas pruebas habían sido invalidadas por lo siguiente:

- a) Contar con un número mayor de sujetos (12) para la aplicación grupal, con el propósito de que ésta se realizara en condiciones semejantes a las que habían tenido los demás sujetos que seguían formando parte de la muestra. Es decir, se pretendía así, controlar variables extrañas.
- b) Asegurar de esta forma, de que si se volvía a invalidar alguna otra prueba, de todos modos se contara con los sujetos que se requerían para la investigación.

De esta segunda aplicación sólo se invalidó una prueba, por lo que el otro sujeto de la misma edad y sexo pasó a formar parte de la muestra de investigación. De los demás sujetos, se seleccionó al azar a uno de los dos con que se contaba. Una vez obtenidos los puntajes brutos o crudos de los 48 sujetos se obtuvieron los puntajes normalizados utilizando la tabla IV, Baremo Córdoba-Adolescentes-Forma colectiva.

Se seleccionó dicho baremo debido a que la muestra de donde se obtuvo coincidía con la de la investigación que nos ocupa en lo siguiente:

- a) La aplicación se realizó en forma grupal.
 - b) Los sujetos eran también estudiantes secundarios.
 - c) Las edades de los sujetos de nuestra investigación (13-15) quedaban incluidas en las que se presentan en el Baremo Córdoba.
3. Se aplicó en forma individual, en el salón de Orientación Vocacional de cada una de las dos escuelas el WISC-R y el WISC-RM.

Para esta aplicación se verificó que se realizara en una hora en que los alumnos no tuvieran clases consideradas por ellos mismos como muy importantes o difíciles (matemáticas o química), solicitándoles incluso a ellos el horario más adecuado, con el propósito de que se encontraran en la mejor disposición posible y se pudiera establecer un adecuado rapport.

Nuevamente se les recordó el objetivo de la investigación y se les pidió su colaboración.

La aplicación de ambas pruebas se llevó a cabo de acuerdo a las instrucciones del manual correspondiente.

Después de realizada la prueba se calificaba, en su totalidad, por la persona responsable de la aplicación. (fueron dos aplicadores).

4. Una vez obtenidas las 48 pruebas de Wechsler, calificadas, se recurrió al acuerdo interjueces para las subpruebas de comprensión y vocabulario, con el propósito

de objetivar y validar las calificaciones ya que en estas subpruebas a pesar de contar con criterios generales establecidos para su calificación, puede intervenir el criterio del examinador.

El acuerdo interjueces se llevó a cabo de la siguiente manera:

- a) Se intercambiaron las pruebas entre los dos examinadores, de tal forma que se volvieron a calificar esas pruebas por el examinador contrario al que se había realizado la aplicación.
 - b) Se solicitó a un tercer juez, ajeno a la aplicación de la presente investigación, experto en la aplicación del instrumento, que realizara la calificación de las subpruebas mencionadas.
5. Se realizó el vaciado de datos en las tablas correspondientes.
6. Se realizó el análisis estadístico de los datos en forma computarizada, a través del paquete estadístico de SPSS/PCT, versión 2.0 de la Unidad de Cómputo, Informática e Instrumentación, de la Facultad de Psicología de la Universidad Nacional Autónoma de México.

CAPITULO V

RESULTADOS Y DISCUSION

Al hacer una evaluación de la capacidad intelectual de los jóvenes estudiantes, es sumamente importante utilizar instrumentos que arrojen resultados válidos para la población mexicana. Otro punto importante es el contar con instrumentos que proporcionen las ventajas de la prueba de Matrices Progresivas de Raven; descritas anteriormente.

En virtud de lo anterior, la investigación que nos ocupa tuvo como propósito determinar:

¿Cuál es el nivel de validez concurrente entre la Escala de Matrices Progresivas de Raven y las escalas de Wechsler, (WISC-R y WISC-RM)?

Con el fin de responder estadísticamente a esta pregunta de investigación se utilizaron diversos tratamientos estadísticos como:

1. La prueba de producto - momento de Pearson (r ; este es un coeficiente de correlación lineal).

$$r = \frac{n \sum Xy - [\sum x] [\sum y]}{\sqrt{n \sum x^2 - [\sum x]^2} \sqrt{n \sum y^2 - [\sum y]^2}}$$

La utilización de esta prueba permite conocer la relación directa entre conjuntos de pares ordenados (WISC-R / Raven; WISC-RM / Raven y WISC / Raven) de esta forma conocemos la dirección y grado de las relaciones.

Este coeficiente de correlación, se basa en el estudio de la variación concomitante de los miembros del conjunto de pares ordenados.

El nivel de correlación lineal permite determinar la validez concurrente que existe entre las escalas del Wechsler y la escala de Raven.

2. La prueba t de diferencia de dos medias para muestras apareadas. El modelo matemático que se utilizó fue:

$$t = \frac{M a_1 - M a_2}{SE M a_1 - M a_2}$$

Esta prueba permite conocer la relación general entre dos variables, se aplicó con la finalidad de apoyar en la interpretación de los resultados obtenidos a través de la prueba de correlación lineal de Pearson.

Es importante mencionar que el tratamiento estadístico fue llevado a cabo por medios automatizados, se utilizó el paquete estadístico de SPSS/PC+ versión 2.0 de la Unidad de Cómputo, Informática e Instrumentación de la Facultad de Psicología de la Universidad Nacional Autónoma de México.

ANÁLISIS DE RESULTADOS.

Descripción general de la muestra.

La muestra seleccionada para la investigación quedó constituida por 48 sujetos, de los cuales 24 fueron del sexo masculino y 24 de sexo femenino. Las edades fluctuaban entre 13 y 15 años de edad y todos se encontraban cursando del 1o. al 3er. año de educación secundaria. Este grupo, para efectos del análisis estadístico, será denominado muestra total.

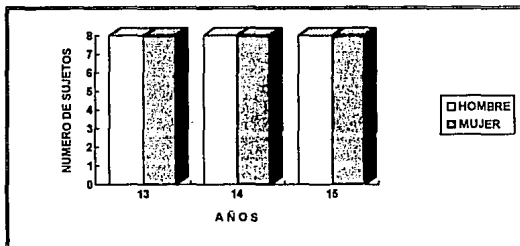
Asimismo es necesario señalar que al trabajar con la muestra total no se hizo diferenciación entre la versión de Wechsler utilizada, por lo que al hacer referencia de la escala únicamente se le denominará WISC, englobando así tanto al WISC-R, WISC-RM.

NOTA: Los datos que son significativos, para facilitar su lectura se presentan sombreados.

EDAD AÑOS/MESES	FRECUENCIA HOMBRES	FRECUENCIA MUJERES	% HOMBRES	% MUJERES
de 13 a 13.11	8	8	33.33	33.33
de 14 a 14.11	8	8	33.33	33.33
de 15 a 15.11	8	8	33.33	33.33
TOTAL	24	24	100	100

DISTRIBUCION GENERAL

	MEDIA	MEDIANA	DESV. ESTÁNDAR	RANGO
AÑOS	14.05	14.05	1.70	3



La muestra total se dividió en dos grupos iguales, los que en lo sucesivo serán designados como: grupo 1, integrado por aquellos sujetos a los que se les aplicó el Raven y el WISC-R y grupo 2, integrado por sujetos a los que se les aplicó el Raven y el WISC-RM.

**GRUPO 1
RAVEN - WISC-R**

EDAD	HOMBRES	MUJERES
13 AÑOS 11 M.	4	4
14 AÑOS 11 M.	4	4
15 AÑOS 11 M.	4	4
TOTAL	12	12

**GRUPO 2
RAVEN - WISC-RM**

EDAD	HOMBRES	MUJERES
13 AÑOS 11 M.	4	4
14 AÑOS 11 M.	4	4
15 AÑOS 11 M.	4	4
TOTAL	12	12

ANÁLISIS DEL COMPORTAMIENTO ESTADÍSTICO DE LA MUESTRA.

El análisis estadístico de los datos se llevó a cabo de la siguiente forma:

En primer lugar se trabajó con puntajes crudos de las tres pruebas utilizadas, con el propósito de conocer los resultados sin la influencia de la normalización de los instrumentos.

Posteriormente se hizo el análisis de las puntuaciones normalizadas, considerando que éstas son las que finalmente se utilizan con fines diagnósticos al aplicar una prueba.

Como se mencionó antes, el propósito de la investigación es conocer el nivel de correlación de las escalas de Wechsler (WISC-R y WISC-RM) con la prueba de Raven.

Los resultados de la investigación se presentan en el siguiente orden:

- Análisis de las distribuciones de frecuencia de:

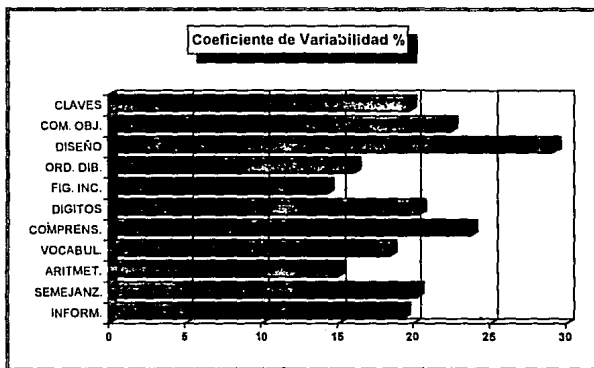
- Puntajes crudos
 - Puntajes normalizados

- Correlación de:

- Puntajes crudos
 - Puntajes normalizados

DISTRIBUCION GENERAL DE PUNTUACIONES CRUDAS POR SUBPRUEBA DEL TOTAL DE LA MUESTRA

VARIABLE	FRECUENCIA	MEDIA	DESVIACION ESTANDARD	COEFICIENTE VARIABILIDAD %
INFORM.	48	20.43	3.95	19.31
SEMEJANZ.	48	20.02	4.05	20.20
ARITMET.	48	13.93	2.08	14.96
VOCABUL.	48	44.79	8.24	18.38
COMPENS.	48	20.27	4.81	23.73
DIGITOS	48	11.98	2.44	20.33
FIG. INC.	48	19.75	2.81	14.23
ORD. DIB.	48	32.75	5.23	15.96
DISEÑO	48	38.92	11.34	29.14
COM. OBJ.	48	23.65	5.30	22.40
CLAVES	48	59.31	11.70	19.71



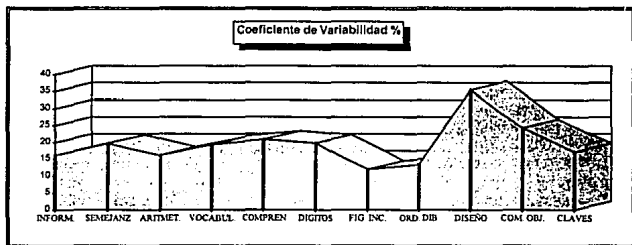
En este análisis se observó que las subpruebas de Semejanzas, Comprensión, Retención de Dígitos, Diseño con Cubos y Composición de Objetos, son las que presentan un coeficiente de variabilidad arriba del 20% con respecto a la media obtenida por el total de la muestra.

DISTRIBUCION GENERAL DE LAS CALIFICACIONES CRUDAS POR SUBPRUEBA WISC-R

Para el grupo 1, al que se le aplicó el WISC-R, se observó que de las once subpruebas estudiadas, las de Comprensión, Diseño con Cubos y Composición de Objetos, presentaron un coeficiente de variabilidad arriba del 20% con respecto a la media del grupo.

Grupo 1

VARIABLE	FRECUENCIA	MEDIA	DESVIACION ESTANDARD	COEFICIENTE VARIAB. %
INFORM.	24	19.95	3.22	16.14
SEMENJANZ.	24	20.33	4.00	19.67
ARITMET.	24	13.70	2.21	16.13
VOCABUL.	24	46.66	9.04	19.37
COMPRES.	24	20.20	4.21	20.84
DIGITOS	24	11.95	2.36	19.75
FIG. INC.	24	21.16	2.53	11.95
ORD. DIB.	24	32.37	4.36	13.47
DISENO	24	35.87	12.73	35.49
COM. OBJ.	24	23.04	5.59	24.28
CLAVES	24	60.04	10.30	17.15



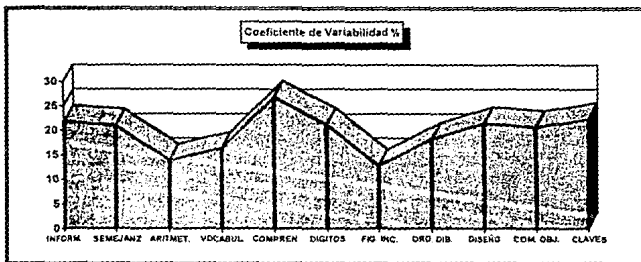
Para el resto de las subpruebas, el desempeño de los sujetos fue homogéneo dado que el 80% de las puntuaciones está cercano a la media obtenida.

DISTRIBUCION GENERAL DE PUNTAJES CRUDOS POR SUBPRUEBA WISC-RM

Para el grupo 2, al que se le aplicó la prueba WISC-RM se observó que se incrementa el número de subpruebas que presentan un coeficiente de variabilidad arriba del 20 % con respecto a la media. Ese coeficiente de variabilidad se presenta en Información, Semejanzas, Comprensión, Retención de Dígitos, Diseño con Cubos, Composición de Objetos y Claves.

Grupo 2

VARIABLE	FRECUENCIA	MEDIA	DESVIACION ESTANDARD	COEFICIENTE VARIAB. %
INFORM.	24	20.91	4.57	21.85
SEMENJANZ.	24	19.70	4.14	21.01
ARITMET.	24	14.16	1.97	13.91
VOCABUL.	24	42.91	7.02	16.36
COMPRESN.	24	20.33	5.45	26.81
DIGITOS	24	12.00	2.55	21.25
FIG. INC.	24	18.33	2.35	12.82
ORD. DIB.	24	33.12	6.04	18.24
DISEÑO	24	41.95	9.02	21.50
COM. OBJ.	24	24.25	6.02	20.70
CLAVES	24	58.58	13.12	22.39



Es importante señalar que solamente en cuatro de las subpruebas los sujetos presentaron un rendimiento homogéneo, en éstos más del 80% de sus puntajes se encuentran cercanos a la media.

**COMPARACION DE COEFICIENTES DE VARIABILIDAD
DE PUNTAJES CRUDOS**

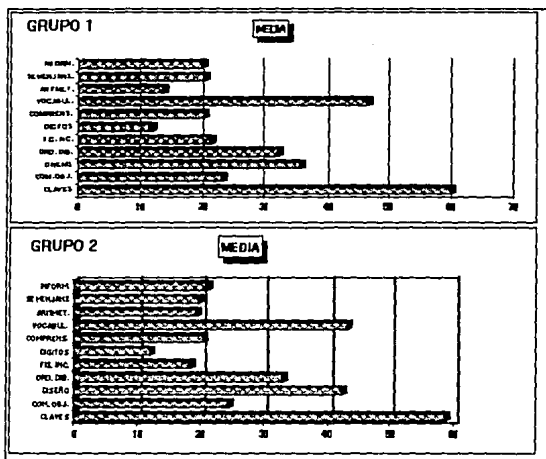
POR GRUPO

VARIABLE	MUESTRA TOTAL WISC	GRUPO 1 WISC-R	GRUPO 2 WISC-RM
Información	19.31	16.14	21.85
Semejanzas	20.20	19.67	21.01
Aritmética	14.98	16.13	13.91
Vocabulario	18.38	19.37	16.36
Comprensión	23.73	20.84	26.81
Dígitos	20.33	19.75	21.25
Fig. Incompletas	14.23	11.95	12.82
Ord. de Objetos	15.96	13.47	18.24
Diseño	29.14	35.49	21.50
Com. Objetos	22.40	24.26	20.70
Claves	19.71	17.15	22.39

Dado que para la selección de la muestra se controlaron el mayor número de variables interventoras o extrañas que pudieran influir sobre los resultados, la variabilidad de calificaciones que se observa en ambas versiones del WISC puede estar representado la diferencia de un sujeto a otro en cuanto al aprovechamiento de las experiencias proporcionadas por la situación escolar y medio ambiente.

Esto en virtud de que el WISC es una escala que evalúa la participación de varias operaciones e incluso una serie de habilidades específicas que al combinarse dan un producto intelectual, Morales (1990).

PUNTUACIONES MEDIAS DE CALIFICACIONES CRUDAS POR SUBPRUEBA



Al observar la diferencia en el número de subpruebas que tienen un coeficiente de variabilidad arriba del 20% con respecto a la media, en los dos grupos se aplicó la prueba de hipótesis llamada comparación de dos medias (t) con el propósito de determinar si existía una diferencia significativa

entre las puntuaciones medias de ambos grupos.

VARIABLE	GRUPO 1	GRUPO 2	VALOR DE t	SIGNIFICANCIA
INFORM.	19.95	20.91	-0.84	0.406
SEMENJANZ.	20.33	19.70	0.53	0.598
ARITMET.	13.70	14.16	-0.76	0.453
VOCABUL.	46.66	42.91	1.60	0.116
COMPRES.	20.20	20.33	-0.09	0.930
DIGITOS	11.95	12.00	-0.06	0.954
FIG. INC.	11.16	18.33	-4.02	0.000
ORD. DIB.	32.37	33.12	-0.49	0.624
DISEÑO	35.87	41.95	-1.91	0.062
COM. OBJ.	23.04	24.25	-0.79	0.435
CLAVES	60.04	58.58	0.43	0.671

Con este análisis se pudo determinar que únicamente existe una diferencia significativa entre la media de puntajes crudos obtenida por ambos grupos en la subprueba de Figuras Incompletas.

Entonces, la variación que se observa en el grupo 2 puede ser atribuida tanto a las diferencias individuales entre los miembros del grupo, a consecuencia del azar o a fluctuaciones de medida que no se pueden identificar (Kerlinger 1975).

DISTRIBUCION GENERAL DE CALIFICACIONES CRUDAS POR ESCALAS

Muestra Total

VARIABLE	FREC.	MEDIA	DEV. ESTAND.	COEF. VARIAB. %
ESCALA VERBAL	48	119.42	16.16	13.53
ESCALA DE EJECUCION	48	174.37	19.11	10.95
TOTAL WISC	48	293.80	31.07	10.59

Grupo 1

VARIABLE	FREC.	MEDIA	DEV. ESTAND.	COEFIC. VARIAB. %
ESCALA VERBAL	24	120.87	16.17	13.37
ESCALA DE EJECUCION	24	172.50	19.19	11.13
TOTAL WISC	24	293.37	31.04	10.58

Grupo 2

VARIABLE	FREC.	MEDIA	DEV. ESTAND.	COEFIC. VARIAB. %
ESCALA VERBAL	24	117.96	16.36	13.87
ESCALA DE EJECUCION	24	176.25	19.24	10.91
TOTAL WISC	24	294.21	31.77	10.80

En lo referente a los puntajes crudos obtenidos en las escalas WISC-R, WISC-RM y WISC, se observó que ninguno presentaba un coeficiente de variabilidad arriba del 20% con respecto a la media obtenida.

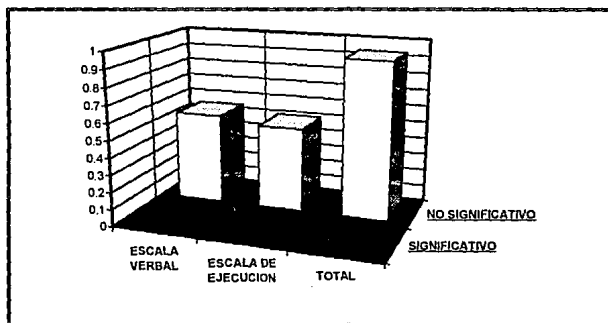
Esto pone de manifiesto la compensación que existe al evaluar la capacidad del individuo a través de diferentes lenguajes, es decir, al analizar por separado cada una de las subpruebas queda clara la diferencia que existe en cada una de las capacidades específicas de los individuos, sin embargo al conjuntarlas para obtener una puntuación general de su capacidad, se compensan esas diferencias.

PUNTUACIONES MEDIAS DE CALIFICACIONES CRUDAS POR ESCALAS

COMPARATIVO POR GRUPOS

Para conocer si existía alguna diferencia entre las puntuaciones crudas a nivel escala de una versión a otra de las escalas de Wechsler, se sometieron a la prueba t las puntuaciones medias.

VARIABLE	GRUPO 1 WISC-R	GRUPO 2 WISC-RM	VALOR DE t	SIGNIFICANCIA
ESCALA VERBAL	120.8750	117.9583	0.62	0.538
ESCALA DE EJECUCION	172.500	176.250	-0.68	0.502
TOTAL	293.375	294.2083	-0.09	0.927



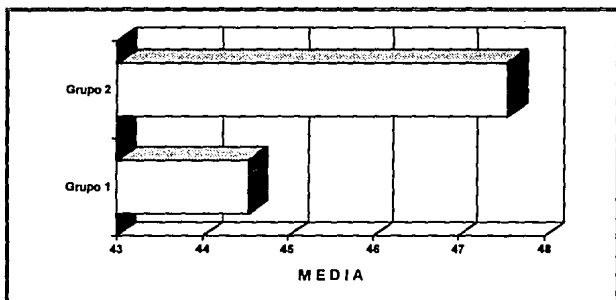
El análisis mostró que no existe ninguna diferencia entre las calificaciones crudas de las escalas de un grupo a otro, lo que señala la homogeneidad de la muestra.

**COMPARACION DE LAS DISTRIBUCIONES DE LOS
PUNTAJES CRUDOS OBTENIDOS A TRAVES DE LA
PRUEBA DE MATRICES PROGRESIVAS DEL RAVEN**

GRUPO	FRECUENCIA	MEDIA	DESV. ESTANDART	% COEFIC. VARIABILIDAD
Grupo 1	24	44.54	5.66	12.70
Grupo 2	24	47.58	6.04	12.68
Muestra total	48	46.06	5.98	12.19

El coeficiente de variabilidad por debajo del 13% observado señala que el 87% de las puntuaciones quedan agrupadas alrededor de la media, por lo que se puede hablar de la homogeneidad de los grupos. Homogeneidad que concuerda con lo observado en las escalas de Wechsler.

PUNTUACIONES MEDIAS DE CALIFICACIONES CRUDAS DE LA PRUEBA DE MATRICES PROGRESIVAS DEL RAVEN



Para determinar si hay significancia en la diferencia de las puntuaciones medias observadas entre un grupo y otro, se sometieron a la prueba t.

GRUPO 1	GRUPO 2	VALOR DE t	SIGNIFICANCIA
44.54	47.58	-1.88	0.078

Este análisis muestra que la diferencia de puntuaciones medias de un grupo a otro no es estadísticamente significativa.

ANALISIS DE LA DISTRIBUCION GENERAL DE PUNTUACIONES NORMALIZADAS

A nivel diagnóstico, la capacidad intelectual de un sujeto es reportada tomando en cuenta su desempeño en comparación con un grupo normativo, o sea que el coeficiente intelectual obtenido se basa en una muestra representativa de la población a cada nivel, llamándose C.I. de desviación.

Por esta razón se creyó conveniente realizar el análisis de frecuencias y distribuciones de las puntuaciones normalizadas.

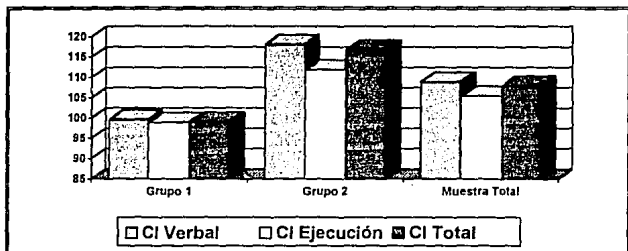
COMPARACION DE LA DISTRIBUCION GENERAL DE PUNTUACIONES NORMALIZADAS

VARIABLE	GRUPO 1				GRUPO 2				MUESTRA TOTAL			
	FREC.	MEDIA	D. EST.	% C.V.	FREC.	MEDIA	D. EST.	% C.V.	FREC.	MEDIA	D. EST.	% C.V.
C.I. VERBAL	24	99.58	10.39	10.43	24	118	13.22	11.20	48	108.80	14.99	13.78
C.I. EJECUCION	24	98.92	12.80	12.94	24	111.83	8.40	7.51	48	105.38	12.54	11.90
C.I. TOTAL	24	99	11.42	11.53	24	116.67	10.87	9.31	48	107.83	14.19	13.16

El análisis de la distribución de los C.I.s, para el grupo 1, grupo 2 y muestra total, demuestra que más del 80% de las puntuaciones quedaron distribuidas alrededor de la media, al observarse en los tres grupos un coeficiente de variabilidad abajo del 20%.

COMPARACION DE PUNTUACIONES NORMALIZADAS DE LA ESCALA DE WECHSLER

Para determinar que tanto pudiera influir en los resultados la normalización de una versión a otra, las puntuaciones medias de los grupos se sometieron a la prueba t.



NIVEL DE SIGNIFICANCIA DE LAS DIFERENCIAS DE MEDIAS

VARIABLE	GRUPO 1	GRUPO 2	VALOR DE t	SIGNIFICANCIA
C.I. VERBAL	99.58	118.00	-5.37	0.00
C.I. EJECUCION	98.92	111.83	-4.13	0.00
C.I. TOTAL WISC	99.00	116.61	-5.49	0.00

Los resultados muestran que las diferencias son altamente significativas de un grupo a otro, situación no esperada debido a que al realizar el análisis con puntajes crudos no se observó ninguna diferencia significativa.

Llama la atención en el caso del WISC-RM que las puntuaciones normalizadas sean más altas que las del grupo al que se le aplicó WISC-R; lo que podría estar reflejando una sobre valoración de dichas calificaciones.

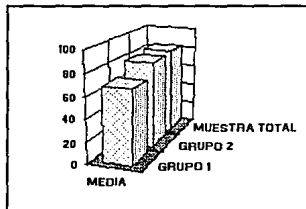
Dato que apoya los resultados reportados por Heredia (1993) "En lo referente a los coeficientes intelectuales se observó que el WISC-RM arroja un promedio tan alto en la escala de ejecución y total que estas se localizan en la segunda desviación a la derecha de la curva, considerando la media y la desviación propuesta por Wechsler. Al medir la capacidad intelectual con esta versión se sobrevalora la misma y por lo tanto las acciones que se tomen a partir de ella, pueden resultar falseadas"

DISTRIBUCION GENERAL DE CALIFICACIONES NORMALIZADAS DE LA ESCALA DE RAVEN

GRUPO	FRECUENCIA	MEDIA	DESV. EST.	% COEF. VARIAB
1	24	68.54	25.60	37.35
2	24	80.83	19.37	23.96
MUESTRA TOTAL	48	82.51	23.30	31.19

El coeficiente de variabilidad de la escala de Raven es alto, ésto se puede entender debido a que los rangos percentiles tienen una distribución rectangular en el que las distribuciones de las calificaciones de las pruebas se aproximan en general a la curva normal, como consecuencia de ello las pequeñas diferencias de las calificaciones brutas cerca del centro de la distribución presentan grandes diferencias en percentiles (Brown 1980).

COMPARACION DE PUNTUACIONES MEDIAS NORMALIZADAS DE LA ESCALA DEL RAVEN



NIVEL DE SIGNIFICANCIA DE LA DIFERENCIA DE MEDIAS

VARIABLE	GRUPO 1	GRUPO 2	VALOR DE t	SIGNIFICANCIA
PERCENTIL RAVEN	68.54	80.83	-1.88	0.067

Estos resultados muestran que no hay una diferencia significativa entre las puntuaciones medias normalizadas de un grupo a otro.

ANALISIS DEL NIVEL DE RELACION DEL WISC CON RAVEN

Las escalas de medición de Wechsler y Raven están basadas en la teoría bifactorial de Spearman, en el sentido de que tanto para Wechsler como para Raven existe un factor fundamental en las funciones intelectuales denominado factor G o común, así como factores específicos.

Wechsler analiza la idea de la inteligencia general, no obstante, para Spearman G era un factor primario de la inteligencia y para Wechsler G es un factor secundario, o sea, hace la función de unificador para la expresión unitaria de una conducta que requiere la participación de varias operaciones, e incluso, de una serie de habilidades específicas que al combinarse dan un producto intelectual.

Las escalas de Wechsler miden diferentes funciones, y por tanto distintos factores.

Raven busca medir el factor G apoyándose en la idea de Spearman de que medir G, significa medir la "Inteligencia General", es decir G es el factor común y fundamental en todas las funciones cognoscitivas del mismo individuo.

Debido a que ambas pruebas tienen el mismo sustento teórico sobre el concepto de inteligencia, diferenciándose únicamente en que una mide exclusivamente el factor G (Raven) y la otra mide tanto factores G como específicos, a continuación se analiza la relación entre ambas escalas.

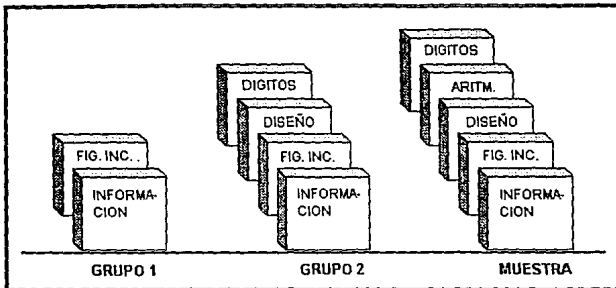
NIVEL DE CORRELACION DE LAS ESCALAS DE WECHSLER CON RAVEN PUNTAJES CRUDOS

VARIABLE	GRUPO 1	GRUPO 2	MUESTRA TOTAL
INFORM.	0.59*	0.61**	0.61**
SEMEJANZ.	0.21	0.42	0.29
ARITMET.	0.41	0.47	0.45*
VOCABUL.	0.16	0.37	0.01
COMPENS.	0.36	0.21	0.26
DIGITOS	0.24	0.54*	0.39*
FIG. INC.	0.63**	0.60*	0.36
ORD. DIB.	0.12	0.46	0.33
DISEÑO	0.42	0.60*	0.52**
COM. OBJ.	0.12	0.30	0.23
CLAVES	0.007	0.18	0.29

Nivel de significancia

* -0.01

** -0.001



La visión global de los grupos pone de manifiesto la diferencia que existe en cuanto al número de subpruebas que correlacionan con Raven, diferencias que pueden ser atribuidas:

A la versión de Wechsler utilizada ya que en la versión estandarizada para la población mexicana se observa un número mayor de subpruebas que correlacionan con Raven. Esto puede significar que al reajustar los reactivos de esta versión, se pone más de manifiesto las capacidades generales del individuo.

Sin embargo, esta aseveración deberá considerarse con reservas dado que las diferencias observadas entre los grupos pueden estar influenciadas por el tamaño de la muestra, por lo que debido a que se trata de una muestra pequeña, cualquier variación en el tamaño de ésta, puede representar diferencias importantes en los resultados.

ANÁLISIS COMPARATIVO DE CORRELACIONES POR SUBPRUEBA. PUNTAJES CRUDOS

El siguiente análisis tiene como objetivo explicar la ausencia o presencia de correlaciones significativas en base a la concordancia de habilidades medidas por el Raven y las escalas de Wechsler. Se fundamenta en las investigaciones que diversos autores han realizado a través del método matemático de análisis factorial.

CORRELACION ENTRE RAVEN Y LA SUBPRUEBA DE INFORMACION

VARIABLE	GRUPO 1	GRUPO 2	MUESTRA TOTAL
INFORM.	0.58 ^{**}	0.41 ^{**}	0.41 ^{**}

El análisis comparativo de las correlaciones significativas para los tres grupos, muestran en forma constante una correlación en la subprueba de Información, la cual es considerada como la segunda mejor medida de factor G por Sattler (1988), y Kaufman (1982). Asimismo, Morales (1990) menciona que en primer lugar esta prueba está influenciada por el factor memoria, operaciones de asociación y organización de la experiencia.

Sattler (1988) señala que la suma de conocimientos que poseen los niños suelen depender de sus dotes naturales y de sus oportunidades y preferencias culturales.

En general, la subprueba de Información muestra el tipo de conocimientos que es común en los niños con oportunidades similares.

La respuesta de los niños proporciona índices sobre la amplitud general de su información, de como aprovechan el medio y de sus antecedentes sociales y culturales.

De acuerdo a la categorización de Bannatyne, en Kaufman (1982), dice que es una capacidad influenciada de conocimientos adquiridos.

Esta correlación significativa, independientemente de su interpretación en relación al factor G, muestra además la homogeneidad de la muestra en estudio en cuanto a las oportunidades culturales y de aprendizaje a las que se encuentran expuestos.

CORRELACION DE SUBPRUEBAS CONSIDERADAS BUENAS MEDIDAS DE G.

VARIABLE	GRUPO 1	GRUPO 2	MUESTRA TOTAL
SEMENJANZ.	0.21	0.42	0.29
VOCABUL.	0.16	0.37	0.01
COMPRES.	0.36	0.21	0.26

Para las subpruebas de Semejanzas, Vocabulario y Comprensión tanto Sattler como Kaufman y Morales los reportan como buenas medidas de G, sin embargo éstas no presentan una correlación significativa con la prueba de Raven.

Correlación que debía esperarse dado que comparten habilidades tales como: grado de pensamiento abstracto, distinguir los detalles esenciales, relación causa efecto, y evaluación.

La baja correlación observada puede deberse a que estas subp.uebas tienen gran influencia de factores externos tales como oportunidades culturales, aprendizaje escolar, intereses y riqueza del medio ambiente; factores que no afectan el desempeño de la prueba de Raven.

CORRELACION DE LA SUBPRUEBA DE ARITMETICA CON RAVEN.

VARIABLE	GRUPO 1	GRUPO 2	MUESTRA TOTAL
ARITMET.	0.41	0.47	0.45**

En la subprueba de Aritmética se observa una correlación significativa al .001 sólo en el caso del estudio de la muestra total. Tomando en cuenta que esta subprueba no sufre ninguna modificación de contenido en la estandarización del WISC-RM, la ausencia de correlación tanto en el grupo 1 como en el grupo 2, esperada en base a lo señalado por Kaufman, Sattler y Morales, puede estar afectada por el tamaño de la muestra.

Dicha subprueba constan de problemas que exigen concentración y atención específica, para su resolución se requieren operaciones de razonamiento con abstracciones, formación de conceptos y retención de procesos aritméticos.

CORRELACION DE LA SUBPRUEBA DE RETENCION DE DIGITOS CON RAVEN.

Para la subprueba de Retención de Dígitos se presenta una correlación significativa al 0.01 tanto para el grupo 2 como para la muestra total.

VARIABLE	GRUPO 1	GRUPO 2	MUESTRA TOTAL
DIGITOS	0.24	0.54*	0.33*

Esta subprueba está considerada como una medida deficiente de G, por Sattler, Kaufman y Morales ya que está fuertemente influida por el factor memoria.

Sin embargo, esta subprueba, así como la de Aritmética están consideradas por Bannantine como integrantes del grupo de atención y concentración, funciones que también son importantes para la adecuada resolución del Raven, lo que explica la correlación existente entre esta última y dichas subpruebas.

Preguntas tales como ¿Porqué no existe correlación en el grupo 1 y sí para el grupo 2 y la muestra total? deberán ser sujetas a nuevas investigaciones en donde deberá de incrementarse el tamaño de la muestra.

CORRELACION DE LA SUBPRUEBA DE FIGURAS INCOMPLETAS CON RAVEN.

VARIABLE	GRUPO 1	GRUPO 2	MUESTRA TOTAL
FIG. INC.	0.83**	0.68*	0.38*

En la subprueba de Figuras Incompletas se observa una correlación significativa para los tres grupos, Morales (1990) menciona que en esta subprueba aparece una situación especial: presenta una carga sobre factores no verbales en la mayoría de los análisis efectuados, y muestra poca influencia del factor G, la más baja de todas las subpruebas de ejecución; sin embargo, señala que surge un nuevo y único factor que parece importante en el comportamiento intelectual y que no ha podido ser identificado con precisión.

Intervienen en él factores como la percepción visual (análisis) y la imaginación visual, por lo anterior, cabe la hipótesis de que este factor sea el responsable de esa correlación significativa con Raven, aunque para Sattler y Kaufman, esta subprueba representa una medida regular de G.

En esta subprueba el sujeto debe saber lo que la imagen representa y apreciar que la parte faltante es, de alguna manera, esencial a la forma total. El Raven plantea problemas de percepción de totalidad y el sujeto debe integrar o cerrar una figura inconclusa y por tanto debe de ser capaz de percibir las semejanzas, diferencias, simetría y continuidad de las partes en relación con la estructura o forma del todo.

Ambas pruebas son, entonces de tipo fuertemente gestáltico.

CORRELACION DE LA SUBPRUEBA DE ORDENAMIENTO DE DIBUJOS CON RAVEN

VARIABLE	GRUPO 1	GRUPO 2	MUESTRA TOTAL
ORD. DIB.	0.12	0.46	0.33

En el caso de la subprueba de Ordenamiento de Dibujos, la correlación existente no llega a ser significativa, lo que concuerda con lo expresado tanto por Kaufman como por Sattler, quienes señalan que esta es una subprueba considerada como una medida regular de G.

Mide la habilidad del niño para comprender y valorar una situación global. La tarea es de razonamiento no verbal, que puede concebirse como una medida de la habilidad para planear, en la que están de por medio la anticipación y la organización visual.

La anticipación que se requiere es una habilidad para prever las consecuencias de los actos o situaciones iniciales e interpretar situaciones sociales. La capacidad para anticipar los posibles antecedentes y consecuencias de los eventos es muy importante para proporcionar a las experiencias de la vida diaria una continuidad que les de significado (Blatty Allison, 1968, citado en Sattler, 1982).

CORRELACION DE LA SUBPRUEBA DE DISEÑO CON CUBOS CON RAVEN.

VARIABLE	GRUPO 1	GRUPO 2	MUESTRA TOTAL
DISEÑO	0.42	0.60*	0.52**

Diseño con Cubos es una subprueba señalada por Sattler, Kaufman y Morales como una buena medida del factor G, esto se confirma en la correlación significativa observada entre las puntuaciones crudas tanto del grupo 2 como de la muestra total con el Raven. Asimismo, se observa una correlación moderada con el grupo 1.

Diseño con Cubos supone la habilidad para percibir y analizar formas, dividiendo el todo en sus partes componentes y ensamblándolas luego para formar un diseño idéntico; a este proceso se le llama análisis y síntesis. La prueba combina la organización visual con aspectos de coordinación visomotora.

El éxito en la prueba involucra la aplicación de la lógica y razonamiento a problemas de relaciones espaciales, por consiguiente, puede concebirse como una formación de conceptos no verbales que requieren habilidades de organización perceptual, visualización espacial y conceptualización abstracta; funciones que son básicas en la resolución de los problemas del Raven.

CORRELACION DE LA SUBPRUEBA DE COMPOSICION DE OBJETOS CON RAVEN

VARIABLE	GRUPO 1	GRUPO 2	MUESTRA TOTAL
COM. OBJ.	0.12	0.30	0.23

Composición de Objetos requiere coordinación visomotora, es una prueba de habilidad para la organización perceptual, la organización de lo que se percibe parece jugar un papel productivo en cuanto que hay que producir algo con piezas que algunas veces no se reconocen de inmediato. A los examinados se les exige que capten todo el conjunto, anticipando las relaciones entre las piezas individuales, por tanto, las tareas suponen cierta habilidad para construir y buena habilidad perceptual.

Es considerada por Kaufman, Sattler y Morales como una medida regular de G, sin embargo, dado que puede llegarse a su resolución a través del ensayo y error, ésta

puede ser la causa por la que las correlaciones observadas en los tres grupos sean bajas.

CORRELACION DE LA SUBPRUEBA DE CLAVES CON RAVEN

VARIABLE	GRUPO 1	GRUPO 2	MUESTRA TOTAL
CLAVES	0.007	0.18	0.29

Claves es una subprueba considerada por Sattler, Kaufman y Morales como una mala medida del factor G, situación que se confirma al encontrar una correlación muy baja con el Raven en las puntuaciones crudas de los tres grupos.

COMPARACION DE CORRELACIONES DE PUNTAJES CRUDOS POR ESCALA CON RAVEN

VARIABLE	GRUPO 1		GRUPO 2		MUESTRA TOTAL	
	CORRELACION	SIGNIFICANCIA	CORRELACION	SIGNIFICANCIA	CORRELACION	SIGNIFICANCIA
ESCALA VERBAL	0.039	0.85	0.42	0.04	0.21	0.20
ESCALA DE EJECUCION	0.43	0.033	0.23	0.27	0.34	0.01
ESCALA TOTAL	0.29	0.16	0.35	0.09	0.32	0.10

De acuerdo a la teoría de Cattell, tanto la inteligencia fluida como la cristalizada representan "inteligencia".

La inteligencia o capacidad fluida representa influencias grandemente constitucionales y fisiológicas, mientras que la inteligencia cristalizada ésta más relacionada a influencias educacional-experiencia y a técnicas y estrategias adquiridas.

La Escala Verbal (excluyendo Retención de Dígitos), puede considerarse como una buena medida del pensamiento cristalizado.

La Escala de Ejecución al parecer, es en gran parte una medida tanto de la capacidad fluida como de la visualización amplia (organización visual), en tanto que sus subpruebas demandan la habilidad para ver relaciones, series de números y letras, clasificación de figuras, analogías de figuras, figuras que deben ser encajadas, diseño de bloques, etc.

Kaufman, (1982) señala que la prueba de Matrices Progresivas de Raven es un instrumento con tareas que son medidas más puras de la inteligencia fluida.

En base a lo anterior, se justifica la correlación que se observa con la escala de ejecución del grupo 1 y de la muestra total.

La discrepancia en cuanto a las correlaciones por escala, de una versión a otro con el Raven, deberán ser objeto de nuevas investigaciones, en donde deberá de incrementarse el tamaño de la muestra.

COMPARACION DE CORRELACIONES DE PUNTAJES NORMALIZADOS POR ESCALA CON RAVEN

VARIABLE	GRUPO 1		GRUPO 2		MUESTRA TOTAL	
	CORRELACION	SIGNIFICANCIA	CORRELACION	SIGNIFICANCIA	CORRELACION	SIGNIFICANCIA
C.I. VERBAL	0.17	0.40	0.15	0.48	0.29	0.45
C.I. DE EJECUCION	0.52	0.005	0.21	0.31	0.49	0.001
C.I. TOTAL	0.46	0.049	0.17	0.42	0.46	0.01

Al realizar el análisis de las puntuaciones normalizadas únicamente se mantiene la correlación con el grupo 1 y con la muestra total para la escala de ejecución, correlacionando en este caso también el C.I. total del grupo 1 y la muestra total con la escala del Raven.

Esta correlación del Raven con la escala de ejecución puede justificarse tomando en cuenta que ambas escalas evalúan la capacidad intelectual con tareas que son medidas más puras de la inteligencia fluida entonces, las correlaciones bajas del Raven con el WISC-RM aparentemente pueden ser atribuidas por efectos de los cambios en las tablas de equivalencias, hecho que deberá ser objeto de nuevas investigaciones.

**RELACION ENTRE CALIFICACIONES NORMALIZADAS Y
CLASIFICACION DE INTELIGENCIA**

	GRUPO 1	GRUPO 2	MUESTRA TOTAL
C.I.	99.58	118	108.80
VERBAL	Normal	Arriba del normal (brillante)	Normal
C.I.	98.92	111.83	105.38
EJECUCION	Normal	Arriba del normal (brillante)	Normal
C.I.	99	116.66	107.83
TOTAL WISC	Normal	Arriba del normal (brillante)	Normal
PERCENTIL	68.54	80.83	74.69
RAVEN	Término medio	Superior al término medio	Término medio

La comparación de los diagnósticos o clasificaciones de inteligencia de la escala del Wechsler y Raven, coinciden o son equivalentes para cada una de estas escalas, por lo que se puede interpretar que aunque no exista una correlación estadística significativa entre cada una de las escalas del Wechsler y e' Raven, al dar un diagnóstico de inteligencia este resulta equivalente, sin importar a través de que prueba se obtenga.

CAPITULO VI

CONCLUSIONES

La Prueba de Raven es una prueba ampliamente difundida en los diferentes ámbitos educativos, donde se requiere valorar la capacidad intelectual de los estudiantes, la cual permite tomar decisiones referentes a la selección y clasificación de los mismos.

Dicho conocimiento proporciona elementos auxiliares para la formación de grupos homogéneos que posibiliten el avance académico paralelo.

Al mismo tiempo forma parte de una batería de pruebas psicológicas encaminadas a determinar: las causas del fracaso escolar, para la orientación vocacional, canalización a programas de educación especial, etc..

Brown (1980) al respecto señala que "el hecho de que estas decisiones puedan modificar drásticamente la vida de una persona, hace que la comprensión del proceso de validez.... sea esencial para el uso de la prueba".

La prueba de Matrices Progresivas de Raven es una escala elaborada inicialmente para determinar la capacidad intelectual de la población inglesa. A través de diferentes estudios, los expertos coinciden en la opinión de que la prueba constituye en el presente un instrumento de trabajo altamente satisfactorio.

En varios países se continúa utilizando en su versión original, y en algunos de ellos se han adaptado las tablas o baremos de calificaciones de acuerdo a las características de la población. Lo que no se ha llevado a cabo en México, por lo que al utilizarse en este país, se compara el rendimiento de los niños mexicanos con el desempeño de niños de otros países, como por ejemplo: Argentina, Uruguay y España.

García (1987) menciona que existen evidencias empíricas de que resulta injusto para los niños mexicanos ser evaluados con instrumentos no estandarizados, ya que pueden ser ubicados en una categoría descriptiva inferior.

De ahí la importancia de determinar la validez del instrumento. Un método adecuado para determinarla es a través de la comparación de calificaciones con una variable criterio.

Se considera como variable criterio aquel que, según se sabe, mide el atributo del cual se trata. Tal es el caso de la escala del WISC-R y en especial, dada su estandarización para la ciudad de México, del WISC-RM.

Ambas pruebas (WISC-R Y WISC-RM) han demostrado, a través de diferentes estudios, su validez para determinar la capacidad intelectual. Por este motivo se consideraron adecuadas para ser utilizadas como variables criterio.

- La presente investigación tuvo como objetivo conocer el nivel de validez concurrente de la prueba de Matrices Progresivas del Raven con las escalas WISC-R y WISC-RM.

Los resultados encontrados indican:

Cuando se analizan los resultados de la muestra total, o sea sin diferenciar la versión del WISC que se aplicó, se observó:

- A nivel de subpruebas con puntajes crudos existe correlación estadísticamente significativa con Información, Aritmética, Retención de Dígitos, Figuras Incompletas y Diseño con Cubos.

Los tres primeros pertenecientes a la escala verbal y los dos restantes a la de ejecución.

Esto apoya lo reportado por Kaufman (1982) referente a que en la escala verbal se encuentran un mayor número de subpruebas consideradas como buenas medidas de g; a pesar de ser subpruebas que tiene gran influencia de factores externos.

- A nivel de escalas con puntajes crudos, únicamente se encontró correlación estadísticamente significativa con la escala de ejecución. Kaufman también considera que dado que en la resolución de las subpruebas de la escala de ejecución se ven involucradas tareas de percepción, de semejanzas, diferencias, simetría, continuidad de las partes, integración y cierre de figuras, organización perceptual, etc., todas estas tareas son consideradas de tipo gestáltico (capacidad fluida) por lo que se esperaría la correlación con el Raven.

Esta aparente discrepancia es justificada también por Kaufman al señalar que la inteligencia es producto de un sin número de factores causales, la mayor parte de las tareas mentales implican la participación no sólo de la capacidad fluida, sino también del pensamiento cristalizado; en razón de ésto, las escalas verbal y de ejecución están muy lejos de ser medidas puras de una sola capacidad.

- Con puntuaciones normalizadas se encontró correlación estadísticamente significativa con el C.I. de ejecución y con el C.I. total. Resultado que concuerda con lo mencionado anteriormente.

Con el propósito de conocer con que versión de la escala de Wechsler existía una mayor relación con el RAVEN, se dividió la muestra conformándose dos grupos.

- A nivel de subpruebas con puntajes crudos, en el grupo 1 (Raven / WISC-R) únicamente se encontró una correlación estadísticamente significativa con Información y Figuras Incompletas.

En el grupo 2 (Raven / WISC-RM) se encontró correlación estadísticamente significativa con los puntajes crudos de información y figuras incompletas -al igual que en el grupo 1- además hubo correlación significativa con diseño con cubos y retención de dígitos.

Esta diferencia en cuanto al número de subpruebas que correlacionan de una versión a otra podría ser atribuida por efecto del tamaño de la muestra, o por la diferencia en las tablas de equivalencia, esta última suposición se apoya en lo reportado por Cuevas (1993) ya que al analizar la estructura interna de las tres versiones del Wechsler para niños menciona que esta es adecuada y que la versión que resulta más confiable, debido a que obtuvo los más altos coeficientes de consistencia interna es el WISC-RM. Sin embargo, de acuerdo a la experiencia profesional, se ha observado que al aplicar estas tres versiones se obtienen coeficientes intelectuales diferentes, por lo que se cuestiona el porque de este hecho, dado que las tres versiones poseen una consistencia interna satisfactoria. Al respecto ella plantea la hipótesis de que las discrepancias de los coeficientes intelectuales obtenidos en las tres versiones son debido a los puntajes escalares.

- A nivel de escalas en el grupo 1 existe correlación estadísticamente significativa con puntajes crudos con la escala de ejecución. Para el grupo 2, la correlación significativa encontrada fue con la escala verbal.

Nuevamente se puede observar diferencias de una versión a otra, situación que deberá ser objeto de otras investigaciones.

- A nivel de puntajes normalizados se observa únicamente una correlación estadísticamente significativa con el C.I. de Ejecución y C.I. Total del grupo al que se le aplicó el WISC-R y Raven.

Heredia (1993) al respecto, señala que las medias de los puntajes crudos obtenidos entre el WISC-R y el WISC-RM son prácticamente iguales, pero el puntaje escalar difiere significativamente favoreciendo al WISC-RM. Circunstancias que pueden estar reflejadas en la ausencia de correlación entre los puntajes escalares del Raven y WISC-RM.

Dado que las escalas de Wechsler están elaboradas tomando en consideración el concepto que su autor tenía de la inteligencia, quien la consideraba como un agregado o capacidad global del individuo, en donde intervienen tanto factores específicos como el factor G, y siendo la escala de Raven un instrumento específico para medir el factor G es de esperarse que el nivel de validez concurrente entre ambas pruebas fuera significativo.

Los resultados de la presente investigación hacen que se acepten las siguientes hipótesis alternas:

- H1 Existe correlación estadísticamente significativa entre las puntuaciones crudas obtenidas con la prueba de Raven y las puntuaciones crudas obtenidas con la escala total de WISC (R / RM).
- H1 Existe correlación estadísticamente significativa entre las puntuaciones crudas obtenidas con la prueba de Raven y las puntuaciones crudas obtenidas con la escala total de WISC - R.
- H1 Existe correlación estadísticamente significativa entre las puntuaciones crudas obtenidas con la prueba de Raven y las puntuaciones crudas obtenidas con la escala total de WISC - RM.
- H1 Existe correlación estadísticamente significativa entre las puntuaciones normalizadas obtenidas con la prueba de Raven y las puntuaciones normalizadas obtenidas con la escala de ejecución de WISC (R / RM).
- H1 Existe correlación estadísticamente significativa entre las puntuaciones normalizadas obtenidas con la prueba de Raven y las puntuaciones normalizadas obtenidas con la escala total de WISC (R / RM).
- H1 Existe correlación estadísticamente significativa entre las puntuaciones normalizadas obtenidas con la prueba de Raven y las puntuaciones normalizadas obtenidas con la escala de ejecución de WISC - R.
- H1 Existe correlación estadísticamente significativa entre las puntuaciones normalizadas obtenidas con la prueba de Raven y las puntuaciones normalizadas obtenidas con la escala total de WISC - R.

Hecho que es de esperarse debido a que en la resolución de las tareas tanto de la escala de ejecución como la del Raven, influyen factores tales como: agudeza y organización visual, habilidad para percibir y analizar formas dividiendo el todo en sus partes componentes y ensamblándolo luego (análisis y síntesis), de razonamiento no verbal, conceptualización abstracta asociación de símbolos, secuenciación, habilidad para comprender y valorar una situación global, capacidad de planeación y memoria visual.

Estas habilidades tienen poca influencia de factores ambientales por lo que permiten determinar una medida de la capacidad intelectual general (factor G), denominador común de la totalidad de las operaciones de la inteligencia.

La correlación observada en el C.I. global del grupo 1 y la muestra total confirman la postura de Wechsler a cerca de la existencia de los dos principales factores por medio

de los cuales se expresan las habilidades del humano (comprensión verbal y organización perceptual).

La correlación observada se justifica debido a que todas las subpruebas, independientemente de su efectividad para medir la inteligencia, sumadas son indispensables para apreciar en forma global la inteligencia. Misma que contiene el factor G, que es un factor cuantitativo de la inteligencia, el común y fundamental de todas las funciones cognoscitivas. Factor de inteligencia general que también es evaluado a través del Raven.

La presencia de correlación en la escala verbal del grupo 2 así como la ausencia de correlaciones con puntuaciones normalizadas del mismo grupo deberá ser objeto de nuevas investigaciones, en donde el tamaño de la muestra deberá ser un factor a tomarse en cuenta.

En lo referente a nivel de clasificación de la inteligencia, se encontró que tanto lo obtenido para la muestra total, para el grupo 1 y para el grupo 2 coinciden en diagnóstico al comparar las escalas de Wechsler con el Raven, o sea:

El C.I. Total del grupo al que se le aplicó WISC-R es Normal y con el Raven es Término Medio.

El C.I. Total del grupo al que se le aplicó WISC-RM es Arriba del Normal y con el Raven es Superior al Término Medio.

El C.I. Total obtenido por la muestra total es Normal y con el Raven es Término Medio.

Por tal motivo se considera que, independientemente de que es conveniente profundizar en la investigación, la Prueba de Matrices Progresivas del Raven nos da una evaluación válida de la capacidad intelectual para la muestra estudiada.

CAPITULO VII

LIMITACIONES Y SUGERENCIAS

La investigación en México la mayoría de las veces se ve limitada debido a la dificultad que representa el poder contar con las muestras requeridas.

Aunque para la presente investigación se contó con todo el apoyo del personal directivo de los colegios, por tratarse de instituciones con una población estudiantil no muy amplia, y dadas las características que se requerirían de los sujetos para un mejor control de variables extrañas (que fueran de una edad determinada, que no se encontraran recurriendo el año escolar, etc.), la muestra estudiada quizá resulte pequeña para lo que se requeriría en una investigación como la presente.

Por lo anterior, y dados los resultados obtenidos se sugiere profundizar en la investigación con una muestra mayor, debido a que "cuanto mayor sea el tamaño de la muestra más se acercarán los estadígrafos a los valores de la población" Kertlinger (1988).

Aunque el análisis de variabilidad de la muestra en estudio, demostró la homogeneidad de dicho grupo, para posteriores investigaciones es recomendable aplicar un cuestionario para ver comparabilidad de un grupo a otro y de esta forma estar completamente seguros de que los resultados no están afectados por características de diferencias individuales.

Otra limitación de la investigación se refiere al horario de aplicación de las pruebas a los diferentes sujetos, debido a que con el propósito de contar con su apoyo, se les aplicó la prueba de Wechsler en los horarios por ellos sugeridos, esto provocó para algunos sujetos se les aplicara al inicio de la jornada escolar y a otros casi al finalizar la misma; circunstancias que pueden influir en los resultados por la incursión de variables extrañas.

Por lo anterior, se sugiere que en nuevas investigaciones se controle este factor.

La estadística es la teoría y el método de analizar datos cuantitativos a partir de muestras de observación, a fin de estudiar y comparar las fuentes de varianza de los fenómenos, tomar decisiones sobre la aceptación o rechazo de relaciones hipotéticas entre los fenómenos y facilitar la obtención de inferencias fidedignas a partir de observaciones empíricas.

Por el momento, los resultados de esta investigación no permiten hacer ninguna inferencia estadística; tiene su importancia porque pretende despertar el interés de nuevas investigaciones.

BIBLIOGRAFIA

- Anastasi, A. Tests psicológicos. Ed. Aguilar, Madrid, 1978
- Bernstein, J. Test de matrices progresivas para la medida de la capacidad intelectual. Ed. Paidós, Argentina, 1988.
- Brody, E. and Brody, N. Intelligence: nature, determinants and consequences. Ed. Academic Press, New York, 1976.
- Brown, F. Principio de la medición en psicología y educación. Ed. El Manual Modero, México, 1989.
- Butcher, H. La inteligencia humana: su naturaleza y evolución. Ed. Marova, Madrid, 1979.
- Castelán, M. y Martínez, S. Normalización de la escala de inteligencia Wechsler para los niveles preescolar y primario (WPPSI) en una muestra de la Cd. de México.. Tesis UNAM, México, 1990.
- Cuevas, A. Análisis interno del WISC en México. Tesis de maestría, UNAM, México, 1993.
- Chávez, A. y Lara G. Correlación entre la escala de inteligencia de Wechsler (WISC) y el Test de Percepción Visual de Frostig. Tesis, UNAM, México, 1982.
- Franco, V. Contribución a dos pruebas de inteligencia: el Raven y el Barsit. Tesis, UNAM, México, 1969.
- Fuentes A. y López G. Adaptación y normalización de la prueba de Raven a la población del Consejo Tutelar de la Ciudad de México, Tesis UNAM, México, 1988.
- García, C. Validación de constructo y utilidad en el diagnóstico diferencial del K-ABC. Tesis de Doctorado, UNAM, México, 1987.
- García, C. y Terrazas, D. Evaluación de niños con problemas de aprendizaje de acuerdo con la categorización de Bannantyne del WISC-RM. Revista Mexicana de Psicología, Vol. 6, No. 2, 1988.
- Heredía, A. Poder de discriminación de las diferentes versiones del WISC. Tesis de Maestría, UNAM, México, 1993.
- Jiménez, O. Baremos de la prueba de J.C. Raven para enfermos psicóticos. Tesis UNAM, México, 1965.

- Jensen, A. Bias in mental testing. Great Britain. Methven E. Co., Ltd., 1980. First Edition.
- Kaufman, A. Psicometría razonada con el WISC-R. Ed. El Manual Moderno, México, 1979.
- Kerlinger, F. Investigación del comportamiento, técnica y metodología. Ed. Interamericana, México, 1985.
- Kreysig, E. Introducción a la estadística matemática. Principios y métodos. Ed. Limusa, México, 1991.
- Lemke, E. and Wiersma, W. Principles of psychological measurement. Chicago. Rand McNally College. Pub. Co., 1976.
- Magnusson, D. Teoría de los Test. Ed. Trillas, México, 1990.
- Matarazzo, J. Medida y valoración de la inteligencia del adulto. Salvat Editores, S.A., España, 1976.
- Morales, M. Psicometría aplicada. Ed. Trillas, México, 1990.
- Nunnally, J. Introducción a la medición psicológica. Ed. Paidós. Buenos Aires, 1970.
- Raven, J. Manual del test de matrices progresivas para la medida de la capacidad intelectual. Ed. Paidos, Argentina, 1988.
- Sattler, J. Evaluación de la inteligencia infantil. Ed. El Manual Moderno, México, 1988.
- Selltiz, C. Wrightsman, L. y Cook, S. Métodos de investigación en las relaciones sociales. Ed. Rialp, España, 1980.
- Siegel, S. Estadística no paramétrica. Ed. Trillas, México, 1990.
- Soberanis, Q. Estudio comparativo entre tres escalas de inteligencia: Terman-Merril, WISC y WPPSI. Tesis, UNAM, México, 1984.
- Sternberg, R. Inteligencia humana I. Sociedad, cultura e inteligencia. Ed. Paidós, Barcelona, 1988.
- Undheim, J. On intelligence II. A Neo-Spearman Model to replace Cattell's Theory of Fluid and crystallized intelligence. Scandinavian Journal of Psychology, 1981b, vol. 22, pp 181-187
- Vernon, P. Ability factors and enviromental influenes. American Psychologist, 1965. Vol. 20. pp 223-233.
- Wechsler, D. La medida de la inteligencia del adulto. Ed. Huascar, Argentina, 1973.

- Wechsler, D. WISC-R Escala de inteligencia revisada para el nivel escolar. Ed. El Manual Moderno, México, 1989.
- Wechsler, D. WISC-RM Escala de inteligencia revisada para el nivel escolar. Ed. El Manual Moderno, México, 1989.
- Zavala, C. Estandarización de las matrices progresivas de J.C. Raven en el Instituto Politécnico Nacional: Baremo para ingenieros, médicos, biólogos, contadores. Tesis de Doctorado. UNAM, México, 1958.