

320825

3
Lej.



UNIVERSIDAD DEL VALLE DE MEXICO
CAMPUS TLALPAN
ESCUELA DE PSICOLOGIA

CON ESTUDIOS INCORPORADOS A LA
UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

**REVALIDACION DE LA
ESCALA DE SINTOMAS
SYMPTOM RATING TEST (SRT)
EN ESTUDIANTES MEXICANOS
DE NIVEL MEDIO SUPERIOR**

T E S I S

QUE PRESENTA:
MARIA ELENA BARRIE GUILLEN
PARA OBTENER EL TITULO DE
LICENCIADA EN PSICOLOGIA

ASESORA DE TESIS: MAESTRA BEATRIZ VAZQUEZ ROMERO

MEXICO, D.F.

**TESIS CON
FALLA DE ORIGEN**

1998

259433



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.

A mi madre y hermanos

A mi amado compañero: Salvador

A mi hijo: Gerardo

A la memoria de mi hijo: Ernesto

A mis amigos: Marisa, Livia, Elena, Marisol,
Lidia, Pepe y Lolita

Universidad del Valle de México
Maestra Beatriz Vázquez Romero
Dra. Guillermina Natera
Maestro David Ayala

Gracias por su orientación, aportaciones, apoyo y paciencia.

Un agradecimiento especial a las autoridades y alumnos de la
Preparatoria de la Universidad del Valle de México.

Índice

	Pág.
Resumen	
Introducción	
1. Los tests o pruebas psicológicas. Orígenes y funciones	
1.1. Función básica y usos de las pruebas psicológicas	2
1.2. Las diferencias individuales en los orígenes de la psicología	4
1.3. Tipos frente a rasgos	4
1.4. La clasificación de los débiles mentales	6
1.5. El inicio de los métodos cuantitativos	7
1.6. Los primeros psicólogos experimentales	8
1.7. Las aportaciones de Francis Galton	8
1.8. Desarrollo del método estadístico	9
1.9. Los primeros "tests mentales"	10
1.10. Los tests de inteligencia	11
1.11. Los tests colectivos de inteligencia	14
1.12. Tests para el periodo infantil	15
1.13. Medición de las aptitudes	18
1.14. Medición de la personalidad	21
1.14.1. Inventarios autodescriptivos	22
1.14.2. Medidas de los intereses	24
1.14.3. Medidas de las actitudes	25
1.14.4. Técnicas proyectivas	26
2. La estructura de los tests o pruebas psicológicas. Teoría psicométrica	
2.1. ¿Qué es un test psicológico?	31
2.2. El concepto de medición	32
2.3. La medición en psicología	33
2.4. Las escalas de medición	35
2.4.1. Escalas nominales	36
2.4.2. Escalas ordinales	36
2.4.3. Escalas de intervalos o distancias iguales	37
2.4.4. Escalas con punto cero absoluto	37
2.5. La teoría del error de medición	38
2.6. Modelos de escalamiento	39
2.7. La construcción de una prueba psicológica	39
2.7.1. Finalidad de la prueba	41
2.7.2. Construcción de la prueba	43
2.8. Análisis técnico de la prueba	46
2.8.1. Estandarización de la prueba	47

	Pág.
2.8.2. Confiabilidad de la prueba	48
2.8.3. Validez de la prueba	55
3. La Escala de síntomas <i>Symptom Rating Test</i> (SRT)	
3.1. ¿Qué es un síntoma?	68
3.2. Construcción de la Escala de síntomas (SRT)	69
3.3. La neurosis	72
3.3.1. El concepto de neurosis	72
3.3.2. Tipos de neurosis	76
3.3.3. Criterios de diagnóstico de los trastornos neuróticos	77
3.4. Estudios de medición de la confiabilidad y validez de la Escala de síntomas (SRT)	83
3.4.1. Confiabilidad (versión semanal)	84
3.4.2. Validez de la lista de control de síntomas (versión semanal)	85
3.4.3. Validez de la SRT (forma semanal) y correlaciones con otras pruebas	85
3.4.4. Estudios con otras formas	89
3.5. Estudios posteriores de medición de la confiabilidad y validez de la SRT	90
3.5.1. Validación y confiabilidad de la SRT y la Escala L-22	90
3.5.2. Validación y confiabilidad de la traducción italiana de la SRT	97
3.5.3. Validación y confiabilidad de la SRT en México	100
4. Metodología	
4.1. Problema de investigación	108
4.2. Planteamiento de objetivos	108
4.2.1. Objetivo general	108
4.2.2. Objetivos específicos	108
4.3. Planteamiento de hipótesis	108
4.4. Variables	109
4.4.1. Variable independiente	109
4.4.2. Variable dependiente	109
4.5. Definición de variables	109
4.5.1. Definición conceptual de la variable independiente	109
4.5.2. Definición operacional de la variable independiente	110
4.5.3. Definición conceptual de la variable dependiente	110
4.5.4. Definición operacional de la variable dependiente	110
4.6. Tipo de estudio	110
4.7. Definición de los sujetos	111
4.8. Muestreo	111
4.9. Instrumento de medición	112
4.9.1. Descripción del instrumento	112
4.9.2. Cuestionario sociodemográfico	113
4.10. Ambiente de investigación	113
4.11. Procedimiento	113
4.12. Análisis estadístico	114

5. Resultados

5.1. Análisis estadístico de las variables sociodemográficas y de las respuestas a los reactivos	116
5.1.1. Sexo	116
5.1.2. Edad	117
5.1.3. Grado escolar	117
5.1.4. Ocupación	118
5.1.5. Número de personas que forman la familia	118
5.1.6. Lugar numérico que se ocupa en la familia	119
5.1.7. Promedio escolar	119
5.1.8. Becados	119
5.1.9. Respuestas a los reactivos	120
5.2. Validez	120
5.3. Confiabilidad	121
5.4. Varianza	121
5.5. Totales	121

6. Discusión y conclusiones

Referencias

Anexo A

Índice de cuadros y figuras

	Pág.
Cuadro 1. Niveles de medición de Stevens, operaciones definitivas básicas, transformaciones permisibles, ejemplos de estadísticas permisibles y ejemplos	36
Cuadro 2. Etapas en la construcción de una prueba	40
Figura 1. Curva de distribución normal	45
Figura 2. Distribución desviada (a la izquierda)	46
Figura 3. Distribución desviada (a la derecha)	46
Cuadro 3. Reactivos de las subescalas originales	71
Cuadro 4. Versión corta de la Escala de síntomas (SRT)	73
Cuadro 5. Medias y desviaciones estándar de los puntajes de la SRT (forma semanal, papel y lápiz) obtenidos, principalmente, por dos grupos de nuevos pacientes externos, neuróticos, en dos clínicas, una en Houston, Texas, y otra en Albuquerque, Nuevo México, y los puntajes de dos grupos de normales	87
Cuadro 6. Rangos porcentuales de sensibilidad, especificidad y mal clasificados en puntajes de la SRT (forma semanal, papel y lápiz) en dos estudios y los rangos porcentuales de los puntajes del MAS y el EPI	87
Cuadro 7. Puntajes de varios grupos en dos escalas de síntomas	93
Cuadro 8. Análisis factorial de la SRT	94
Cuadro 9. Análisis factorial de la Escala L-22	95
Cuadro 10. Medias, confiabilidades e intercorrelaciones de la Escala L-22, la SRT y las subescalas de la SRT	96
Cuadro 11. Características sociodemográficas de la muestra	101
Cuadro 12. Resultados del análisis factorial de la SRT en el estudio de México	104
Cuadro 13. Subescalas de síntomas psicológicos y físicos de la SRT	105
Cuadro 14. Porcentajes de sexo por edad y ocupación	116

	Pág.
Cuadro 15. Porcentajes, media, desviación estándar y mediana de edad por sexo y ocupación	117
Cuadro 16. Porcentajes, media, desviación estándar y mediana de grado escolar por sexo, edad y ocupación	117
Cuadro 17. Porcentajes de ocupación por sexo, edad y ocupación	118
Cuadro 18. Porcentajes, media, desviación estándar y mediana de número de integrantes de la familia por sexo, edad y ocupación	118
Cuadro 19. Porcentajes de lugar numérico que se ocupa en la familia por sexo, edad y ocupación	119
Cuadro 20. Porcentajes, media, desviación estándar y mediana de promedio escolar por sexo, edad y ocupación	119
Cuadro 21. Porcentajes de becados por sexo, edad y ocupación	119
Cuadro 22. Porcentajes, media, desviación estándar y mediana de los puntajes de la SRT	120
Cuadro 23. Valores Eigen de los factores extraídos de la rotación varimax	120
Cuadro 24. Valores del coeficiente alpha de Cronbach de los factores obtenidos	121
Cuadro 25. Porcentaje de varianza que explica cada factor	121
Cuadro 26. Factores, cargas factoriales de cada reactivo, valores Eigen y el coeficiente alpha de Cronbach de cada factor	122
Cuadro 27. Media, mediana y desviación estándar de los puntajes de la SRT, totales y por sexo, de cada factor y la SRT completa, donde “nunca” corresponde a 1, “algunas veces” a 2 y “a menudo” a 3	123

Resumen

Con el propósito de analizar el comportamiento psicométrico de la Escala de síntomas *Symptom Rating Test* (SRT) se aplicó a 456 sujetos estudiantes de la Preparatoria de la Universidad del Valle de México para observar si se obtenía el mismo número y tipo de factores, extraídos del análisis factorial, en comparación con la validación de la misma Escala que se llevó a cabo en 1994, en México, por Mora, Natera y Andrade-Palos, con una población de estudiantes que asistía a una escuela pública. Asimismo, se compararon los resultados del Coeficiente Alpha de Cronbach obtenidos en ambos estudios con el objetivo de observar si aparecían los mismos valores para cada factor y la escala total.

Se obtuvieron seis factores extraídos del análisis factorial: somatización, ansiedad, inadecuación, depresión, alteraciones cardiacas y alteraciones del sueño, aunque sólo los dos primeros -somatización y ansiedad- con un valor Eigen mayor a uno, por lo que son las áreas que mejor definen a la escala y que explican el 29% de la varianza. Estos resultados fueron diferentes a los obtenidos en 1994, donde del análisis factorial resultaron cuatro factores con un valor Eigen mayor a uno.

Asimismo, los coeficientes alpha de Cronbach fueron diferentes en ambos estudios, sin embargo, en esta investigación la escala total tuvo una confiabilidad de .90. Los datos obtenidos permiten confirmar a la SRT como una escala válida y confiable.

Se realizó un ajuste de la SRT a dos factores -síntomas somáticos y síntomas psicológicos- con lo cual aumentó su confiabilidad y coincide con los resultados obtenidos en esta investigación donde dos factores fueron los válidos, por lo que se sugiere su uso y calificación en esta forma.

Introducción

Uno de los problemas fundamentales que ha enfrentado la psicología como ciencia desde su creación en el siglo XIX es el de contar con sistemas de medición objetivos de las variables psicológicas.

El problema de la medición en psicología es considerablemente más complejo que en los campos donde se emplean instrumentos físicos comunes de medición, dado que los fenómenos humanos a los que dirige su interés son de múltiple determinación. (Nunnally en Bernstein, 1995).

Las pruebas psicológicas (tests) que son en esencia medidas objetivas y tipificadas de una muestra de conducta (Anastasi, 1967) son los instrumentos de medición que los psicólogos han creado para obtener datos confiables en las diversas áreas hacia las cuales enfocan sus investigaciones, como son: la educación, la detección de patología, la psicoterapia, la personalidad, las relaciones sociales, etc.

Por tanto, para la investigación es importante la construcción y validación de pruebas psicológicas. En México, como en toda Latinoamérica existe la necesidad de contar con escalas de medición válidas y confiables que detecten psicopatología moderada por medio de encuestas comunitarias (Mora, Natera y Andrade-Palos, 1994), que ayuden a identificar de una manera rápida y eficiente la posible relación entre síntomas que manifiestan malestar psicológico con otros trastornos y conductas como son el abuso de sustancias, el pobre desempeño escolar y la delincuencia, entre otros. Este tipo de escalas pueden utilizarse, también, como filtros para identificar casos con un número importante de síntomas psicológicos con el fin de atender las necesidades de salud de las poblaciones abordadas.

De lo anterior se desprende el interés de esta investigación, cuyo objetivo fue comprobar los resultados de la validación en México de la Escala de síntomas *Symptom Rating Test* (SRT) en estudiantes de enseñanza media superior (Mora, Natera y Andrade-Palos, 1994), a fin de revalidar su aplicabilidad en nuestra cultura, o en su caso, contribuir a mostrar sus límites. La revalidación de la escala se llevó a cabo a través de su aplicación a estudiantes de bachillerato que asisten a una institución escolar privada. Los resultados obtenidos se compararon con los que se encontraron en la investigación realizada en 1994, para validar la SRT, con una población diferente, dado que asistía a una escuela pública.

La Escala de síntomas (SRT) permite explorar brevemente la presencia de rasgos psicopatológicos, no sólo en la práctica clínica sino también entre la población general. Este cuestionario de síntomas fue diseñado originalmente para medir el malestar psicológico. El malestar se define como un estado transitorio y cambiante. Según Kellner, creador de la escala, la función del SRT es análoga a un termómetro clínico, ya que permite detectar los cambios en el estado clínico de un paciente (Kellner y Sheffield, 1973). En una investigación posterior, Cochrane (1980) propuso localizar las propiedades psicométricas de esta escala con el fin de usarla para detectar psicopatología en población comunitaria, y en México, en 1994, Mora, Natera y Andrade-Palos realizaron un estudio para localizar su validez y confiabilidad como instrumento de medición de síntomas en población juvenil.

La presente investigación se divide en seis capítulos.

En el capítulo uno se abordan los orígenes y funciones de los tests, las razones por las que sus creadores los construyeron, sus funciones y los estudios de validación y confiabilidad que han experimentado incluyendo las que se han hecho en México.

En el capítulo dos se analiza la estructura psicométrica de los tests, comenzando por la definición de la medición en psicología, las escalas de medición, los pasos en la construcción de una prueba psicológica hasta su análisis técnico.

En el capítulo tres se recorren las etapas de construcción de la Escala de síntomas (SRT), los estudios de confiabilidad y validación que ha tenido desde su creación hasta el estudio en México en 1994, así como una somera revisión del concepto de síntoma neurótico.

El capítulo cuatro enfoca la metodología usada en este estudio.

En los capítulos cinco y seis se exponen y discuten los resultados así como las conclusiones que se derivan de la presente investigación. Finalmente se presentan referencias y anexos.

Capítulo 1
Los tests o pruebas psicológicas
Orígenes y funciones

1. Los tests o pruebas psicológicas Orígenes y funciones

1.1. Función básica y usos de las pruebas psicológicas.

La singularidad de los individuos es una de las cuestiones que más interrogantes ha planteado a los hombres de todas las épocas. Este hecho, fundamental y característico de la vida, se ha constituido en un vasto campo del conocimiento humano, pues implica describir, definir y explicar las diferencias entre los seres que pueblan nuestro mundo, ya sean plantas, animales o seres humanos, con el fin de comprender y explicar lo que constituye su singularidad. La filosofía, la ciencia, la técnica y las artes lo han abordado desde múltiples vías, dependiendo de los objetivos que guían a cada una.

El problema de las variaciones individuales se consideraba ya importante en el mundo antiguo, cuya población _en el comienzo de la era cristiana_ no sobrepasaba los 250 millones de personas. Pero la importancia de este fenómeno es aún mayor hoy en día, cuando la población mundial supera con mucho los 5000 millones de habitantes. En todos los períodos de la historia humana el hombre ha observado y descrito las diferencias entre los individuos. Los educadores, los políticos y los gobernantes, por sus mismas funciones, han instaurado métodos que les permiten organizar o sistematizar la complejidad de los hombres. Para la psicología éste ha sido uno de los problemas capitales que ha guiado su acción, al grado de que se puede afirmar que la historia de su desarrollo como ciencia está marcado por la forma en que ha respondido a las cuestiones acerca del conocimiento de los sujetos y la medición de los atributos psicológicos que los diferencian, con el fin de enfrentar los efectos que estos atributos producen.

Para la ciencia conocer un fenómeno significa, en un primer nivel, el poder describirlo de una forma precisa para, en un segundo nivel de conocimiento, poder explicar sus causas y sus efectos, o sea construir teorías (Nunnally en Bernstein, 1995). Un fenómeno se puede describir cuantitativa o cualitativamente. Las descripciones cuantitativas son una de las exigencias de la ciencia porque implican precisión, objetividad y la posibilidad de la comprobación de dicha descripción.

El objetivo de conocer la singularidad de los sujetos humanos partiendo de la descripción de sus diferencias y similitudes dio como resultado la creación de una rama de la psicología,

la psicología diferencial, además de la invención de variados métodos e instrumentos que le permiten llevar a cabo este objetivo.

Entre los instrumentos de medición que la psicología ha creado se destacan los tests o pruebas psicológicas cuya función básica “consiste en medir las diferencias entre los individuos o entre las reacciones del mismo individuo en distintas ocasiones”. (Anastasi, 1967), y que según el criterio de Brown (1980) representan el método más desarrollado para medir las características psicológicas de los individuos, en relación a otros métodos como serían la observación, las escalas de estimación o las medidas de ejecución.

En la definición de Anastasi (1967) acerca de la función básica de los tests aparece como clave el concepto de medición, con lo que se alude a un método cuantitativo, el método de la psicología científica, para la cual uno de sus mayores problemas es el de saber cómo descubrir la naturaleza y la amplitud de las características psicológicas individuales en forma precisa y exacta. Este problema surge a partir de que la psicología se planteó como meta el conocimiento objetivo de las características humanas, lo cual constituyó una de sus conquistas, pues hace sólo un siglo era impensable cuantificar algún aspecto de la vida mental humana.

En la actualidad, a partir de esa conquista, los tests han tenido un amplio desarrollo y son utilizados en múltiples áreas de nuestra vida. Uno de los primeros problemas que estimularon la creación de los tests psicológicos, y hoy en día es una de sus aplicaciones importantes, fue el de la identificación de los débiles mentales y su diferenciación de los alienados, además de que en la psicología clínica son ampliamente usados para el diagnóstico y el pronóstico de las personas con problemas emocionales, los delincuentes, los psicóticos, en síntesis, para la descripción de la personalidad.

Los problemas educativos también impulsaron el desarrollo de los tests. La clasificación de los estudiantes en relación a su aptitud escolar, la identificación de los intelectualmente retrasados o los superdotados, el diagnóstico de los fracasos escolares, la selección de los aspirantes a escuelas profesionales, son algunos de los usos en el área de la educación.

En la industria y en el ejército, la selección y clasificación del personal constituyen otra de las funciones importantes de los tests. Asimismo, se les utiliza para el conocimiento de las actitudes de las poblaciones acerca de problemas tan variados como son la opinión acerca de los gobernantes, de modificaciones o reformas políticas y para el conocimiento de la salud

física o psicológica de los pueblos. Es pues evidente que en la actualidad los tests se emplean en la solución de una enorme variedad de problemas prácticos, por lo que se ha constituido en un campo esencial de la investigación psicológica.

Las conquistas obtenidas en este terreno se deben al esfuerzo de múltiples investigadores que a lo largo de la historia de la psicología se han dado a la tarea de resolver el problema del conocimiento de las afinidades y las diferencias individuales de los sujetos humanos, aplicando para ello los descubrimientos que se iban obteniendo en otras áreas del conocimiento humano. Al hacer un recorrido histórico se puede presenciar un interés por las diferencias individuales ya presente desde los filósofos griegos.

1.2. Las diferencias individuales en los orígenes de la psicología.

Anastasi (1977) considera que uno de los primeros ejemplos del reconocimiento explícito de las diferencias individuales se encuentra en la *República*, de Platón; obra donde este filósofo griego plantea que uno de los objetivos fundamentales del Estado ideal es la asignación de tareas a los individuos que se encuentran dotados para realizarlas.

Aristóteles, en sus múltiples obras, también consideró las diferencias de grupo, incluyendo las diferencias de especie, racial, social y de sexo, en cuanto a sus características mentales y morales. A través de las varias *Éticas* aparecen pasajes que suponen la variación individual. Su inclusión de las diferencias individuales se descubre también en sus descripciones de los hombres a partir de la posesión o no de ciertos rasgos como la irascibilidad, la audacia, la impudicia, etc.

Santo Tomás de Aquino y San Agustín en la Edad Media también tomaron en cuenta las diferencias individuales al proponer la psicología de las facultades: la memoria, la imaginación, la voluntad.

1.3. Tipos frente a rasgos.

Tyler (1973), ubica como una de las primeras estrategias que se les ocurrieron a los hombres, en su interés por imponer un orden al complicado problema de las diferencias individuales, la de clasificar a la gente por tipos. En el siglo V a. C., Hipócrates clasificó a las personas en dos tipos basados en la estructura del cuerpo: el *habitus apoplecticus* y el *habitus phthisicus*. Galeno, médico griego, propuso una cuádruple clasificación basada en la

predominancia de los diferentes “humores” o fluidos del cuerpo. Sus cuatro tipos eran el sanguíneo, el colérico, el flemático y el melancólico. Variantes de estas clasificaciones se mantienen hasta nuestros días.

Spranger (1928) en su libro *Types of men*, analizó los valores humanos que se presentan como básicos en la vida, y llegó a la conclusión de que había seis tipos de valores fundamentales: el teórico, el económico, el estético, el social, el político y el religioso. Jaensch (1938) en su obra *Der Gegentypus*, basando su sistema en observaciones de la imagen eidética, describió lo que llamó tipo-T, o persona integrada, opuesta al tipo-B, o persona desintegrada. Kretschmer (1925) en la obra *Constitución y carácter* construyó un sistema con las dos principales variedades de psicosis: los estados esquizofrénicos y los maniaco-depresivos. Jung (1923) en el texto llamado *Psychological types* clasificó a los sujetos en extrovertidos e introvertidos.

La otra estrategia importante para imponer orden sobre el campo de las diferencias individuales fue estudiar los rasgos aisladamente y en combinación con otros. En vez de adscribir cada individuo a su clase, como hacen los tipólogos, se dirige la atención a las cualidades o rasgos que se manifiestan en su conducta. Los conceptos de rasgo han dominado la psicología diferencial americana durante la primera mitad del siglo XX, y continúan siendo un campo importante en la investigación de los tests.

Brown (1980) señala que el concepto de rasgo se puede definir como un conjunto de conductas interrelacionadas o intercorrelacionadas, o sea, que un rasgo es un término que describe un grupo de conductas que tienden a producirse juntas. Se trata de una abstracción más que de una realidad tangible y objetiva. El procedimiento típico para identificar los rasgos es hacer que un grupo de personas respondan a un gran número de reactivos, analizando a continuación, estadísticamente, los datos, para ver cuáles son los reactivos que se agrupan. Con el fin de que resulten útiles como construcciones descriptivas, los rasgos tienen que ser relativamente estables en el tiempo y en diferentes situaciones. Aunque la mayoría de los rasgos cognoscitivos de las habilidades son muy estables, las características de la personalidad tienden a fluctuar de manera más amplia.

Debido a la estabilidad relativa de la mayoría de los rasgos algunos psicólogos han sostenido que se determinan genéticamente como Burt (1972) en el artículo “Inheritance of general intelligence” en la revista *American Psychologist*, y Jensen (1974) en el libro *The*

effect of race of examine on the mental test scores of white and black pupils. Un punto de vista alternativo es el de Ferguson (1954) que en su artículo “ On learning and human ability” en la revista *Canadian Journal of Psychology*, concibe a 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.

1.4. La clasificación de los débiles mentales.

Anastasi (1967) por su parte señala que en el siglo XIX, el interés por el adecuado tratamiento de los débiles mentales y los alienados llevó a la creación de criterios uniformes para la identificación y clasificación de estos casos. En primer lugar se hacía necesario distinguir entre alienado y débil mental. Los primeros manifestaban trastornos emocionales que podían ir o no acompañados de deterioro intelectual desde un nivel inicial normal, los segundos se caracterizaban esencialmente por un defecto intelectual ya presente desde el nacimiento o la primera infancia. La primera afirmación explícita de esta diferencia se halla en una obra de dos volúmenes, *Des maladies mentales considérées sous les rapports médical, hygiénique, et médico-légal*, publicada en 1838 por el médico francés Esquirol, en la que indicó que existen muchos grados de debilidad mental, los cuales varían, a lo largo de un continuo, desde la normalidad hasta el grado más agudo de la idiocia. Intentó varios procedimientos, en su esfuerzo por constituir algún sistema para la clasificación de los diferentes grados y variedades de debilidad mental, pero llegó a la conclusión de que el mejor criterio para establecer el nivel intelectual del individuo lo constituye su uso del lenguaje. Partiendo de este criterio estableció dos grados de imbecilidad y tres de idiocia. Sostenía que en el primer grado de imbecilidad, el más cercano a la normalidad, se emplea el discurso con facilidad y en el segundo grado el discurso es más difícil y el vocabulario más limitado. Los idiotas del grado superior se valen solamente de unas pocas palabras o de frases muy cortas, en el segundo nivel son capaces de emitir monosílabos y gritos, y el idiota de nivel inferior carece completamente del lenguaje. En la actualidad, un criterio para el diagnóstico de la debilidad mental es el lingüístico y los tests de inteligencia consideran el desempeño verbal como una parte esencial para determinarla, por lo que la obra de Esquirol es considerada como un importante precursor de este tipo de tests.

En la educación de los débiles mentales, las contribuciones de Seguin, poseen particular

importancia. Seguin, otro médico francés, en su obra *Idiocy: its treatment by the psychological method*, de 1866, rechazando la noción de incurable aplicada a los débiles mentales estableció “el método fisiológico” de educación. En 1837 creó la primera escuela dedicada a la educación de los niños mentalmente deficientes. En 1848 emigró a los Estados Unidos, donde sus ideas tuvieron gran aceptación y son el origen de la mayoría de las técnicas para el adiestramiento de los sentidos y el adiestramiento fisioterápico que se aplican en las instituciones para débiles mentales. Siguiendo estos métodos, en la actualidad, se dan ejercicios intensivos a los deficientes mentales más graves sobre discriminación sensorial y control motor. Algunos de estos procedimientos creados por Seguin se han incorporado a las subescalas de ejecución o no verbales de los tests de inteligencia; como el Test de ajuste de formas de Seguin, en el que se le pide al sujeto que inserte tan rápidamente como pueda piezas de diferentes formas en sus respectivos lugares.

1.5. El inicio de los métodos cuantitativos.

Uno de los aspectos más sobresalientes de la labor que los psicólogos han realizado en torno a las diferencias individuales es la importancia concedida a los métodos cuantitativos. El descubrimiento de que se podía representar numéricamente el comportamiento de las funciones del sistema nervioso fue hecho casualmente, y por un astrónomo, no por un psicólogo. Tyler (1973) relata que Bessel, hojeando en 1816 una historia del Observatorio Astronómico de Greenwich, quedó sorprendido por un incidente registrado allí. Un joven auxiliar había sido despedido en 1796 por anotar continuamente el tiempo del paso aparente de las estrellas a través de una línea en el campo de visión del telescopio, aproximadamente con un segundo de desfase respecto al tiempo anotado por su maestro. Bessel se preguntaba por qué el joven había sido tan lento, sabiendo que su puesto estaba en peligro a causa de ello. Empezó a poner a prueba a sus colegas astrónomos, y descubrió que había una variación considerable entre los individuos en la rapidez con que reaccionaban ante un estímulo visual. A esta variación la bautizó con el nombre de ecuación personal. La importancia de esto para la psicología consistió en que se había llegado a la demostración de que al menos una característica mental podía ser medida.

1.6. Los primeros psicólogos experimentales.

Anastasi (1967) dice que los primeros psicólogos experimentales del siglo XIX estuvieron más interesados en descubrir las leyes generales de la naturaleza humana, válidas para todos, que en explorar las diferencias entre individuos. Su objetivo fue formular descripciones generalizadas de la conducta humana, por tanto, su interés se centraba en las afinidades, más que las diferencias; cuando éstas sucedían en el transcurso de la observación de dos sujetos sometidos a condiciones idénticas se les consideraba una forma de error. La presencia de tal error, o variabilidad individual, hacía que las generalizaciones fueran sólo aproximadas. Esta concepción prevaleció en los laboratorios como el de Wundt, en Leipzig, en 1879, donde se formaron la mayoría de los primeros psicólogos experimentales. Asimismo, la elección de sus temas de trabajo reflejaban su formación original como médicos o fisiólogos: estudiaron ampliamente la sensibilidad a los estímulos visuales, auditivos y de otros sentidos y los tiempos de reacción, exigiéndose un control riguroso de las condiciones en que se hacían las observaciones; con lo que se demostró la importancia de hacer observaciones sobre todos los sujetos bajo condiciones normalizadas. Esta normalización de los procedimientos es la principal aportación a la teoría de los tests, por lo que hoy constituye una de las características especiales de los tests psicológicos.

1.7. Las aportaciones de Francis Galton.

Sir Francis Galton, explorador, meteorólogo y biólogo, uno de los más eminentes seguidores de Darwin, fue el primero que intentó aplicar los principios evolucionistas de la variación, selección y adaptación al estudio de los individuos humanos, de la misma manera, fue el principal promotor del movimiento a favor de los tests.

Las investigaciones de Galton fueron muchas y muy variadas aunque tenían un eje: la herencia. En 1869 publicó un libro titulado *Hereditary Genius*, en el cual, mediante la aplicación del ahora famoso método de la historia familiar, trató de demostrar la herencia de talentos específicos en varios campos. En su interés por la investigación de la herencia humana, comprendió la necesidad de medir las características de las personas emparentadas y no emparentadas en grandes números, a fin de descubrir los grados de parecido que entre ellos existan. Con este fin ideó numerosos *tests* y medidas, y en 1882 estableció un laboratorio antropométrico en el South Kensington Museum, de Londres, donde, relata

Anastasi (1967), mediante el pago de una pequeña cantidad, cualquier individuo podía conocer algunos de sus rasgos físicos y pasar por tests de agudeza visual y auditiva, energía muscular, tiempo de reacción y otras funciones psicomotoras. El mismo ideó la mayoría de las pruebas que se aplicaban en su laboratorio, muchas de las cuales se utilizan hoy en día, en su forma original o modificadas, como son la *barra de Galton* que sirve para medir la discriminación visual de la longitud, y el *silbato de Galton*, para la determinación del tono más agudo que puede percibirse; también las series graduadas de pesos para medir la discriminación cinestésica, y las pruebas dinamométricas de velocidad de reacción.

Pensaba que las pruebas de discriminación le servirían para medir la inteligencia, idea que tomó de Locke. En su obra: *Inquiries into human faculty and its development* de 1883, 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”. Por tanto, afirmó que los idiotas tienden a ser deficientes en su capacidad de discriminar el calor, el frío y el dolor, con lo que confirmaba su tesis de que la capacidad discriminativa “sería, en conjunto, mayor entre los mejor dotados intelectualmente”, enunciado que se encuentra en la misma obra.

1.8. Desarrollo del método estadístico.

Costa Neiva (1996) menciona que otra contribución de Galton es el desarrollo que impuso a los métodos estadísticos para los análisis de datos de las diferencias individuales. Adaptó y seleccionó ciertas técnicas que habían sido creadas por matemáticos como Quetelet, matemático belga, que fue el primero en descubrir que la teoría matemática de la probabilidad podría ser aplicada a las mediciones humanas. Entre los principales problemas o análisis estadísticos por los que Galton se interesó destacan el de la curva de distribución normal y el de la correlación. Sobre este último trabajó intensamente y desarrolló un índice que llegó a conocerse con el nombre de coeficiente de correlación. Sin embargo, relata Anastasi (1967) que fue uno de sus alumnos, Karl Pearson, quien más tarde resolvió los detalles matemáticos de la teoría de la correlación. Pearson fue también autor del desarrollo y sistematización de lo que, hasta hace poco, constituía casi todo el campo de la estadística.

Otro estadístico británico cuyas contribuciones afectaron de forma significativa el curso

de la estadística, fue R. A. Fischer, quien, trabajando dentro del área de la investigación agrícola, dedujo nuevas técnicas estadísticas que han demostrado su utilidad en muchos otros campos, incluyendo a la psicología, y que han abierto nuevas y vastas posibilidades para el análisis de datos. Su nombre se ha asociado con el análisis de las varianzas, técnica que permite la investigación simultánea de los efectos de varias variables en un solo experimento. (Anastasi, 1967).

1.9. Los primeros “tests mentales”.

Uno de los primeros alumnos del laboratorio del psicólogo alemán Wundt se interesó por las diferencias entre los sujetos, que aparecían en los estudios que se realizaban en ese laboratorio y por la posible significación de esas diferencias. Era un norteamericano: James McKeen Cattell.

Anastasi (1967) subraya que para su doctorado en Leipzig, concluyó una disertación sobre las diferencias individuales a pesar de la resistencia de Wundt a este tipo de investigación. Su contacto con Galton, en 1888, durante su estancia en Cambridge, reforzó su interés en la medición de las diferencias individuales. A su regreso a Norteamérica, creó laboratorios de psicología experimental y difundió ampliamente los tests. En su artículo de 1890, “Mental tests and measurements” aparecido en la revista *Mind*, se empleó por primera vez el nombre “test mental” en la literatura psicológica. En este artículo describía una serie de tests que se aplicaban anualmente a los estudiantes universitarios con el fin de establecer su nivel intelectual. Cattell compartía la convicción de Galton de que era posible obtener una medida de las funciones intelectuales mediante tests de discriminación sensorial y tiempo de reacción. El entusiasmo de Cattell se fomentaba porque la medición de funciones simples puede realizarse con precisión y exactitud, y porque pensar en medir objetivamente funciones más complejas, en ese momento aparecía como imposible.

Aun así, dice Tyler (1973), en Europa aparecieron investigadores que crearon tests para medir funciones más complejas, como Kraepelin que en su artículo “Der psychologische Versuch in der Psychiatrie”, de la revista *J. Psycho-Astenics, Monogr. Suppl.*, de 1912, presentó una larga serie de tests para medir lo que él consideraba factores básicos en la caracterización de un individuo, lo que le era muy útil dado su interés en el examen clínico de pacientes psiquiátricos. Los tests estaban destinados a medir los efectos prácticos, la

memoria y la susceptibilidad a la fatiga y a la distracción, a través de la resolución de operaciones aritméticas simples.

Otro psicólogo alemán, Ebbinghaus, aplicó a escolares tres tests: de cálculo aritmético, de memoria, y de completamiento de frases. El más complejo de los tres, el de completamiento de frases, fue el que mostró una clara correspondencia con el rendimiento escolar de los niños.

En este panorama, Binet y Henry critican en su artículo “La psychologie individuelle” aparecido en 1895, en la revista *Année psychol.*, a la mayoría de los tests existentes por ser medidas sensoriales y concentrarse indebidamente en aptitudes especializadas sencillas. Sostenían que al medir funciones más complejas no era necesaria una gran precisión dado que las diferencias individuales son mayores en ellas y propusieron una lista de tests que abarcaban funciones como la memoria, la imaginación, la atención, la comprensión, la sugestibilidad, la apreciación estética, que serían las tendencias que se expresarían en las escalas de inteligencia de Binet.

1.10. Los tests de inteligencia.

Morales (1993) dice que el nombre más célebre en torno al tema de los tests de inteligencia es el de Alfred Binet, quien, en 1905, con la colaboración de Teodore Simon, en Francia, elaboró la primera escala de inteligencia, la Escala Binet-Simon. En el curso de su trabajo puso en práctica criterios completamente distintos de los de sus colegas al mantener que la suma de las aptitudes elementales no expresa la inteligencia, por lo que en esta escala se presentan 30 problemas o tests destinados a cubrir una amplia variedad de funciones, haciendo hincapié en el juicio, la comprensión y el razonamiento que Binet consideraba componentes esenciales de la inteligencia.

En 1908 crearon una segunda escala, a partir de la primera, donde se agruparon todos los tests en niveles de edad. Así, en el nivel de los tres años se colocaron todos aquellos que podían resolver los niños normales de tres años, en el de cuatro los que podían resolver niños de cuatro años, y así sucesivamente, hasta la edad de trece años. La puntuación del niño en el test podía, entonces, expresarse como una edad mental, es decir, la edad de los niños normales cuya ejecución igualaba. En 1911, apareció una tercera revisión, que se extendió hasta el nivel adulto. Esta escala atrajo tanto la atención de los psicólogos de todo el mundo,

por su uso relativamente sencillo, que se realizaron traducciones y adaptaciones en muchos idiomas.

En 1916, Terman, de la Universidad de Stanford, revisa la escala e introduce el cociente de inteligencia (CI) que es un concepto que enriquece el de edad mental y que consiste en dividir la edad mental obtenida por la edad cronológica real con el fin de obtener un índice de la razón a la que se ajusta el desarrollo mental en un individuo dado. Esta escala lleva el nombre de Stanford-Binet. Esta revisión introdujo tantos cambios que representó virtualmente un nuevo test. Más de un tercio de los elementos eran nuevos, y algunos elementos descartados. La escala entera fue tipificada nuevamente sobre una muestra americana de un millar de niños y cuatrocientos adultos aproximadamente.

La segunda revisión del Stanford, aparecida en 1937, constaba de dos formas equivalentes, la L y la M. En esta revisión se extendió la escala de manera considerable, y se volvió a tipificar sobre una nueva muestra cuidadosamente escogida de la población americana.

La tercera revisión, realizada en 1960, presentaba una única forma LM, que incluía los mejores reactivos de las formas L y M publicadas anteriormente. Las modificaciones realizadas consistieron en la eliminación de algunos reactivos y reubicación de otros, proporcionando una nueva distribución de éstos en los diferentes niveles de edad. (Terman y Merrill, 1972; citados por Costa Neiva, 1996).

En 1972 se llevó a cabo una reestandarización sistemática de la forma L-M. El material se mantuvo prácticamente el mismo, fueron cambiados el diseño de una tarjeta (muñeca) y una palabra de la subprueba Semejanzas. Se utilizó una muestra representativa de la población norteamericana de cerca de 100 sujetos para cada nivel de edad. (Costa Neiva, 1996).

Morales (1993) señala que esta escala tiene un amplio uso en México, aunque no ha sido estandarizada todavía.

Otras dos escalas de similar importancia a las de Binet, son las escalas creadas por David Wechsler, una para adultos y otra para niños. Aunque se aplican como tests individuales y se destinan a muchos de los mismos usos que el Stanford-Binet, difieren de éste en la inclusión de subtests verbales y de ejecución, a partir de los cuales se calculan los CI verbales y de ejecución por separado.

La primera forma de las escalas de Wechsler, conocida como la Escala de inteligencia de Wechsler-Bellevue, se publicó en 1939. Uno de los objetivos primarios en su preparación fue

proporcionar un test de inteligencia apropiado para adultos, pues Wechsler consideraba que las existentes sólo eran adaptaciones de las escalas para escolares. Esta escala se revisó transformándose en la Wechsler Adult Intelligence Scale (Escala de inteligencia para adultos de Wechsler, WAIS).

El WAIS, publicado en 1955, comprende once subtests. Seis de ellos se agrupan en una escala verbal: información, comprensión, aritmética, semejanzas, memoria inmediata de dígitos y vocabulario; los otros cinco se agrupan en una escala de ejecución: clave, completamiento de figuras, dibujo con cubos, disposición de cuadros y ensamblaje de objetos. (Anastasi, 1967).

Costa Neiva (1996) menciona que en 1981 el WAIS fue objeto de una revisión y pasó a llamarse WAIS-R. El objetivo de esta última revisión fue actualizar algunos contenidos y proveer nuevas normas basadas en las puntuaciones obtenidas por muestras más actuales de población. Asimismo, dice que el WAIS ha sido traducido y adaptado a varios idiomas. La versión para el idioma español se llama WAIS-español y fue presentada en 1981. En México se utiliza esta versión pero todavía no existen normas estandarizadas para la población mexicana. Según Morales (1993) desde 1964 vienen siendo realizados los primeros pasos para la estandarización de la escala, bajo la dirección del doctor Rogelio Díaz Guerrero.

La escala de inteligencia Wechsler para niños (WISC) se publicó en 1949. Se preparó como una extensión hacia abajo del Wechsler-Bellevue. Muchos de sus elementos se tomaron de esta escala, añadiéndose a cada test elementos más sencillos del mismo tipo. La WISC consta de doce subtests, dos de los cuales han de usarse como tests alternativos o de complemento si el tiempo lo permite. Como en las otras escalas de Wechsler, los subtests se agrupan en una escala verbal: información general, comprensión general, aritmética, semejanzas, vocabulario y el alternativo, memoria inmediata de dígitos; y en una escala de ejecución: completamiento de cuadros, disposición de imágenes, dibujos de Kohs, ensamblaje de objetos, claves y el alternativo, laberintos.

En 1974 surgió el WISC-R que es fundamentalmente similar, pero técnicamente superior a la original. Esta escala fue traducida a varios idiomas. El WISC y el WISC-R fueron presentadas en español en 1981. En México, desde 1964, esta siendo realizada su estandarización para la población mexicana (WISC-RM), por el equipo de investigadores del INCCAPAC (Instituto Nacional de Ciencias del Comportamiento y de la Actitud Pública,

A.C.), bajo la dirección del doctor Rogelio Díaz Guerrero. Para elaborar esta versión se tomó la traducción de Puerto Rico, se le comparó con el Wechsler original, se hicieron correcciones a la traducción y se adaptaron aquellos reactivos considerados no adecuados a la cultura mexicana, cuidando de no alterar ni el contenido ni el grado de dificultad de los mismos. (Morales, 1993). Costa Neiva (1996) subraya que para la versión mexicana hay que tomar en cuenta que la muestra utilizada se limitó a la ciudad de México, por lo que no es representativa de toda la población mexicana.

1.11. Los tests colectivos de inteligencia.

Costa Neiva (1996) señala que mientras las escalas individuales de inteligencia, como las de Binet y todas sus revisiones, hallan su principal aplicación en la clínica, los tests colectivos de inteligencia se usan primordialmente en la pedagogía, en la industria y en el ejército. Las escalas individuales no se adaptan para su aplicación a un grupo por varias razones: sólo se pueden aplicar a una persona a la vez, muchas de estas escalas requieren respuestas orales del sujeto o necesitan la manipulación de materiales, algunas exigen la medida individual del tiempo invertido en responder, además, en muchos casos, requieren de un examinador de gran experiencia.

Los tests colectivos de inteligencia fueron creados para satisfacer una urgente necesidad práctica. En 1917, cuando los Estados Unidos decidieron participar en la Primera Guerra Mundial, la American Psychological Association designó un comité para evaluar de qué manera la psicología podía colaborar. Este comité, bajo la dirección de Robert M. Yerkes, se dio cuenta de la necesidad de clasificar rápidamente a millón y medio de reclutas en relación a su nivel intelectual general. Esta información tuvo una gran utilidad, pues a través de ella se tomaron decisiones acerca de quién podía ir al campo de batalla, quién debería estar exento del servicio militar, la asignación a diversos tipos de servicio, o la admisión a la formación de oficiales. Así el primer test de inteligencia colectivo se formó recurriendo a todos los tests disponibles y especialmente a un grupo de tests no publicados preparados por Arthur S. Otis.

Los tests aceptados finalmente por los psicólogos del ejército se conocen como el Alfa y Beta del Ejército. El primero era un test verbal cuya finalidad consistía en hacer una selección general para destinar a los militares. El segundo era un test no verbal dedicado a los individuos a quienes no se podía aplicar el Alfa a causa de su lengua extranjera o su

analfabetismo.

Costa Neiva (1996) señala que la versión original (Beta I) fue revisada en 1934 por Kellog y Morton con el objetivo de adaptarla a poblaciones civiles (Examen Beta Revisado). Posteriormente, apareció la edición utilizada actualmente (Beta II-R). (Kellog y Morton, 1974; citados por Costa Neiva, 1996).

El Beta II-R es una prueba colectiva de inteligencia no verbal diseñada especialmente para adultos que tienen problemas de lectura y escritura. Está compuesta de seis tareas, cada una de ellas mide cómo una persona puede resolver determinados problemas. Las tareas son: Laberintos, Claves, Figuras geométricas, Figuras incompletas, Pares iguales vs. desiguales y Objetos equivocados.

Aunque esta prueba es muy utilizada en México, las tablas presentadas en el manual son adecuadas solamente para la población norteamericana, de donde fueron extraídos los datos normativos (Costa Neiva , 1996).

1.12. Tests para el periodo infantil.

Anastasi (1967) señala que una de las más extensas investigaciones de conducta infantil fue la llevada a cabo en la Clinic of Child Development, de Yale, bajo la dirección de Arnold Gesell. En 1927, Gesell y sus colaboradores empezaron un estudio longitudinal del curso normal del desarrollo de la conducta en el infante humano. Los datos principales se obtuvieron mediante repetidas observaciones de 107 infantes que constituían una muestra relativamente *normal* y homogénea. En el examen solo se incluyeron niños sanos, libres de todo defecto conocido. Se eligieron cuidadosamente, de forma que los padres fueran de estado socioeconómico medio y cercanos al promedio respecto a la educación y nivel de ocupación. Todos los padres eran americanos de nacimiento y de origen nordeuropeo.

Se examinó a los infantes de este grupo a las cuatro, seis y ocho semanas de edad, y después, con intervalos de cuatro semanas, hasta la edad de cincuenta y seis semanas. Estudios continuados, posteriores, se hicieron a los dieciocho meses, y a los dos, tres, cuatro, cinco y seis años. Las repeticiones de los exámenes se continuaron durante un periodo de diez años en todos los casos donde fue posible. Sólo se examinaba una parte del grupo en cualquiera de los niveles de edad, variando las cifras desde 28 hasta 60, hasta la edad de cinco años; a los seis años sólo se incluyeron 18 casos. En cada edad se observó,

aproximadamente, el mismo número de niños que de niñas. El estudio de las muestras normativas se completó con extensas observaciones con gran número de casos clínicos.

Las Developmental Schedules (Escalas de desarrollo) de Gesell se prepararon a partir de los resultados obtenidos en las muestras normativas, así como de los datos complementarios de los otros casos. Consisten en descripciones de conducta que se emplean para determinar el nivel de desarrollo de la conducta que el niño ha alcanzado en cuatro áreas principales: conducta motora, conducta adaptativa, conducta de lenguaje y la conducta personal-social. Las investigaciones de las que se derivaron las descripciones representan una contribución importante en la psicología del niño. Sin embargo, Anastasi (1967), considera que como instrumentos de medición las Escalas de desarrollo de Gesell son relativamente toscas, pues en varios puntos del examen se infiltra cierta cantidad de subjetividad, además de que no se da ningún análisis estadístico sobre confiabilidad y validez.

Morales (1993) señala que esta prueba es ampliamente usada en México, bajo el nombre de *Diagnóstico del desarrollo*, desde que su manual fue traducido al castellano. (Gesell y Amatruda, 1957; citados por Morales, 1993).

En 1973, en México, según Morales (1993), en el Instituto Nacional de Ciencias del Comportamiento y de la Actitud Pública (INCCAPAC) se llevaron a cabo una serie de investigaciones en el campo de la psicología evolutiva, educativa y social especialmente, que han determinado que en ellas se dedique especial atención a los procesos de desarrollo de la etapa preescolar, sobre la cual se habían realizado muy pocos estudios hasta entonces. Estas investigaciones han conducido a la construcción de tests especiales para medir tanto el desarrollo perceptual como la evolución de los procesos cognoscitivos y motores, a partir de un marco de referencia sociocultural que permita detectar las estrategias que aprende el sujeto mexicano para su supervivencia, dado su ambiente general. Hasta el momento se han construido 17 tests psicológicos que miden el rendimiento intelectual, la percepción, las habilidades específicas y el desarrollo del lenguaje, en los cuales se enfatiza fundamentalmente la formación de conceptos. Estos tests ya han sido evaluados mediante “estudios piloto”, que incluyen análisis de reactivos, consistencia interna, coeficiente de estabilidad temporal (confiabilidad), etc.

A continuación se mencionan los nombres con los cuales se conocen esos instrumentos de medición:

Para la investigación sobre rendimiento intelectual:

1. Test de conocimientos generales.
2. Test de números.
3. Test de relaciones.

Para la investigación sobre percepción:

1. Test de habilidad para clasificar.
2. Test de clasificación de habilidades.
3. Test de comparación.
4. Test de eliminación.
5. Test de figuras ocultas.

Para la evaluación del desarrollo del lenguaje:

1. Test de comprensión oral.
2. Test de designación de objetos.
3. Test de reconocimiento.
4. Test de memorización de oraciones.
5. Test de comprensión de oraciones.
6. Test de letras.
7. Test de clasificación de objetos para niños preescolares.
8. Test de memorización de objetos.
9. Test de creatividad verbal.

Costa Neiva (1996) destaca la Escala de inteligencia para los niveles preescolar y primario (WPPSI). La escala WPPSI (Wechsler Preschool and Primary Scale of Intelligence), fue presentada en 1963. La idea original era elaborar una escala que fuera una extensión descendiente del WISC. Sin embargo, después de un estudio preliminar, y por razones teóricas y prácticas, se consideró que el WPPSI “era una escala independiente, adaptada de manera óptima al examen mental de los niños de 4 a 6 y medio años” (Wechsler, 1981; citado por Costa Neiva, 1996).

Asimismo, ocho de los once subtests del WPPSI y casi la mitad de sus reactivos provienen directamente del WISC. Las modificaciones básicas que se hicieron en los subtests provenientes del WISC son: adición de elementos más fáciles y eliminación de otros más difíciles, algunos cambios en el contenido de reactivos y en la forma de la aplicación.

Posteriormente se presentó una versión revisada (WPPSI-R) que puede ser utilizada con niños de 3 a 7 años de edad. La versión en español -WPPSI español- surgió en 1981 (Weschler, 1981). En México fue realizado un estudio para la traducción, adaptación, confiabilidad y validez del WPPSI en un grupo de 60 niños de nivel socioeconómico medio alto. (Coto Villa y Llano de la Torre, 1981; citados por Costa Neiva, 1996).

1.13. Medición de las aptitudes.

Bennett, Seashore y Wesman (1990; citados por Costa Neiva, 1996) acudiendo al Diccionario de Psicología de Warren (1934) definen el concepto de *aptitud* como “la condición o conjunto de características consideradas sintomáticas de la capacidad de un individuo de adquirir mediante entrenamiento algún conocimiento, técnica o conjunto de respuestas (generalmente especificadas) como la habilidad para hablar un idioma, producir música...” (p. 1).

Esta definición indica que las aptitudes son producto de una interacción entre la herencia y el ambiente y que pueden ser desarrolladas.

Brown (1980) diferencia los términos habilidad y aptitud señalando que la habilidad indica “el poder para realizar una tarea” y la aptitud “el poder para aprender a realizar una tarea” (p. 282). El primer término se refiere a un estado actual y el segundo a un estado futuro. En el término aptitud se incluye la idea de previsión, de probabilidad de que se adquieran o aprendan conductas.

Costa Neiva (1996) menciona que, desde 1928, con la publicación del libro de Kelley, *Crossroads in the Mind of Man*, ya se observaba la preocupación por investigar los diferentes factores que componen la inteligencia. El modelo multifactorial, defendido sobre todo por Thurstone, profundizó esta idea. A partir de varias investigaciones, él propuso 12 factores grupales a los cuales denominó de *capacidades mentales primarias*. De éstos, los que más se confirmaron en los estudios posteriores fueron los siguientes:

1. Comprensión verbal.
2. Fluencia verbal.
3. Número.
4. Espacio.
5. Memoria asociativa.

6. Rapidez perceptual.

7. Inducción (o razonamiento general).

Después, Guilford (1967, citado por Costa Neiva, 1996), propuso un modelo para la estructura intelectual, donde llegó a calcular la existencia de 120 habilidades.

Antes de que fueran construidas baterías de aptitudes múltiples, se hicieron tentativas de comparar la posición relativa de un individuo en diferentes subtests o grupo de reactivos de pruebas de inteligencia. Como estas pruebas no fueron desarrolladas con este objetivo, generalmente, los subtests no permitían una comparación intraindividual adecuada.

El desarrollo de baterías de aptitud múltiple fue estimulado por las limitaciones de las pruebas de inteligencia que, en general, se concentraban en la medida de la capacidad verbal, no proporcionando medidas de otras capacidades (abstracta, mecánica, etc.). Además, la creciente actividad de los psicólogos en las áreas de selección y clasificación de personal y de orientación vocacional llevaron a la necesidad de disponer de instrumentos más específicos que facilitaran estas tareas. Pero, fue sobre todo el desarrollo del método estadístico del análisis factorial que impulsó la investigación en esta área y permitió el desarrollo de los primeros instrumentos de medida de las aptitudes.

Hay pruebas de aptitudes construidas para medir aptitudes específicas o especiales que pueden ser de tipo cognoscitivo (habilidad para matemáticas, creatividad), de tipo vocacional (velocidad y precisión perceptual, destreza manual) y de tipo musical o artístico.

Existen también las baterías de pruebas de aptitudes múltiples que se componen de una serie de pruebas de aptitudes específicas. Siguiendo a Costa Neiva (1996), a continuación mencionaremos dos baterías de aptitud múltiple de amplio uso en México:

1. Pruebas de Aptitud Diferencial (DAT).

2. Pruebas para la Clasificación de Aptitudes de Flanagan (FACT).

Las Pruebas de Aptitud Diferencial (DAT), fueron desarrolladas en Estados Unidos por G. Bennett, H. Seashore y A. Wesman, y publicadas en sus formas originales (A y B) en 1947. Fueron diseñadas para atender la demanda de los orientadores educacionales y vocacionales, que necesitaban de un instrumento que permitiera medir rigurosamente las aptitudes múltiples de los alumnos de secundaria y preparatoria. (Bennett, Seashore y Wesman, 1990; citados por Costa Neiva, 1996).

Estas pruebas fueron revisadas y reestandarizadas en 1962 (Formas L y M), en 1972

(Formas S y T), en 1980 (Formas V y W) y en 1990 (Formas C y D). En la primera revisión, el objetivo principal fue facilitar la administración y calificación de las pruebas y para tal efecto se cambió el formato y extensión de las mismas. En la segunda, se conservaron los cambios anteriores y se actualizaron algunas pruebas. La tercera, tuvo como meta principal elaborar reactivos con el mismo nivel de dificultad para varones y mujeres. Se hicieron también cambios en las instrucciones para volverlas más comprensibles. La última revisión (5a. edición) incluye reactivos nuevos y utiliza dos formas paralelas (C y D) para dos niveles diferentes (Nivel 1 y Nivel 2), aumentando así su utilidad y confiabilidad.

En México, fue publicada en 1990 la versión en español del DAT basada en la revisión de 1980 (Bennett y cols., 1990, citado por Costa Neiva, 1996). En esta versión no se incluyeron las pruebas de Razonamiento verbal, Ortografía y Uso del lenguaje.

En 1994, surgió la Prueba Diferencial de Aptitudes PROUNAM, basada en la quinta versión del DAT (DAT 1990), desarrollada mediante la colaboración de la Universidad Nacional Autónoma de México, el Instituto de Evaluación en Gran Escala y The Psychological Corporation (Universidad Nacional Autónoma de México [UNAM], 1994a; citado por Costa Neiva, 1996). Esta prueba está compuesta de las ocho subpruebas del DAT, siendo que tres de ellas fueron elaboradas totalmente en México (Razonamiento verbal, Uso del lenguaje y Ortografía). La elaboración de las dos primeras siguió los principios establecidos en la prueba original, aunque en el caso de la prueba Uso del lenguaje se utilizaron las reglas del idioma español. Sin embargo “la prueba de Ortografía fue diseñada en su totalidad para ‘PROUNAM’ siguiendo modelos escolares utilizados en México para este tipo de evaluación” (Universidad Nacional Autónoma de México [UNAM], 1994b, p. 5; citado por Costa Neiva, 1996), en sustitución de la prueba original de “Spelling” que no era aplicable a México.

Aunque las pruebas de la batería DAT son independientes, éstas fueron desarrolladas y reestandarizadas como una batería integrada, la estandarización de las distintas pruebas fue realizada con una misma muestra. Como el rango de niveles es el mismo para todas las pruebas de la batería, se pueden hacer comparaciones intra e interindividuales consistentes.

La batería DAT se compone de 8 pruebas: Razonamiento verbal, Habilidad numérica, Razonamiento abstracto, Velocidad y exactitud secretariales, Razonamiento mecánico, Relaciones espaciales, Uso del lenguaje y Ortografía. En el PROUNAM están incluidas las

ocho pruebas.

Anastasi (1977), señala que las pruebas para la Clasificación de Aptitudes (FACT) fueron desarrolladas por J.C. Flanagan y dieron continuidad a su proyecto de construcción de pruebas de clasificación para la Fuerza Aérea durante la Segunda Guerra Mundial. A partir de análisis sistemáticos de puestos se identificaron elementos comunes a ciertas actividades profesionales así como las aptitudes necesarias para la buena ejecución de las mismas. Se construyeron entonces pruebas para evaluar las aptitudes necesarias para que se realicen con éxito tareas ocupacionales específicas (Brown, 1980).

Costa Neiva (1996) dice que las pruebas FACT fueron publicadas por primera vez en 1953, su publicación en español fue realizada en 1970, y en 1989 fue presentada la primera edición mexicana.

La batería está compuesta de catorce pruebas, cada una destinada a medir una habilidad específica. Las pruebas son las siguientes: Inspección, Claves, Memoria, Precisión, Ensamblajes, Escalas, Coordinación, Comprensión y discernimiento, Aritmética, Modelos, Componentes, Tablas, Comprensión mecánica y Expresión idiomática.

Las pruebas FACT pueden ser útiles para la orientación profesional y para la selección de personal pues permiten evaluar las habilidades del sujeto y proveer una cierta predicción con respecto al éxito en determinadas tareas ocupacionales. Sin embargo, tienen varias restricciones como son:

1. Que los resultados sobre la confiabilidad y la validez de la batería parecen poco consistentes e inclusive en el manual de la edición mexicana no se proporcionan datos al respecto.

2. La representatividad de la muestra normativa de la versión original es considerada discutible y en cuanto a la edición mexicana, no se presentan en el manual datos sobre la muestra de estandarización utilizada.

3. La edición mexicana del FACT considera solamente 30 ocupaciones, lo que es un número relativamente reducido comparado con el número de ocupaciones existentes en la actualidad. (Costa Neiva, 1996).

1.14. Medición de la personalidad.

Anastasi (1967) dice que los tests de personalidad, en general, fueron creados para

efectuar medidas de características tales como la adaptación emocional, las relaciones sociales, la motivación y los intereses y actitudes. El número de tests de personalidad se eleva a varios centenares.

Un precursor de los tests de personalidad fue Kraepelin (1892) quien utilizó el test de asociación libre con pacientes normales. En este test se le dan al sujeto palabras-estímulo seleccionadas y se le requiere para que responda a cada una exactamente con la primera palabra que se le ocurra. Kraepelin empleó también esta técnica para estudiar los efectos psicológicos de la fatiga, el hambre y las drogas, y llegó a la conclusión de que todos estos agentes aumentan la frecuencia relativa de las asociaciones superficiales.

Sommer, autor de la última década del siglo XIX, sugirió que el test de asociación libre podía usarse para distinguir las diversas formas del trastorno mental. La técnica de la asociación libre ha sido utilizada después para una gran variedad de fines en la aplicación de tests, y todavía se emplea corrientemente.

Anastasi (1967) clasifica a los tests de personalidad en tres apartados: inventarios autodescriptivos, medidas de intereses y actitudes y técnicas proyectivas.

1.14.1 Inventarios autodescriptivos.

El prototipo de los inventarios autodescriptivos de personalidad fue la hoja de datos personales, de Woodworth, puesta en uso durante la primera guerra mundial. Este inventario fue esencialmente un intento de tipificar una entrevista psiquiátrica y adaptar el procedimiento al examen de masas. Woodworth reunió información respecto a los síntomas comunes neuróticos y preneuróticos, consultando la literatura psiquiátrica y a los propios psiquiatras directamente. Las preguntas del inventario se formularon originalmente referidas a estos síntomas; trataban de desviaciones de la conducta, tales como miedos o fobias anormales, obsesiones y compulsiones, pesadillas y otros trastornos del sueño, fatiga excesiva y otros síntomas psicósomáticos, sentimientos de irrealidad y trastornos motores tales como tics y temblores. (Anastasi, 1967).

El Minnesota Multiphasic Personality Inventory (Inventario Multifásico de Personalidad de Minnesota, MMPI), es un inventario cuya aplicación ha alcanzado proporciones sin precedente. Desde su aparición en 1940 y la publicación de su primer manual oficial en 1943, ha estimulado una corriente continua de investigación. El MMPI se creó originalmente para

valorar aquellos rasgos que son comúnmente característicos de la anormalidad psicológica inutilizadora. El inventario consta de 550 enunciados afirmativos, y se pide al sujeto que los clasifique en tres categorías: verdadero, falso y no lo sé. Se destina a adolescentes y adultos desde los dieciséis años en adelante. (Anastasi, 1967).

Los reactivos se han clasificado en 26 áreas, por ejemplo, salud general, sistema gastrointestinal, sistema marital y familiar, religión, afectos, fobias, intereses masculinos, intereses femeninos, etc. Los reactivos han sido agrupados en escalas separadas para calificar nueve rasgos de personalidad, a saber: hipocondriasis, depresión, histeria, desviaciones psicopáticas, intereses masculino-femeninos, paranoia, psicastenia, esquizofrenia e hipomanía. Además, contiene una escala denominada escala social, la cual se desarrolla posteriormente para detectar las relaciones sociales del sujeto.

Este inventario se construyó con el fin, casi exclusivo, de establecer diagnósticos diferenciales de problemas patológicos. La validez del MMPI se ha estudiado profundamente, no sólo en los Estados Unidos, sino también en un gran número de países, ya que ha tenido una considerable difusión. (Morales, 1993).

Pérez y Farías (1995), señala que en 1989 el MMPI fue revisado en forma exhaustiva, y la versión revisada de la prueba fue llamada MMPI-2, que es en algunos aspectos muy similar al original.

En México, se han realizado numerosos estudios como lo es el de Rivera y Ampudia (1976; citados por Pérez y Farías, 1995), cuyos resultados sugieren que existe la necesidad de contar con técnicas psicológicas que faciliten la detección temprana de la presencia de psicopatología en estudiantes universitarios con el fin de ofrecer ayuda clínica oportuna.

Navarro (1971, citado por Pérez y Farías, 1995), en una investigación reporta la influencia del sexo, la edad y el nivel de inteligencia sobre los puntajes T de las 13 escalas básicas del MMPI (Español) en una muestra de jóvenes mexicanos relativamente normales.

El MMPI forma parte de las baterías de Diagnóstico Clínico del Centro de Orientación Psicológica de la Universidad Iberoamericana (UIA). Asimismo se emplea por la misma Universidad en estudios de cambio de carrera (requeridos por la UIA para que un alumno pueda pasar de una carrera a otra dentro de la misma Universidad). (Pérez y Farías, 1995).

El MMPI-2 en español fue estandarizado y normalizado en un grupo de estudiantes universitarios en 1992 (Lucio y Reyes Lagunes; citados por Pérez y Farías, 1995). Pérez y

Fariás (1995) a su vez, reporta que el MMPI-2 es confiable para la muestra de estudiantes mexicanos con los que él mismo realizó el estudio.

1.14.2. Medidas de los intereses .

La fuerza y dirección de los intereses, actitudes, motivos, valores y otras variables representan un importante aspecto de la personalidad del individuo. Estas características afectan materialmente su adaptación educativa y profesional, sus relaciones interpersonales, sus diversiones en el ocio y otras facetas de su vida cotidiana. Por esta razón se han creado inventarios de valores, intereses y actitudes.

Dentro de los tests de intereses se puede mencionar la Vocational Interest Blank (Hoja de intereses profesionales, VIB), elaborada por E. K. Strong, del Instituto de Tecnología Carnegie, en 1920. Esta Hoja ha sido objeto de continua investigación, revisión y ampliación. (Anastasi, 1967).

Morales (1993) señala que en México, el Inventario de intereses vocacionales de Strong, “es de uso común sin que aún se haya hecho una buena adaptación, cuando menos para utilizarlo en todas aquellas situaciones en que sea necesario” (p. 139). El objetivo de este inventario es detectar gustos y desagrados sobre ocupaciones, materias escolares, actividades cotidianas y rasgos peculiares de personalidad. Este inventario se ha diseñado en dos formas, uno para hombres y otro para mujeres, y es aplicable desde los 17 años. Cada ocupación se califica por separado y los registros obtenidos se consideran como los “indicadores” de los intereses del sujeto en cada ocupación. Puede agruparse también en seis grupos ocupacionales, más que por ocupaciones específicas. Además, califica los intereses no ocupacionales que pueden ser de utilidad en la orientación del sujeto, tales como maduración en el área de intereses, masculinidad-femineidad, nivel ocupacional, nivel de escolaridad, etc.

Otro registro de preferencias muy usado en México es el Inventario de Kuder y Strong que se aplica desde los 15 años de edad. Está constituido por 168 reactivos distribuidos en tríadas, es decir, en cada respuesta se debe indicar cuál de los tres reactivos agrada más y cuál menos.

Este inventario cubre un rango de actividades muy amplio y su calificación produce un perfil que representa diez áreas: mecánica, científica, persuasiva, artística, literaria, musical, de servicio social, de oficina, trabajo al aire libre y de computación o cálculos numéricos.

“La confiabilidad y validez de estos inventarios son satisfactorias para los Estados

Unidos, en donde se emplean frecuentemente. En México queda en duda el que los coeficientes de confiabilidad y validez sean tan altos como en los Estados Unidos (0.85 para Kuder y 0.95 para Strong), ya que no hay estudios de este tipo". (Morales, 1993, p. 140).

1.14.3. Medidas de las actitudes.

Nadelsticher (1983), dice que la palabra actitud significa la predisposición hacia un objeto y/o hacia una situación. El cuestionario de actitudes mide precisamente esta predisposición.

Anastasi (1967), a su vez, menciona que las escalas de actitudes tienen como fin proporcionar una medida cuantitativa de la posición relativa del individuo a lo largo de un continuo de actitud unidimensional.

La medición de las actitudes se basa principalmente en la teoría del juicio comparativo de Thurstone con la que se establece la posibilidad de cuantificar toda experiencia subjetiva. Las escalas de actitudes adquirieron mucho auge dado que permitían al investigador elaborar un cuestionario confiable con una precisión elevada. Hay muchas formas de hacer un cuestionario de actitudes, todo depende de la escala que se utilice. Han existido muchas escalas de actitudes, entre las principales tenemos:

1. Método de intervalos aparentemente iguales de Thurstone.
2. Método de intervalos sucesivos de Thurstone.
3. Método de rangos sumariados de Likert.
4. Método de diferencial semántico de Osgood.
5. Método del escalograma de Guttman.

Dentro de los métodos mencionados el más utilizado por su sencillez y facilidad es el de Rangos sumariados de Likert. (Nadelsticher, 1983).

A diferencia de las escalas de tipo Thurstone, el procedimiento de Likert no requiere que un grupo de jueces haga la clasificación de los elementos. Estos se seleccionan solamente basándose en las respuestas de los sujetos a quienes se aplican en el curso del desarrollo del cuestionario. A menudo el criterio para la selección de elementos es la consistencia interna, aunque se pueden emplear criterios externos cuando se dispone de ellos.

La escala del tipo Likert exige una respuesta graduada a cada elemento, y se expresa corrientemente en función de las cinco alternativas siguientes:

- a) Totalmente de acuerdo (TA),

- b) Acuerdo (A),
- c) Indiferente (I),
- d) Desacuerdo (D),
- e) Totalmente en desacuerdo (TD).

Las opciones se califican del 0 al 4 o del 1 al 5, respectivamente, desde el extremo favorable al desfavorable. La suma de las calificaciones de los elementos representa la puntuación total del individuo, que debe interpretarse en función de las normas establecidas empíricamente. (Nadelsticher, 1983).

1.14.4. Técnicas proyectivas.

La principal característica de las técnicas proyectivas se encuentra en su dedicación a una tarea relativamente no estructurada; es decir, una tarea que permite una variedad casi ilimitada de respuestas posibles. A fin de conceder libertad de juego a la imaginación del sujeto, sólo se dan breves instrucciones generales. Por la misma razón, los estímulos del test son generalmente vagos y ambiguos. La hipótesis subyacente reside en que el modo en que el individuo perciba e interprete el material del test o estructure la situación, reflejará aspectos fundamentales de su funcionamiento psicológico. En otras palabras, se espera que los materiales del test sean como una especie de pantalla sobre la que el sujeto proyecta sus ideas características, actitudes, esfuerzos, temores, conflictos, agresiones, etc.

Típicamente, los instrumentos proyectivos representan también procedimientos de examen disfrazados, en la medida en que el sujeto rara vez se da cuenta del tipo de interpretación psicológica que se hará de sus respuestas. Las técnicas proyectivas se caracterizan, también, por el método global de apreciar la personalidad. La atención se centra sobre el conjunto de la personalidad en lugar de hacerlo sobre la medida de rasgos específicos. Los métodos proyectivos se originaron en los medios clínicos, y siguen siendo predominantemente un instrumento clínico. (Anastasi, 1967).

Según Morales (1993), en México, dentro de estos tests destacan por su uso:

Las manchas de tinta de Rorschach, que es la técnica proyectiva más famosa y de mayor uso. Esta técnica, creada por el psiquiatra suizo Hermann Rorschach, fue descrita por primera vez en 1921. Para desarrollarla, experimentó con un gran número de manchas de tinta que aplicó a distintos grupos psiquiátricos. Mediante el proceso de ensayos sucesivos, se fueron

incorporando gradualmente en el sistema de puntuación aquellas características de respuesta que establecían diferencias entre los diversos síndromes psiquiátricos. Los procedimientos de puntuación se fueron enriqueciendo gradualmente al aplicar tests a deficientes mentales, seres normales, artistas, intelectuales y otras personas de características conocidas.

El Rorschach utiliza diez láminas, en cada una de las cuales hay impresa una mancha de tinta simétrica. Cinco de las manchas son en gris y negro solamente; dos contienen toques adicionales de rojo brillante, y las tres restantes combinan varias tonalidades. A medida que se le muestra cada mancha de tinta al sujeto, se le pide que diga lo que ve, lo que la mancha podría representar.

La prueba original de Rorschach no tiene tiempo límite, ni número específico de respuestas para cada tarjeta. El examinador debe anotar el comportamiento del sujeto cuando éste es enfrentado a cada estímulo: expresiones, registro de la verbalización de la respuesta, tiempo de reacción (es decir, el tiempo transcurrido desde la presentación de la tarjeta, hasta el principio de la respuesta del sujeto), longitud del tiempo entre las respuestas, total de tiempo para responder a cada lámina; posición en que colocó la lámina para dar la respuesta, indicando la extensión o minuciosidad con que el sujeto exploró la situación estímulo; movimientos extraños del sujeto, y otras conductas significativas.

Según Rorschach, los tres registros del tiempo sirven para analizar los bloqueos emocionales o la resistencia a responder de cada individuo, frente a la percepción de cada mancha en particular. Todo esto constituye la primera parte de la aplicación. La segunda parte la constituye la "encuesta", es decir, preguntarle al sujeto en qué parte de la mancha centró su respuesta; qué parte le hizo pensar o imaginar lo que respondió, su tamaño, color, sombreado, movimiento, etc.; todo aquello que pueda servir para dar una calificación.

Los datos obtenidos de las respuestas del sujeto (es decir, la calificación) tienen como fin principal conocer la adaptación general del mismo: si es la adecuada o si existen dificultades psicológicas. Morales (1993) subraya que los especialistas del Rorschach dicen que si bien es difícil la administración del test, más difícil es su interpretación, lo cual es cierto puesto que en la utilización del instrumento no existe una regla o método específico que permita objetivizar los resultados.

En contraposición con el test de manchas de tinta de Rorschach, se puede hablar de la técnica de manchas de tinta de Wayne H. Holtzman (Morales, 1993). En 1954 se inició en la

Universidad de Texas una amplia investigación con el fin de detectar, con la mayor exactitud, las limitaciones psicométricas del Rorschach, por lo que se inició la producción de una serie de láminas que contenían manchas de tinta. Las 135 manchas de tinta que finalmente se consideraron como las mejores, después de los análisis pertinentes y requeridos para elaborar un instrumento de este tipo, fueron retocadas por un dibujante profesional para que intencionalmente variaran en color, sombreado, forma y simetría. Además, esto permitió elaborar dos formas paralelas (es decir, sustituibles) que no sólo tuvieran el mismo número de manchas, sino también que fueran correlativas unas de otras. De este modo, quedó elaborado un test con formas paralelas A y B, conteniendo cada una 45 láminas que el sujeto tendría que examinar.

En este test, el sujeto, a diferencia del Rorschach, sólo puede dar una respuesta; inmediatamente después se hace la encuesta sobre la localización y sobre las razones que el sujeto tuvo para emitir su respuesta.

En esta nueva forma de investigación sobre rasgos de la personalidad se tienen en consideración, para cada una de las respuestas, no sólo el calificarlas a través de 22 variables que constituyen los aspectos que se intentan medir con ella, sino también el desarrollar una nueva forma de calificación objetiva, es decir, dimensionalizando o cuantificando cada variable, para superar las deficiencias del Rorschach y de otros sistemas anteriores y para obtener resultados válidos y confiables.

Se han llevado a cabo numerosos estudios con este test. En México, los primeros estudios con la prueba de manchas de tinta de Holtzman se iniciaron en 1964, en lo que originalmente fue el Centro de Ciencias del Comportamiento de la UNAM. Primero se realizaron estudios con el fin de adaptar la prueba a México y después estudios de estandarización, confiabilidad y validación para aplicarlo a sujetos mexicanos.

Mediante la prueba de Holtzman se realizó un seguimiento longitudinal para observar el desarrollo de la personalidad del escolar mexicano, enriquecido con un gran número de otros instrumentos, entre los cuales figuran tests de desarrollo perceptual, de desarrollo conceptual, que han permitido investigar, mediante intercorrelaciones, análisis factoriales y análisis de varianza, tanto el desarrollo de los rasgos de personalidad que mide el instrumento aludido, como las interacciones entre las diversas variables medidas por éste y los demás instrumentos.

Las muestras originales sobre las cuales se llevaron a cabo los primeros estudios quedan circunscritas, en primer lugar, para escolares mexicanos del Distrito Federal. Además se han hecho estudios con muestras de sujetos que presentan diversas patologías. La muestra de escolares del Distrito Federal cubre 12 años de desarrollo de la personalidad del escolar mexicano. (Morales, 1993).

Como se puede observar, muchas de las pruebas psicológicas que hoy son usadas en México, y en varios países de Latinoamérica, se han originado en otras lenguas y su estandarización, confiabilidad y validez en lengua española han permitido su uso en nuestros países; lo que hace necesario y valioso que los psicólogos mexicanos continúen realizando este trabajo en forma sistemática, lo que implica el conocimiento de su estructura psicométrica. Con ese objetivo y después de este recorrido por los orígenes y funciones de los tests, en el siguiente capítulo se abordarán los aspectos psicométricos que los constituyen como instrumentos de medida.

Capítulo 2

La estructura de los tests o pruebas psicológicas

Teoría psicométrica

2. La estructura de los tests o pruebas psicológicas Teoría psicométrica

2.1. ¿Qué es un test psicológico?

El concepto de test psicológico ha sido definido de diversas maneras. A continuación se presentan algunas definiciones:

En 1890 el psicólogo americano Catell creó el término de test mental (“mental test”) para designar una serie de pruebas psicológicas utilizadas en el examen de las diferencias individuales de estudiantes universitarios. (Cfr. cap. 1 de esta tesis).

La Asociación Internacional de Psicotécnica adoptó la siguiente definición, utilizada nuevamente en el *Vocabulaire de psychologie* de H. Piéron (1952, citado por Pichot, 1960): “Es una prueba definida, que implica una tarea a realizar, idéntica para todos los sujetos examinados, con una técnica precisa para la apreciación del éxito o el fracaso, o para la puntuación numérica del éxito. Esta tarea puede poner en juego ya sea conocimientos adquiridos (test pedagógico), ya sea funciones sensoriomotrices o mentales (test psicológico)” (Pichot, 1960, p. 11).

Pichot (1960) propone la siguiente definición: “Se llama test mental a una situación experimental estandarizada que sirve de estímulo a un comportamiento. Tal comportamiento se evalúa por una comparación estadística con el de otros individuos colocados en la misma situación, ya sea cuantitativamente, ya sea tipológicamente” (p. 11).

Anastasi (1967) considera que “un test psicológico constituye esencialmente una medida objetiva y tipificada de una muestra de conducta” (p. 19).

Brown (1980) define una prueba psicológica como “un procedimiento sistemático para medir una prueba de conducta” (p. 6).

A partir de estas definiciones y de las propuestas de Brown y Székely se puede inferir que un test psicológico reúne los siguientes criterios:

1. Un instrumento de medición de la psicología.
2. Una medida objetiva y tipificada, esto significa que implica un procedimiento sistemático de medición que según Brown (1980) indica que una prueba se construye, se administra y califica de acuerdo con reglas preestablecidas. “Los reactivos se seleccionan

sistemáticamente para que se ajusten a las especificaciones de la prueba, se administran los mismos reactivos u otros equivalentes a todas las personas y las instrucciones y los límites de tiempo son iguales para todos los que se someten a dicha prueba. El uso de reglas predeterminadas para la evaluación (calificación) de las respuestas, asegura un acuerdo entre diferentes personas que tengan que calificar la prueba” (p. 6).

3. Mide una muestra de conducta. Brown (1980) señala que “en el sentido más estricto, una prueba mide sólo la conducta registrada por la prueba, esto es, las respuestas dadas por una persona a los reactivos de la prueba. No se mide directamente a una persona, sino que inferimos sus características (rasgos) a partir de sus respuestas dadas a los reactivos de una prueba” (p. 7).

4. Es una medida estadística. Székely (1970), dice que el método psicométrico introduce la apreciación cuantitativa de los fenómenos psicológicos y con ello la manipulación estadística de los datos obtenidos en una prueba.

Como se puede observar, dado que los tests son instrumentos de medición, para el conocimiento de su estructura es necesario abordar, primero, el problema de la medición en psicología y, después, el método que esta ciencia ha creado para construirlos de manera que sean válidos y confiables.

2.2. El concepto de medición.

El concepto de medición ha recibido varias definiciones. A continuación se presentan algunas:

Bertrand Russell (1938, citado por Reidl, 1990), define el “medir magnitudes como cualquier método por medio del cual se establece una correspondencia única y recíproca entre todas o algunas de las magnitudes de algún tipo y todos o algunos de los números, integrales racionales o reales” (p.76).

Torgerson (1967, citado por Reidl, 1990), manifiesta que la medición de una propiedad implica asignar números a los sistemas que representan esta propiedad. Para poder representar esta propiedad se debe obtener una relación isomórfica entre ciertas características del sistema numérico y las relaciones entre las varias cantidades de la propiedad a medirse. Lo importante del procedimiento es asignar los números de manera tal

que reflejen esta equivalencia de formas entre las características de los números y las relaciones correspondientes entre las cantidades.

Reidl (1990) señala que “medir quiere decir introducir a las matemáticas en la ciencia, aumentando así su cientificidad. Las matemáticas son un lenguaje universal que cualquier ciencia o tecnología puede usar para su conveniencia. Son también un sistema deductivo, un conjunto de reglas para la manipulación de símbolos internamente consistentes que hace referencia a símbolos, no al sistema real. Los números sólo son un conjunto de símbolos específicos y la medición siempre se refiere a números.

Por medio de la medición se trata de averiguar qué tanto de cualquier atributo existe, y tiene como propósito el cuantificar los atributos de objetos y personas reales. Esto se logra mediante la realización de una operación: es decir, hacer algo de acuerdo a un conjunto de reglas para obtener medidas. Su validez o utilidad va a depender del carácter de los datos empíricos” (p. 2).

2.3. La medición en psicología.

Reidl (1990) dice que la medición en psicología surge de dos fuentes principales:

1. La tradición psicofísica que trató de establecer la relación que existía entre los estímulos físicos y las sensaciones despertadas por éstos en el hombre, para lo cual desarrolló los llamados métodos psicofísicos: el del error promedio, el de los cambios mínimos, y el de los estímulos constantes.

2. La tradición de las pruebas mentales o tests que tiene sus inicios a finales del siglo pasado y principios de éste, con investigadores como Galton, Catell, Binet, Terman, Otis, Thurstone, etc.

“La medición en psicología tiene por objeto registrar diferencias entre individuos o entre reacciones del mismo individuo en diferentes ocasiones utilizando instrumentos especiales denominados pruebas psicológicas o tests” (p. 1).

Nunally (en Bernstein, 1995) por su parte, señala que la medición es el principal problema de la psicología, porque hay muchas teorías, pero una teoría sólo puede ser probada en la medida en que sus atributos hipotetizados pueden ser medidos de manera adecuada. Para este autor la medición consiste en reglas para asignar símbolos a objetos de manera que:

1. Representen cantidades o atributos de forma numérica (escala de medición) o

2. Definan si los objetos caen en las mismas categorías o en otras diferentes con respecto a un atributo determinado (clasificación).

Con el término “reglas” indica que la asignación de números debe establecerse de manera explícita. Las reglas son, a su vez, continúa Bernstein, un aspecto importante de la estandarización. Una medida se estandariza hasta el punto en que a) sus reglas sean claras, b) su aplicación sea práctica, c) no requiera una gran habilidad de parte de los administradores más allá de la necesaria para su entrenamiento inicial, y d) sus resultados no dependan del administrador específico. El punto esencial sobre la estandarización es que los usuarios de un instrumento determinado deben obtener resultados similares. Por ejemplo, una prueba de inteligencia está bien estandarizada si diferentes examinadores obtienen puntajes similares al evaluar a un niño en particular en un momento determinado.

El término “atributo” en la definición indica que la medición siempre implica alguna característica particular de los objetos. Los objetos no pueden medirse, se miden sus atributos, esto es, que la medición requiere de un proceso de abstracción, por ejemplo, medimos la inteligencia de un sujeto, un atributo del sujeto, lo que requiere un proceso de definición de la inteligencia, de ese atributo.

Reidl (1990) subraya que en las ciencias sociales, la elección, localización y descripción de las propiedades estratégicas del objeto a estudiar constituyen, en sí mismas, un problema esencial. Estas propiedades reciben a veces el nombre de atributos, pero a menudo se recurre al término matemático de variable. Este término que procede de las matemáticas y de la física teórica se refiere a una medida o clasificación sometida a determinadas reglas formales. En las ciencias sociales puede ser el resultado de la división de conjuntos de acuerdo con uno o varios criterios específicos: por ejemplo, sexo, nivel de educación y edad. La primera corresponde a una clasificación nominal, la segunda a una clasificación ordenada, sólo la tercera es cuantitativa y del mismo tipo que las variables utilizadas por la física.

Las variables, que son los elementos de las ciencias sociales, son a veces la traducción en términos operativos de nociones usadas corrientemente (inteligencia, moral, costo de la vida). En otros casos proceden de consideraciones teóricas (introversión, integración, cohesión). Pero sea cual fuere su origen son obtenidas siempre mediante la aplicación, a los conceptos y nociones, de un método definido e inmutable, que consiste en cuatro reglas generales de traducción que permiten pasar del vocabulario de los conceptos al de las variables. Las cuatro

reglas son las siguientes:

1. La primera se refiere a la actividad intelectual y al análisis de la literatura relacionada con el objeto de estudio, de manera que el investigador pueda tener un primer esbozo, una construcción abstracta, una imagen. Con este paso se logra obtener una definición conceptual de la variable que se pretende investigar.

2. La segunda regla consiste en analizar los componentes de esta primera noción. A estos componentes se les ha llamado aspectos o dimensiones. Dichos componentes se pueden deducir analíticamente a partir del concepto general que los engloba, o empíricamente a partir de la estructura de sus intercorrelaciones.

3. La tercera regla consiste en seleccionar los indicadores de las dimensiones definidas. Uno de los principales resultados a los que se enfrenta el investigador es el hecho de que el indicador está relacionado con el concepto tan sólo en términos de probabilidad y no de certeza, y debido a esto, es absolutamente necesario utilizar, en la medida de lo posible, un gran número de indicadores. Al llegar a este paso se obtiene una definición indicativa, la cual define a la variable haciendo referencia al índice que la mide. Desde aquí se puede arribar a una definición operacional, en la que se hace referencia a operaciones o conductas de investigación definidas específicamente. En la medida en que se tenga una variable o concepto más sencillo, es más fácil llegar a una definición operacional.

4. La cuarta regla consiste en sintetizar los datos elementales obtenidos para construir un índice de medición.

Székely (1970) señala que al aplicar la medición en el campo de la psicología es necesario considerar las propiedades de los fenómenos o atributos psicológicos que se desea medir para saber qué tipo de escala es la más adecuada en relación con esas propiedades. Este conocimiento permite establecer los índices estadísticos que pueden ser calculados partiendo de los puntos obtenidos con diferentes escalas.

2.4. Las escalas de medición.

Stevens (1946, citado en Nunally en Bernstein, 1995), propuso que las mediciones entran en cuatro tipos de escalas principales: nominal, ordinal, de intervalo y de razón. Estos niveles permitieron que se ejecutaran en las medidas procedimientos cuantitativos cada vez más complicados. (Ver cuadro 1).

Cuadro 1. Niveles de medición de Stevens, operaciones definitorias básicas, transformaciones permisibles, ejemplos de estadísticas permisibles, y ejemplos.

Escala	Operación básica.	Transformaciones permisibles.	Estadísticas permisibles.	Ejemplos
Nominal	igualdad vs desigualdad	Cualquiera uno a uno	Número de casos, moda	Números telefónicos
Ordinal	Mayor que vs. menor que	Incremento monotónico	Mediana, percentiles, estadística de orden	Dureza de los minerales, rango de clase
Intervalo	Igualdad o diferencias de los intervalos	Lineal general	Media aritmética, varianza, correlación de Pearson	Puntajes de pruebas
Razón	Igualdad de razones	Multiplicativa	Media geométrica	Temperatura (Kelvin)

Fuente: Adaptada de Stevens (1951), en Nunally (en Bernstein, 1995), p. 13.

Nunally (*opus cit.*) al definir las escalas menciona lo siguiente:

2.4.1. Escalas nominales.

Las escalas nominales contienen reglas para decidir si dos objetos son equivalentes o no equivalentes para una categorización. Equivalencia significa que dos objetos tienen una propiedad crítica en común, por ejemplo, dos personas son mujeres. El resultado de una escala nominal es una serie de clases a las cuales podría darse una designación numérica. El número desempeña un papel meramente nominal. Las escalas nominales operan en el nivel más inferior de la medición, y admiten solamente la clasificación.

2.4.2. Escalas ordinales.

Constituyen un nivel de medición superior al comprendido en la escala nominal. En la escala ordinal, el número sirve para establecer un ordenamiento entre las distintas categorías, sobre la base del grado de posesión de un atributo. Si se ordena de menor a mayor, o viceversa, una serie de categorías sobre la base de las variaciones de un atributo, se habrá hecho una escala ordinal. En cada una de las categorías, la frecuencia debiera ser igual a la unidad. Eso permitiría una discriminación tal, que establecería una categoría para cada sujeto. En la práctica se acepta que intervenga mayor número de casos o sujetos. La práctica consistente en incluir varios sujetos en una misma categoría está determinada por la falta de discriminación de nuestros medios de observación o medida. La única condición que se exige

-cuando ordenamos de mayor a menor- es que la categoría primera represente un nivel de posesión de un atributo mayor que el representado en la categoría segunda; que la segunda sea mayor que la tercera y así sucesivamente. En las escalas ordinales no se exige que la distancia que separa la categoría primera de la segunda sea igual a la distancia que separa la categoría segunda de la tercera, etc. En realidad, la escala ordinal procede a una clasificación cuantitativa de las categorías a lo largo de un continuo que representa el atributo medido u observado.

2.4.3. Escalas de intervalos o distancias iguales.

Representan un nivel de medición superior al comprendido por la escala ordinal. Su característica fundamental es que las distancias numéricamente iguales en la escala representan distancias también iguales en el atributo o cualidad medido. Supongamos que, sobre la base de la posesión de un atributo determinado, establecemos cuatro categorías: *a*, *b*, *c*, *d*. La categoría *a* representa tres unidades, la *b* seis unidades, la *c* nueve unidades y la *d* doce unidades. La distancia que media entre *a* y *b* es igual a la distancia entre *b* y *c*; la distancia entre *a* y *c* es igual a la distancia entre *b* y *d*. Puede afirmarse también que la diferencia entre *b* y *d* es igual a la diferencia entre *a* y *c*, etcétera.

Aunque se da a entender que la suma de intervalos o distancias es posible, es necesario aclarar que la posibilidad de la adición no se da en un sentido absoluto. La razón es que en esas escalas el punto cero se ha establecido arbitrariamente. Conociendo estas limitaciones podemos tratarlas como si pertenecieran a una escala de nivel de medición superior con unidades iguales y punto cero absoluto.

2.4.4. Escalas con punto cero absoluto y unidades iguales.

Admiten todas las operaciones aritméticas y permiten establecer la magnitud absoluta de cualquier aspecto mensurable de un atributo. Las escalas de unidades iguales sólo permiten establecer la magnitud relativa de los aspectos mensurables de un atributo, con referencia a un cero arbitrario. No hay en psicología instrumentos de medida que admitan el uso de escalas con cero absoluto.

Las escalas de medida, cualquiera que sea su naturaleza, están afectadas por un margen de error. Las escalas de medida que se aplican en psicología, al carecer de cero absoluto y

disponer de unidades sólo relativamente constantes, dan márgenes de error mayores que otros tipos de escalas, como las utilizadas en las ciencias exactas.

2.5. La teoría del error de medición.

2.5.1. El concepto de error de medición.

Reidl (1990) introduce brevemente que el error de medición puede ser de dos tipos: sistemático y aleatorio. El error sistemático es aquel que se comete cuando el instrumento de medición arroja datos que subestiman o sobrestiman la cantidad real de atributo poseído por los sujetos o eventos que se están midiendo. Si este es el caso, y las mediciones del atributo en cuestión se realizan con el mismo instrumento, el problema no es tan grave, ya que el error es constante y todos los eventos y objetos lo tienen en la misma cantidad; en este caso, aún se pueden obtener diferencias entre eventos o personas.

Nunnally (en Bernstein, 1995) menciona que cuando el error es sistemático puede:

- a) Afectar a todas las observaciones por igual y ser un error constante, o
- b) Afectar a ciertos tipos de observaciones de manera diferente que a otras y ser un sesgo.

“Por ejemplo, un termómetro mal calibrado que siempre marque tres grados de más ilustra un error constante en las ciencias físicas. Si el termómetro fuera sensible a algún atributo irrelevante tal como el color o la densidad de lo que se está midiendo, el error sería un sesgo. Por último, el error aleatorio sería introducido si la persona que lee el termómetro transpusiera los dígitos de vez en cuando mientras registraba las observaciones.” (p. 237).

Reidl (1990) dice que el error aleatorio se refiere al hecho de que algunas veces se sobrestima la cantidad de atributo poseída por el evento u objeto, y otras se subestima; además estas sobre o subestimaciones no son iguales; es decir el error de medición es el azar: algunas veces por encima, otras por debajo de la cantidad real, y nunca en la misma cuantía. Esto es lo que sucede con los instrumentos de medición que se emplean en las ciencias sociales. Esto trae como resultado que no se puede estar seguro de las diferencias reportadas por un instrumento en cuanto a si la cantidad de atributo poseído por el evento o sujeto son reales. Aunque los errores aleatorios nunca se pueden eliminar, se deben minimizar hasta donde sea posible.

Nunnally (en Bernstein, 1995) subraya que un instrumento de medición es más confiable cuanto menor sea el error de medición. La confiabilidad de un instrumento de medición es el

grado hasta el cual las medidas sean repetibles:

1. Cuando diferentes personas hacen las mediciones,
2. Con instrumentos alternativos que intentan medir lo mismo, y
3. Cuando hay variación incidental en las condiciones de la medición.

2.6. Modelos de escalamiento.

Un modelo de escalamiento es un plan internamente consistente para desarrollar un nuevo instrumento de medición.

Reidl (1990) dice que el propósito de cualquier modelo escalar es el de generar un continuo en el que se puedan localizar personas u objetos.

Uno de los aspectos más importantes que se debe tomar en cuenta al construir una escala de medición es el de la especificación del atributo a medir; esto es, qué es lo que se quiere medir, cuál es la propiedad que pueden tener los objetos, en qué cantidad la tienen y qué se desea representar numéricamente. Especificar un atributo significa definir la propiedad.

Cuando se habla de escalar y de métodos escalares, por lo general se hace referencia al escalamiento de estímulos. Cuando se habla de escalar personas respecto a un atributo se hace referencia a la construcción de pruebas.

Para lograr construir escalas se requiere que los sujetos den respuestas frente a ciertos estímulos. Las respuestas que se pueden dar son de dos tipos: juicios y sentimientos. Los juicios se refieren a aquellas respuestas que se pueden comprobar en forma empírica, si son correctas o no. Los sentimientos se referirán a las reacciones, preferencias o actitudes, opiniones o intereses del sujeto, donde por supuesto no existe una respuesta correcta.

El escalamiento de estímulos tiene que ver con respuestas de sentimiento, mientras que el escalamiento de personas tiene que ver con respuestas de juicio.

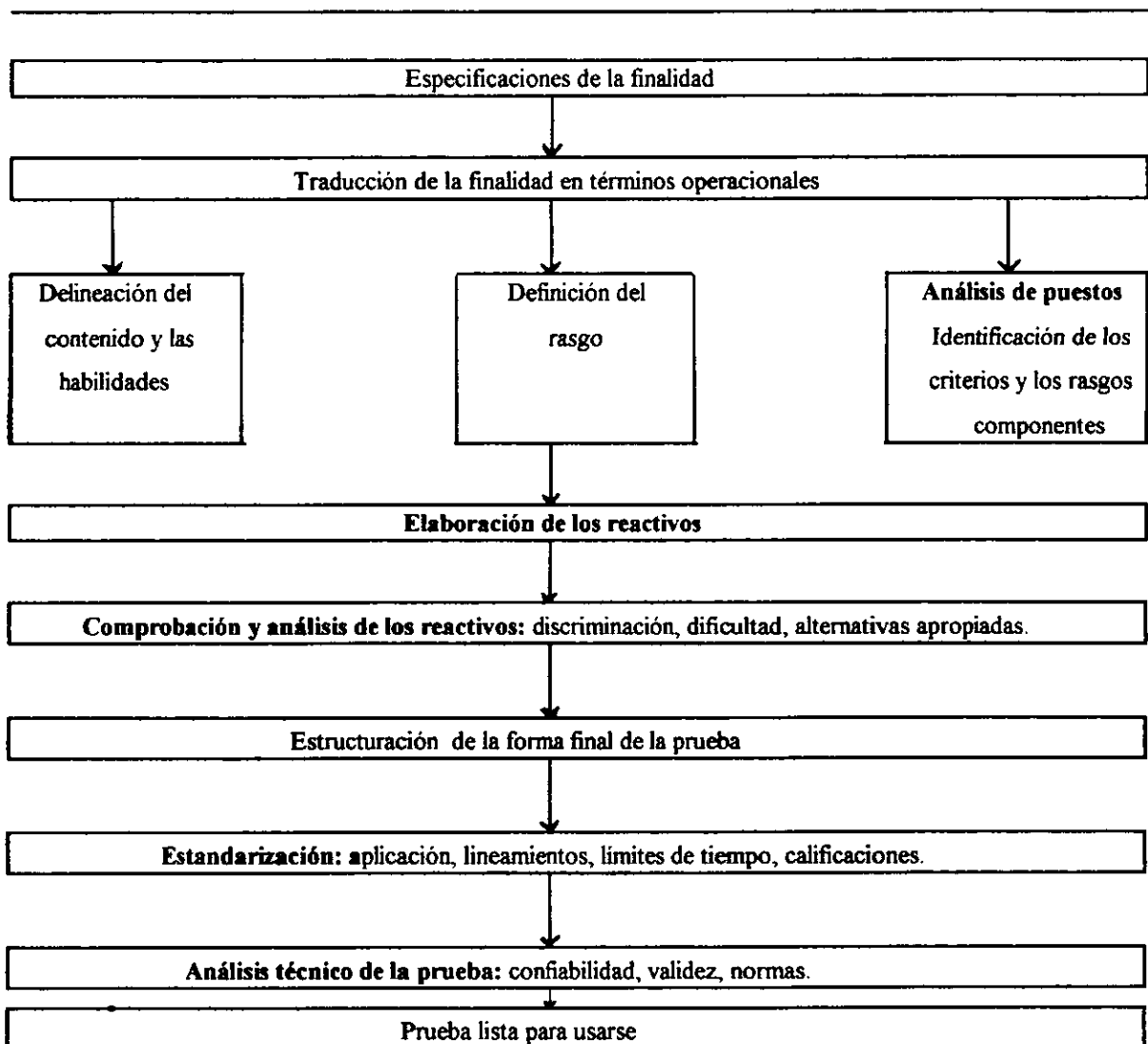
2.7. La construcción de una prueba psicológica.

La construcción de una prueba psicológica responde siempre a la necesidad de aplicar algún nivel de medida a una función psicológica o aspecto de la conducta. Brown (1980) señala que aunque el proceso exacto varía, dependiendo de la prueba psicológica, se puede establecer una secuencia general en los pasos de construcción de las pruebas psicológicas, que incluyen:

1. La especificación de su finalidad,
2. La construcción y la comprobación de reactivos,
3. El ensamblaje de la forma final de la prueba,
4. El análisis y estandarización de las calificaciones de la prueba psicológica y
5. El análisis técnico de la prueba.

En la construcción de cualquier prueba específica se pueden omitir algunas de las etapas, puede variar su orden, sin embargo, es posible identificar una secuencia general. Ver cuadro 2.

Cuadro 2. Etapas en la construcción de una prueba.



Fuente: Brown (1980), p.21.

Al hablar de la construcción de pruebas psicológicas Brown (1980) comienza por

clasificarlas de acuerdo con sus funciones, y las distingue en dos :

1. Pruebas representativas, que son las que contienen reactivos similares a las conductas que les interesa medir y

2. Pruebas predictoras, que están constituidas por reactivos que no enfocan alguna conducta, sus calificaciones permiten predecir el modo en que se desempeñará una persona en una situación cualitativamente diferente.

Según Brown (1980) y Reidl (1990) los pasos a seguir en la construcción de un instrumento son los siguientes:

2.7.1. Finalidad de la prueba.

Brown (1980) afirma que el constructor de pruebas tiene que tomar, al inicio, dos decisiones importantes: determinar el contenido de la prueba y su formato. O sea que debe determinar las conductas, los conocimientos, las actitudes, etc. que cubrirá ésta y cómo se presentarán los reactivos. Sin embargo, antes de poder tomar esas decisiones, deberá hacerse dos preguntas: ¿para qué fines servirá la prueba? y ¿qué grupo de sujetos serán sometidos al examen? Las respuestas a estas dos preguntas establecerán límites y sugerirán cómo seguir adelante en el proceso de construcción de la prueba.

Reidl (1990) se refiere a esta etapa con el término de relevancia de la prueba. Este término connota que el investigador debe tener muy claro el tipo de datos que requiere su estudio. Específicamente el investigador debe explicitar las razones que lo llevan a incluir cada una de las preguntas o reactivos, y no sólo por qué se hace cada pregunta, sino qué se va a hacer con la información que ésta proporcione. Esto requiere decidir de antemano ciertos aspectos de codificación y análisis.

2.7.1.1. Contenido de las pruebas.

Para elaborar la prueba, dice Brown (1980) el constructor deberá traducir sus propósitos en términos operacionales. Es decir que debe especificar las operaciones que utilizará para su construcción. En el caso de las pruebas representativas, si el instrumento se desarrolla para medir rendimiento se deberá bosquejar el contenido y las habilidades que cubrirá la prueba; y si el instrumento lo que pretende medir es algún rasgo o característica psicológica dada, el constructor deberá definir, de modo tan explícito como sea posible, el rasgo que se medirá. A

continuación, tendrá que indicar y describir las conductas, habilidades o aptitudes que sean indicadoras del rasgo.

En el caso de las pruebas psicológicas con fines predictivos, la primera etapa consiste en efectuar un análisis sistemático de la ejecución que se trata de predecir. Los resultados de ese análisis deben indicar una medida apropiada de criterio e identificar los rasgos y las conductas que se necesitan para que la ejecución sea apropiada.

A continuación, hay que construir un plan de la prueba psicológica que sirve para dos fines:

a) En la etapa de elaboración de los reactivos, indica la cantidad y el tipo de ellos que es preciso redactar;

b) En la etapa terminal, se puede comparar la distribución real de los reactivos en la forma final de la prueba con las proporciones indicadas en el plan, para determinar si los reactivos sirven de hecho para efectuar un muestreo adecuado del universo que se trate.

2.7.1.2. Formato de la prueba.

El formato es el modo en que se presentarán los reactivos de una prueba. A continuación se mencionan algunos de los formatos más comunes, señalados por Brown (1980):

1. Respuesta alternativa vs. libre. En un reactivo de respuesta alternativa, el examinando elige la respuesta apropiada de entre varias alternativas, como en los reactivos de comparación, elección múltiple o verdadero y falso. En los reactivos de respuesta libre, el sujeto proporciona una respuesta, como en completamiento de frases, respuestas cortas o preguntas de ensayo.

2. Pruebas de velocidad vs. pruebas de poder. En una prueba de velocidad, los reactivos son sencillos con un tiempo límite de respuesta, así, la calificación es un índice de la velocidad de las respuestas. Una prueba de poder se compone de reactivos de dificultad variable y tiene un límite de tiempo que permite dar respuesta a todos ellos. La calificación expresa el nivel de dificultad de los reactivos a los que puede responder el sujeto.

3. Ejecución máxima vs. ejecución típica. En las pruebas de ejecución máxima, el sujeto recibe instrucciones de tratar de obtener la mejor calificación que pueda. Se exige del examinado su máximo rendimiento en las distintas tareas que componen la prueba. Este es el

caso de las pruebas de inteligencia, habilidades, aptitudes específicas y de rendimiento. En las pruebas de ejecución típica, lo que interesa es conocer el comportamiento habitual o normal del sujeto. Estas también son llamadas pruebas de clasificación tipológica o cualitativa. En esta clasificación se sitúan las escalas o inventarios de personalidad, de intereses o preferencias, de valores, de actitudes y técnicas proyectivas.

4. Papel y lápiz vs. ejecución. Esta distinción se refiere al modo en que se presentan los reactivos de una prueba y cómo se dan las respuestas. Las pruebas de ejecución implican, con frecuencia, la manipulación de algún aparato u objeto.

5. Aplicación colectiva vs. individual. Las pruebas colectivas se pueden aplicar a más de un solo individuo a la vez, por lo que suelen ser de papel y lápiz. Las pruebas individuales se pueden aplicar sólo a una persona cada vez y pueden ser de ejecución o verbales.

6. Pruebas estructuradas vs. proyectivas. En una prueba estructurada se especifican con claridad los estímulos y las tareas del sujeto; en una prueba proyectiva, los estímulos y las tareas son ambiguos.

2.7.2. Construcción de la prueba.

Reidl (1990) dice que una vez que se han definido la o las variables desde el punto de vista conceptual, indicativa u operacional, se procede a llevar a cabo la redacción de los reactivos. La definición puede contemplar una o varias dimensiones, lo que determinará si se elaborará un instrumento homogéneo o heterogéneo. La definición de la variable pudo haber provenido de una o todas las siguientes fuentes: literatura especializada, literatura común u ordinaria, o de los mismos sujetos que serán investigados o para los cuáles se elaborará el instrumento.

2.7.2.1. Elaboración de reactivos.

Reidl (1990) señala que los tipos generales de reactivos o preguntas que un investigador de las ciencias sociales emplea se pueden clasificar en tres grandes clases:

1. Información de identificación;
2. Datos factuales tipo censo o antecedentes sociales;
3. Reactivos o preguntas acerca del objeto de estudio, que pueden ser de diferentes formas, dependiendo de los objetivos del estudio. Existen las siguientes formas:

3.1. Preguntas abiertas. Esta forma se usa en las etapas iniciales, exploratorias o piloto de un estudio, ya sea de investigación o de elaboración de un instrumento. Estas preguntas proporcionan gran libertad en las respuestas del sujeto investigado.

3.2. Método de caso o historias libres. Esta manera de obtener información se realiza por medio de extensas conversaciones con el o los sujetos.

3.3. Preguntas dicotómicas. Estas preguntas son aquellas que piden que se responda con sí o no, de acuerdo o desacuerdo, cierto o falso, correcto o incorrecto, bueno o malo.

3.4. Listas de cotejo. Este tipo de preguntas consiste en la presentación del problema o pregunta seguida por una lista de tres a quince posibles respuestas de entre las cuales se le pide al sujeto que elija una.

3.5. Ordenación de reactivos. Se le pide al sujeto que ordene una lista de afirmaciones, palabras, frases, dibujos u otras cosas en orden de su preferencia.

3.6. Reactivos de opción múltiple. Los reactivos de opción múltiple se presentan de manera que el respondiente tiene que escoger entre las varias posibles respuestas, aquella que mejor represente su opinión, creencia, actitud, etc. Este tipo de preguntas es particularmente útil cuando el tema en cuestión no puede quedar bien representado por una pregunta dicotómica. En este formato, se da la oportunidad de expresión o manifestación de todos los grados de opinión, actitud o rasgo. Esta técnica fue desarrollada por Likert en 1932, como escala de actitudes. Las preguntas de opción múltiple pueden adoptar diversas presentaciones:

- a) escalas de clasificación de tres puntos,
- b) escalas de clasificación de cuatro puntos,
- c) escalas de clasificación de cinco puntos, y
- d) escalas gráficas de clasificación.

Nadelsticher (1983), por su parte, señala que en el Método de rangos sumariados de Likert, que así se llama la técnica que creó Likert en 1932, en primer lugar se elabora un cuestionario piloto con un mínimo de 70 preguntas por cada dimensión: 35 son favorables y 35 desfavorables al objeto medido, con una escala de clasificación de cinco puntos o alternativas que son las siguientes:

1. Totalmente de acuerdo (TA).
2. Acuerdo (A)
3. Indiferente (I)

4. Desacuerdo (D)
5. Totalmente en desacuerdo (TD).

La redacción de las alternativas puede variar, por ejemplo puede ir de Muy perfecto a Imperfecto; lo importante es que la redacción exprese una escala de preferencias para el sujeto.

2.7.2.2. Análisis de reactivos.

Reidl (1990) menciona que el primer paso estadístico que se lleva a cabo al inicio de la construcción de una prueba es el análisis de reactivos, que consiste en aplicar la prueba a una muestra de personas, similares a las que se pretenden examinar con la misma, para verificar la reacción de éstas ante el instrumento. Con esta aplicación puede verificarse:

1. El nivel de dificultad de la prueba,
2. El poder de discriminación de los reactivos.

1. *El nivel de dificultad de la prueba* depende de sus reactivos y de su distribución en la prueba (Anastasi, 1967). Costa Neiva (1996) propone que éste puede ser verificado a través de la distribución de las puntuaciones totales de una muestra de sujetos. Un nivel de dificultad normal proporciona una curva de distribución de las puntuaciones semejante a la curva normal, donde hay una acumulación de sujetos en el centro (puntuaciones medianas) y una disminución gradual de sujetos al acercarse a los extremos de la curva (puntuaciones altas o bajas). (Figura 1).

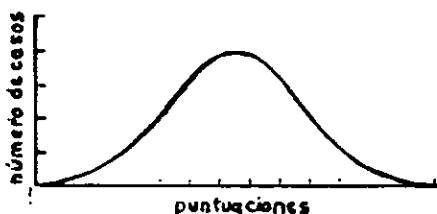


Figura 1. Curva de distribución normal. Fuente: Costa Neiva (1996), p.28.

Si la prueba es muy difícil, la curva de distribución de las puntuaciones de los sujetos presenta una acumulación en el extremo inferior o sea un gran número de sujetos con puntuaciones bajas (figura 2). Esto indica que faltan a la prueba suficientes reactivos fáciles. Sin embargo, cuando la prueba es muy fácil, la acumulación de puntuaciones se da en el extremo superior (puntuaciones altas) y por lo tanto le faltan reactivos difíciles (figura 3).



Figura 2. Distribución desviada (a la izquierda)



Figura 3. Distribución desviada (a la derecha)

Fuente: Costa Neiva (1996), p. 29.

Cuando sucede uno de los dos casos arriba mencionados suele modificarse el nivel de dificultad de la prueba hasta que la gráfica de distribución de las puntuaciones se aproxime a la curva normal. Para tal, se pueden añadir reactivos más fáciles o más difíciles, cambiar la posición de los reactivos, eliminar o modificar reactivos, revisar pesos o ponderaciones de las puntuaciones asignadas a determinadas respuestas, etc. (Anastasi, 1967).

2. *La verificación del poder discriminativo de los reactivos* se hace a través de un análisis estadístico que “indica la amplitud con la que el reactivo mide lo que debe medir” (Brown 1980, p. 31). El procedimiento más común es el que, utilizando la calificación de la prueba como medida de criterio, correlaciona las respuestas a los reactivos individuales con las calificaciones totales obtenidas en la prueba. Se supone que las personas que obtuvieron calificaciones totales altas respondieron a un reactivo de forma correcta con mayor frecuencia que aquellas que obtuvieron calificaciones totales bajas. Si esto ocurre, el reactivo es considerado “bueno”, o sea que discrimina relativamente bien lo que se desea medir.

2.8. Análisis técnico de la prueba.

El análisis técnico de la prueba, como un todo, supone diversos procedimientos, que objetivan la minimización de los efectos de variables que pudieron intervenir en la medición. En otras palabras supone el control de los errores. Este análisis contiene los siguientes procedimientos:

1. Estandarización.
2. Confiabilidad.
3. Validez. (Costa Neiva, 1996).

2.8.1. Estandarización de la prueba.

Brown (1980) indica que en la medición al administrar una prueba, la meta es obtener una estimación tan precisa como sea posible de la ejecución de los sujetos. La estimación precisa en las pruebas psicológicas, igual que en otros procedimientos científicos, depende del control de los errores, esto se logra haciendo que la situación de prueba sea lo más semejante posible para todos los individuos. El proceso de desarrollo de esos controles se llama estandarización.

Los diferentes autores consultados, y ya mencionados, definen la estandarización como la existencia de reglas específicas para la administración de la prueba, a partir de que a cada sujeto le son presentados los mismos reactivos o equivalentes.

En segundo lugar, señalan que las condiciones de aplicación también deben ser, en la medida de lo posible, idénticas. Para tal, las instrucciones de aplicación, las demostraciones previas, los límites de tiempo, las formas de resolver dudas, la disposición del material, etc. deben ser rigurosamente las mismas para todos los sujetos. Por ello, el manual de una prueba debe proveer estas informaciones con el máximo de detalles.

Finalmente, el procedimiento de calificación debe ser lo más uniforme posible. Se necesita para tal un alto grado de objetividad en la puntuación y calificación de las respuestas de los sujetos.

Otro aspecto importante de la estandarización del procedimiento de puntuación es el establecimiento de *normas*. “Una norma es la actuación media o normal” (Anastasi, 1967, p. 22) de un grupo de personas similar al que se destina la prueba. Al aplicarse la prueba a este grupo llamado *norma de estandarización, muestra de tipificación o grupo normativo*, se obtiene el promedio de actuación del grupo, así como los grados de desviación por arriba y por abajo del promedio. La comparación del resultado de un sujeto con la norma, permite determinar la posición del mismo con respecto a este grupo. Muchas veces se necesita comparar el resultado de un sujeto con diferentes grupos (según el sexo, la edad, el nivel de escolaridad, la ocupación, etc.) y por lo tanto son necesarios varios grupos normativos. Además las normas pueden estar expresadas en diferentes unidades de calificación: CI de desviación, percentil, estantina, puntuación T, puntuación Z, etc.

2.8.2. *Confiabilidad de la prueba.*

2.8.2.1. *Concepto de confiabilidad.*

La confiabilidad es definida por los diversos autores como la ausencia relativa de errores de medición en un instrumento. Así se podría considerar sinónimo de seguridad, consistencia, predictibilidad y exactitud. A continuación se presentan varias definiciones:

Anastasi (1967) afirma que “el término fiabilidad de un test, según se usa en psicometría, significa siempre estabilidad o consistencia. La fiabilidad de un test es la consistencia de las puntuaciones obtenidas por las mismas personas cuando se les aplica otra vez el mismo test o una forma equivalente” (p. 26).

Pick y López (1979) definen a la confiabilidad como la estabilidad o consistencia de los resultados obtenidos. Por medio de la confiabilidad puede advertirse que tan consistentes, exactos y estables son los resultados alcanzados al aplicar los instrumentos.

Para Reidl (1990), “ por confiabilidad se entiende el hecho de poder “confiar” en que las mediciones realizadas sean correctas. Que las magnitudes de aquellos aspectos medidos, sean de confiar, esto es, que la confiabilidad de un instrumento se refleja en que el mismo instrumento empleado en dos o más ocasiones para medir los mismos atributos de los mismos objetos o sujetos, arroja magnitudes iguales. La confiabilidad tiene que ver con el error de medición. La cantidad de error que contenga un instrumento no debe sobrepasar de cierta magnitud. La mayoría de los investigadores han optado por aceptar como confiables a los instrumentos que tengan 15% o menos de error sistemático, que es el que se tiene que determinar de forma empírica. En otras palabras, se dice que un instrumento es confiable si su coeficiente arroja un valor de 0.85 o más.” (Cap. VIII, p. 1)

Brown (1980) considera que la confiabilidad “es el grado de consistencia de una medición. 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 más 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.” (p. 80).

Como ya se dijo en el apartado 2.5 toda medida implica un cierto error. El objetivo de una prueba psicológica es reducir, al mínimo, el error de medición. Al reducirlo se eleva el nivel de confiabilidad de la prueba.

Thorndike (1973, citado por Costa Neiva, 1996) apunta cuatro fuentes principales de error

que resultan de:

1. Variaciones surgidas dentro del procedimiento mismo de medición, que pueden estar relacionadas con las condiciones de aplicación de la prueba, tales como: la situación física, las instrucciones, los errores en el registro del tiempo, los factores de distracción, etc.

2. Variaciones en el sujeto mismo, a corto plazo (salud, atención, motivación, ansiedad, experiencia con la prueba, etc.) o a largo plazo (maduración, aprendizaje, cambios en su ambiente, etc.)

3. Variaciones procedentes de la muestra específica de tareas (reactivos) elegidas para representar el área de la conducta que mide la prueba.

4. Variaciones en la velocidad de trabajo del sujeto.

Existen diferentes procedimientos para examinar la cantidad de error en la medición, o sea, estimar el grado de consistencia de la misma. Brown (1980) diferencia dos tipos de consistencia, que aunque están relacionados, considera útil separar para fines didácticos. La *confiabilidad* verifica la consistencia de la medida a través del tiempo (estabilidad) y a lo largo de formas equivalentes de la prueba (equivalencia). La *homogeneidad o consistencia interna* verifica si todos los reactivos de la prueba miden o no el mismo rasgo. Serán analizados, por separado, los distintos procedimientos.

2.8.2.2. Procedimientos para verificar la confiabilidad de una prueba psicológica.

Los autores consultados, nombrados anteriormente, mencionan diferentes procedimientos que permiten verificar el nivel de confiabilidad de una prueba psicológica:

2.8.2.2.1. Repetición de la misma prueba (test-retest).

Este procedimiento consiste en aplicar la misma prueba dos veces. La medida de confiabilidad es la correlación entre las calificaciones obtenidas en la primera aplicación y en la reaplicación. Esta medida se denomina *coeficiente de estabilidad* y puede ser realizado de dos maneras distintas:

1. *Sin intervalo de tiempo* entre las dos mediciones, o sea, la reaplicación es realizada en seguida de la aplicación. En este caso son controlados dos tipos de errores provenientes de:

a) variaciones en el procedimiento mismo de medición y

b) variaciones en la velocidad de trabajo del sujeto.

2. *Con intervalo de tiempo* entre las mediciones, o sea, la reaplicación se hace después de cierto tiempo, que puede ser de días o de años. Tres errores son controlados en este procedimiento, que resultan de:

- a) variaciones en el procedimiento mismo de medición,
- b) variaciones en el sujeto de una a otra aplicación y
- c) variaciones en la velocidad de trabajo del sujeto.

Este tipo de procedimiento no debe ser utilizado en el caso de rasgos que sean, por principio, inestables o cuando se produzca algún tipo de aprendizaje diferencial entre las dos aplicaciones (por ejemplo un programa de estudios o entrenamiento). Otro problema es el del efecto de la práctica, sobre todo cuando la aplicación y reaplicación son realizadas sin intervalo de tiempo o con un intervalo muy corto entre las dos. En este caso “algunas personas pueden recordar sus respuestas o aprender algo sobre la técnica de pasar una prueba” (Brown, 1980, p. 85).

2.8.2.2.2. *Aplicación de formas paralelas (o equivalentes) de la prueba.*

Para este procedimiento hay que disponer de formas equivalentes de la prueba, que sigan la misma especificación pero que sean compuestas de muestras separadas de la conducta que se pretende medir. Las formas paralelas deben por tanto cubrir el mismo contenido, utilizar el mismo número de reactivos y poseer el mismo grado de dificultad. La medida de confiabilidad es la correlación entre las puntuaciones obtenidas en las dos formas de la prueba. Este procedimiento también puede ser realizado de dos maneras:

1. *Sin intervalo de tiempo* entre la aplicación de las dos formas equivalentes de la prueba. En este caso la correlación entre las dos medidas ofrece un *coeficiente de equivalencia*. Los tipos de errores tomados en cuenta son los que provienen de:

- a) variaciones en el procedimiento mismo de medición,
- b) variaciones en la velocidad de trabajo del sujeto y
- c) variaciones procedentes de la muestra de tareas.

2. *Con intervalo de tiempo* entre la aplicación de las dos formas paralelas de la prueba. La correlación entre las dos medidas ofrece, en este caso un *coeficiente de estabilidad y equivalencia*. Con este procedimiento pueden controlarse los cuatro tipos de errores mencionados anteriormente, que provienen de:

- a) variaciones en el procedimiento mismo de medición,
- b) variaciones en el sujeto de una vez a otra,
- c) variaciones en la velocidad de trabajo y
- d) variaciones procedentes de la muestra de tareas.

El procedimiento que utiliza la aplicación de formas paralelas con intervalo de tiempo es el que proporciona el coeficiente de confiabilidad más preciso y riguroso pues permite controlar las distintas fuentes de error. Sin embargo, él plantea problemas prácticos importantes, puesto que exige que se disponga de dos formas paralelas de la prueba y del tiempo necesarios para aplicarlas por separado.

2.8.2.2.3. *Confiabilidad por mitades.*

Este procedimiento, también llamado de test subdividido (split half), es el más ampliamente utilizado para estimar la confiabilidad a partir de una sola administración de la prueba (Thorndike, 1973, citado por Costa Neiva, 1996). Según Murphy y Davidshofer (1994, citado por Costa Neiva, 1996), este procedimiento soluciona los dos problemas prácticos del método de formas paralelas: la dificultad de desarrollar formas paralelas y la necesidad de tiempo para las aplicaciones.

Para realizarlo se aplica la prueba y después se divide en dos mitades supuestamente equivalentes y equilibradas en relación al contenido y nivel de dificultad. Para la división puede procederse de distintas maneras. La más común es la que utiliza los reactivos pares, en una de la formas, y los impares en la otra. Sin embargo, este tipo de división no puede ser manejada cuando existe alguna desviación sistemática en la prueba como por ejemplo, los tipos alternativos de reactivos o reactivos en que la respuesta depende de la anterior (Brown, 1980).

La correlación entre las calificaciones de las dos mitades de la prueba proporciona una estimación de la confiabilidad. Mientras tanto, como cada conjunto de calificaciones está basado en la mitad de los reactivos de la prueba original completa y la confiabilidad de una prueba depende de su longitud, la correlación entre las dos mitades proporciona una estimación de la confiabilidad inferior a la que podría esperarse de la prueba original. Para estimar la confiabilidad de la prueba original completa se utiliza la siguiente fórmula, propuesta por Spearman-Brown:

$$r_{11} = \frac{2r_{1/2 \ 1/2}}{1 + r_{1/2 \ 1/2}}$$

en donde r_{11} = confiabilidad de la prueba original completa

$r_{1/2 \ 1/2}$ = correlación entre las mitades de la prueba.

Este procedimiento toma en cuenta solamente dos tipos de errores provenientes de variaciones en el procedimiento mismo de medición y variaciones procedentes de la muestra de tareas. Además, presenta algunas limitaciones como:

a) dos medias pruebas pueden ser más semejantes que dos formas paralelas separadas; por ejemplo, cuando los reactivos están basados en un mismo material de referencia -texto, experimento- pueden provocar distorsiones en el coeficiente de confiabilidad y

b) este tipo de coeficiente no tiene significado en pruebas de tiempo limitado donde influya la rapidez; el factor velocidad tenderá a inflar el coeficiente de confiabilidad (Brown, 1980).

Sin embargo, la utilización de este procedimiento presenta ventajas prácticas importantes relacionadas a la aplicación de un mismo material de prueba, sólo una vez. Hay que tomar también en cuenta que cuanto mayor sea la muestra de conducta de la persona, tanto más confiable será la medida. Además, entre mayor sea la muestra de conducta registrada menos dependerá la medida de elementos fortuitos en la conducta del sujeto o en la muestra específica de tareas. Por ello, la utilización de un gran número de reactivos es un medio utilizado para garantizar la confiabilidad de una prueba.

2.8.2.3. Interpretación de coeficientes de confiabilidad.

Costa Neiva (1996) dice que en la interpretación de un coeficiente de confiabilidad es importante tomar en cuenta que éste es específico de la muestra utilizada, de la situación de prueba y del método utilizado para su cálculo. Por eso hay que verificar la confiabilidad de una prueba en una gran variedad de situaciones y con diferentes muestras y procedimientos. Brown (1980) propone tres modos de interpretación de los coeficientes de confiabilidad:

2.8.2.3.1. Correlación entre calificaciones reales y obtenidas.

Se puede interpretar un coeficiente de confiabilidad tomando como base que la proporción de la variancia en las calificaciones obtenidas representa la variación de las calificaciones reales. Por ejemplo, si $r = 0.80$, entonces 80% de la variabilidad en las calificaciones se debe a diferencias en las calificaciones reales y 20% se debe a errores de medición.

2.8.2.3.2. Comparación con la confiabilidad de otras pruebas.

Este es un criterio muy utilizado para decidir si la prueba puede o no ser considerada confiable. En este caso se compara el coeficiente de confiabilidad de la prueba con el nivel de confiabilidad de pruebas similares ya existentes.

2.8.2.3.3. Errores en las calificaciones individuales.

Puede interpretarse el coeficiente de confiabilidad evaluando el cambio que podrá pasar en la calificación de un sujeto al volver a contestar la prueba. Para eso hay que verificar cuál es el error estándar de la medición. Thorndike (1973, p. 204; citado por Costa Neiva, 1996), presenta un cuadro del error estándar de medición para diferentes valores del coeficiente de confiabilidad.

2.8.2.4. Procedimientos para verificar la homogeneidad de una prueba psicológica.

Los procedimientos para verificar la *homogeneidad* o *consistencia interna* de una prueba buscan determinar el grado de las relaciones recíprocas entre los reactivos que la componen. Brown (1980), afirma que “se puede definir la homogeneidad como la consistencia en la ejecución de todos los reactivos de una prueba psicológica”. (p. 102). Una correlación positiva entre las calificaciones de los distintos reactivos de una prueba indica que ésta es homogénea. Si la prueba está compuesta de subpruebas heterogéneas, se procede al estudio de la homogeneidad de cada subprueba, cuyos reactivos deben por tanto ser homogéneos.

Los procedimientos más comunes para verificar la homogeneidad o consistencia interna de una prueba son:

2.8.2.4.1. Coeficiente alfa.

Este coeficiente, propuesto por Cronbach (1951, citado por Costa Neiva, 1996), es utilizado con bastante frecuencia. Su utilización es limitada a las pruebas cuyos reactivos no son calificados de forma dicótoma (correcto o incorrecto). Su fórmula es:

$$\text{Alfa} = \frac{n}{n - 1} \left[1 - \frac{\sum_i s_i^2}{s_x^2} \right]$$

en donde: n = número de reactivos de la prueba
 Σ = sumatoria
 s_i = desviación estándar de cada reactivo
 s_x = desviación estándar de la prueba

2.8.2.4.2. Fórmula de Kuder-Richardson (KR-20).

Esta fórmula, propuesta por Kuder y Richardson en 1937, es utilizada cuando se califican los reactivos de la prueba de forma dicótoma y suele ser considerada como un caso especial del coeficiente alfa:

$$r_{11} = \frac{n}{n - 1} \left[\frac{\sum_x s_x^2 pq}{s_x^2} \right]$$

en donde:

n = número de reactivos de la prueba
 s_x = desviación estándar de la prueba
 Σ = sumatoria
 p = proporción de individuos que contestan bien un reactivo
 q = $1 - p$ = proporción de individuos que fallan en este reactivo.

Esta fórmula no se aplica a pruebas psicológicas de velocidad, pues, para el cálculo del componente Σpq , es necesario que cada reactivo sea probado por todas las personas, lo que no siempre ocurre en las pruebas de velocidad.

Estos mismos autores propusieron una segunda fórmula (KR-21) que supone que todos los reactivos de la prueba tienen el mismo grado de dificultad. Esta fórmula es menos utilizada puesto que tal situación raramente ocurre.

2.8.2.4.3 Análisis factorial.

Este procedimiento es uno de los más utilizados y en general es considerado como el mejor por varios autores (Brown, 1980). Es un proceso complejo que supone la utilización de varios métodos de cálculo. Estos métodos utilizan la matriz de intercorrelaciones entre los reactivos de la prueba y permiten verificar si existe un único factor o varios que expliquen estas intercorrelaciones. Si existe un sólo factor la prueba es considerada homogénea. Si son varios factores que explican las intercorrelaciones significa que, aunque la prueba es heterogénea, existen subpruebas compuestas por conjuntos de reactivos homogéneos. Este procedimiento permite también detectar los reactivos que no tienen relación con el factor principal, ni con ninguno de los factores que componen la prueba.

2.8.3. Validez de la prueba.

2.8.3.1. Concepto de validez.

A continuación se presentan algunas definiciones del término validez:

Brown (1980) menciona que “la característica más importante de una prueba es su validez, o sea, la extensión con la que la prueba debe medir. Sin pruebas de la validez de una prueba no sabremos lo que mide en realidad; así, no será posible interpretar o dar un significado a las calificaciones.” (p. 40).

Reidl (1990) define al concepto de validez señalando que “un instrumento es válido cuando mide lo que pretende medir”. (p. 6)

Nunnally (en Bernstein, 1995) indica que “el término validez denota la utilidad científica de un instrumento de medida, en el que puede establecerse ampliamente qué tan bien mide lo que pretende medir”. (p. 92).

2.8.3.2. Procedimientos para medir la validez.

Existen tres procedimientos que permiten verificar la validez de una prueba. La utilización de uno u otro depende del objetivo y tipo de prueba. (Costa Neiva, 1996). Estos

tres procedimientos son los siguientes:

1. Validez de contenido.
2. Validez de criterio.
3. Validez de construcción.

2.8.3.2.1. Validez de contenido.

Este tipo de validez es utilizado sobre todo en pruebas de rendimiento/aprovechamiento, aunque también se aplica a otros tipos de pruebas psicológicas (Brown, 1980). A través de la validez de contenido se verifica si los reactivos de la prueba realmente constituyen “una muestra representativa del universo de las conductas o de los contenidos que se quiere medir” (*op. cit.*, p. 154). El procedimiento utilizado para verificar la validez de contenido de una prueba consiste en:

1. Precisar las habilidades, conocimientos y comprensiones que corresponden al contenido enseñado y que ha sido el objetivo de un curso o entrenamiento.
2. Examinar cuidadosamente la prueba (muestra de tareas) para ver qué conocimientos, habilidades y comprensiones requiere.
3. Verificar qué tanto contenido de la prueba representa el contenido y los objetivos del curso o entrenamiento.

Una prueba válida debe estar compuesta de tareas que representen al máximo los contenidos y metas de la enseñanza. Este procedimiento no utiliza ningún índice cuantitativo para verificar la adecuación de la muestra de tareas; la evaluación se da a través de un examen detallado y cuidadoso de la misma y por lo tanto implica un proceso de juicio y razonamiento por parte del evaluador. El autor antes mencionado presenta algunos de los inconvenientes que puede presentar este tipo de validez:

- a) no poder utilizar un índice cuantitativo para verificar el grado de relación,
- b) distintos evaluadores pueden no estar de acuerdo en cuanto a la validez del contenido de la prueba y
- c) la falta de claridad en la precisión del dominio dificultará los juicios de validez de contenido. Murphy y Davidshofer (1994, citado por Costa Neiva, 1996), hacen además hincapié en el hecho de que “los estudios de validez de contenido no pueden, por ellos

mismos, garantizar la validez de la medida” (p. 113).

2.8.3.2.2. *Validez de criterio.*

Este tipo de validez está relacionado con la predicción de un resultado específico futuro, por ejemplo, el éxito en una futura ocupación o en un futuro programa de capacitación. Por ello, se aplica a pruebas utilizadas para pronosticar resultados por medidas de criterios muy claras y definidas (tests de selección y clasificación de personal). Se elige así una medida como criterio (como desempeño en el empleo) y se verifica cuánto la medida en la prueba (por ejemplo, una prueba de habilidad numérica) se relaciona con la medida del criterio. Existen varios métodos para verificar la validez de criterio. El más utilizado supone el siguiente procedimiento:

1. Aplicar la prueba a un grupo, antes de iniciar el desempeño en un trabajo o de ingresar en un programa de capacitación o curso.

2. Definir un criterio que sirva de valoración del éxito en el empleo, programa de capacitación o curso y obtener más tarde, junto al mismo grupo, una medida de este criterio.

3. Calcular la correlación entre la calificación en la prueba y la calificación en el criterio de valoración del éxito. El coeficiente de correlación puede variar de -1 a +1. Cuando el valor es cercano a 0 indica que no existe correlación entre las dos medidas. Tanto un coeficiente positivo cuanto un coeficiente negativo indican una relación entre las medidas. Un coeficiente positivo indica que ambas medidas varían en la misma dirección (ambas altas o ambas bajas), mientras que un coeficiente negativo indica que si una medida es alta, la otra es baja o viceversa. Tanto mayor es esta correlación, más efectiva será la prueba para predecir este criterio y consecuentemente más válida.

La gran dificultad de este procedimiento está en la elección de un criterio adecuado de valoración. Thorndike (1973, citado por Costa Neiva, 1996), propone cuatro cualidades que debe tener un criterio de valoración:

- a) Adecuación: Un criterio es adecuado cuando la calificación de éste está determinada por los mismos factores que determinan el éxito en el trabajo/curso/tarea. No hay ninguna prueba que permita verificar la adecuación o no del criterio. Hay que confiar en el juicio del profesional para elegirlo.

b) Exención de prejuicios: La medida del criterio debe proporcionar, a cada persona, la misma oportunidad de obtener una buena calificación. Por ello, debe estar exenta de prejuicios del evaluador.

c) Confiabilidad: La medida del criterio debe ser estable y reproducible para que una prueba cualquiera pueda predecirla. Si la medida del criterio varía de una situación a otra o de tiempo en tiempo, ésta no podrá relacionarse con otras medidas y por lo tanto no habrá prueba que la pueda predecir.

d) Disponibilidad: En la elección del criterio debe tomarse en cuenta los problemas prácticos de su utilización: disponibilidad, tiempo, costo, etc.

Un segundo método para determinar la validez de criterio es “ver si las calificaciones predictoras diferencian a los grupos específicos por su ejecución en el criterio” (Brown, 1980, p.133). Por ejemplo, para determinar si una prueba de habilidad numérica es un buen predictor del éxito de alumnos en un curso de álgebra, se debe proceder de la siguiente manera:

1. Dividir a los alumnos en dos grupos: los aprobados y reprobados en el curso de álgebra.

2. Comparar las calificaciones, de esos dos grupos, en la prueba de habilidad numérica para verificar si hay una diferencia estadísticamente significativa entre sus calificaciones. El procedimiento generalmente utilizado para verificar si la diferencia es significativa es la obtención de la estadística t . Esta estadística “compara la diferencia en las medias con la medida del error de muestreo” (Brown, 1980, p. 132). Si el valor t obtenido es significativo (según las tablas), quiere decir que los dos grupos tienen calificaciones significativamente diferentes en la prueba. Un valor t no significativo indica que las calificaciones de la prueba no discriminan entre los grupos definidos según el criterio y consecuentemente, que la prueba no es válida.

Un tercer método es la utilización de índices de exactitud en la toma de decisiones (Brown, 1980). El procedimiento, en este caso, consiste en:

1. Clasificar las calificaciones de la prueba en dos o más categorías independientes (ejemplo: fracaso vs éxito o aceptable vs inaceptable).
2. Clasificar, según las mismas categorías, las calificaciones del criterio.
3. Comparar el conjunto de datos.

4. Obtener la proporción de decisiones correctas (Pdc). Costa Neiva (1996) presenta un ejemplo en el que, a través de un cuadro, se indican las predicciones correctas (aciertos) e incorrectas (fallas). Se consideran en el cuadro cuatro grupos formados a través de la comparación de dos categorías de criterio y dos categorías de la prueba predictiva (éxito vs fracaso).

	Ejecución del criterio	
Predicción de la prueba psicológica	Éxito	Fracaso
Éxito	Aciertos (A)	Falla (B)
Fracaso	Falla (C)	Aciertos (D)

Fuente: Costa Neiva (1996), p. 35.

Cuando los grupos obtienen éxito o fracaso en ambas medidas, la de la prueba y la del criterio, se considera una predicción o decisión correcta o sea un acierto (grupos A y D). Los demás casos son considerados una falla (grupos B y C). Un índice importante en la toma de decisiones es la proporción de decisiones correctas (Pdc) en el total de decisiones. Esta proporción se obtiene a través de la siguiente fórmula:

$$Pdc = \frac{\text{Aciertos}}{\text{Aciertos} + \text{Fallas}} = \frac{A + D}{A + D + B + C} = \frac{\text{Aciertos}}{N}$$

Los valores A,B,C y D corresponden al número de sujetos de cada uno de los cuatro grupos y N al total de sujetos. La proporción de decisiones correctas (Pdc) indica la exactitud de las decisiones y puede ser utilizada como índice de validez de la prueba.

Brown (1980) comenta sobre las variaciones en este procedimiento, como la que supone la determinación de la *calificación límite óptima*, calificación que separa a los grupos para producir el número máximo de decisiones correctas.

2.8.3.2.3. Validez de construcción (o validez interna o estructural).

Este tipo de validez verifica el grado en que una prueba mide un constructo o rasgo teórico. En realidad, no existe un índice cuantitativo único o sólo un procedimiento de validez de construcción. La validez se evalúa mediante un conjunto de evidencias y métodos. Brown (1980) agrupa en cinco categorías la variedad de técnicas y procedimientos utilizados

para establecer la validez de construcción:

2.8.3.2.3.1. *Métodos intrapruebas.*

En esta categoría se encuentran las técnicas utilizadas para estudiar la estructura interna de una prueba: su contenido, los procesos utilizados para contestar a las preguntas, la relación entre sus reactivos, la relación entre las subescalas. Se incluyen en esta categoría desde procedimientos de validez de contenido hasta estudios de homogeneidad de la prueba (coeficiente alfa, coeficiente Kuder-Richardson). Estos índices indican si la prueba mide uno solo o varios rasgos.

2.8.3.2.3.2. *Métodos interpruebas.*

Estos métodos están diseñados para verificar las relaciones entre varias pruebas o sea, si éstas miden o no el mismo rasgo. Algunos de los procedimientos usados son:

2.8.3.2.3.2.1. *Validez congruente.* En este caso se verifica la correlación entre las calificaciones de una nueva prueba y las obtenidas en otra, ya validada, que mida el mismo rasgo. Si la correlación es alta significa que ambas pruebas miden el mismo rasgo y que pueden interpretarse los resultados de la nueva prueba como los de aquella ya validada.

2.8.3.2.3.2.2. *Validez factorial.* Este es considerado uno de los tipos más importantes de validez. Es verificada a través de un procedimiento estadístico -el análisis factorial- que permite determinar “cuantos factores (construcciones) se necesitan para explicar las intercorrelaciones entre un conjunto de calificaciones de varias pruebas” (Brown, 1980. p.165). El análisis factorial permite obtener:

- a) el número de factores que explican las intercorrelaciones entre las pruebas,
- b) los factores que componen cada prueba y
- c) la cantidad de varianza en las calificaciones de las pruebas que explican los factores.

Así, pueden determinarse las pruebas que comparten una varianza común y por lo tanto miden el mismo factor o sea la misma construcción. Además, la proporción de varianza total en las calificaciones de la prueba que es atribuida al factor, puede ser utilizada como índice de validez de construcción. De esta forma, para que un instrumento sea válido, la

varianza acumulada deberá alcanzar valores de 70% o más.

2.8.3.2.3.2.2.1. La técnica del análisis factorial.

Skékely (1970) dice que todo estudio factorial consta de cuatro fases:

1. Preparación,
2. Factorización,
3. Rotación y
4. Interpretación.

1. Preparación. Como en toda investigación científica, ésta es la fase más importante. En ella se estudia el campo de la actividad humana en cuestión, se recoge la información existente y se formulan las hipótesis oportunas. Estas hipótesis se refieren al número e índole de los factores que se buscan. Según estas hipótesis, el investigador idea o selecciona los reactivos que sean capaces de ponerlas de manifiesto, procurando cubrir todo el campo que se estudia, de modo que cada factor hipotético esté representado por varios reactivos. Estas reactivos son aplicados a la población conveniente y calculadas sus intercorrelaciones. Con éstas se forma la llamada matriz de correlaciones.

2. Factorización. Así se llama la segunda etapa. El objeto de ella es averiguar el número de factores comunes necesarios para explicar las correlaciones halladas. El número de factores está determinado por las correlaciones y se resuelve por medios matemáticos.

Una sencilla explicación geométrica ayudará a comprender el método. Es fácil mostrar que la matriz de correlaciones puede representarse por una configuración de vectores. Cada reactivo se representa por un vector. La longitud de cada vector está determinada por un valor de la matriz de correlaciones que se denomina comunidad. La posición de cada vector está determinada por las correlaciones del reactivo correspondiente con los demás. El conjunto o configuración de vectores proporciona la misma información que la matriz de correlaciones y equivale punto por punto a ella.

El número de dimensiones del espacio en que se extiende la configuración vectorial es el número de factores que se buscan.

Si todos los vectores se sitúan en una misma recta, es decir, en un espacio de una dimensión, todas las correlaciones se explican por un solo factor común y general.

Si los vectores se extienden en un espacio de dos dimensiones, es decir en un

plano, se necesitarán dos factores comunes.

Si los vectores se extienden en un espacio de tres o cuatro o n dimensiones, se necesitarán tres, cuatro o n factores comunes. Es claro que cuando se necesitan más de tres factores el espacio vectorial no corresponde a ningún espacio intuitivo ni puede representarse físicamente, pero puede manejarse matemáticamente con igual sencillez y rigor que en los casos anteriores.

Analizar factorialmente una matriz de correlaciones equivale a introducir en la configuración vectorial tantos ejes de coordenadas como dimensiones.

Los ejes pueden disponerse de cualquier manera con tal de que todos pasen por el origen de los vectores. Lo más sencillo y corriente es situarlos perpendicularmente entre sí.

Una vez colocados los ejes, o realizadas algebraicamente las operaciones equivalentes, se hallan las coordenadas de cada vector en cada eje. Cada vector es un reactivo; cada eje, un factor; cada coordenada el coeficiente factorial de un reactivo en un factor.

El conjunto de estas coordenadas constituye la matriz factorial, con tantas columnas como factores y tantas filas como reactivos. En cada fila se registran los coeficientes factoriales de cada reactivo, es decir las coordenadas de cada vector en los ejes.

Para averiguar lo que significa cada factor hay que examinar, ahora, los coeficientes que en él tienen los diversos reactivos. Cada factor será aquello común a los reactivos que tengan coeficientes en él y que esté presente en pequeño grado en los que tengan coeficientes bajos y en alto grado en los que los tengan elevados.

Es claro que si los coeficientes de los reactivos dependen del lugar en que se hayan situados los ejes, y si, al mismo tiempo, los ejes se han situado en cualquier lugar, por mera comodidad matemática, no hay ninguna razón para suponer que estos coeficientes tengan significación psicológica alguna. No pueden ser, pues, interpretados en esta fase.

3. Rotación. La rotación, es precisamente, la operación por la cual los ejes se giran hasta una posición que tenga, por razones teóricas generales, una significación psicológica. Hay dos formas de rotación:

- a) ortogonal, que incluye a su vez tres técnicas (en la computadora): la quartimax, la varimax, y la equimax; y
- b) oblicua.

La rotación consiste en girar los ejes hasta una posición tal que cada vector tenga proyecciones significativas en el mínimo número de ejes. La lógica de este procedimiento es sencilla. Si en el campo que se estudia operan de hecho diversos factores distintos y hemos ideado diversos reactivos que los ponen de manifiesto clara y discriminativamente, de modo que unos reactivos dependen de un factor y otros de otro, pero a ninguno o casi a ninguno de todos los factores, tendremos una configuración en la que los vectores se agrupan en constelaciones distintas.

Cada constelación estará formada, en el caso ideal, por los reactivos que dependen de un factor y son independientes de los demás. Si pasamos los ejes por esos grupos de vectores tendremos una estructura simple que reflejará la hipótesis supuestamente verdadera.

Si no puede hallarse esta estructura es que la hipótesis era falsa o que se ha sido incapaz de ponerla de manifiesto por los reactivos ideados o por los sujetos examinados. Si los ejes pueden girarse hasta una posición de estructura simple se hallarán las coordenadas de los vectores en los nuevos ejes y con ellas se formará una nueva matriz factorial. Ella mostrará los coeficientes factoriales de los reactivos en los nuevos factores.

4. Interpretación. Al obtener la nueva matriz factorial se puede interpretar, lo que quiere decir que se puede averiguar la significación de cada factor examinando los procesos que intervienen en todos los reactivos que dependen de él. La definición psicológica del factor se hará en función de lo que parezca ser común a todos en relación a sus coeficientes factoriales en él. La interpretación de cada factor debe ser coherente asimismo con el hecho de que otros reactivos tengan coeficientes nulos en él.

Actualmente el análisis factorial se realiza por medio de las computadoras. A continuación se presentan las seis matrices que proporciona un análisis factorial completo utilizando el programa computacional SPSS (Paquete Estadístico para las Ciencias Sociales, Nie, Hull, Jeankins, Steinbrenner y Bent, 1975; citados por Reidl, 1990):

1. Una matriz de correlaciones de las variables analizadas.
2. Cargas factoriales iniciales.
3. Pesos para estimar las variables a partir de los factores (*factor pattern matrix*-matriz del patrón factorial). Esta matriz contiene los pesos o coeficientes de regresión

de los factores comunes y por lo tanto señala la composición de una variable en términos de los factores hipotéticos.

4. Pesos para estimar factores a partir de las variables (*factor estimate-estimación de factores*; o *factor score coefficient matrix*-matriz de coeficientes de calificaciones factoriales). Esta matriz proporciona un medio de estimar puntajes factoriales a partir de variables observadas. Es decir, son los pesos o coeficientes de regresión que se emplean para estimar las calificaciones factoriales a partir de las variables observadas expresadas en unidades o puntajes z.

5. Correlación entre los factores y las variables o cargas factoriales (*factor structure matrix*-matriz de la estructura factorial). Esta matriz está constituida por los coeficientes de correlación (o cargas factoriales) entre cada variable y cada factor. Esta es la matriz que se emplea para interpretar (nombrar) a los factores extraídos u obtenidos. Cuando se emplea un método ortogonal de rotación, la matriz de patrón factorial y la matriz de la estructura factorial, son iguales por lo que aparece únicamente la segunda. Cuando el procedimiento de rotación es el oblicuo, aparecen ambas, con sus nombres respectivos.

6. Matriz de intercorrelaciones de los factores terminales. Esta matriz sólo se obtiene cuando la solución terminal tuvo una rotación oblicua. Está constituida por las intercorrelaciones de las dimensiones (o factores) subyacentes, y puede servir para análisis factoriales de un orden más alto (someter a análisis factorial los puntajes que los sujetos obtienen en los factores extraídos en el primer nivel o primer paso).

Las variables pueden introducirse por medio de los puntajes crudos que los sujetos obtuvieron en cada una de ellas, o por medio de una matriz cuadrada de correlaciones (igual número de columnas, k y de hileras, r).

Someter a análisis factorial a un conjunto de variables significa, en el nivel más general, expresar a una variable como la combinación lineal de ciertas variables (factores) independientes, ya sea definidos o inferidos.

La importancia de los factores está dada, primero, por el valor eigen que obtiene cada uno de los factores extraídos. El valor eigen es la raíz de la ecuación (polinomio) que explica la matriz de varianza reducida (la que se trabaja a partir de la extracción de los factores iniciales). Un segundo aspecto que habla de la importancia de los factores es el porcentaje de varianza, en la matriz reducida, que cada factor explica; así, son

más importantes aquellos factores que tienen un valor eigen y un porcentaje de varianza explicada mayores.

2.8.3.2.3.2.3. Validez discriminante.

En este caso el objetivo es verificar que las calificaciones de una nueva prueba no se correlacione con las obtenidas en otras pruebas que miden otros rasgos distintos. Se demuestra así que la medida de esta nueva prueba es independiente de la medida de otras construcciones o rasgos. (Costa Neiva, 1996).

2.8.3.2.3.3. Estudios relacionados con criterios.

Este método se basa en la predicción acerca de diferencias entre grupos, naturalmente separados o diseñados experimentalmente. Por ejemplo, una teoría sugiere que algunas clases de grupos deben obtener puntuaciones elevadas o bajas respecto al rasgo medido por una nueva prueba. Al aplicársela se verificará cuánto las calificaciones en esta prueba separan a estos grupos, o sea, pueden predecir las diferencias entre ellos. (Costa Neiva, 1996).

2.8.3.2.3.4. Manipulación experimental

En este procedimiento se manipula experimentalmente una variable y se observan sus efectos sobre las calificaciones de la prueba o la relación de éstas con algún criterio. Por ejemplo, una teoría puede sugerir que la expresión de un rasgo o conducta se modificará a consecuencia de algunas condiciones o tratamientos experimentales. Se aplica entonces la nueva prueba en la condición experimental para verificar si las modificaciones previstas ocurren. “Sería el caso de aplicarse una prueba de ansiedad a personas que están por someterse a una operación quirúrgica para verificar si esta prueba puede detectar el nivel de ansiedad que supuestamente debe encontrarse aumentado por la inminencia de la cirugía”. (Costa Neiva, 1996, p. 38).

2.8.3.2.3.5. Estudios de capacidad de generalización.

Según Brown (1980) se encuentran en esta categoría los estudios que “analizan sistemáticamente la prueba sobre una gama de condiciones o dimensiones -por ejemplo- en

una variedad de poblaciones (transferibilidad) o con diferentes condiciones de administración” (p. 167).

En el siguiente capítulo, a partir de este conocimiento de la estructura psicométrica de los tests, podremos observar cómo se construyó, se validó y se hicieron estudios de confiabilidad para la Escala de síntomas *Symptom Rating Test* (SRT) desde su creación.

Capítulo 3
La Escala de síntomas *Symptom Rating Test*
(SRT)

3. La Escala de síntomas *Symptom Rating Test* (SRT)

3.1. ¿Qué es un síntoma?

Esta es la primera pregunta que surge ante el nombre de esta Escala, por lo que antes de abordar la forma en que se construyó y se validó, se hará una somera descripción del significado de la palabra síntoma:

El *Diccionario de la lengua española* define el término síntoma como: “Fenómeno revelador de una enfermedad” y en una segunda acepción como: “Señal, indicio de una cosa que está sucediendo o va a suceder”.

Así entonces, el conocimiento, ordenamiento y descripción de los síntomas es un área por la que el ser humano se ha interesado profundamente desde la antigüedad pues desde este conocimiento se puede distinguir al sano del enfermo (la salud de la patología), el tipo de enfermedad que se está sufriendo y por tanto las formas de curar ese sufrimiento.

La ciencia médica y por ende la psiquiatría se han esforzado por obtener detalladamente la descripción de los síntomas pues es desde esas señales que el enfermo relata que se puede realizar un diagnóstico (el tipo de enfermedad que se padece) y un pronóstico (la forma de desarrollo de dicha enfermedad), así como la manera de enfrentar a los síntomas para su curación. Así se creó la nosología, que es una parte de la medicina que tiene por objeto describir, diferenciar y clasificar las enfermedades.

La Asociación Psiquiátrica Americana (American Psychiatric Association) ha creado el *Manual diagnóstico y estadístico de los trastornos mentales*, que es sometido a continuas revisiones, donde a través de listados de síntomas psicológicos, ampliamente desglosados, se pueden obtener criterios de diagnóstico de los trastornos que la psiquiatría norteamericana considera existentes.

Por tanto, hablar de síntomas es hablar de psicopatología, la cual desde la psicología se ha enfocado desde diversos ángulos, definidos por la teoría que sustenta quien describa el complejo fenómeno del sufrimiento psicológico. Así entonces, “para los teóricos conductistas, los trastornos psicológicos se conciben como un complicado patrón de respuestas a una presión ambiental y se les describe fundamentalmente a nivel de conductas. Para la teoría fenomenológica, se los veía como expresiones de inquietud y angustia

personales. Considerados desde un punto de vista fisiológico, cabría interpretarlos como series de actividades neurales y químicas complejas, en tanto que para los teóricos del psicoanálisis y las teorías intrapsíquicas los síntomas se organizan en procesos inconscientes que defienden contra la ansiedad y el conflicto". (Millon, 1974, p. 2).

3.2. Construcción de la Escala de síntomas (SRT).

La Escala de síntomas (SRT) fue diseñada por Robert Kellner, de la Universidad de Liverpool, Inglaterra, en 1967 (a), (citado en Kellner y Sheffield, 1973), para medir cambios en los síntomas de pacientes neuróticos que participaban en experimentos terapéuticos y ensayos con drogas. En el artículo se reporta que es una escala que mide un estado temporal y cambiante, el sufrimiento o malestar psicológico derivado de la existencia de síntomas y que no es una escala que mide rasgos, o sea, disposiciones duraderas constituyentes del sujeto. "Las respuestas a los reactivos en un inventario de personalidad deben ser estables a lo largo del tiempo, pero las respuestas a los elementos en una escala de medición de síntomas deberían cambiar con el tiempo y medir cambios en el estado clínico del paciente." (Citado en Kellner y Sheffield, 1973, p. 88). "La función del SRT es análoga a la de un termómetro clínico, más que a la de una medición por cinta" (Artículo citado, p. 89).

La Escala de síntomas (SRT), en su versión original, consiste en una lista de control de 38 síntomas, 15 de ellos son somáticos y 23 psicológicos, en un formato de papel y lápiz. El paciente mide cada síntoma que marca en la lista con una escala de automedición de cuatro respuestas. Hay varias formas de la SRT: la forma diaria, donde el paciente marca los síntomas que ha experimentado en el día, la forma semanal, donde marca los síntomas experimentados la semana pasada, la forma mensual, donde se refiere a síntomas experimentados en el mes o los meses anteriores y la forma anual.

La lista de síntomas fue compilada a partir de las quejas de 100 pacientes neuróticos. El objetivo era compilar una breve, pero comprensiva lista de control de síntomas, usando expresiones de fácil entendimiento para la mayoría de los pacientes. La lista de control contiene 38 reactivos, pero cubre un rango más amplio de síntomas, esto se logró al construir reactivos que consisten en más de una queja similar. Cuenta con un espacio para síntomas idiosincráticos que no son incluidos en la lista de control. Esta lista contiene la mayoría de los síntomas neuróticos más comunes.

Se construyó también una versión de cartas. La versión de cartas de prueba de la SRT

consiste de una entrevista semi-estructurada, basada en la lista de control. (Dado que en este estudio sólo se investigará la lista de síntomas, la versión de cartas no será descrita con más detalle).

Inicialmente el paciente medía cada síntoma en tres dimensiones: intensidad, frecuencia y duración. Se probaron muchas opciones de respuesta de automedición con diferentes tipos de indicaciones. Sólo aquellas indicaciones que los pacientes neuróticos letrados podían medir sin dificultad fueron mantenidas. Las opciones que los pacientes encontraron difíciles de usar fueron descartadas. El resto de las opciones de respuesta fueron validadas en dos formas: fueron administradas a los pacientes neuróticos y normales, y a pacientes neuróticos antes y después de tratamiento. La opciones de respuesta que:

- 1) discriminaban entre pacientes neuróticos y normales, y
- 2) variaban en los pacientes neuróticos después del tratamiento en la dirección esperada, fueron consideradas válidas y sensibles al cambio. Se llevó a cabo una estimación de la confiabilidad de estabilidad temporal, y se aplicó el coeficiente de correlación de Spearman. (Sheffield y Kellner, 1970, citados en Kellner y Sheffield, 1973). Después, las opciones fueron acortadas a cuatro respuestas eliminando indicaciones que eran raramente usadas por los pacientes. La prueba fue entonces convertida en una versión más económica de papel y lápiz.

Los autores de la prueba mencionan que a partir de que Battle, Imber, Hoehn-Saric, Stone, Nash y Frank (1966) encontraron que más de la mitad de los pacientes prefieren una “escala compactada”, en la SRT las indicaciones se concentraron en respuestas a través de la disposición parcial de las palabras en columnas y se pide a los pacientes dibujar un círculo alrededor de la afirmación que describe su condición.

La construcción de tres subescalas (ansiedad, depresión y somatización) se basó en varios análisis de factores de síntomas publicados. La cuarta subescala (inadecuación) no es un factor singular y consiste principalmente de sentimientos de inferioridad y síntomas cognoscitivos. Las subescalas son enumeradas en el cuadro 3.

En 1972, Kellner construyó una versión corta donde el número de reactivos se redujo a 30 al encontrar que no había ventajas con el número anterior de reactivos. Ninguna diferencia significativa se encontró entre el número de pacientes que prefieren la escala de automedición original y los que prefieren la versión compactada, por tanto la escala compactada se adoptó en 1973 (Kellner y Sheffield, 1973). Las indicaciones de esta escala

Cuadro 3. Reactivos de las subescalas originales. (Traducción de la prueba existente en español).

Núm. de reactivo	Núm. de reactivo
<p style="text-align: center;">Anxiety (Ansiedad o angustia)</p> <p>3 Nervous (Nervioso)</p> <p>5 Scared, frightened (Con temores o miedos)</p> <p>9 Restless, jumpy (Inquieto o asustadizo)</p> <p>16 Trembling, shaking (Temblores o sacudimientos)</p> <p>19 Tense, wound up (Sentimientos de tensión o como si me hubiesen dado cuerda (como un reloj))</p> <p>23 Thoughts you could not push out of your mind (Con pensamientos que no se pueden quitar de la cabeza)</p> <p>26 Attacks of panic (Ataques de pánico)</p> <p>29 Difficulty in falling asleep, restless sleep (Dificultad para conciliar el sueño, o sueño desasosegado, o pesadillas)</p>	<p style="text-align: center;">Depression (Depresión)</p> <p>2 Tired, lack of energy (Cansado o sin energía)</p> <p>6 Poor appetite (Con poco apetito)</p> <p>8 No hope (Sentimientos de que no hay esperanza)</p> <p>12 Feeling guilty (Sentimientos de culpa)</p> <p>18 Unworthy, failure (Sentimientos de poca valía o fracaso)</p> <p>24 Lost interest (Pérdida de interés en casi todas las cosas)</p> <p>25 Unhappy, depressed (Descontento o deprimido)</p> <p>30 Awakening early (Me despierto demasiado temprano y no puedo conciliar el sueño otra vez)</p>
<p style="text-align: center;">Somatic (Síntomas somáticos)</p> <p>1 Dizzy, faint (Mareado o débil)</p> <p>4 Feeling of pressure or tightness (Sensación de presión o tensión en la cabeza)</p> <p>7 Heart beats quickly or strongly (Palpitaciones rápidas o fuertes sin razón aparente)</p> <p>11 Chest pains (Dolores en el pecho o dificultad al respirar o sensación de no tener suficiente aire)</p> <p>14 Muscle pains, aches (Dolores musculares o reumatismo).</p> <p>21 Feeling numb or tingling (Sensaciones de cosquilleo o adormecimiento en partes de mi cuerpo)</p> <p>27 Parts of body feel weak (Sensaciones de debilidad en partes del cuerpo)</p>	<p style="text-align: center;">Inadequacy (Inadecuación)</p> <p>10 Poor memory (Con poca memoria)</p> <p>13 Worrying (Preocupaciones)</p> <p>15 People look down on you (Sentimientos de que la gente me considera menos o piensa mal de mí)</p> <p>17 Difficulty in thinking clearly or making up your mind (Dificultad para pensar claramente o para tomar decisiones)</p> <p>20 Inferior to others (Sentimientos de ser inferior a otras personas)</p> <p>22 Irritable (Irritabilidad)</p> <p>28 Could not concentrate (Sin capacidad de concentración)</p>

son, de izquierda a derecha: “Para nada”, “Un poco, levemente”, “Mucho, bastante” y “Extremadamente, no podría ser peor”, que son similares a aquellas usadas en varias escalas de automedición de síntomas. (Clyde, 1963, McNair y Lorr, 1964, Lipman, Covi, Rickels, Uhlenhuth, y Lazar, 1968; citados por Kellner y Sheffield, 1973). Ver cuadro 4.

Los autores mencionan que se adoptó una sola forma en la versión corta, de lápiz y papel y su única diferencia radica en las instrucciones iniciales al aplicarla. Por ejemplo, en la forma semanal, la instrucción es enunciada así: “Describa cómo se ha sentido la semana pasada.” En la forma trimestral se enuncia así: “Describa cómo se ha sentido en los últimos tres meses”, etc.

La prueba se califica marcando un valor para cada automedición, que va de izquierda a derecha de la escala, con 1 (Para nada), 2 (Un poco, levemente), 3 (Mucho, bastante) y 4 (Extremadamente, no podría ser peor). La calificación total es la suma de todas las automediciones. (Ver cuadro 4).

La calificación de cada subescala la constituye la suma de las automediciones como sigue:

Subescala de ansiedad con los reactivos: 3,5,9,16,19,23,26,29.

Subescala de depresión con los reactivos: 2,6,8,12,18,24,25,30.

Subescala de somatización con los reactivos: 1,4,7,11,14,21,27.

Subescala de inadecuación con los reactivos: 10,13,15,17,20,22,28.

3.3. La neurosis.

Dado que la Escala de síntomas (SRT) es una escala cuyos reactivos se construyeron a partir de las quejas de cien personas neuróticas, es necesario hacer una somera revisión de la literatura que se refiere a este trastorno.

3.3.1. El concepto de neurosis.

Sarason (1981) dice que el término neurosis fue usado primeramente a fines del siglo XVIII por William Cullen. Lo empleó para referirse a padecimientos que no vienen acompañados de fiebre o de patología localizada. En la raíz de todos los padecimientos neuróticos, según Cullen, debiera haber una especie de colapso fisiológico: “El miedo y la melancolía o una disposición tímida y desolada pueden surgir en ciertos estados, o en ciertas ocasiones de mera debilidad... La enfermedad de la melancolía... depende manifiestamente del temperamento general del cuerpo” (Cullen 1781, citado en Sarason, 1981, p. 210).

Cuadro 4. Versión original, corta, de la Escala de síntomas (SRT).

Short version of the Symptom Rating Test (pencil and paper)*

Describe how you have felt during the PAST WEEK

If you have not had the symptom at all make a check mark () in the box on the left like this:

	Not at all	A little, Slightly	A great deal, Quite a bit	Extremely, could not have been worse
Headaches or head pains	✓			

If you have had the symptom describe how much it has bothered you or troubled you, for example, like this:

	Not at all	A little, Slightly	A great deal, Quite a bit	Extremely, could not have been worse
Headaches or head pains			✓	

Please answer all questions. Do not think long before answering.

	Not at all	A little, Slightly	A great deal, Quite a bit	Extremely, could not have been worse
1 Feeling dizzy or faint				
2 Feeling tired or a lack of energy				
3 Nervous				
4 Feelings of pressure or tightness in head or body				
5 Scared or frightened				
6 Poor appetite				
7 Heart beating quickly or strongly without reason (throbbing or pounding)				
8 Feeling that there was no hope				
9 Restless or jumpy				
10 Poor memory				

	Not at all	A little, Slightly	A great deal, Quite a bit	Extremely, could not have been worse
11 Chest pains or breathing difficulties or feeling of not having enough air				
12 Feeling guilty				
13 Worrying				
14 Muscle pains, aches, or rheumatism				
15 Feeling that people look down on you or think badly of you				
16 Trembling or shaking				
17 Difficulty in thinking clearly or difficulty in making up your mind				
18 Feeling unworthy or a failure				
19 Feeling tense or "wound up"				
20 Feeling inferior to other people				
21 Parts of body feel numb or tingling				
22 Irritable				
23 Thoughts which you cannot push out of your mind				
24 Lost interest in most things				
25 Unhappy or depressed				
26 Attacks of panic				
27 Parts of your body feel weak				
28 Cannot concentrate				
29 It takes a long time to fall asleep, or restless sleep or nightmares				
30 Awakening too early and not being able to fall asleep again				

* The original pencil and paper version can be obtained from the authors.

Fuente: Kellner y Sheffield, 1973, p. 98.

Su tratamiento de la neurosis comprendía cambios de dieta, fisioterapia, ejercicios, sangrados y vómitos; las medidas tomadas en aquellos tiempos para combatir los problemas fisiológicos.

Con el paso de los años, continúa Sarason, han ido surgiendo connotaciones variadas asociadas con el término de neurosis. En un tiempo este padecimiento implicaba un trastorno grave; sin embargo, hoy en día sugiere una perturbación moderada con un nivel mediano de pesar y ansiedad psicológicos. Como lo aplican los psiquiatras, el término implica un desorden que restringe, en algún grado, el juicio total del individuo, su habilidad para lograr buen contacto con la realidad y su capacidad para relacionarse con los demás que participan de su ambiente. El diagnóstico psiquiátrico de neurosis debe aplicarse fundamentalmente a aquella persona que desconfía fundamentalmente de su propio y general nivel de competencia, así como de su mérito básico, y de la amistosidad y naturaleza de aceptación fundamentales del ambiente. En general, la conducta neurótica es impropia, inadecuada e insatisfactoria. Puede ser también bastante infantil. Produce discrepancias subjetivas y objetivas entre el potencial de una persona y su conducta y logros sociales. Los neuróticos tienden a conducirse de la manera que Dollard y Miller (1950, citado por Sarason, 1981), llamó “estúpida”, es decir, que tienden a responder repetitiva y rígidamente en las relaciones interpersonales, al parecer, y no aprovechan sus experiencias. La conducta neurótica no se vuelve, típicamente, por completo incapacitante, aunque las incapacidades menores son notables tanto para el individuo mismo como para los demás. La percepción que el neurótico tiene del mundo está distorsionada y experimenta sentimientos de miedo y culpa. Puede sentirse persistentemente desesperado, frustrado y enojado, parecer que se está autocastigando o puede ser particularmente adepto a aislarse, a molestar y a disgustarse con los demás.

Ey (1987) menciona que el neurótico es siempre, o casi siempre, un ser cuyo carácter es patológico, es decir cuya manera de ser está fijada y estereotipada (compulsiones de repetición) en forma de disposición fundamental a la angustia, a la introversión, a los escrúpulos o, por el contrario, a la exaltación imaginativa y apasionada. Es sobre este fondo de trastorno caracterológico que se desarrolla la existencia neurótica. De manera que el Yo neurótico aparece con un desequilibrio afectivo esencial, profundamente enraizado en el desarrollo mismo del ser psíquico. Este desequilibrio llama la atención por los elementos de extrañeza, de disarmonía, de originalidad. La alteración conflictiva de la personalidad

característica de la persona neurótica agrega al cuadro clínico sus síntomas fundamentales. El neurótico es un sujeto para quien los problemas de la coexistencia con el otro son insolubles (por lo menos no resueltos), pues no alcanza su propia identificación. La patología de esta autoconstrucción del personaje está profundamente enraizada, como Freud y el psicoanálisis lo han observado bien, en las primeras relaciones de objeto y las identificaciones que ellas facilitan o inhiben. Pero se prolonga y continúa en el curso del desarrollo de la existencia. Ésta, desde entonces, está profundamente trastornada, especialmente en sus relaciones con el otro y en particular con el medio familiar. Este malestar del individuo frente a sí mismo y a los otros es experimentado como una angustia en la que se mezclan los sentimientos de vergüenza, de culpabilidad, el deseo de castigarse, la decepción de las frustraciones, los complejos de inferioridad, etc.

3.3.2. Tipos de neurosis.

La ansiedad es la causa y es el síntoma fundamental de la neurosis, tal como ésta se define en el *Manual diagnóstico y estadístico de los trastornos mentales*, DSM-III-R, de la American Psychiatric Association (APA):

“La ansiedad es la característica primordial de la neurosis. Puede percibirse y expresarse, o puede controlarse inconsciente y automáticamente por conversión, desplazamiento y varios otros mecanismos psicológicos. En general, estos mecanismos producen síntomas que se experimentan como desconfianza subjetiva, de la que el paciente quisiera librarse.

La neurosis, en contraste con la psicosis, no manifiesta distorsión pronunciada ni mala interpretación de la realidad externa, tampoco desorganización muy acentuada de la personalidad. La posible excepción sería el caso de la neurosis histérica, que según creen algunos puede ir acompañada de alucinaciones y otros síntomas que se encuentran en la psicosis.

Tradicionalmente, los pacientes neuróticos, a pesar de estar gravemente impedidos por sus síntomas, no se clasifican como psicóticos, por el hecho de que están conscientes de que su desenvolvimiento intelectual está perturbado”. (APA, 1968, citado por Sarason, 1981, p. 212.)

La siguiente es una lista de las categorías más ampliamente usadas de neurosis, según Sarason (1981):

1. Neurosis de ansiedad. El individuo experimenta aprehensividad y miedo más allá del

que se podría esperar de la mayoría de la gente. El neurótico ansioso es acosado por tensiones y preocupaciones difusas. En algunos casos, su ansiedad puede alcanzar proporciones de pánico.

2. Neurosis fóbica. Es este un temor muy concreto (por ejemplo, miedo a la sífilis, a la suciedad, a los lugares altos), en situaciones que no plantean peligro real.

3. Neurosis disociativa. Puede haber conducta llena de estupor y de amnesia y despersonalización (pérdida del sentido de identidad personal). Algunas reacciones disociativas no se distinguen fácilmente de las correspondientes a la psicosis. El individuo parece “congelarse” o sumirse en una amnesia sin objeto, sonambulismo, fuga y actividad de personalidad múltiple.

4. Neurosis histérica. Hay síntomas orgánicos sin que haya de por medio ningún desorden orgánico observable, bajo control voluntario. Esta agrupación incluye algunos casos de ceguera, sordera y parálisis, atribuibles más bien a condiciones psicológicas que corporales.

5. Neurosis obsesivo-compulsiva. Ideas y actos ocurren repetitivamente y al parecer de manera inexplicable (por ejemplo, incapacidad para dejar de pensar en una palabra en particular, lavarse las manos reiteradamente).

6. Neurosis depresiva. Hay un estado de depresión, de autodesprecio y de culpa. La depresión neurótica suele seguir a la pérdida de un ser o una posesión querida.

3.3.3. Criterios de diagnóstico de los trastornos neuróticos.

El *Manual diagnóstico y estadístico de los trastornos mentales*, DSM-III-R (1988), expone criterios para el diagnóstico de los trastornos a través de listados de síntomas que en caso de que se presenten juntos permitan diagnosticar dicho trastorno. A continuación se presentarán los síntomas que para este *Manual* definen a los trastornos por angustia, por depresión y somatización, lo que permitirá, por otro lado, tener un punto de comparación con el listado de síntomas de las subescalas del SRT. Los síntomas de la subescala de inadecuación se realizó, como se vio en el apartado de la construcción de la prueba, bajo el criterio de Kellner, en la cual agrupó síntomas cognoscitivos y sentimientos de inferioridad.

3.3.3.1. Criterios para el diagnóstico de trastorno por angustia.

“A. En algún momento durante la alteración se han presentado una o más crisis de angustia (períodos concretos de miedo o molestias intensas) que han sido 1) inesperados, es

decir, que no se han presentado inmediatamente antes o durante la exposición a una situación que casi siempre causa ansiedad, y 2) no han sido desencadenados por situaciones en las cuales el individuo ha sido el foco de la atención de los demás.

B. Por lo menos cuatro crisis tal como se definen en el criterio A durante un período de cuatro semanas, o bien una o más crisis seguidas de miedo persistente o presentar otro ataque por lo menos durante un mes.

C. Por lo menos, cuatro de los síntomas siguientes durante alguna de las crisis:

- 1) falta de aliento (disnea) o sensaciones de ahogo;
- 2) mareo, sensación de inestabilidad, sensación de pérdida de conciencia;
- 3) palpitaciones o ritmo cardiaco acelerado (taquicardia);
- 4) temblor o sacudidas;
- 5) sudoración;
- 6) sofocación;
- 7) náuseas o molestias abdominales;
- 8) despersonalización o desrealización;
- 9) adormecimiento o sensaciones de cosquilleo en diversas partes del cuerpo (parestias);
- 10) escalofríos;
- 11) dolor o molestias precordiales;
- 12) miedo a morir;
- 13) miedo a volverse loco o a perder el control.

Nota: Las crisis que incluyen cuatro o más síntomas se las denomina crisis de angustia; aquellas que tengan menos de cuatro síntomas son crisis de síntomas limitados.

D. En alguna de estas crisis, por lo menos cuatro de los síntomas C han aparecido rápidamente y han aumentado en intensidad durante los diez minutos siguientes al comienzo del primer síntoma.

E. No es posible establecer la presencia de algún factor orgánico que haya iniciado o mantenido la alteración, por ejemplo, intoxicación por anfetamina o cafeína, o hipertiroidismo” (p. 284).

3.3.3.2. *Criterios para el diagnóstico del trastorno por ansiedad generalizada.*

“A. Ansiedad y preocupación no realistas o excesivas en torno a dos o más circunstancias vitales. El trastorno ha durado por lo menos seis meses o más, durante los que el individuo ha estado preocupado por estos problemas la mayor parte de los días.

B. La ansiedad y la preocupación no tiene nada que ver con el miedo a tener una crisis de angustia (como ocurre en el trastorno por angustia).

C. El trastorno no debe presentarse sólo en el curso de un trastorno del estado de ánimo o de un trastorno psicótico.

D. Por lo menos, hay seis de los siguientes dieciocho síntomas cuando el individuo está ansioso (no deben incluirse aquellos síntomas que sólo se presentan durante una crisis de angustia):

Tensión motora

- 1) temblor, contracciones o sacudidas;
- 2) tensión o dolor muscular;
- 3) inquietud;
- 4) fatigabilidad excesiva.

Hiperactividad vegetativa

- 5) falta de aliento o sensación de ahogo;
- 6) palpitaciones o ritmo cardiaco acelerado (taquicardia);
- 7) sudoración o manos frías y húmedas;
- 8) boca seca;
- 9) mareo o sensación de inestabilidad;
- 10) náusea, diarrea u otros trastornos abdominales;
- 11) sofocaciones o escalofríos;
- 12) micción frecuente;
- 13) dificultades para tragar o sensación de tener “un nudo en la garganta”.

Hipervigilancia

- 14) sentirse atrapado o al borde de un peligro;
- 15) exageración de la respuesta de alarma;
- 16) dificultad para concentrarse o episodios en los que la mente se queda en blanco debido a la ansiedad;
- 17) dificultades en dormir o en mantener el sueño;

18) irritabilidad.

E. No se ha podido demostrar que un factor orgánico haya iniciado y mantenido la alteración; por ejemplo, un hipertiroidismo o la intoxicación por cafeína” (p. 303).

3.3.3.3. *Criterios para el diagnóstico de episodio depresivo mayor.*

“Nota: Un “síndrome depresivo mayor” se define según el criterio A siguiente:

A. Como mínimo, cinco de los síntomas siguientes han estado presentes durante un mismo período de dos semanas y representan un cambio de la conducta previa y, al menos, uno de los síntomas es o 1) estado de ánimo deprimido, o 2) pérdida de interés o de capacidad para el placer. (No se incluyen los síntomas que se deben claramente a causas físicas, ideas delirantes o alucinaciones no congruentes con el estado de ánimo, incoherencia, o notable pérdida de asociaciones):

1) estado de ánimo deprimido (o puede ser irritable en niños y adolescentes) durante la mayor parte del día, casi cada día y, tal como se indicó, sugerido por la experiencia subjetiva o por la observación de los demás;

2) notable disminución de placer o interés en todas o casi todas las actividades habituales, casi cada día (como se indicó, por referencia subjetiva o por la observación de los demás de apatía durante la mayor parte del tiempo);

3) aumento o pérdida significativa de peso sin hacer ningún régimen (por ejemplo, más de un 5% del peso corporal en un mes), o disminución o incremento del apetito casi cada día (en los niños hay que tener en cuenta la dificultad para alcanzar el peso necesario);

4) insomnio o hipersomnia casi cada día;

5) agitación o enlentecimiento psicomotor casi cada día (observable para los demás, no simplemente el sentimiento subjetivo de inquietud o lentitud);

6) fatiga o pérdida de energía casi cada día;

7) sentimientos excesivos o inadecuados de inutilidad o culpa (que pueden ser delirantes) casi cada día (no simplemente autorreproches o sentimientos de culpa por estar enfermo);

8) disminución de la capacidad para pensar o concentrarse o indecisión, casi cada día (manifestada por el paciente u observada por los demás);

9) ideas de muerte recurrentes (no el simple miedo a morir), ideas de suicidio recurrentes sin un plan específico, o un intento de suicidio o un plan específico para suicidarse.

B. 1) No puede establecerse que una causa orgánica inició y mantiene la alteración.

2) La alteración no es una reacción normal a la muerte de un ser querido (duelo no complicado).

Nota: Preocupación mórbida con sentimientos de inutilidad, ideas de suicidio, notable deterioro funcional o enlentecimiento psicomotor, o una duración prolongada sugieren que el duelo se ha complicado con una depresión mayor.

C. Fuera del contexto de la alteración anímica (antes de la aparición de los síntomas y después de que se han remitido) no ha habido en ningún momento ideas delirantes ni alucinaciones que hayan durado más de dos semanas.

D. El trastorno no está superpuesto a una esquizofrenia, a un trastorno esquizofrenizante, a un trastorno delirante o a un trastorno psicótico no especificado” (p. 266).

3.3.3.4. *Criterios para el diagnóstico del trastorno por somatización.*

“A. Historia de múltiples síntomas físicos o creencia de estar enfermo, que empieza antes de los 30 años y persiste durante varios años.

B. Por lo menos, 13 de los que se exponen a continuación. Para que un síntoma se considere significativo deben cumplirse los siguientes criterios:

1) el síntoma no puede explicarse por ninguna patología orgánica o mecanismo fisiopatológico (trastorno físico, efectos de una herida, medicación, fármacos o alcohol) y cuando hay una patología orgánica, los síntomas o el deterioro social o profesional resultante son mucho mayores de lo que cabría esperar por los hallazgos físicos;

2) los síntomas no aparecen únicamente durante las crisis de angustia;

3) el síntoma ha obligado a la toma de fármacos (no sólo analgésicos), a acudir con frecuencia a las consultas médicas o a alterar el estilo de vida.

Lista de síntomas:

Síntomas gastrointestinales:

1) vómitos (no durante el embarazo);

2) dolor abdominal (distinto del dolor menstrual);

3) náuseas (al margen de las que aparecen por mareo al viajar);

4) distensión (gases);

5) diarrea;

6) intolerancia a ciertos alimentos.

Síntomas dolorosos:

- 7) **Dolor en las extremidades;**
- 8) dolor en la espalda;
- 9) dolor articular;
- 10) dolor al orinar;
- 11) otros dolores (con exclusión de las cefaleas).

Síntomas cardiopulmonares:

- 12) **falta de aliento (sin hacer esfuerzos);**
- 13) palpitaciones;
- 14) dolor torácico;
- 15) mareo o inestabilidad;

Síntomas de conversión o pseudoneurológicos:

- 16) **amnesia;**
- 17) **dificultad para deglutir;**
- 18) pérdida de la voz;
- 19) sordera;
- 20) visión doble;
- 21) visión borrosa;
- 22) ceguera;
- 23) mareo o pérdida de conciencia;
- 24) crisis o convulsiones;
- 25) dificultades en la marcha;
- 26) parálisis o debilidad muscular;
- 27) retención o dificultad para orinar;

Síntomas sexuales durante la mayor parte de la vida de un individuo con oportunidades para la actividad sexual:

- 28) **sensación de quemazón en los órganos sexuales o en el recto (en momentos distintos del coito);**
- 29) indiferencia sexual;
- 30) dolor durante el coito;
- 31) impotencia;

Síntomas sexuales femeninos que la paciente considera que se presentan con más frecuencia o gravedad que en la mayor parte de las mujeres:

- 32) **menstruación dolorosa;**
- 33) ciclos menstruales irregulares;
- 34) excesivas pérdidas menstruales;
- 35) vómitos durante el embarazo.

Nota: Los siete ítems en negrita pueden utilizarse para detectar el trastorno. La presencia de dos o más ítems sugiere una alta posibilidad de padecer el trastorno” (p.315).

Como se puede observar, el SRT en sus treinta reactivos menciona algunos de los principales síntomas mencionados para estos trastornos en la literatura consultada acerca de las neurosis y también coincide con la literatura en su ordenamiento, esto es, el DSM-III y el SRT concuerdan en considerar síntomas similares bajo el mismo rubro, por ejemplo, bajo el nombre de ansiedad o angustia como trastorno o como subescala, se encuentran casi los mismos síntomas.

3.4. Estudios de medición de la confiabilidad y validez de la Escala de síntomas (SRT).

En 1971, Kellner (citado en Kellner y Sheffield, 1973), propuso que la medición de la confiabilidad y validez de una escala diseñada para medir cambios es de alguna manera distinta de aquellas diseñadas para medir un rasgo de personalidad estable. Postuló que una prueba válida que pudiera medir cambios en el malestar psicológico debería:

1. distinguir entre personas afligidas y normales,
2. medir cambios en la aflicción o sufrimiento psicológico comparables a aquellos obtenidos por las mediciones de los psiquiatras,
3. discriminar entre las respuestas a una droga que es conocida por reducir la angustia y las respuestas al placebo, es decir, la escala debe ser validada contra el “criterio químico externo” (Clyde, 1960, citado en Kellner y Sheffield, 1973).

Con el propósito de establecer la confiabilidad y la validez de la SRT, Kellner y Sheffield (1973), reportan que se realizaron varios estudios, descritos abajo, en los que se aplicó la SRT junto con tres escalas de automedición que han sido extensamente validadas y han demostrado ser mediciones válidas de la personalidad: la Escala de Ansiedad Manifiesta de Taylor (MAS) (Taylor, 1953), el Inventario de Personalidad de Maudsley (MPI) (Eysenck, 1959) y el similar Inventario de Personalidad de Eysenck (EPI) (Eysenck y Eysenck, 1964).

La Escala de Ansiedad Manifiesta de Taylor (MAS) (1953) fue derivada del Inventario de Personalidad Multifásica de Minnesota y consta de 50 reactivos extraídos de aquellos enunciados del MMPI que se consideran como francas admisiones de ansiedad, más 175 reactivos amortiguadores, extraídos en gran parte de las tres escalas de validez del MMPI (L, F, K).

El Inventario de Personalidad de Maudsley (MPI) producido por Eysenck (1959) comprende 48 preguntas. La mitad de los elementos miden neuroticismo (N) definido por el autor como “el estado general emocional de una persona, su sobre-respuesta emocional, y su probabilidad de crisis neurótica bajo estrés”. Los restantes 24 reactivos se relacionan a la extraversión (E), “las proclividades desinhibidas, sociables y externadas de una persona”. (Bedford, McIver, y Pearson, 1978).

3.4.1. Confiabilidad (versión semanal).

Se llevó a cabo una confiabilidad test-retest. La confiabilidad test-retest de un inventario de personalidad es evidencia de su estabilidad temporal (ver cap. 2 de esta tesis), una prueba que apunte a medir rasgos de personalidad estables debería tener resultados casi inalterados con el paso del tiempo. Ya que la SRT mide el malestar psicológico -un estado cambiante- el retest después de un largo intervalo de tiempo mediría un estado alterado en una gran proporción de pacientes.

A 28 pacientes neuróticos externos se les readministró la SRT, versión semanal (cartas de prueba) después de un intervalo de 24 horas, la correlación test-retest fue de 0.94. El estudio se repitió con 40 pacientes internos con la versión de papel y lápiz y la correlación fue de 0.92. La MAS se administró junto con la versión de papel y lápiz con objeto de comparar, la correlación test-retest de la MAS fue de 0.82.

Los autores mencionan que la correlación de Split-half o mitades (que es un tipo de confiabilidad convencional) (ver cap. 2 de esta tesis) no fue llevada a cabo porque los reactivos no son psicométricamente equivalentes. La confiabilidad de split-half de los cambios en los puntajes de la SRT en 40 pacientes neuróticos externos fue de 0.89.

Cuatro respuestas descriptivas para la medición de cada síntoma fueron encontradas adecuadas, incrementar el número de respuestas disminuye, al parecer, la confiabilidad de la escala. (Kellner, 1967, Morton, 1968, Kellner, 1972b, citados en Kellner y Sheffield, 1973).

3.4.2. Validez de la lista de control de síntomas (versión semanal).

La lista de control de síntomas se validó administrándola a 40 pacientes neuróticos externos y a 43 sujetos normales como una entrevista semiestructurada, y con la versión de papel y lápiz, autoadministrada, con pacientes neuróticos (externos, internos y diarios externos) y sujetos normales en seis estudios separados. En todos los estudios la lista de control discriminó significativamente entre los pacientes neuróticos y normales ($P < 0.001$) (Kellner, 1967a, Kellner y Sheffield, 1967, Morton, 1968, Kellner, 1972b; citados por Kellner y Sheffield, 1973).

La lista de control SRT se proporcionó a 100 pacientes neuróticos (principalmente internos y diarios) y 100 normales. Treinta y seis reactivos discriminaron entre ambos grupos con $P < 0.001$, y dos reactivos no discriminaron significativamente. Este estudio se realizó en Inglaterra y fue repetido en Nuevo México. En el segundo estudio, en Nuevo México, todos los reactivos discriminaron significativamente entre pacientes neuróticos externos y normales (un reactivo con $P < 0.05$ y el resto con $P < 0.001$). Cuando las automediciones fueron comparadas (al contrario de contar las respuestas en las listas de control) todos los reactivos discriminaron significativamente (Kellner, 1972b; citado en Kellner y Sheffield, 1973).

En un ensayo cruzado de drogas con diazepam, 5mg. t.i.d., hydroxyne pamoate, 50 mg. t.i.d. y placebo con 17 neuróticos ansiosos, la lista de control (papel y lápiz) discriminó significativamente entre los efectos de tratamientos ($P < 0.005$), mientras que la forma de lista de control del MAS y la Escala de Neurosis de la MPI fallaron en discriminar a un nivel significativo (Kellner, et. al., 1968; citado en Kellner y Sheffield, 1973).

3.4.3. Validez de la SRT (forma semanal) y correlaciones con otras pruebas.

La literatura sobre ensayos con drogas en pacientes neuróticos fue revisada y aquellas escalas de medición y automedición que aparecieron como efectivas en detectar diferencias entre una droga activa y un placebo fueron seleccionadas (Kellner, 1971; citado en Kellner y Sheffield, 1973). Estas pruebas fueron comparadas con la SRT (cartas de prueba y papel y lápiz) en cuatro ensayos con drogas. Los resultados obtenidos con la SRT fueron similares a los obtenidos por las mediciones de psiquiatras utilizando la Escala de Medición de Ansiedad de Hamilton (Hamilton, 1959; citado en Kellner y Sheffield, 1973), en las mediciones globales (Kellner, et. al., 1968, Kellner y Sheffield, 1968, Kellner y Claghorn, 1970, Kellner, Gervais y Pathak, 1972; citados en Kellner y Sheffield, 1973). La SRT fue tan sensitiva como

las mediciones de los psiquiatras en detectar las diferencias entre droga/placebo (o entre dos drogas).

En siete estudios la SRT (cartas de prueba y papel y lápiz) fue administrada a sujetos normales y pacientes neuróticos y todos los puntajes discriminaron significativamente entre las dos poblaciones ($P < 0.001$). La SRT discriminó con más precisión que la Escala de Neurosis de la MPI (o de la EPI) y casi tan precisamente como la MAS (Kellner, 1967a, Kellner y Sheffield, 1967, Morton, 1968, Kellner, 1972b; citados en Kellner y Sheffield, 1973).

La validación del método de puntuación (cartas de prueba) fue realizada comparando los puntajes de 16 pacientes neuróticos externos y 24 normales que tenían el mismo número de síntomas. Los neuróticos tenían puntajes totales significativamente más altos que los normales incluso cuando el número de síntomas era el mismo ($P < 0.001$) (Kellner y Sheffield, 1967; citado en Kellner y Sheffield, 1973).

La sensibilidad de la SRT (cuántos “casos” falla en detectar la prueba) y su especificidad (cuántos normales están mal clasificados como “casos”) (Reid, 1960; citado en Kellner y Sheffield, 1973) fueron comparadas con la sensibilidad y especificidad de la MAS y la Escala N de la EPI. Las tres pruebas fueron administradas a 50 pacientes externos, con un grupo distinto de 50 normales. Los promedios y desviaciones estándar de la SRT y subescalas son enumerados en el cuadro 5. La especificidad (número de normales identificados correctamente por los puntajes de las pruebas dividido entre el número total de normales) y la sensibilidad (el número de pacientes correctamente identificados por los puntajes dividido entre el número total de pacientes) están enlistados en el cuadro 6. La mayoría de los pacientes y normales fueron mal clasificados por los resultados de la prueba en la subescala de somatización de la SRT, seguida de la N de EPI. Los puntajes de la MAS fueron similares a los puntajes totales de la SRT y a los de las otras tres subescalas SRT, con el total del puntaje SRT y dos de las subescalas SRT mostrando una menor proporción de sujetos mal clasificados. Los rangos de especificidad y sensibilidad fueron similares a los de la mala clasificación, hubo pocas diferencias entre los rangos de especificidad y sensibilidad de las pruebas en los dos estudios. (Ver cuadro 6).

Cuadro 5. Medias y desviaciones estándar de los puntajes de la SRT (forma semanal, papel y lápiz) obtenidos, principalmente, por dos grupos de nuevos pacientes externos, neuróticos, en dos clínicas, una en Houston, Texas, y otra en Albuquerque, Nuevo México, y los puntajes de dos grupos de normales.

	Subescalas											
	Lista de control		Total de puntajes SRT		Angustia		Depresión		Inadecuación		Somatización	
	x	DE	x	DE	x	DE	x	DE	x	DE	x	DE
Clinica 1 Houston (N 50)	24.87	8.32	51.08	25.14	11.68	6.63	11.62	5.81	13.62	6.62	2.88	5.49
Clinica 2 Albuquerque (N 50)	24.42	7.88	49.82	27.21	11.30	6.10	11.6	6.6	13.44	8.54	5.7	5.0
Normales 1 (N 50)	7.02	5.69	8.98	8.33	1.5	1.9	1.5	2.1	2.26	2.82	1.6	1.8
Normales 2 (N 50)	7.76	6.86	11.88	13.64	2.4	2.9	2.5	2.9	3.18	4.19	1.6	2.3

Fuente: Kellner y Sheffield, 1973, p. 90.

Cuadro 6. Rangos porcentuales de sensibilidad, especificidad y mal clasificados en los puntajes de la SRT (forma semanal, papel y lápiz) en dos estudios y los rangos porcentuales de los puntajes del MAS y el EPI.

	Puntajes de la Escala de síntomas SRT						Taylor MAS	Escala N EPI
	Lista de control	Totales SRT	Angustia	Depresión	Inadecuación	Somatización		
Estudio 1 Principalmente nuevos pacientes neuróticos externos (N 50) y normales (N 50)								
Sensibilidad	88	84	86	82	76	60	76	72
Especificidad	90	92	96	94	90	84	96	86
Mal clasificados	11	12	11	12	17	28	16	21
Estudio 2 Pacientes neuróticos externos (N 50) de otra clínica y otro grupo de normales (N 50)								
Sensibilidad	90	84	82	84	86	68	86	78
Especificidad	82	84	90	88	80	82	80	82
Mal clasificados	14	15	14	14	17	25	17	20

Fuente: Kellner y Sheffield, 1973, p. 91.

Hubo un subgrupo de 11 sujetos normales que habían sido tratados por su médico con tranquilizantes o sedantes en los pasados cinco años, los cuales tuvieron un puntaje significativamente más alto (cartas de prueba) que otros normales ($P < 0.001$) pero menor que

los pacientes neuróticos externos ($P < 0.001$), la Escala N de la MPI no discriminó significativamente entre el subgrupo y los pacientes neuróticos externos (Kellner y Sheffield, 1967; citados en Kellner y Sheffield, 1973).

Se manifestó un marcado descenso en los puntajes de la lista de control de síntomas y los totales SRT (cartas de prueba) en 40 pacientes neuróticos externos después de un mes de tratamiento ($P < 0.001$). Hubo un descenso posterior en los puntajes después de dos meses, pero la diferencia entre un mes y dos meses no fue significativa. Después de dos meses no hubo diferencias significativas entre los puntajes de pacientes neuróticos externos y el subgrupo de normales, los puntajes permanecieron significativamente más altos que aquellos del grupo normal total ($P < 0.001$).

Los 35 pacientes neuróticos probados en su primera atención tenían puntajes SRT (papel y lápiz) significativamente más altos que los 40 pacientes neuróticos internos y diarios ($P < 0.05$), muchos de los cuales habían aparentemente mejorado y estaban listos para ser dados de alta (Kellner, 1967a; citado en Kellner y Sheffield, 1973).

En un estudio longitudinal de 20 pacientes esquizofrénicos ansiosos externos hubo una tendencia significativa en los puntajes SRT (papel y lápiz) y la Escala de Ansiedad de Hamilton en cambiar en la misma dirección ($P < 0.001$). Los puntajes SRT de los pacientes fueron significativamente más altos que los de los normales ($P < 0.001$).

Los puntajes SRT (papel y lápiz) de 16 pacientes internos sufriendo de depresión endógena fueron comparados con los puntajes de 26 pacientes neuróticos internos y de 51 normales. Los pacientes deprimidos tuvieron puntajes SRT significativamente más altos que los pacientes neuróticos internos ($P < 0.01$). Después de tres semanas de tratamiento con antidepresivos su puntaje disminuyó significativamente ($P < 0.001$) pero permaneció significativamente más alto que el de los normales ($P < 0.001$) (Kellner, 1972b; citado en Kellner y Sheffield, 1973).

Los puntajes totales SRT (papel y lápiz) y los puntajes de las subescalas para 52 pacientes alcohólicos internos antes de tratamiento y 31 pacientes alcohólicos después de seis meses de tratamiento con terapia de grupo y medicación (principalmente tranquilizantes menores) fueron comparados con los puntajes de 72 nuevos pacientes neuróticos externos y de 51 normales. Todos los puntajes SRT de pacientes alcohólicos antes de tratamiento fueron significativamente más altos que los de los normales ($P < 0.001$), el puntaje total SRT y los puntajes de las subescalas de ansiedad, depresión e inadecuación fueron significativamente

más bajos que los de pacientes neuróticos externos (ansiedad, depresión e inadecuación con $P < 0.001$, puntaje total SRT con $P < 0.05$). Después del tratamiento todos los puntajes de pacientes alcohólicos fueron significativamente más bajos que antes del tratamiento ($P < 0.001$) y no difirieron significativamente de los de normales (Schenkel y Kellner, 1971; citados en Kellner y Sheffield, 1973).

Los descubrimientos en un ensayo de drogas con neuróticos ansiosos sugieren que la Escala de Neurosis de la MPI, la MAS y la SRT difieren en su estabilidad (Kellner, et. al., 1968; citado en Kellner y Sheffield, 1973). La Escala de Neurosis tiende a medir el rasgo de personalidad de neurosis (o emocionalidad) (Bendig, 1960; citado en Kellner y Sheffield, 1973) y los resultados son más estables que en las otras dos escalas. La SRT tiende a medir angustia recientemente experimentada, un estado temporal y cambiante. La MAS al parecer es intermedia entre las dos.

En un ensayo de drogas con pacientes neuróticos sufriendo reacciones ansiosas dos psiquiatras midieron a los pacientes en cuatro ocasiones utilizando la Escala de Medición de Ansiedad de Hamilton. El coeficiente de correlación de Kendall Tau (Siegel, 1956; citado en Kellner y Sheffield, 1973) entre la SRT y las mediciones promedio de los psiquiatras va de 0.58 a 0.83 (Kellner, 1967a; citado en Kellner y Sheffield, 1973).

En siete estudios con pacientes neuróticos y normales la correlación entre los puntajes SRT (papel y lápiz y cartas de prueba) y la Escala N de la MPI (o EPI) va de 0.26 a 0.74, media de 0.56. Las correlaciones fueron significativas, excepto por una (0.26, puntajes de 40 nuevos pacientes neuróticos externos en cartas de prueba) (Kellner, 1967a, Kellner y Sheffield, 1967, Morton, 1968, Kellner, 1972b; citados en Kellner y Sheffield, 1973). La correlación entre la SRT y la Escala de Automedición de Zung (Zung, 1965; citado en Kellner y Sheffield, 1973) en 69 pacientes neuróticos externos deprimidos y ansiosos fue de 0.66 (Kellner, 1972b; citado en Kellner y Sheffield, 1973).

3.4.4. Estudios con otras formas.

Kellner y Sheffield (1973), reportan que varios de los estudios realizados con la forma semanal fueron repetidos con la forma diaria y los resultados obtenidos fueron similares (Kellner, 1967a, Morton, 1968; citados en Kellner y Sheffield, 1973).

Inicialmente había pocas diferencias entre las indicaciones de las pruebas de automedición en las formas semanal y diaria. Se encontró que estas diferencias no son importantes y la

forma diaria usada en recientes estudios ha sido idéntica a la semanal, excepto por la instrucción inicial. (Ver apartado 3.2 de este capítulo).

En un ensayo con drogas el puntaje promedio con la forma diaria discriminó significativamente entre los tratamientos, mientras que la forma semanal y otras mediciones fallaron (Kellner y Sheffield, 1968; citados en Kellner y Sheffield, 1973).

La forma diaria, al parecer, es adecuada cuando son necesarias varias mediciones durante un bloque corto de tratamiento, por ejemplo, cuando los cambios de un día al otro son estudiados, o para el establecimiento de respuestas tempranas a un tratamiento.

Kellner y Sheffield (1973) mencionan que también se realizaron varios estudios con la forma horaria, por ejemplo, se usó en un estudio de variaciones diurnas en pacientes deprimidos (Kellner 1967a; citado en Kellner y Sheffield, 1973). En otros estudios se utilizaron formas que enfocaban un pasado más distante tales como: ¿Cómo se ha sentido durante los últimos años?. En general, entre más reciente el enfocamiento -por ejemplo, la pasada hora, ahora,- mayor la tendencia a cambios en los puntajes de una prueba a la siguiente, aparentemente reflejando fluctuaciones en la severidad del sufrimiento en los pacientes psiquiátricos. Las formas enfocadas a un pasado más distante son más estables, tienden a mostrar una mayor confiabilidad test-retest, pero son menos adecuadas para la medición de cambios luego de un corto intervalo de tiempo como una o dos semanas.

La forma semanal de la versión corta (papel y lápiz) fue comparada con la original en dos estudios con 70 pacientes neuróticos externos y 70 normales y en un ensayo paralelo con drogas de benzocetamina y placebo. Las diferencias entre las medidas de especificidad y sensibilidad (Reid, 1960; citado en Kellner y Sheffield, 1973), y mala clasificación de las dos versiones fueron pequeñas. Similarmente, las diferencias entre los cambios en el puntaje durante el ensayo con drogas fueron pequeñas y no hubo ventaja en usar la versión original que consume más tiempo. (Kellner, 1972b; citado en Kellner y Sheffield, 1973).

Kellner y Sheffield, 1973, afirman, después de reportar todos los estudios anteriores, que se puede concluir que la SRT es una medida de malestar psicológico válida y confiable.

3.5. Estudios posteriores de medición de la confiabilidad y validez de la SRT.

3.5.1. Validación y confiabilidad de la SRT y la Escala L-22.

En 1980, Raymond Cochrane, de la Universidad de Birmingham, Inglaterra, reporta un estudio comparativo de la Escala de Síntomas (SRT) y la Escala Langner de 22-elementos,

cuyo objetivo fue realizar una evaluación sistemática de la confiabilidad, validez y propiedades psicométricas de la SRT y la Escala L-22 para determinar si son adecuadas para sondeos comunitarios de epidemiología psiquiátrica. Cochrane afirma que “por razones prácticas y de economía, los sondeos a gran escala tienen que usar mediciones breves y de fácil comprensión de los síntomas psicológicos que puedan ser administradas por entrevistadores legos y que no contengan elementos que son percibidos como amenazantes o inaceptables por los entrevistados potenciales. Obviamente, continúa, los objetivos de un estudio que use un instrumento de este tipo serán derrotados si la escala no puede demostrarse como válida, confiable y de importancia psicométrica en todos aspectos” (p. 115).

La Escala Langner de 22 Elementos (Escala L-22) fue desarrollada como un método rápido y barato de medición de la extensión del desorden psicológico en subgrupos de la población (Langner, 1962; citado en Cochrane, 1980). Aunque la escala ha sido objeto de extensas investigaciones metodológicas y algunas críticas, continúa siendo ampliamente utilizada, por lo regular con la asunción de que es unidimensional.

Los puntos a favor de la SRT son que aunque no fue diseñada originalmente como un instrumento para sondeos epidemiológicos, su simplicidad, brevedad y aceptabilidad la proponen como la de mejor elección. Además, la ventaja potencial principal que ofrece la SRT sobre la L-22 es que fue diseñada con una estructura de subescalas distinguible, lo cual significa que se pueden derivar puntajes separados de ansiedad, depresión, inadecuación y desórdenes somáticos así como de un desorden psicológico general. Si esta estructura de subescalas puede ser demostrada empíricamente, representa entonces un logro notable en un total de 30 reactivos.

La muestra la constituyeron 667 miembros de la comunidad en grandes ciudades de Inglaterra y la República de Irlanda que fueron entrevistados en sus hogares por entrevistadores capacitados. En todos los casos los entrevistadores leyeron el cuestionario a los entrevistados. El procedimiento de muestra utilizado es una variedad de la técnica de “camino al azar” donde cada entrevistador empieza en un punto al azar y recibe instrucciones específicas sobre cómo seleccionar las moradas en la ruta que se le entrega. Se dan instrucciones detalladas sobre cuántas moradas deben saltarse entre las entrevistas, cómo enfrentar edificios con ocupación múltiple y sobre cómo doblar esquinas. De esta manera la responsabilidad de seleccionar a los entrevistados potenciales es removida del entrevistador y

se pueden esperar resultados análogos a los de un muestreo azaroso. Chequeos independientes se realizaron en al menos 15% de las entrevistas conducidas y donde aparecían discrepancias éstas eran ya sea registradas o todo el trabajo de ese entrevistador se rechazaba. Ya que el anonimato de las respuestas fue garantizado, muestras separadas fueron requeridas para checar la validez test-retest de las escalas. Se utilizaron muestras separadas para el chequeo de la Escala L-22 (N38) y el chequeo de la SRT (N 50).

Finalmente, con el objeto de comprobar si las escalas discriminaban entre normales y neuróticos, una de las dos escalas fue administrada a dos grupos separados de pacientes que recibían atención en un gran hospital mental de Birmingham, mientras eran pacientes externos o diarios de las clínicas. Se encontró una variedad de diagnósticos con predominancia de condiciones neuróticas.

La SRT se utilizó en la forma sugerida por Kellner y Sheffield (1973) con una modificación. Debido a que el trabajo piloto había mostrado que el formato de opción múltiple con 4 opciones era demasiado pesado para administrarse oralmente, sólo se ofrecieron tres alternativas, es decir, se pidió a los entrevistados decir si habían experimentado cada síntoma “frecuentemente”, “a veces” o “nunca”. Así, sus respuestas fueron medidas como 2, 1, ó 0 respectivamente para derivar los puntajes de la escala. Como no había intención de comparación con el monto limitado de trabajos previos con la SRT, la reducción en el número de alternativas de respuestas no tiene consecuencias.

La única modificación que se hizo a la Escala Lagner fue alterar 1 ó 2 elementos de la primera a la tercera persona para que fueran adecuados a la administración verbal.

El análisis principal se llevó a cabo utilizando la rutina de análisis factorial incluida en la versión 6 del SPSS. El método alpha de análisis factorial (Nie et. al. 1970; citado en Cochrane, 1980) se llevó a cabo en cada escala para la muestra completa de 667 sujetos. El peso original de los factores y los valores Eigen de los factores principales fue registrado. Después, los factores fueron rotados para alcanzar la mejor discriminación entre factores. Se escogió un rotación oblicua en lugar de una rotación ortogonal porque un punto principal en esta parte del estudio era determinar si la estructura de subescalas a priori de la SRT emergía empíricamente. Como no había ninguna razón en particular para creer que las subescalas no estarían relacionadas entre ellas, y de hecho emergió evidencia de que estaban altamente correlacionadas, se prefirió la rotación oblicua como más realista que una rotación ortogonal que habría sido más fácil de interpretar.

Los resultados obtenidos para la validez muestran que ambas escalas fueron válidas, ya que las medias en los puntajes de pacientes y no pacientes fueron muy diferentes. Aunque los grupos de pacientes y no pacientes eran de tamaños diferentes, su edad y composición de sexo eran muy similares. (Ver cuadro 7).

Cuadro 7. Puntajes (medias \pm desviaciones estándar) de varios grupos en dos escalas de síntomas.

Variables	N	SRT	Escala L-22.
Pacientes	40	24.63 \pm 11.95	7.67 \pm 5.12
Comunidad	667	8.03 \pm 9.03	3.02 \pm 3.35
P	----	< 0.001	< 0.001
Hombres	351	6.59 \pm 7.97	2.41 \pm 2.92
Mujeres	316	9.62 \pm 8.81	3.70 \pm 3.66
P	-----	< 0.001	< 0.001

Fuente: Cochrane, 1980, p. 117.

Los cuadros 8 y 9 muestran los resultados principales de este estudio.

Para cada reactivo de la SRT (ver cuadro 8) y de la Escala L-22 (ver cuadro 9) se da la siguiente información:

1. La columna 1 contiene los pesos del primer factor sin rotar en el análisis factorial alpha. Después de las iteraciones, este factor tuvo un valor eigen de 13.97 para la Escala L-22 y de 21.82 para la SRT, que significa que para escalas de 22 y 30 elementos respectivamente los primeros factores fueron responsables de 63.5% y 77.9% de la varianza común atribuible a factores significativos (aquellos con un valor eigen mayor de 1.0). Estos resultados pueden ser interpretados como que cada escala tuvo un solo tema dominante a través de ella que en gran parte determina cómo las personas responderán a los reactivos individuales. No hubo factores subsidiarios importantes para ninguna de las dos escalas.

2. Las columnas 1-4 contienen los pesos de los reactivos de los primeros cuatro factores que emergieron después que se impuso una rotación oblicua. Estos pesos representan la correlación actual entre cada elemento y cada factor. Como los factores mismos están correlacionados, tal como permite el uso de una rotación oblicua, los pesos de los factores contenidos en el cuadro representan la suma de la contribución directa de cada factor a los puntajes de los reactivos y su contribución indirecta a través de esta correlación con otros factores. El cuadrado de los pesos de los factores da la proporción de la varianza del puntaje por reactivo de la cual es responsable ese factor. La correlación actual entre estos cuatro factores varió de -0.47 a +0.38 para la SRT y de -0.34 a + 0.49 para la Escala L-22.

Cuadro 8. Análisis factorial de la SRT (N 667).

Columnas de pesos factoriales	1	2	3	4	5
	Factor sin rotación oblicua		Factores resultantes después de la rotación oblicua		
Reactivos*	Factor 1	Factor 1	Factor 2	Factor 3	Factor 4
Anxiety scale					
3 Nervous	- 0.72	- 0.53	0.45	- 0.37	0.64
5 Scared	- 0.65	- 0.51	0.41	- 0.31	0.48
9 Restless	- 0.73	- 0.61	0.44	- 0.44	0.58
16 Trembling	- 0.53	- 0.30	0.50	- 0.34	0.38
19 Tense	- 0.74	- 0.60	0.44	- 0.55	0.53
23 Recurrent thoughts	- 0.64	- 0.57	0.30	- 0.39	0.51
26 Panic attacks	- 0.54	- 0.47	0.33	- 0.40	0.56
29 Restless sleep	- 0.60	- 0.27	0.35	0.76	0.57
Depression scale					
2 Tired	- 0.65	- 0.45	0.43	- 0.51	0.58
6 Poor appetite	- 0.45	- 0.23	0.36	- 0.26	0.28
8 No hope	- 0.66	- 0.41	0.33	- 0.37	0.47
12 Feeling guilty	- 0.48	- 0.43	0.33	- 0.45	0.28
18 Unworthy, failure	- 0.59	- 0.49	0.19	- 0.36	0.44
24 Lost interest	- 0.68	- 0.41	0.32	- 0.44	0.51
25 Unhappy, depressed	- 0.70	- 0.55	0.27	- 0.42	0.67
30 Waking early	- 0.59	- 0.17	0.35	- 0.43	0.62
Somatic scale					
1 Dizzy, faint	- 0.50	- 0.27	0.63	- 0.35	0.32
4 Pressure in head	- 0.63	- 0.36	0.52	- 0.51	0.48
7 Heart beats fast	- 0.47	- 0.22	0.57	- 0.32	0.34
11 Chest pains	- 0.51	- 0.20	0.50	- 0.35	0.44
14 Aches and pains	- 0.42	- 0.25	0.24	- 0.65	0.33
21 Numb, tingling	- 0.42	- 0.19	0.31	- 0.52	0.23
27 Weakness	- 0.66	- 0.33	0.41	- 0.63	0.49
Inadequacy scale					
10 Poor memory	- 0.43	- 0.44	0.41	- 0.25	0.24
13 Worrying	- 0.69	- 0.56	0.35	- 0.43	0.68
15 People look down on	- 0.53	- 0.42	0.25	- 0.26	0.41
17 Thinking difficult	- 0.56	- 0.78	0.23	- 0.29	0.53
20 Inferior	- 0.50	- 0.35	0.35	- 0.68	0.23
22 Irritable	- 0.65	- 0.58	0.32	- 0.54	0.53
28 Can't concentrate	- 0.66	- 0.63	0.30	- 0.39	0.35

* Reactivos transcritos sin traducir, como aparecen en el cuadro original. Fuente: Cochrane, 1980, p. 118.

Cuadro 9. Análisis factorial de la Escala L-22 (N 667).

Columnas de pesos factoriales	1	2	3	4	5
	Factor sin rotación oblicua		Factores resultantes después de la rotación oblicua		
Reactivos*	Factor 1	Factor 1	Factor 2	Factor 3	Factor 4
1 Feel weak	-0.59	-0.49	0.27	0.33	-0.56
2 Can't get going	-0.56	-0.46	0.19	0.44	-0.49
3 Low spirits	-0.58	-0.46	0.39	0.47	-0.45
4 Feel hot	-0.48	-0.43	0.29	0.13	-0.39
5 Heart beating hard	-0.45	-0.28	0.29	0.27	-0.43
6 Poor appetite	-0.32	-0.14	0.24	0.57	-0.19
7 Restless	-0.46	-0.51	0.17	0.27	-0.28
8 Worrying	-0.45	-0.52	0.10	0.16	-0.30
9 Short of breath	-0.41	-0.22	0.21	0.14	-0.59
10 Nervous	-0.62	-0.54	0.44	0.44	-0.40
11 Fainting	-0.29	-0.16	0.50	0.21	-0.13
12 Sleep trouble	-0.54	-0.41	0.22	0.30	-0.58
13 Stomach trouble	-0.32	-0.20	0.11	0.20	-0.20
14 Poor memory	-0.21	-0.12	0.29	0.11	-0.15
15 Cold sweats	-0.24	-0.16	0.19	0.05	-0.19
16 hands tremble	-0.37	-0.28	0.20	0.57	-0.19
17 Fullness in head	-0.32	-0.21	0.27	0.12	-0.21
18 Worries	-0.54	-0.55	0.14	0.30	-0.44
19 Feel apart	-0.52	-0.51	0.29	0.22	-0.32
20 Nothing turns out right	-0.41	-0.50	0.22	0.09	-0.17
21 Headache	-0.38	-0.35	0.25	0.22	-0.26
22 Nothing worthwhile	-0.51	-0.57	0.21	0.24	-0.31

* Reactivos transcritos sin traducir, como aparecen en el cuadro original.

Fuente: Cochrane, 1980, p. 119.

El cuadro 10 muestra algunas estadísticas descriptivas de la Escala L-22, la SRT y sus subescalas. Estas incluyen las medias de la muestra total y las medias de los coeficientes de correlación inter-reactivos. Es de notarse que el coeficiente es más bajo para la Escala L-22 que para la SRT y las subescalas. Esto indica que quizá hay menos consistencia en la composición de la Escala L-22 porque, aunque un muy alto coeficiente promedio sería indicativo de redundancia de elementos, este índice es propuesto como una medida de unidimensionalidad (Roberts, et. al., 1976; citado en Cochrane, 1980).

Cuadro 10. Medias, confiabilidades e intercorrelaciones de la Escala L-22, la SRT y las subescalas de la SRT. (N = 667).

	L-22	SRT				
		Total	Ansiedad	Depresión	Somatización	Inadecuación
Medias de los puntajes	3.02	8.03	2.51	2.02	1.37	2.12
Medias de las correlaciones entre los reactivos	0.18	0.35	0.42	0.39	0.31	0.37
Confiabilidad (alpha)	0.83	0.94	0.86	0.80	0.75	0.77
Confiabilidad (test-retest)* (una semana)	0.88	0.91	0.85	0.85	0.87	0.82
Correlación con						
Ansiedad	0.80	0.94	----	----	----	----
Depresión	0.80	0.91	0.81	----	----	----
Somatización	0.70	0.82	0.69	0.69	----	----
Inadecuación	0.72	0.88	0.79	0.73	0.60	----
L-22	1.00	0.85	0.80	0.80	0.70	0.72

* Basado en N = 50 (SRT), y N = 38 (Escala L-22).

Fuente: Cochrane, 1980, p. 120.

Se realizaron dos estimaciones de la confiabilidad. El coeficiente alpha de Cronbach, y la confiabilidad test-retest que fue calculada de los puntajes de dos muestras separadas que completaron las mediciones dos veces en un intervalo de una semana. Como las subescalas de la SRT son considerablemente más cortas que las dos escalas completas, es de esperarse que su confiabilidad será menor, pero todavía están dentro de límites aceptables.

Cochrane (1980) concluye que la validez de ambos instrumentos fue confirmada. “Dado que las escalas fueron examinadas por su utilidad como medidas continuas de desorden psicológico en vez de como indicadoras de “casos”, se puede concluir que miden muchos de los síntomas de psicopatología leve” (p. 121), y por tanto son útiles en detectar sufrimiento psicológico en grandes muestras de población. No obstante, señala que los resultados revelan a la SRT como una escala unidimensional, enfatizando que los síntomas psicológicos y somáticos están altamente correlacionados y, en consecuencia, la distribución de los reactivos en las cuatro subescalas son imposibles de discernir.

Acerca de la confiabilidad, quedó demostrado por los datos presentados en el cuadro 10 que la SRT (0.94) obtuvo una más alta confiabilidad por lo que será la prueba a elegir en posteriores estudios, aunque la confiabilidad de la L-22 (0.83) es todavía aceptable.

“Por lo anterior, tanto la SRT como la L-22 son instrumentos potencialmente útiles para medir psicopatología indiferenciada en sondeos comunitarios. Ambas son simples y

económicas, aparecen aceptables a los respondientes, han demostrado confiabilidad y validez. Mucho de la superioridad de la SRT es atribuible a su mayor extensión y método para puntuación más sensitivo. Es claro que psicométricamente la extensión del uso de la SRT en contextos clínicos a no clínicos es totalmente justificada” (p. 123).

3.5.2 Validación y confiabilidad de la traducción italiana de la SRT.

En 1983, Fava, Kellner (et. al.) llevaron a cabo la validación de la traducción italiana de la SRT y del Cuestionario de Síntomas (SQ). Ambas escalas fueron administradas en varios estudios para pacientes psiquiátricos (neuróticos y depresivos), grupos control y pacientes sufriendo de varias enfermedades orgánicas (desórdenes dermatológicos, hipertensión, amenorrea secundaria y pacientes sometidas a amniocentesis).

Los autores mencionan que varias escalas de autoevaluación de malestar psicológico han sido diseñadas en las décadas pasadas, sin embargo, muy pocas han sido usadas y validadas en poblaciones de habla no inglesa. Aunque es imaginable que las escalas de autoevaluación basadas en síntomas son menos influidas por factores culturales que, por ejemplo, otras escalas de autoevaluación que tienen que ver con peculiaridades de personalidad, poco es sabido acerca de su validez en otros idiomas y otras culturas. Existen muchas dificultades semánticas inherentes a la traducción de tests psicológicos, y son necesarias pruebas adicionales para establecer su validez. El propósito del estudio es reportar la validez transcultural de las dos escalas de autoevaluación de malestar psicológico.

Asimismo mencionan que se eligió la SRT pues desde su publicación en 1973, ha sido utilizada en un gran número de estudios, los cuales confirman que es una medida válida, confiable y sensitiva de malestar. Acerca del SQ reportan que evolucionó de la SRT después de varios estudios validatorios cuya meta era hacer escalas más sensitivas. El SQ ha sido utilizado en varios estudios validatorios subsecuentes y en tratamientos de drogas anti-ansiedad y placebo. En dos estudios de drogas, algunas de las escalas del SQ fueron más sensitivas que las de la SRT y otras escalas de medición de síntomas. En un estudio que comparaba malestar en diferentes poblaciones, el SQ mostró ser más sensitivo que la SRT en discriminar entre poblaciones.

La SRT y el SQ fueron traducidos por uno de los autores (Giovanni A. Fava, 1983). Las traducciones italianas fueron dadas a cinco americanos bilingües residentes en Bologna que las retradujeron al inglés, sin conocer los originales, para confirmar su exactitud.

La SRT fue administrada en la forma de papel y lápiz, que contiene una lista de 30 síntomas y cada uno es calificado por el propio paciente en una escala de cuatro respuestas. El SQ es un cuestionario de sí-no con 92 preguntas; contiene cuatro escalas que al igual que en la SRT están basadas en análisis factoriales de síntomas: ansiedad, depresión, síntomas somáticos y hostilidad. Cada escala puede ser dividida en dos subescalas: una subescala de síntomas que consiste en objetos que representan síntomas y una subescala de bienestar que consiste en objetos que indican bienestar. El tiempo de observación de las dos escalas puede ser variado. Por ejemplo, se le puede decir a un paciente que describa cómo se siente ahora (forma de ahora), como se sintió el día de hoy (forma de día), o cómo se sintió durante la semana pasada (forma de semana). En todas las muestras la forma de semana de la SRT y el SQ fue usada excepto para pacientes de obstetricia, en quienes la forma de ahora fue utilizada. Aunque la SRT y el SQ son ambas escalas de malestar válidas y tienen un contenido común, las dos fueron traducidas porque en cierta medida tienen diferentes propiedades psicométricas.

La muestra se constituyó con 464 sujetos provenientes de seis estudios diferentes:

1. Pacientes neuróticos. 20 pacientes neuróticos y un grupo control de 20 empleados universitarios fueron combinados por edad (en décadas), sexo, y estado civil. La edad promedio de los neuróticos fue 28.1 años (± 6.3); 12 pacientes eran hombres y 8 mujeres; 8 eran casados y 12 solteros. Todos los pacientes habían sido remitidos a la clínica psiquiátrica del Instituto de Psicología de la Escuela de Medicina de la Universidad de Bologna. Todos fueron entrevistados por los mismos psiquiatras e incluidos en el estudio si estaban dentro del Criterio de Diagnóstico de Feighner para neurosis ansiosa, neurosis obsesivo compulsiva, neurosis fóbica o histeria. La SRT y el SQ fueron administrados después de las entrevistas.

2. Pacientes depresivos. 40 pacientes depresivos y un grupo control de 40 empleados de hospital fueron combinados por edad (en décadas), sexo, estado civil y clase social. Todos los pacientes habían sido remitidos a la clínica psiquiátrica de la Escuela de Medicina de la Universidad de Padua. Todos tenían depresiones primarias y estaban dentro del criterio para desórdenes depresivos mayores, de acuerdo al Criterio de Investigación Diagnóstica. La SRT y el SQ fueron administrados después de una entrevista psiquiátrica y antes de tratamiento antidepresivo.

3. Pacientes de endocrinología. Para examinar la aflicción experimentada por pacientes con hiperprolactinemia, 20 pacientes hospitalizadas con amenorrea secundaria fueron

estudiadas. Las pacientes fueron divididas en dos grupos: 10 pacientes con amenorrea e hiperprolactinemia; 10 pacientes con niveles normales de plasma prolactina. Un grupo control de 10 mujeres empleadas de hospital fue combinado por edad y clase social. A todas las pacientes les fue administrado el SQ.

4. Pacientes de obstetricia. 44 mujeres embarazadas que habían sido remitidas para amniocentesis. La edad promedio de las mujeres era de 35.2 años (± 6.2). Todas las pacientes eran casadas. Al principio del estudio, ellas estaban en los primeros tres meses de embarazo. El SQ les fue administrado en tres fases entre la octava y la semana 22 del embarazo.

5. Pacientes de dermatología. A 60 internos en la Clínica Dermatológica de la Escuela de Medicina de la Universidad de Padua, sufriendo de desórdenes dermatológicos (20 pacientes, con psoriasis, 20 con urticaria crónica y 20 con infecciones de hongos en la piel) se les aplicó la SRT.

6. Sujetos hipertensos. 57 hombres nuevos hipertensos y un grupo control de 66 hombres normales de población general fueron incluidos en este estudio. A todos se les aplicó la SRT.

Los métodos estadísticos usados fueron la prueba t y un análisis de varianza para evaluar las diferencias entre los grupos. La correlación producto-momento de Pearson fue el método usado para computar las correlaciones.

Dentro del apartado de resultados, los autores reportan que todas las subescalas de la SRT y del SQ discriminaron entre los pacientes neuróticos y los grupos de control a un gran nivel de significancia: todas las subescalas discriminaron con $P < .001$ excepto la subescala de síntomas somáticos del SQ, que discriminó con $P < .01$.

Los depresivos reportaron puntajes más altos que el grupo control en todas las subescalas de la SRT y del SQ. Todas discriminaron entre depresivos y normales a un nivel de significancia de .001, excepto la subescala "bienestar somático" del SQ que discriminó con $P < .02$ y la subescala "amistosa" que no discriminó.

Para los pacientes de endocrinología la mayoría de las escalas del SQ discriminaron significativamente entre los tres grupos.

Para los pacientes de obstetricia cuando las madres supieron del resultado favorable a la amniocentesis, esto llevó a un alivio de ansiedad y depresión y un descenso en los síntomas somáticos, que se reflejó en la SRT y el SQ.

Para los pacientes de dermatología hubo diferencias significativas en los puntajes de ansiedad, depresión, inadecuación y aflicción total entre pacientes con psoriasis y urticaria

crónica y aquéllos con infecciones de hongos.

En los pacientes hipertensos los puntajes de aflicción psicológica fueron más bajos que el grupo control de normotensos.

Los resultados reflejan su utilidad en investigación psiquiátrica y psicosomática, excepto en el caso de la hipertensión, además sugieren que las traducciones italianas son válidas y sensitivas para la medición de aflicción psicológica y pueden ser usadas en investigación tan efectivamente como las originales, asimismo, que tienen un potencial considerable para investigación transcultural.

3.5.3. Validación y confiabilidad de la SRT en México.

En 1994, Mora, Natera y Andrade-Palos, reportan un estudio de validación y confiabilidad de la SRT en estudiantes mexicanos de nivel medio superior. Las autoras afirman que “en México, al igual que en otros países latinoamericanos, existe la necesidad de contar con mediciones válidas y confiables, que son indispensables para la evaluación de psicopatología por medio de encuestas comunitarias. Sin embargo, continúan, los instrumentos usualmente empleados para esta finalidad, son muy largos y su aplicación requiere de mucho tiempo, lo cual los hace inaccesibles para este tipo de población. La Escala de síntomas (SRT), en cambio, permite explorar brevemente la presencia de rasgos psicopatológicos, no sólo en la práctica clínica, sino también incluye algunos síntomas comunes entre la población general, por lo que es valiosa su confiabilidad y validación psicométricas” (p. 15).

La muestra de sujetos quedó integrada por 338 estudiantes de una institución pública de enseñanza media superior, ubicada en una zona popular del sur de la ciudad de México. El instrumento se aplicó a estudiantes del primero al sexto semestre. En promedio, 54 % de los participantes fueron mujeres y el resto hombres. El rango de edades osciló entre los quince y los veintitrés años de edad; el promedio fue de dieciocho años y solamente dos de los sujetos eran mayores de 24 años. El 67% eran estudiantes y el resto, además de estudiar, desempeñaban, primordialmente, labores de empleados o auxiliares, carpinteros, cocineros, choferes, meseros, entre otras actividades (32%). (Ver cuadro 11).

Cuadro 11. Características sociodemográficas de la muestra (N = 338).

Sexo	%	N
Mujeres	54	184
Hombres	46	154
Edad		
15 - 17	31	107
18 - 20	59	199
21 - 34	10	32
Ocupación		
Estudia	68	228
Estudia y trabaja (empleado, vendedor, chofer, mesero)	32	87

Fuente: Mora, Natera y Andrade-Palos, 1994, p. 16.

A esta población se le aplicó la SRT en su versión corta de treinta reactivos, y el formato de respuesta sugerido por Cochrane (1980), que incluye tres opciones de respuesta tales como “Nunca” (0), “Algunas veces” (1) y “A menudo” (2).

La SRT se tradujo al español y después la traducción española se entregó a cinco americanos bilingües que la retradujeron al inglés, sin conocer la original, para confirmar su exactitud.

Se realizó un estudio evaluativo de campo en la población estudiantil de enseñanza media superior y se pidió la colaboración voluntaria de los estudiantes, explicándoles que se trataba de una encuesta para conocer el estado de salud general de la población estudiantil. Asimismo, se les indicó que respondieran con base en las tres opciones de respuesta si habían experimentado cada uno de los treinta síntomas contenidos en la escala, durante los últimos tres meses.

El cuestionario fue autoaplicable, se les indicó a los entrevistados que sus respuestas deberían ser contestadas individualmente y que eran confidenciales. La aplicación se llevó a cabo en los salones de clase y tardó veinte minutos.

Con el fin de obtener la confiabilidad y validez del SRT las autoras reportan la realización de los siguientes estudios:

1. Análisis de discriminación. Dado que se pretendía lograr la variabilidad de los reactivos, la población se dividió en dos grupos; el 25 % que obtuvo el puntaje más alto y el 25% que obtuvo el puntaje más bajo en la escala, eliminando los casos intermedios (indiferentes). A través de la prueba Chi cuadrada, se analizaron las respuestas entre ambos

grupos, encontrándose diferencias significativas en todos los reactivos, por lo cual se incluyeron todos ellos en el análisis factorial.

2. **Consistencia interna.** Se realizó una correlación ítem-escala considerando el supuesto de que toda la escala constituye una sola dimensión, utilizando el modelo alpha de Cronbach para determinar los coeficientes de confiabilidad.

El coeficiente alpha de Cronbach para la escala global de la SRT fue de 0.93, de lo cual se deduce la solidez de la misma.

3. **Análisis factorial.** En la rotación varimax se extrajeron seis factores, que tuvieron un valor Eigen mayor a uno. Los tres primeros factores explicaron el 47.8% de la varianza; el factor 4, el 4.8%; el 5, el 4.2% y, por último, el 6 explicó el 3.8% de la varianza.

Posteriormente se llevó a cabo el análisis de la congruencia conceptual de los reactivos y sus respectivos factores. Se localizaron cuatro áreas principales: inadecuación, somatización, ansiedad y depresión, identificados como los primeros cuatro factores.

El factor 5, que incluyó dos reactivos relacionados con síntomas cognoscitivos, se eliminó del análisis por no conformar una dimensión, y los reactivos del factor 6 se distribuyeron con base en las cargas factoriales en los cuatro primeros factores que los explicaban mejor conceptualmente. Así, el reactivo “descontento o deprimido” se cambió al área de ansiedad con una carga de .43. El reactivo “dificultades para dormir” se cambió a somatización con una carga de .47. Finalmente, el reactivo “pensamientos que no se me pueden quitar de la cabeza” se eliminó debido a que tuvo una carga factorial muy baja en el factor de inadecuación. Por tanto, la escala quedó integrada por veintisiete reactivos divididos en cuatro subescalas (ver cuadro 12), de la siguiente manera:

3.1. Inadecuación: Este factor agrupó siete reactivos. Cuatro de ellos correspondieron a la subescala original de inadecuación, tal es el caso de: 13. preocupaciones, 17. dificultades para pensar claramente, 20. sentimientos de inferioridad y 15. sentimientos de que la gente me considera menos. Asimismo, se incluyeron en esta área los reactivos 12. sentimientos de culpa, 18. sentimientos de poca valía o de fracaso, con cargas factoriales mayores a .60, que en la escala original corresponden a depresión; por último, 19. sentimientos de tensión tuvo una carga de .66 y originalmente se encontraba en ansiedad. El factor obtuvo una confiabilidad de .86.

3.2. Somatización: Está integrado por siete reactivos, de los cuales cuatro de ellos coincidieron con la subescala original, tales como: 1. mareado o débil, 14. dolores musculares

o reumatismo, 21. sensaciones de cosquilleo o adormecimiento en partes del cuerpo y 27. debilidad en partes del cuerpo. Se incluyeron tres reactivos más tales 2. cansado o sin energía con una carga de .62, 6. poco apetito con una carga de .47 que en la escala original corresponden a depresión y, finalmente, 29. dificultades para dormir, con una carga de .47, que originalmente se encuentra en ansiedad. Este factor tuvo una confiabilidad de .82.

3.3. *Ansiedad.* Está integrado por nueve reactivos, de los cuales cinco correspondieron con la subescala original de ansiedad; los reactivos son los siguientes: 3. nervioso, 5. temores o miedos, 9. inquieto o asustadizo, 16. temblores o sacudimientos, 26. ataques de pánico. Se incluyeron cuatro reactivos más, como 4. sensación de presión en la cabeza, que en la escala original corresponde con somatización; mientras que los reactivos 8. sentimientos de que no hay esperanza, 24. pérdida de interés en las cosas, y 25. descontento o deprimido tuvieron cargas mayores a .40 en este factor, y en la escala original se incluyeron en el área de depresión. La confiabilidad alpha de este factor fue de .85.

3.4. *Depresión.* Se integró únicamente por cuatro reactivos como 7. palpitaciones rápidas o fuertes sin razón aparente, 22. irritabilidad, 11. dolores en el pecho y 30. me despierto demasiado temprano y ya no puedo conciliar el sueño otra vez. Este último fue el único reactivo que coincidió con la subescala original, la cual agrupaba inicialmente nueve reactivos, los cuales sufrieron modificaciones durante la adaptación del instrumento y se distribuyeron en las tres subescalas anteriores. Esta subescala fue la más indefinida de las cuatro, tuvo una confiabilidad de .71 y se refiere básicamente a síntomas físicos asociados con la depresión.

Cuadro 12. Resultados del análisis factorial de la SRT (rotación varimax) del estudio de México en 1994.

Reactivos	Inadecuación	Somatización	Ansiedad	Depresión
1. Mareado o débil		.70388		
2. Cansado o sin energía		.62042		
3. Nervioso			.71782	
4. Sensación de presión o tensión en la cabeza			.42197	
5. Con temores o miedos			.65420	
6. Con poco apetito		.47766		
7. Palpitaciones rápidas o fuertes sin razón aparente				.63719
8. Sentimientos de que no hay esperanza			.56400	
9. Inquieto o asustadizo			.48073	
10. Con poca memoria*				
11. Dolores en el pecho, dificultad para respirar o sensación de no tener suficiente aire				.55092
12. Sentimientos de culpa	.66440			
13. Preocupaciones	.39872			
14. Dolores musculares o reumatismo		.65353		
15. Sentimientos de que la gente me considera menos o piensa mal de mi	.72078			
16. Temblores o sacudimientos			.66008	
17. Dificultades para pensar claramente o para tomar decisiones	.54131			
18. Sentimientos de poca valía o de fracaso	.67573			
19. Sentimientos de tensión	.60984			
20. Sentimientos de ser inferior a otras personas	.69706			
21. Sensaciones de cosquilleo o adormecimiento en algunas partes del cuerpo		.59714		
22. Irritabilidad				.59253
23. Con pensamientos que no se pueden quitar de la cabeza*				
24. Pérdida de interés en casi todas las cosas			.50838	
25. Descontento o deprimido			.43538	
26. Ataques de pánico			.42598	
27. Sensaciones de debilidad en algunas partes del cuerpo		.60283		
28. Sin capacidad de concentración*				
29. Tener dificultad para conciliar el sueño o tener pesadillas		.47414		
30. Me despierto demasiado temprano y no puedo conciliar el sueño otra vez				.68715
Alpha de Cronbach	.86	.82	.85	.71
Alpha total = .93				

*Reactivos eliminados del análisis.

Fuente: Mora, Natera y Andrade-Palos, 1994, p. 18.

3.5.3.1. Escala de síntomas por medio de dos subescalas.

Asimismo, las autoras realizaron un análisis factorial, con rotación varimax, ajustando la escala a dos factores con el fin de establecer comparaciones con investigaciones anteriores que han reportado únicamente dos subescalas (Welch, McColl y Peace, 1989; Matson, 1992,

citados en Mora, Natera y Andrade-Palos, 1994)), identificando un factor sólido que explicó el 36% de la varianza, y otro que explicó el 7%.

En el cuadro 13 se muestra la distribución de los reactivos con sus respectivas cargas factoriales, en los dos factores llamados “Síntomas psicológicos” y “Síntomas físicos”. Las confiabilidades obtenidas fueron de 0.91 para la subescala de síntomas psicológicos y de 0.87 para la subescala de síntomas físicos.

Cuadro 13. Subescalas de síntomas psicológicos y físicos de la SRT.

Síntomas psicológicos*	Cargas factoriales	Síntomas físicos**	Cargas factoriales
18. Poca valía	0.74943	22. Irritabilidad	0.65849
20. Inferioridad	0.73115	27. Debilidad en el cuerpo	0.65352
17. Dific. pensar	0.68300	11. Dolores en el pecho	0.65015
19. Tensión	0.67318	6. Poco apetito	0.64887
15. Me consideran menos	0.67202	21. Adormecimiento en el cuerpo	0.64216
12. Sentimientos de culpa	0.64593	1. Mareado	0.63186
8. No hay esperanza	0.62097	2. Cansado	0.59064
13. Preocupaciones	0.58461	3. Nervioso***	0.56274
25. Deprimido	0.55160	29. Dificultades para dormir	0.53857
24. Pérdida de interés	0.54436	14. Dolores musculares	0.52286
23. Pensamientos fijos	0.54370	30. Despertar temprano	0.49907
28. Sin concentración	0.51144	4. Presión en la cabeza	0.49577
9. Inquieto	0.49877	16. Temblores	0.36071
5. Temores	0.48983		
10. Poca memoria	0.47913		
3. Nervioso***	0.43779		
26. Ataques de pánico	0.38061		

* alpha = .91

** alpha = .87.

*** El reactivo número 3 se encuentra reportado en ambas subescalas, lo que puede ser un error de imprenta, pues no aparece el reactivo número 7.

Fuente: Mora, Natera y Andrade-Palos, 1994, p. 19.

Las autoras concluyen que los resultados revelan una alta consistencia interna y demuestran la validez de constructo de la SRT aunque mostraron una serie de variaciones con respecto a la distribución de los reactivos en las subescalas propuestas originalmente (Kellner y Sheffield, 1973), que posiblemente se deben, en primer lugar, al tipo de población considerada en este estudio, y en segundo lugar a una imprecisión metodológica de sus autores quienes elaboraron teóricamente el diseño de la escala sin efectuar la validez de constructo por medio de un análisis factorial, abocándose únicamente a la comparación entre poblaciones con y sin problemas psiquiátricos para validar el instrumento, lo que afecta la

estructura interna de las cuatro subescalas, que fue puesto de relieve en otras investigaciones (Cochrane, 1980).

No obstante, los resultados indican que la SRT puede trabajarse, en población comunitaria, particularmente en la población juvenil, por medio de 27 reactivos divididos en cuatro subescalas.

Partiendo de lo anterior, la presente investigación tiene por objeto estudiar la confiabilidad y validez de la Escala de síntomas (SRT) en México, con otra población, con el fin de comprobar su utilidad en encuestas comunitarias.

Capítulo 4
Metodología

4. Metodología

4.1. Problema de investigación.

¿Cuál es el comportamiento psicométrico de la Escala de síntomas (SRT) con un grupo de población similar al que se aplicó para su validación en México?

4.2. Planteamiento de objetivos.

4.2.1. *Objetivo general.*

Analizar el comportamiento psicométrico de la Escala de síntomas (SRT), en su aplicación a adolescentes mexicanos inscritos en una institución privada de educación media superior (Preparatoria de la Universidad del Valle de México, Plantel Tlalpan), en comparación con el análisis psicométrico que se aplicó con una institución pública.

4.2.2. *Objetivos específicos.*

1. Definir el número y tipo de factores extraídos en el análisis factorial que se realizó con los puntajes obtenidos en la aplicación de la SRT, en estudiantes mexicanos, de la Preparatoria de la Universidad del Valle de México, Plantel Tlalpan.

2. Identificar la confiabilidad alpha de Cronbach de la escala y de cada uno de los factores de la SRT, en su aplicación a estudiantes mexicanos, de la Preparatoria de la Universidad del Valle de México, Plantel Tlalpan.

4.3. Planteamiento de hipótesis.

1. Se encontrará el mismo número y tipo de factores de la SRT, extraídos del análisis factorial, que contiene la Escala original validada en México.

2. El coeficiente de confiabilidad alpha de Cronbach de la escala y de cada uno de los factores extraídos de la SRT, aplicada a estudiantes de una institución privada, es el mismo que el obtenido en la escala original validada en México.

4.4. Variables.

4.4.1. Variable independiente.

Estudiantes inscritos en una institución privada de educación media superior (Preparatoria de la Universidad del Valle de México, Plantel Tlalpan).

4.4.2. Variable dependiente.

Número y tipo de factores, así como el coeficiente de confiabilidad alpha de Cronbach obtenidos por la SRT.

4.5. Definición de variables.

4.5.1. Definición conceptual de la variable independiente.

1. *Institución educativa.* “En sentido lato, toda educación promovida y realizada en el seno de alguna agencia educativa. En sentido restringido, la educación promovida exclusivamente en el centro educativo de cualquier nivel. Tiene dos objetivos: por un lado, como colectividad organizada que persigue unos fines (escuelas) y por el otro, se ocupa de la socialización del alumno y de la transmisión de la herencia cultural de una sociedad de una generación a la otra.” (Diccionario de las Ciencias de la Educación, 1987, citado por Magaña, 1990).

2. *Escuela.* “Conjunto organizado de recursos humanos y físicos bajo la autoridad de un director o responsable, comprometido en impartir enseñanza a alumnos de un mismo nivel.” (Diccionario de las Ciencias de la Educación, 1987, citado por Magaña, 1990).

3. *Escuela privada.* “Es la desarrollada por la iniciativa privada, de carácter particular, de personas privadas, instituciones, padres de familia, etc. financiada totalmente por ellos y/o subvencionada en mayor o menor medida por el Estado y siempre bajo los lineamientos de la Constitución y la Ley.” (Diccionario de las Ciencias de la Educación, 1987, citado por Magaña, 1990).

4. *Educación media superior.* “En los servicios educativos de este nivel se imparte el bachillerato, cuyo objetivo es preparar al estudiante en todas las áreas del conocimiento para poder elegir y cursar estudios de nivel superior; es de carácter propedéutico.” (Estadística Básica del Sistema Educativo Mexicano, 1988, citado por Magaña, 1990).

4.5.2. Definición operacional de la variable independiente.

1. *Institución privada de educación media superior*: se considera el lugar donde los adolescentes acuden a recibir instrucción educativa. Este rubro se refiere al lugar físico en donde los adolescentes contestaron la Escala de Síntomas (SRT): la Preparatoria de la Universidad del Valle de México, Plantel Tlalpan,

4.5.3. Definición conceptual de la variable dependiente.

1. Análisis factorial. Es un método estadístico para obtener la validez de construcción de un instrumento de medición, del cual se derivan los factores o dimensiones que componen una prueba o test psicológico. Con él se obtienen las cargas factoriales de cada factor, sus valores Eigen y la cantidad de varianza que cada factor explica. (Ver capítulo 2 de esta tesis).

2. Coeficiente alpha de Cronbach. Es un método estadístico para obtener la confiabilidad de consistencia interna de los reactivos que constituyen a un instrumento. Entre más cerca de uno esté la puntuación del coeficiente alpha de Cronbach, más alto es el grado de confiabilidad. (Ver capítulo 2 de este estudio).

4.5.4. Definición operacional de la variable dependiente.

Número y tipo de factores, sus cargas factoriales, sus valores Eigen, la cantidad de varianza que cada uno explica, y el coeficiente de confiabilidad alpha de Cronbach de la SRT en su totalidad y de cada factor que la conforma.

4.6. Tipo de estudio.

Según la clasificación de Rodrigues (1981), se trata de una investigación *ex post facto*, porque estudia una situación en la que ya tuvieron lugar las variables independientes y dependientes y es un estudio de campo, ya que se llevó a cabo en el ambiente natural de los sujetos y porque se recopilaron los datos una vez pasado el efecto de la variable independiente. Asimismo, según la clasificación de Hernández, Fernández y Baptista (1993), se trata de un diseño de investigación transeccional descriptivo porque se recolectaron los datos en un solo momento, en un tiempo único; y porque su objetivo fue indagar la incidencia y los valores en que se manifiestan las variables, por lo tanto las hipótesis son descriptivas.

4.7. Definición de los sujetos.

Los sujetos que participaron en la investigación fueron los 456 alumnos de la Preparatoria de la Universidad del Valle de México, Plantel Tlalpan, inscritos en el Plan SEP, que asistieron y entraron al salón de clases el día que se aplicó el instrumento.

La muestra quedó constituida por 456 sujetos de ambos sexos, 272 hombres y 184 mujeres; la edad quedó comprendida entre los 15 y los 26 años (media = 17.4, d.e. = 1.5)

Esta institución privada ofrece a sus alumnos dos planes de estudios para cursar su bachillerato: el Plan UNAM y el Plan SEP. Los estudiantes pueden elegir el plan que más les convenga. Ambos planes se realizan en el turno matutino, a través de un pago mensual a la Universidad del Valle de México. Las diferencias entre ambos planes son las siguientes:

Plan UNAM:

- Incorporado a la Universidad Nacional Autónoma de México.
- Ciclo escolar anual, distribuido en tres años.
- La UNAM rige los programas y planes de estudio, por lo que la Preparatoria del Valle de México no puede hacer reestructuraciones. Desde hace 27 años es el mismo programa.

Plan SEP:

- Incorporado a la Secretaría de Educación Pública.
- Ciclo escolar semestral, distribuido en seis semestres (tres años), por lo que un alumno que reprueba un semestre puede recursarlo en el próximo, bajo un reglamento que regula estos casos.
- La SEP permite que las escuelas incorporadas hagan cambios a los programas de estudio en colaboración con la Secretaría, por lo que la Preparatoria del Valle de México realiza los cambios que considere pertinentes al programa y nombra profesores bajo la aprobación de la SEP.

La población total de los alumnos de la Preparatoria de la Universidad del Valle de México fue, en el momento de la investigación, de 1450, distribuidos de la siguiente manera:

- Plan SEP = 500 alumnos.
- Plan UNAM = 950 alumnos.

4.8. Muestreo.

Se seleccionó una muestra no probabilística, acorde con la índole de esta investigación,

pues como señalan Hernández, Fernández y Baptista (1993), la muestra no probabilística o muestra dirigida consiste “en una cuidadosa y controlada elección de sujetos con ciertas características especificadas previamente en el planteamiento del problema” (p. 231); así entonces fue una muestra de la forma sujetos-tipos (Hernández, Fernández y Baptista, 1993), que se utiliza en estudios exploratorios y en investigaciones de tipo cualitativo, donde el objetivo es la riqueza, profundidad y calidad de la información, y no la cantidad, como lo es esta investigación cuyo objetivo fue observar el comportamiento psicométrico de la Escala de síntomas con un grupo de población similar al que se aplicó para su validación en México.

Así entonces, de la población total de estudiantes que acuden a la Preparatoria de la Universidad del Valle de México, se eligieron los 500 alumnos inscritos en el plan de estudios proporcionado por la Secretaría de Educación Pública, dado que es una población similar, en número, a la abordada en 1994 para la validación de la SRT.

Finalmente, la muestra quedó conformada por 456 estudiantes, de ambos sexos, 272 hombres y 184 mujeres, dado que esa era la población que se encontraba presente en los salones de clase el día que se realizó la aplicación de la Escala de síntomas (SRT). Los 44 alumnos faltantes para la conformación de la totalidad de la población no se encontraban en los salones de clase donde se aplicó la SRT por razones desconocidas a la realizadora de este estudio.

4.9. Instrumento de medición.

Se utilizó la Escala de síntomas, *Symptom Rating Test* (SRT), versión corta de treinta reactivos, forma trimestral, en su traducción al español realizada para la investigación de Mora, Natera y Andrade-Palos (1994). (Ver capítulo 3, de este estudio, donde se encuentra ampliamente desglosada).

4.9.1. Descripción del instrumento.

En la parte superior se encuentran las instrucciones debidamente detalladas para contestar la escala SRT. En la parte inferior se encuentran los 30 reactivos que conforman el instrumento. Cada uno de los reactivos consta de tres opciones de respuesta, de acuerdo con el formato sugerido por Cochrane (1980): “Nunca”, “Algunas veces”, “A menudo”. La respuesta elegida se marca con una “palomita”. En el extremo derecho aparece un paréntesis para la codificación del análisis estadístico. Los valores de cada respuesta son: Nunca (1),

Algunas veces (2), A menudo (3). (Ver anexo A)

4.9.2. Cuestionario sociodemográfico.

Se realizó un cuestionario que consta de 8 preguntas, con la finalidad de conocer las variables sociodemográficas: sexo de los adolescentes, edad, datos familiares, ocupación y datos escolares. (Ver anexo A).

4.10. Ambiente de investigación.

La aplicación del instrumento se efectuó en las aulas de la Preparatoria de la Universidad del Valle de México en el transcurso de la mañana y durante las horas que la coordinación de la institución educativa designó para la realización de la investigación.

Se procuró que los alumnos estuvieran sentados en sus pupitres y permanecieran en silencio, tanto para escuchar las instrucciones como para contestar las preguntas.

4.11. Procedimiento.

En primer lugar se acudió con las autoridades correspondientes de la institución para solicitar la autorización de la aplicación de esta investigación. Los alumnos no fueron previamente notificados de la aplicación del instrumento, con el objeto de evitar el ausentismo y la predisposición.

Una vez obtenida la autorización, se acudió a las aulas y se aplicó el cuestionario sociodemográfico, así como la Escala de síntomas, con el nombre de Cuestionario.

Se usó la misma consigna que utilizaron las investigadoras (Mora, Natera, Andrade-Palos, 1994), que validaron la SRT en una institución pública de educación media superior, por lo que se comentó a los alumnos que los cuestionarios tenían como finalidad la realización de una encuesta para conocer el estado de salud general de la población estudiantil, con las siguientes frases:

“Buenos días, se está haciendo una investigación para conocer el estado de salud general de la población estudiantil, por lo que desearía que cooperaran contestando el cuestionario que se les va a entregar ahora. El cuestionario es anónimo y confidencial. Gracias por su cooperación”.

Una vez que se les proporcionaron los cuestionarios se les dieron las siguientes instrucciones:

“Lean cuidadosamente las instrucciones que se encuentran en la parte superior de su cuestionario. En la hoja número 1 son preguntas sobre algunos datos personales, necesarios para esta investigación, pongan, por favor, una palomita en la respuesta que responda a su situación o contesten la pregunta que se les hace. En la hoja número 2 aparecen preguntas acerca de cómo se han sentido en los últimos tres meses, de la misma forma, pongan una palomita en la respuesta que mejor exprese cómo se han sentido, como aparece en el ejemplo al inicio de su cuestionario. Si tienen dudas, levanten la mano y acudiré a explicarles”.

4.12. Análisis estadístico.

1. Se obtuvieron frecuencias y porcentajes de los datos sociodemográficos de los adolescentes, así como de las respuestas a los reactivos.

2. Con el fin de de realizar la validez de construcción se llevó a cabo, por medio del Paquete Estadístico para las Ciencias Sociales (SPSS), (Nie, Hull, Jeankins, Steinbrenner y Bent, 1975; citados por Reidl, 1990), un análisis factorial de los puntajes obtenidos por los adolescentes en la SRT, del cual se derivaron los factores que según esta investigación conforman a la SRT y sus respectivas cargas factoriales.

3. Para conocer la confiabilidad de la SRT se aplicó el coeficiente alpha de Cronbach a la escala en su totalidad y a cada uno de los factores obtenidos.

Capítulo 5 Resultados

5. Resultados

En este capítulo se presentan los resultados del análisis estadístico que se aplicó con el fin de probar las hipótesis. Los resultados de este estudio se conformaron con la ayuda del Paquete Estadístico Aplicado a las Ciencias Sociales (SPSS) de Nie, Hull, Steinbrenner y Jenkins (1985).

5.1. Análisis estadístico de las variables sociodemográficas y de las respuestas a los reactivos.

Los resultados que a continuación se describirán fueron obtenidos con un grupo de 456 sujetos.

El primer análisis que se llevó a cabo en esta investigación fue desarrollar medidas de tendencia central obteniendo: porcentajes de la base, medias, desviaciones estándar y medianas para las variables sociodemográficas: sexo, edad, grado escolar, ocupación, número de personas que forman la familia, lugar numérico que se ocupa en la familia, promedio escolar, y becados; así como de las respuestas de los adolescentes a los reactivos.

Los resultados son los siguientes: *

5.1.1. Sexo: 272 varones (60%), 184 mujeres (40%).

Cuadro 14. Porcentajes de sexo por edad y ocupación.

Sexo	Total	Sexo		Edad			Ocupación	
		Varón	Mujer	17 años o menos	18 a 20 años	21 años o más	Solamente estudias	Estudias y trabajas
Varón	60%	100%		52%	67%	79%	55%	87%
Mujer	40%		100%	48%	33%	21%	45%	13%
Base	456	272	184	252	184	19	395	62
Resp. válidas	456	272	184	252	184	20	394	62

* La base de los cuadros se constituyó a través de las respuestas válidas a los cuestionarios, por lo que se presentan algunas discordancias en la base, por ejemplo, en el cuadro de edad, un alumno varón no dio su edad, por tanto, en ese rubro, la base es de 455 sujetos.

5.1.2 *Edad*: Comprende de 15 a 26 años, siendo la media total de 17.4 años.

Cuadro 15. Porcentajes, media, desviación estándar y mediana de edad por sexo y ocupación.

Edad	Total	Sexo		Edad			Ocupación	
		Varón	Mujer	17 años o menos	18 a 20 años	21 años o más	Solamente estudias	Estudias y trabajas
15	8%	7%	9%	14%			8%	3%
16	19%	18%	20%	34%			20%	10%
17	28%	23%	36%	52%			31%	15%
18	29%	31%	26%		72%		29%	31%
19	9%	11%	4%		21%		8%	15%
20	3%	4%	3%		7%		2%	7%
21	3%	4%	1%			63%	2%	8%
22	1%	2%				27%		8%
23	0%	0%				5%		2%
26	0%		1%			5%		1%
Media	17.4	17.5	17.2	16.4	18.4	21.6	17.2	18.6
Desv. estándar	1.5	1.6	1.4	.7	.6	1.2	1.3	2.1
Mediana	17.0	18.0	17.0	17.0	18.0	21.0	17.0	18.0
Base	456	272	184	252	184	19	395	62
Resp. válidas	455	271	184	252	184	19	394	61

5.1.3 *Grado escolar*: Comprende desde el primer semestre hasta el sexto semestre. La mayoría de respondientes (31%) se ubicó en el sexto semestre.

Cuadro 16. Porcentajes, media, desviación estándar y mediana de grado escolar por sexo, edad y ocupación.

Grado escolar	Total	Sexo		Edad			Ocupación	
		Varón	Mujer	17 años o menos	18 a 20 años	21 años o más	Solamente estudias	Estudias y trabajas
1o. semestre	9%	8%	9%	14%	3%		10%	3%
2o. semestre	15%	17%	11%	21%	7%	5%	15%	13%
3o. semestre	12%	14%	10%	12%	10%	42%	11%	19%
4o. semestre	25%	27%	22%	31%	18%	16%	27%	16%
5o. semestre	8%	9%	8%	3%	15%	11%	8%	11%
6o. semestre	31%	25%	40%	19%	47%	26%	29%	38%
Media	4.0	3.9	4.3	3.4	4.8	4.1	4.0	4.3
Desv. estándar	1.7	1.6	1.7	1.6	1.4	1.4	1.7	1.6
Mediana	4.0	4.0	4.0	4.0	5.0	4.0	4.0	4.0
Base	456	272	184	252	184	19	395	62
Resp. válidas	456	272	184	252	184	20	394	62

5.1.4. *Ocupación.* De los 456 sujetos, 395 (86%) solamente estudian y 62 sujetos (14%) estudian y trabajan.

Cuadro 17. Porcentajes de ocupación por sexo, edad y ocupación.

Ocupación	Total	Sexo		Edad			Ocupación	
		Varón	Mujer	17 años o menos	18 a 20 años	21 años o más	Solamente estudias	Estudias y trabajas
Solamente estudias	86%	80%	96%	93%	83%	37%	100%	
Estudias y trabajas	14%	20%	4%	7%	17%	63%	100%	
Base	456	272	184	252	184	19	395	62
Resp. válidas	457	273	184	253	184	20	395	62

5.1.5. *Número de personas que forman la familia.* La mayoría de los sujetos (31%) pertenecen a una familia de cuatro miembros.

Cuadro 18. Porcentajes, media, desviación estándar y mediana de número de integrantes de la familia por sexo, edad y ocupación

Integrantes	Total	Sexo		Edad			Ocupación	
		Varón	Mujer	17 años o menos	18 a 20 años	21 años o más	Solamente estudias	Estudias y trabajas
Ninguna	6%	8%	2%	6%	6%		6%	6%
1.0	1%	1%	1%		2%		1%	
2.0	6%	4%	9%	7%	5%	5%	7%	2%
3.0	13%	13%	14%	14%	12%	11%	13%	13%
4.0	31%	29%	35%	27%	38%	26%	32%	27%
5.0	27%	28%	27%	31%	24%	16%	28%	26%
6.0	9%	11%	7%	9%	8%	21%	8%	15%
7.0	4%	4%	3%	4%	3%	5%	4%	5%
8.0	3%	2%	1%	2%	1%	11%	1%	3%
9.0	0%		1%	0%			0%	
12	0%	0%				5%		2%
15	0%	0%			1%			1%
Media	4.2	4.2	4.2	4.2	4.1	5.4	4.1	4.7
Desv. estándar	1.8	2.0	1.4	1.7	1.8	2.3	1.6	2.4
Mediana	4.0	4.0	4.0	4.0	4.0	5.0	4.0	5.0
Base	456	272	184	252	184	19	395	62
Resp. válidas	456	272	184	252	184	20	394	62

5.1.6. *Lugar numérico que se ocupa en la familia.* En este rubro la mayoría de los sujetos, 35%, es segundo hijo y el 34% es primer hijo.

Cuadro 19. Porcentajes de lugar numérico que se ocupa en la familia por sexo, edad y ocupación.

Lugar	Total	Sexo		Edad			Ocupación	
		Varón	Mujer	17 años o menos	18 a 20 años	21 años o más	Solamente estudias	Estudias y trabajas
Hijo único	6%	8%	4%	8%	5%		7%	3%
Primer hijo	34%	32%	36%	32%	35%	37%	34%	31%
Segundo hijo	35%	33%	37%	37%	34%	32%	36%	27%
Tercer hijo	15%	15%	15%	15%	14%	16%	14%	19%
Cuarto hijo	6%	7%	5%	5%	7%	11%	6%	10%
Quinto hijo	1%	1%	1%	0%	1%	4%	1%	3%
Sexto hijo	0%		1%	0%	1%		1%	
Ultimo hijo	3%	4%	1%	3%	3%		1%	7%
Base	456	272	184	252	184	19	395	62
Resp. válidas	456	272	184	252	184	20	394	62

5.1.7. *Promedio escolar.* La media del promedio escolar es 8.3, siendo el 38% de los sujetos con un promedio de 8.1 a 9.0.

Cuadro 20. Porcentajes, media, desviación estándar y mediana de promedio escolar por sexo, edad y ocupación.

Promedio	Total	Sexo		Edad			Ocupación	
		Varón	Mujer	17 años o menos	18 a 20 años	21 años o más	Solamente estudias	Estudias y trabajas
6 - 7	9%	16%	2%	14%	7%	6%	10%	10%
7.1 - 8	38%	38%	35%	36%	38%	30%	37%	34%
8.1 - 9	38%	38%	41%	35%	43%	42%	38%	44%
9.1 - 10	15%	8%	22%	15%	12%	22%	15%	12%
Media	8.3	8.2	8.6	8.3	8.3	8.5	8.3	8.3
Desv. estándar	.7	.7	.7	.8	.7	.7	.8	.7
Mediana	8.3	8.0	8.5	8.0	8.4	8.5	8.2	8.5
Base	456	272	184	252	184	19	395	62
Resp. válidas	437	255	182	240	180	17	379	58

5.1.8. *Becados.* El 11% de los sujetos están becados por la institución.

Cuadro 21. Porcentajes de becados, por sexo, edad y ocupación.

Beca	Total	Sexo		Edad			Ocupación	
		Varón	Mujer	17 años o menos	18 a 20 años	21 años o más	Solamente estudias	Estudias y trabajas
Sí	11%	8%	16%	13%	8%	6%	12%	10%
No	89%	92%	84%	87%	92%	94%	88%	90%
Base	456	272	184	252	184	19	395	62
Resp. válidas	455	271	184	252	184	19	393	62

5.1.9. *Respuestas a los reactivos:* Donde 1 es “Nunca”, 2 es “Algunas Veces” y 3 es “A menudo”.

Cuadro 22. Porcentajes, media, desviación estándar y mediana de los puntajes de la SRT.

Reactivo	Nunca %	Algunas veces %	A menudo %	Media	Desviación estándar	Mediana
1. Mareado o débil	37	57	6	1.7	.6	2.0
2. Cansado o sin energía	22	71	7	1.8	.5	2.0
3. Nervioso	26	60	14	1.9	.6	2.0
4. Sensación de presión...	39	51	9	1.7	.6	2.0
5. Con temores y miedos	58	38	4	1.5	.6	1.0
6. Con poco apetito	52	40	8	1.6	.6	1.0
7. Palpitaciones rápidas...	73	21	5	1.3	.6	1.0
8. Sent. de que no hay esperanza	65	30	5	1.4	.6	1.0
9. Inquieto o asustadizo	50	46	4	1.5	.6	1.0
10. Con poca memoria	52	43	5	1.5	.6	1.0
11. Dolores en el pecho...	65	30	5	1.4	.6	1.0
12. Sentimientos de culpa	66	31	2	1.4	.5	1.0
13. Preocupaciones	11	77	13	2.0	.5	2.0
14. Dolores musculares...	51	45	3	1.5	.6	1.0
15. Sent. que la gente me considera menos	68	28	4	1.4	.6	1.0
16. Temblores o sacudimientos	76	22	2	1.3	.5	1.0
17. Dific. para pensar...	58	40	2	1.4	.5	1.0
18. Sent. de poca valía	75	24	2	1.3	.5	1.0
19. Sent. de tensión...	61	36	3	1.4	.6	1.0
20. Sent. de ser inferior...	80	19	1	1.2	.4	1.0
21. Sensaciones de cosquilleo...	47	49	5	1.6	.6	2.0
22. Irritabilidad	38	51	11	1.7	.6	2.0
23. Con pensamientos que no se pueden quitar de la cabeza	23	64	13	1.9	.6	2.0
24. Pérdida de interés...	58	37	5	1.5	.6	1.0
25. Descontento o deprimido	26	67	7	1.8	.5	2.0
26. Ataques de pánico	94	6	0	1.1	.3	1.0
27. Sensaciones de debilidad...	59	39	2	1.4	.5	1.0
28. Sin capac. de concentración	40	55	5	1.6	.6	2.0
29. Dific. para conciliar sueño...	55	37	8	1.5	.6	1.0
30. Me despierto temprano...	66	28	5	1.4	.6	1.0

5.2. Validez .

Para obtener la validez de construcción se llevó a cabo un análisis factorial con los puntajes obtenidos por los 456 sujetos que respondieron la SRT. Se obtuvieron 6 factores de la rotación varimax con los siguientes valores Eigen: (La distribución de los reactivos en cada factor y sus respectivas cargas factoriales se presentan en el cuadro 26).

Cuadro 23. Valores Eigen de los factores extraídos de la rotación varimax.

Factores	Valor Eigen
1. Somatización	7.13
2. Ansiedad	1.46
3. Inadecuación	0.80
4. Depresión	0.63
5. Alteraciones cardíacas	0.58
6. Alteraciones del sueño	0.44

5.3. Confiabilidad.

Para obtener la confiabilidad de toda la escala, así como la de cada factor, se aplicó el coeficiente alpha de Cronbach. En el cuadro 24 aparecen los valores de este coeficiente para cada reactivo. La escala en su totalidad obtuvo un valor de 0.90, de lo que se puede deducir la solidez de la misma. Los factores tuvieron los siguientes valores para el coeficiente alpha de Cronbach:

Cuadro 24. Valores del Coeficiente alpha de Cronbach de los factores obtenidos.

Factores	Coeficiente alpha de Cronbach
1. Somatización	0.74
2. Ansiedad	0.76
3. Inadecuación	0.71
4. Depresión	0.76
5. Alteraciones cardiacas	0.59
6. Alteraciones del sueño	0.53
Coeficiente total: 0.90	

5.4. Varianza.

El porcentaje de la varianza que explican los factores quedó distribuida de la siguiente manera:

Cuadro 25. Porcentaje de varianza que explica cada factor

Factores	Porcentaje de varianza
1. Somatización	23.8%
2. Ansiedad	4.9%
3. Inadecuación	2.7%
4. Depresión	2.1%
5. Alteraciones cardiacas	1.9%
6. Alteraciones del sueño	1.5%

5.5. Totales.

En el cuadro 26 se presentan los resultados de la distribución de los reactivos en cada factor, sus respectivas cargas factoriales, los valores Eigen y el coeficiente alpha de Cronbach que obtuvo cada factor.

En el cuadro 27 se presentan la media, la desviación estándar y la mediana, totales y por factor, de las respuestas a los reactivos de cada factor y de la SRT completa.

Cuadro 26. Factores, cargas factoriales de cada reactivo, valores Eigen y el coeficiente alpha de Cronbach de cada factor.

Reactivos	Factores					
	Somatización Eigen: 7.13	Ansiedad Eigen: 1.46	Inadecuación Eigen: .80	Depresión Eigen: .63	Alteraciones cardíacas Eigen: .58	Alteraciones del sueño Eigen: .44
1. Mareado o débil	0.48					
2. Cansado o sin energía	0.46					
4. Sensación de presión o tensión en la cabeza.	0.33					
14. Dolores musculares o reumatismo.	0.43					
21. Sensaciones de cosquilleo o adormecimiento en partes de mi cuerpo.	0.42					
22. Irritabilidad	0.45					
27. Sensaciones de debilidad en partes del cuerpo.	0.49					
3. Nervioso.		0.33				
5. Con temores o miedos		0.52				
9. Inquieto o asustadizo.		0.54				
12. Sentimientos de culpa.		0.44				
13. Preocupaciones		0.42				
19. Sentimientos de tensión o como si me hubiesen dado cuerda (como un reloj).		0.47				
25. Descontento o deprimido		0.36				
10. Con poca memoria			0.4			
17. Dificultad para pensar claramente o para tomar decisiones.			0.44			
23. Con pensamientos que no se pueden quitar de la cabeza.			0.41			
24. Pérdida de interés en casi todas las cosas.			0.49			
28. Sin capacidad de concentración.			0.6			
8. Sentimientos de que no hay esperanza.				0.38		
15. Sentimientos de que la gente me considera menos o piensa mal de mi.				0.66		
18. Sentimientos de poca valía o fracaso.				0.5		
20. Sentimientos de ser inferior a otras personas.				0.78		
7. Palpitaciones rápidas o fuertes sin razón aparente.					0.51	
11. Dolores en el pecho o dificultad al respirar o sensación de no tener suficiente aire.					0.46	
16. Temblores o sacudimientos					0.38	
26. Ataques de pánico.					0.36	

Continuación Cuadro 26.						
	Somatización	Ansiedad	Inadecuación	Depresión	Alteraciones cardíacas	Alteraciones del sueño
29. Dificultad para conciliar el sueño, o sueño desasosegado, o pesadillas.						0.4
30. Me despierto demasiado temprano y no puedo conciliar el sueño otra vez						0.71
Confiabilidad	0.74	0.76	0.71	0.76	0.59	0.53
Confiabilidad total de la Escala de Síntomas: 0.90						

Cuadro 27. Media, mediana y desviación estándar de las puntuajes de la SRT, totales y por sexo, de cada factor y la SRT completa, donde “nunca” corresponde a 1, “algunas veces” a 2 y “a menudo” a 3.

	Total	Varón	Mujer
Factor: Somatización			
Media	1.6	1.6	1.6
Desviación estándar	.4	.3	.4
Mediana	1.6	1.6	1.7
Factor: Ansiedad			
Media	1.6	1.6	1.7
Desviación estándar	.4	.4	.3
Mediana	1.6	1.6	1.7
Factor: Inadecuación			
Media	1.6	1.6	1.7
Desviación estándar	.4	.4	.4
Mediana	1.6	1.6	1.6
Factor: Depresión			
Media	1.3	1.3	1.4
Desviación estándar	.4	.4	.4
Mediana	1.3	1.0	1.3
Factor: Alteraciones cardíacas			
Media	1.3	1.2	1.3
Desviación estándar	.3	.3	.4
Mediana	1.3	1.0	1.3
Factor: Alteraciones del sueño			
Media	1.5	1.4	1.5
Desviación estándar	.5	.5	.5
Mediana	1.5	1.5	1.5
Escala de Síntomas (completa)			
Media	1.5	1.5	1.6
Desviación estándar	.3	.3	.3
Mediana	1.5	1.4	1.5
Base	456	272	184

Como se puede observar sólo los dos primeros factores obtuvieron un valor Eigen mayor a uno, por lo que resultan los más significativos, los válidos, y explican la varianza que es de 29%. Dado que la escala obtiene valores más altos en somatización y ansiedad, son éstas las áreas que mejor la definen.

Capítulo 6

Discusión y conclusiones

6. Discusión y conclusiones

La discusión de los resultados de esta investigación se hará en función a la presentación de las hipótesis del estudio, lo que permitirá analizar el comportamiento psicométrico de la Escala de síntomas (SRT) en su aplicación a adolescentes mexicanos inscritos en una institución privada de educación media superior (Preparatoria de la Universidad del Valle de México, Plantel Tlalpan), en comparación con el análisis psicométrico que se aplicó en una institución pública, por lo que se analiza la validez y confiabilidad.

En lo que respecta a la primera hipótesis que enuncia que se encontrará, en este estudio, el mismo número y tipo de factores de la SRT, extraídos del análisis factorial - el cual se considera como uno de los análisis más importantes para medir la validez de construcción, dentro de los métodos interpruebas, y que permite determinar “cuántos factores (construcciones) se necesitan para explicar las intercorrelaciones entre un conjunto de calificaciones de varias pruebas” (Brown, 1980) -; en comparación con el número y tipo de factores que se obtuvieron en la escala original validada en México; los resultados muestran que:

1. De la rotación varimax, que es una de las técnicas de la rotación ortogonal, se obtuvieron, en ambas investigaciones, seis factores.

2. En el estudio de 1994 estos resultados se modificaron pues sus autoras realizaron, después de la rotación varimax, un análisis de congruencia conceptual de los reactivos que conformaban los seis factores, produciendo reacomodos de los reactivos en los factores, basándose en las cargas factoriales que éstos presentaron en otros factores, con lo que redujeron la escala a cuatro factores: inadecuación, somatización, ansiedad y depresión, y 27 reactivos en total para toda la escala.

3. En esta investigación no se realizó un análisis de congruencia conceptual de los reactivos que conformaron los seis factores pues los resultados no lo permitieron, al no existir valores a los cuales acudir para conformar una escala de cuatro factores, por tanto se mantuvieron los valores que cada reactivo obtuvo de la rotación varimax, con lo que la escala aparece conformada por seis factores: somatización, ansiedad, inadecuación, depresión,

alteraciones cardíacas y alteraciones en el sueño, en los que se distribuyen los 30 reactivos que constituyen la escala original, versión corta, creada por Kellner en 1972.

4. Así entonces, y partiendo de lo anterior, se realizó una comparación de los reactivos y sus cargas factoriales que conformaron los factores finales de ambas investigaciones, con el fin de observar sus similitudes y diferencias, y así comprobar o rechazar la primera hipótesis de este estudio. De esta comparación resultó lo siguiente:

4.1. En el factor de somatización:

Factor: Somatización

1994: Escuela pública.

1. Mareado o débil = .70

2. Cansado o sin energía = .62

6. Con poco apetito = .48

14. Dolores musculares o

reumatismo = .65

21. Sensaciones de cosquilleo o

adormecimiento en partes de

mi cuerpo = .60

27. Sensaciones de debilidad

en partes del cuerpo = .60

29. Dificultad para conciliar el sueño

o sueño desasosegado,

o pesadillas = .47

Factor: Somatización

1996: Escuela privada

1. Mareado o débil = .48

2. Cansado o sin energía = .46

4. Sensación de presión o

tensión en la cabeza = .33

14. Dolores musculares o

reumatismo = .43

21. Sensaciones de cosquilleo o

adormecimiento en partes de

mi cuerpo = .42

22. Irritabilidad = .45

27. Sensaciones de debilidad en

partes de mi cuerpo = .49

Como se puede observar, cinco reactivos, de los seis de escuela pública y los siete de escuela privada, aparecieron en ambas poblaciones aglutinados en el mismo factor aunque sus valores factoriales son muy diferentes, por lo que se puede afirmar que la SRT en este factor se comportó de una forma similar con las dos poblaciones.

4.2. En el factor de ansiedad los resultados presentaron la siguiente forma:

Factor: Ansiedad

1994: Escuela pública

3. Nervioso = .72

4. Sensación de presión o

tensión en la cabeza = .42

Factor: Ansiedad

1996: Escuela privada

3. Nervioso = .33

5. Con temores o miedos = .52

- | | |
|---|---|
| 5. Con temores o miedos = .65 | 9. Inquieto o asustadizo = .54 |
| 8. Sentimientos de que no hay
esperanza = .56 | 12. Sentimientos de culpa = .44 |
| 9. Inquieto o asustadizo = .48 | 13. Preocupaciones = .42 |
| 16. Temblores o sacudimientos = .66 | 19. Sentimientos de tensión o
como si me hubiesen dado
cuerda (como un reloj) = .47 |
| 24. Pérdida de interés en casi todas
las cosas = .50 | 25. Descontento y deprimido = .36 |
| 25. Descontento o deprimido = .44 | |
| 26. Ataques de pánico = .43 | |

En este factor cuatro reactivos, en ambas poblaciones, son los mismos, de nueve que contiene el de escuela pública y siete el de escuela privada. Las diferencias en las cargas factoriales son importantes, por lo que se puede considerar que en este factor la SRT no se comportó en forma similar.

4.3. Para el factor de inadecuación los resultados se presentaron de la siguiente manera:

Factor: Inadecuación	Factor: Inadecuación
1994: Escuela pública	1996: Escuela privada
12. Sentimientos de culpa = .66	10. Con poca memoria = .40
13. Preocupaciones = .40	17. Dificultad para pensar claramente o tomar decisiones = .44
15. Sentimientos de que la gente me considera menos o piensa mal de mi = .72	23. Con pensamientos que no se me pueden quitar de la cabeza = .41
17. Dificultades para pensar claramente o tomar decisiones = .54	24. Pérdida de interés en casi todas las cosas = .49
18. Sentimientos de poca valía = .68	28. Sin capacidad de concentración = .60
19. Sentimientos de tensión o como si me hubiesen dado cuerda (como un reloj) = .61	
20. Sentimientos de ser inferior a otras personas = .70	

En este factor sólo un reactivo resultó ser el mismo para las dos investigaciones, de los

siete de escuela pública y los cinco de escuela privada, y aunque la carga factorial para ese reactivo es similar, se puede afirmar que la SRT se comportó en forma diferente hacia ambas poblaciones.

4.4. Para el factor de depresión se obtuvieron los siguientes resultados:

Factor: Depresión

1944: Escuela pública

7. Palpitaciones rápidas o

fuertes sin razón aparente = .64

11. Dolores en el pecho o dificultad al

respirar o sensación de no tener

suficiente aire = .55

22. Irritabilidad = .59

30. Me despierto demasiado temprano

y no puedo conciliar el sueño

otra vez = .69

Factor: Depresión

1996: Escuela privada

8. Sentimientos de que no hay

esperanza = .38

15. Sentimientos de que la gente

me considera menos o piensa

mal de mi = .66

18. Sentimientos de poca valía o

fracaso = .50

20. Sentimientos de ser inferior

a otras personas = .78

En este factor la SRT se comportó de forma totalmente diferente con ambas poblaciones.

4.5. Factores: alteraciones cardiacas y alteraciones del sueño:

Factor: Alteraciones cardiacas

1996: Escuela privada

7. Palpitaciones rápidas o fuertes sin

razón aparente = .51

11. Dolores en el pecho o dificultad al

respirar o sensación de no tener

suficiente aire = .46

16. Temblores o sacudimientos = .38

26. Ataques de pánico = .36

Factor: Alteraciones del sueño.

1996: Escuela privada

29. Dificultad para conciliar el sueño,

o sueño desasosegado,

o pesadillas = .40

30. Me despierto demasiado temprano y no puedo conciliar el sueño otra vez = .71

Como se puede observar, en este estudio se obtuvieron dos factores más: alteraciones cardíacas y alteraciones del sueño, que fueron nombradas de esa forma dado que los reactivos que obtuvieron mayor carga factorial en cada una de esas dimensiones aluden a síntomas relativos al corazón y al sueño. Factores de esta índole no son mencionados en el estudio realizado en México en 1994. Aunque hay que hacer notar que las autoras de dicho estudio, al igual que en esta investigación, extrajeron seis factores, de los cuales el factor cinco, que incluyó dos reactivos relacionados con síntomas cognoscitivos, fue eliminado del análisis por no conformar una dimensión, y tres reactivos, como fueron: “descontento o deprimido” que se cambió al área de ansiedad con una carga de .43; “dificultades para dormir” que se cambió a somatización con una carga de .47 y “pensamientos que no se me pueden quitar de la cabeza” que se eliminó debido a que tuvo una carga factorial muy baja en el factor de inadecuación. Entonces, aunque ambas investigaciones discriminaron seis factores, como se puede ver por los datos anteriores, fueron diferentes, por lo que se puede concluir que se rechaza la primera hipótesis de esta investigación por que no se obtuvieron los mismos factores.

En relación a la segunda hipótesis de este trabajo que enuncia que el coeficiente alpha de Cronbach -que es un procedimiento utilizado con bastante frecuencia para verificar la consistencia interna de una prueba, con el objeto de medir su confiabilidad (Cronbach, 1951)- de la Escala de síntomas (SRT) y de cada uno de los factores extraídos será el mismo para ambas poblaciones (escuela pública y escuela privada), se obtuvieron los siguientes resultados:

Año	Escuela	Confiabilidad por factores					
		Somatización	Ansiedad	Inadecuación	Depresión	Alter. card.	Alter. sueño
1994	Escuela pública	.82	.85	.86	.71		
1996	Escuela privada	.74	.76	.71	.76	.59	.53

Confiabilidad de la Escala total

1994	Escuela pública	.93
1996	Escuela privada	.90

De acuerdo con estos resultados la segunda hipótesis también se rechaza por que hubo valores diferentes en la confiabilidad de cada uno de los factores y en la escala total.

Como se puede observar, son varias las diferencias entre ambos estudios, por lo que a continuación se analizarán esas diferencias y se propondrán algunas posibles explicaciones:

Mora, Natera y Andrade-Palos (1994), por su lado, en la población adolescente de escuela pública, extraen cuatro factores, de los cuales se identificó un factor principal, designado como inadecuación, similar en cuatro reactivos al creado por Kellner (1972) que enfatiza el aspecto social; y se refiere particularmente a la presencia de sentimientos de inferioridad y de un cierto desajuste para funcionar adecuadamente en situaciones sociales, extrayendo del análisis a los síntomas cognoscitivos que no fueron representativos de esa área y que por esa razón fueron eliminados. Las autoras del estudio concluyen que es probable que por tratarse de una población de estudiantes, los síntomas observados se hallaran enfocados hacia el aspecto social, más que a la presencia de síntomas cognoscitivos y hacia otras áreas que designan psicopatología.

Asimismo, reportan y analizan el cambio que se produjo en algunos reactivos como, por ejemplo, el reactivo "sentimientos de culpa" que aparece en inadecuación y que teóricamente pertenece al área de depresión, diciendo que quizás estén presentes ciertos rasgos culpígenos frente a situaciones sociales que se explican mejor en el área de inadecuación. Con respecto a los cambios de reactivos del área de ansiedad donde se incluyeron algunos síntomas depresivos como "pérdida de interés en casi todas las cosas", "sentimientos de que no hay esperanza" y "descontento o deprimido", sugieren la posibilidad de unir ambas subescalas en una sola, como ansiedad-depresión.

Con respecto al factor de somatización mencionan que al igual que el de ansiedad no se deben a padecimientos orgánicos comprobables, y son conceptualmente congruentes con los criterios diagnósticos definidos por el *Manual para el diagnóstico de enfermedades mentales*, en su versión revisada (DSM-III-R, 1980).

Finalmente, dicen, la subescala de depresión se integró únicamente con cuatro reactivos y aclaran que no mide la depresión clínica, sino que evalúa únicamente los síntomas físicos

asociados con ese padecimiento. “La irritabilidad, que se incluyó en esta área, es un síntoma característico asociado con la depresión durante el periodo de la adolescencia; mientras que los síntomas como “dolores en el pecho o sensación de no tener suficiente aire” y “palpitaciones rápidas o fuertes sin razón aparente” se asocian con las subescalas de somatización y ansiedad, respectivamente, con cargas factoriales mayores a .30” (p. 21).

En este estudio, con población adolescente de escuela privada, en cambio, se identificó como factor principal el de somatización, que enfatiza los síntomas somáticos, congruentes teóricamente con los criterios expuestos por el *Manual diagnóstico y estadístico de los trastornos mentales*, DSM-III-R (1988), con cinco de los siete reactivos propuestos para esta subescala por Kellner (1972) y con cinco de los siete reactivos encontrados por Mora, Natera y Andrade-Palos (1994), de la siguiente forma:

Kellner, 1972.

Mora, Natera, Andrade-Palos, 1994

Barrié, 1996.

1. Mareado o débil	1. Mareado o débil	1. Mareado o débil
4. Sensación de presión...	2. Cansado o sin energía	2. Cansado o sin energía
7. Palpitaciones rápidas	6. Con poco apetito	4. Sensación de presión...
11. Dolores en el pecho	14. Dolores musculares o reumatismo	14. Dolores musculares o...
14. Dolores musculares	21. Sensaciones de cosquilleo...	21. Sensaciones de cosquilleo...
21. Sensaciones de cosquilleo...	27. Sensaciones de debilidad en...	22. Irritabilidad
27. Sensaciones de debilidad...	29. Dificultad para conciliar sueño...	27. Sensaciones de debilidad...

Es probable que la identificación del factor somatización como principal, en este estudio, se deba a las características de la población adolescente abordada, 15 a 26 años, con una media de 17.4 años, pues como menciona Bloss (1971) esta es una etapa en la que se pueden presentar estados de angustia y somatizaciones debido a la enorme cantidad de cambios corporales y sociales que vive el adolescente. Asimismo, Aberastury (1992) señala que “los cambios psicológicos que se producen en la adolescencia son el correlato de cambios corporales que llevan a una nueva relación con los padres y con el mundo” (p. 15).

Es posible pensar, también, que este resultado no contradice los resultados de Mora, Natera y Andrade-Palos (1994) cuyo factor principal resalta preocupaciones de la población hacia el aspecto social, incluso se puede decir que los hallazgos de ambos estudios concuerdan teóricamente con las preocupaciones adolescentes en relación a su cuerpo y a las relaciones con el mundo como subraya Aberastury (1992), tomando en cuenta que también en 1994 la subescala de somatización es la que aparece, en importancia, después de inadecuación. Sin embargo, es posible concluir, que en la población de esta escuela privada,

los jóvenes están más aquejados por sus sensaciones corporales, por su propia persona, que por los aspectos hacia el grupo social, como lo están los de aquella escuela pública. Por tanto, los diferentes resultados en este aspecto pueden ser explicados por las diferencias psico-sociales de ambas poblaciones.

Por último, la presencia del reactivo "irritabilidad", que en la escala original creada por Kellner (1972) corresponde a inadecuación, en el factor de somatización, puede explicarse por la relación que existe en los adolescentes entre cambios psicológicos y cambios corporales.

Con respecto a la comparación entre la escala original de Kellner y los dos estudios de validación en México, se puede concluir que la subescala de somatización aparece como la más constante y sólida.

En la subescala de ansiedad presenciamos que los reactivos en los tres estudios tienen la siguiente forma:

Kellner, 1972.	Mora, Natera, Andrade-Palos, 1994.	Barrié, 1996.
3. Nervioso	3. Nervioso	3. Nervioso
5. Con temores o miedos	4. Sensación de presión o tensión en la cabeza.	5. Con temores o miedos
9. Inquieto o asustadizo	5. Con temores o miedos	9. Inquieto o asustadizo
16. Temblores o sacudimientos	8. Sentimientos de que no hay esperanza	12. Sentimientos de culpa
19. Sentimientos de tensión...	9. Inquieto o asustadizo	13. Preocupaciones
23. Con pensamientos que...	16. Temblores o sacudimientos	19. Sentimientos de tensión...
26. Ataques de pánico...	24. Pérdida de interés en casi todas las cosas	25. Descontento o deprimido
29. Dific. conciliar sueño...	25. Descontento o deprimido	
	26. Ataques de pánico	

En este estudio, esta subescala fue la segunda válida pues tuvo un valor Eigen mayor a uno, y hay que notar que en la comparación entre las tres subescalas, aparece mucha similitud en la distribución de los reactivos, lo que permite concluir que es un factor diferenciable en la escala, a pesar de las variaciones en los estudios de 1994 y 1996, que se pueden adjudicar a la diferente población abordada.

La presencia del reactivo "descontento o deprimido" en esta subescala, que teóricamente pertenece a síntomas de depresión, en esta investigación, como ocurrió en el estudio de 1994, permite sugerir, como ya lo hicieron las autoras en 1994, la posibilidad de unir ambas subescalas en una sola, como ansiedad-depresión, sin embargo, esto deberá ser clarificado en futuras investigaciones.

Con respecto a la subescala de inadecuación, tercera en importancia en esta investigación,

aunque no representativa pues no tuvo un valor Eigen mayor a uno, los síntomas que detecta corresponden, en una parte, con lo planteado teóricamente por Kellner cuando construyó la escala, quien menciona que “la cuarta subescala, inadecuación, no es un factor singular y consiste principalmente de sentimientos de inferioridad y síntomas cognoscitivos”; en lo que respecta a síntomas cognoscitivos con los reactivos “poca memoria”, “sin capacidad de concentración” aunque contiene dos reactivos que corresponden uno a ansiedad y el otro a depresión como son: “pensamientos que no se pueden quitar de la cabeza” y “pérdida de interés en casi todas las cosas”, que de nuevo son coherentes con el tipo de población adolescente abordada.

Cabe mencionar que aunque fue la subescala que, en el estudio mexicano de 1994, adquirió el lugar principal, es la que menor consistencia teórica tiene pues se creó a partir de la especificación que Kellner le dió, sin tomar en cuenta, como él mismo lo reconoce, los factores de síntomas publicados en aquel entonces, lo que sí hizo con las otras tres subescalas (ansiedad, depresión y somatización), quizá por esa razón, algunos de los reactivos corresponden a síntomas que el DSM-III-R incluye en los criterios para diagnosticar depresión, y que en esta investigación aparecieron en ese factor, el factor depresión, como son: “sentimientos de que no hay esperanza”, “sentimientos de que la gente piensa mal de mi”, “sentimientos de poca valía o fracaso” y “sentimientos de ser inferior a otras personas”; por lo que se puede concluir que esta subescala, aunque no es de las que mejor definen a la escala, pues tiene un valor Eigen menor a uno, es congruente teóricamente con los resultados de este estudio.

Con respecto a los otros dos factores -alteraciones cardíacas y alteraciones del sueño- encontrados en esta investigación, cuyos reactivos corresponden a las áreas de somatización, de ansiedad y de depresión en la escala de Kellner (1972), en este estudio resultaron los síntomas que menos aparecieron en la población adolescente y se debe quizá a que son los síntomas más intensos que contiene la escala y que parece que casi no aquejan a estos jóvenes de institución privada.

Por tanto, se puede concluir, como ya se mencionó en el capítulo de Resultados, que las subescalas que obtuvieron valores Eigen mayores a uno son las de somatización y ansiedad, lo que produce que sean ellas las que mejor definen a la Escala de síntomas (SRT). Esta conclusión se puede sustentar también con el ajuste a dos factores que Mora, Natera y Andrade-Palos llevaron a cabo con la SRT y que también se realizó en este estudio,

obteniendo los siguientes resultados:

1994 : Escuela pública
Subescalas de síntomas físicos y psicológicos del SRT

Síntomas psicológicos*	Cargas factoriales	Síntomas físicos**	Cargas factoriales
18. Poca valía	0.74943	22. Irritabilidad	0.65849
20. Inferioridad	0.73115	27. Debilidad en el cuerpo	0.65352
17. Dific. pensar	0.68300	11. Dolores en pecho	0.65015
19. Tensión	0.67318	6. Poco apetito	0.64887
15. Me consideran menos	0.67202	21. Adormecimiento en el cuerpo	0.64216
12. Sentimientos de culpa	0.64593	1. Mareado	0.63186
8. No hay esperanza	0.62097	2. Cansado	0.59064
13. Preocupaciones	0.58461	3. Nervioso	0.56274
25. Deprimido	0.55160	29. Dificultades para dormir	0.53857
24. Pérdida de interés	0.54436	14. Dolores musculares	0.52286
23. Pensamientos fijos	0.54370	30. Despertar temprano	0.49907
28. Sin concentración	0.51144	4. Presión en la cabeza	0.49577
9. Inquieto	0.49877	16. Temblores	0.36071
5. Temores	0.48983		
3. Nervioso	0.43779***		
10. Poca memoria	0.47913		
26. Ataques de pánico	0.38061		

* alpha = .91

** alpha = .87

*** El reactivo número 3 aparece reportado en ambas subescalas, lo que puede ser un error de imprenta, pues no aparece el reactivo número 7. Fuente: Mora, Natera y Andrade-Palos, 1994, p.19.

1996: Escuela privada
Subescalas de síntomas físicos y psicológicos del SRT

Síntomas psicológicos	Cargas factoriales	Síntomas somáticos	Cargas factoriales
18. Poca valía	.68	1. Mareado	.63
15. Me consideran menos	.63	7. Palpitaciones rápidas	.57
8. No hay esperanza	.61	21. Adormecimiento cuerpo	.53
12. Sentimientos de culpa	.60	27. Debilidad en el cuerpo	.53
20. Inferioridad	.60	2. Cansado	.50
5. Temores	.51	29. Dificultades para dormir	.47
17. Dificultades pensar	.46	16. Temblores	.45
25. Deprimido	.43	11. Dolores en el pecho	.43
23. Pensamientos fijos	.42	4. Presión en la cabeza	.42
13. Preocupaciones	.41	3. Nervioso	.40
19. Tensión	.40	28. Sin concentración	.40
24. Pérdida de interés	.40	22. Irritabilidad	.39
9. Inquieto	.36	14. Dolores musculares	.30
10. Poca memoria	.35		
Coeficiente alpha = .86		Coeficiente alpha = .82	
Coeficiente alpha de Cronbach para la escala total = .93			

Nota: Los reactivos números 6, 26 y 30 no aparecen porque pertenecen a otro factor o sea que no cargan en ninguno de los dos factores arriba mencionados o porque la carga resultó por debajo de .30.

Como se puede observar, cuando la Escala de síntomas se ajustó a dos factores, la

similitud en la distribución de los reactivos en ambas poblaciones es muy alta (a pesar del posible error de imprenta del estudio de 1994, que hace desconocer la verdadera colocación del reactivo 3 y lo que sucedió con el reactivo 7), pues en el factor de síntomas psicológicos sólo dos reactivos (el 26, eliminado en este estudio, y el 28, que aparece en la otra subescala) no aparecieron en la investigación de 1996, y en el factor de síntomas somáticos sólo dos reactivos (30 y 6, que fueron eliminados en este estudio) no aparecieron colocados de la misma forma que en 1994. Asimismo, como afirma la teoría, a mayor número de reactivos en cada factor, aumentó la confiabilidad, así como la de la escala en su totalidad. También, hay que mencionar que en este ajuste a dos factores, resultaron 27 reactivos los que, por su valor factorial, conforman la escala, lo que constituye una similitud con el número de reactivos que obtuvieron las investigadoras en 1994, en la conformación de la escala en cuatro factores.

Estos resultados permiten sugerir que sería recomendable, como ya lo habían propuesto otras investigaciones (Matson, 1992, citado por Mora, Natera y Andrade-Palos, 1994) utilizar la escala, en México, como conformada por dos subescalas, una que mide síntomas psicológicos y otra que mide síntomas somáticos.

Este aspecto de las áreas o dimensiones que conforman la escala es el que, en todos los estudios, se presenta como el más problemático, desde la creación de la escala, pues fue dividida en cuatro factores, como ya se mencionó, a partir de criterios teóricos, uno de los cuales, el del factor inadecuación, bastante cuestionable pues sólo depende de los criterios de Kellner. La versión original de 38 síntomas y la versión corta no fueron sometidas a análisis psicométricos con el fin de probar su estructura factorial, pues el interés en ese momento era ver su poder de discriminación para detectar pacientes de sujetos normales, así como probar los efectos de los tratamientos psiquiátricos en pacientes que recibían medicamentos y observar los cambios en la sintomatología, también en estudios de correlación con otras pruebas, y fue en este sentido donde adquirió validez y confiabilidad. Fue hasta 1980 que Cochrane, en Inglaterra, propuso valorar sus propiedades psicométricas y sometió la escala a análisis factorial con una rotación oblicua, en un formato Likert (1932), con tres opciones de respuesta, para ver si emergían *a priori* los factores que constituyen a la escala, en población comunitaria; los resultados le hicieron sostener la unidimensionalidad de la escala, señalando que los síntomas psicológicos y somáticos están altamente correlacionados y, en consecuencia, la distribución de los reactivos en las cuatro subescalas son imposibles de discernir, asimismo, propuso que la escala puede ser utilizada en encuestas comunitarias para

evaluar psicopatología indiferenciada. Los estudios en México, el de Mora, Natera y Andrade-Palos en 1994, como la presente investigación permiten pensar que esta escala es altamente confiable como para considerar que sólo mide sintomatología de manera tan ambigua, como ya lo dijeron las autoras del estudio de 1994. Aún así es importante destacar que la actual investigación no confirma las cuatro escalas obtenidas por la investigación de 1994 y manifiesta una alta relación entre las subescalas. Estas diferencias, parecen deberse, en todos los casos a divergencias socioculturales de las diversas poblaciones abordadas, así como a la dificultad inherente a gran parte de los padecimientos psicológicos, los cuales no se presentan de manera aislada, sino que tienden a asociarse unos con otros; este fenómeno es conocido en el campo de la psiquiatría como comorbilidad, y en el caso específico de la depresión, este padecimiento comúnmente se asocia con trastornos por somatización, ansiedad, desórdenes en la alimentación y alcoholismo. (Berlanga, 1993). Estas hipótesis deberán ser probadas en posteriores investigaciones con el fin de obtener datos comparativos, con las poblaciones que se aborden, para establecer la conformación definitiva de las subescalas que conforman esta escala.

Respecto de los índices de psicopatología obtenidos en los puntajes de las seis subescalas de este estudio se observa que la población de adolescentes de escuela privada permanece en una media total de 1.5 (ver cuadros 22 y 27), o sea en la respuesta de “algunas veces”, que corresponde a psicopatología moderada, por lo que se puede concluir que la SRT es un instrumento que detecta malestar psicológico en grandes poblaciones, como ya lo habían afirmado Cochrane (1980) y las autoras de la investigación mexicana en 1994. Para comprobar que diferencia población con problemas psicológicos de población normal habría que aplicar la Escala a población de hospitales psiquiátricos; aún así, si comparamos las medidas de tendencia central, halladas en esta investigación, con las obtenidas por Reid en 1960 en su estudio de comparación de la EPI y la MAS con la SRT en dos clínicas (en Houston y Albuquerque) con 100 pacientes en relación a cien normales, podemos notar que las medias en cada factor para los normales se mantuvieron en psicopatología moderada (ansiedad = 1.5, depresión = 1.5, inadecuación = 2.26, somatización = 1.6), lo que no sucedió con los pacientes cuyas medias mostraban patología severa (ansiedad = 11.68, depresión = 11.62, inadecuación 13.44, somatización = 2.88, para el grupo de Houston) (ver cuadro 5). Asimismo el estudio de Cochrane, que si comparó población con y sin problemas psiquiátricos obtuvo medias muy diferentes: comunidad = 8.03 (N 660), pacientes = 24.63

(N 40) (ver cuadro 7), por lo que se puede derivar que la población adolescente estudiada está dentro de la normalidad, de acuerdo con esos estudios.

Finalmente, antes de llegar a conclusiones definitivas, es importante considerar en futuras investigaciones ambos tipos de población, con y sin problemas psiquiátricos.

De cualquier manera, los datos arrojados por esta investigación proporcionan evidencias de que la Escala de síntomas es confiable, que es un instrumento útil, accesible y práctico para su aplicación en México, ya sea por medio de una entrevista personal o bien puede ser autoaplicable para su utilización en encuestas comunitarias. Los reactivos están planteados en un lenguaje cotidiano, lo cual facilita su aplicación incluso en poblaciones marginales de escasos recursos verbales. Asimismo, este estudio revalida y confirma la confiabilidad de la escala para su aplicación en encuestas comunitarias, en particular para población juvenil.

La ventaja que ofrece esta Escala es que puede ser usada en investigaciones cuyo objetivo sea preventivo, pues así como muestra Berlanga (1993), muchos de los síntomas que detecta se encuentran asociados con consumo de drogas, problemas alimenticios, delitos, etc. por lo que puede ser utilizada, una vez detectadas las poblaciones con psicopatología, para instrumentar acciones que prevengan aquellos efectos concomitantes a dichos síntomas. Estos estudios a su vez permitirían seguir validando la Escala.

Referencias

Referencias

- ABERASTURY, A., Knobel, M., La adolescencia normal, Paidós, México, 1992.
- AMERICAN Psychiatric Association, DSM-III-R, Manual diagnóstico y estadístico de los trastornos mentales, Masson, Barcelona, 1988.
- ANASTASI, A., Tests psicológicos, Aguilar, México, 1967.
- ANASTASI, A., Psicología diferencial, Aguilar, Madrid, 1977.
- BEDFORD, A., McIver, D. y Pearson, P.R., "A further test of Foulds' personality and personal illness differentiae in a psychiatric group", Psychological Medicine, núm 8, 1978, pp. 467-470.
- BERLANGA, C. "La comorbilidad en los trastornos depresivos", Salud Mental, vol. 1, núm 16, 1993, p. 1-9.
- BLOSS, P.C., Psicología de la adolescencia, Joaquín Mortiz, Buenos Aires, 1971.
- BROWN, F.G., Principios de la medición en psicología y educación, El Manual Moderno, México, 1980.
- COCHRANE, R., "A comparative evaluation of the Symptom Rating Test and the Langner 22-item Index for use in epidemiological surveys", Psychological Medicine, núm. 10, 1980, pp. 115-124.
- COSTA Neiva K.M., Manual de pruebas de inteligencia y aptitudes, Plaza y Valdes Editores, México, 1996.
- EY, H., Tratado de psiquiatría, Masson, México, 1987.
- FAVA, G.A., Kellner, R., Perini, G.I., Fava, M., Michelacci, L., Munari, F., Evangelisti, L.P., Grandi, S., Bernardi, M., Mastrogiamomo, I. "Italian validation of the Symptom Rating Test (SRT) and Symptom Questionnaire (SQ)", Can. J. Psychiatry, núm. 28, 1983, pp. 117-123.
- HERNANDEZ, S.R., Fernández, C. C. y Baptista, L.P., Metodología de la investigación, McGraw - Hill, México, 1993.
- KELLNER, R., and Sheffield, B.F., "A self-rating scale of distress", Psychological Medicine, núm.3, 1973, pp. 88-100.
- KERLINGER, F.N., Investigación del comportamiento, Nueva Editorial Interamericana, México, 1975.

- MAGAÑA, M., "Estudio comparativo de la presencia y severidad de los síntomas depresivos en adolescentes hijos de padres divorciados e hijos de padres no divorciados a nivel de educación media superior", Tesis de licenciatura, Universidad del Valle de México, México, 1990.
- MILLON, T., Psicopatología y personalidad, Interamericana, México, 1974.
- MORA, J., Natera, G., Andrade Palos, P., "Escala de síntomas Symptom Rating Test (SRT) en estudiantes mexicanos de nivel medio superior", Salud Mental, vol. 17, núm. 3, 1994, pp. 15-23.
- MORALES, M. L. Psicometría aplicada, tercera reimpresión, Editorial Trillas, México, 1993.
- NADELSTICHER, A., Técnicas para la construcción de cuestionarios de actitudes y opción múltiple, Instituto Nacional de Ciencias Penales, México, 1983.
- NIE, H.N., Hull, C.H., Steinbrenner y Jenkins, Statistical package for social sciences, versión 4, Mc. Graw Hill, Estados Unidos, 1985.
- NUNNALLY, C.J. en Bernstein, H.I., Teoría psicométrica, McGraw-Hill, México, 1995.
- PEREZ y Farias, J.M., Un estudio de confiabilidad con el MMPI-2 en un grupo de estudiantes universitarios, Tesis de maestría, Universidad Nacional Autónoma de México, México, 1995.
- PICHOT, P., Los tests mentales, Paidós, México, 1960.
- PICK, S. y López, A.L., Cómo investigar en ciencias sociales, Trillas, México, 1979.
- REAL Academia Española, Diccionario de la lengua española, Espasa Calpe, Madrid, 1970.
- REIDL, L., Medición en ciencias sociales, Inédito, UNAM, México, 1990.
- RODRIGUES, A., Psicología social, Trillas, México, 1981.
- SARASON, I.G., Psicología anormal, Trillas, México, 1981.
- SZEKELY, B., Los tests, Kapeluz, Buenos Aires, 1970.
- TYLER, L., Psicología de las diferencias humanas, Biblioteca Marova de Estudios del Hombre, Madrid, 1973.

Anexo A

1
CUESTIONARIO

INSTRUCCIONES.

Este cuestionario tiene como finalidad recolectar información para la elaboración de una investigación. Por tal motivo, los datos que proporciones serán una contribución muy importante para este trabajo. Por favor, contesta todas las preguntas con el mayor cuidado y sinceridad posibles. El cuestionario es ANONIMO y CONFIDENCIAL. Gracias por tu cooperación.

Pon una palomita en la respuesta con la que te identifiques, o contesta la pregunta que se te hace.

A continuación te mostramos un ejemplo de respuesta:

Pregunta: Durante el invierno la Cd. de México tiene un clima frío.

Respuesta: Nunca ----- Algunas veces ----- A menudo ----- ()

DATOS PERSONALES.

1. Tú eres:

Hombre ----- Mujer ----- ()

2. ¿Cuántos años tienes? ----- ()

3. ¿Qué grado escolar cursas?

1er. sem. ---- 2o. sem. ---- 3er. sem. ---- 4o. sem. ---- 5o. sem. ----- 6o. sem.--- ()

4. ¿Cuál es tu ocupación?

Solamente estudias ----- Estudias y trabajas ----- ¿ En qué?----- ()

5. ¿Cuántas personas forman tu familia directa? (con la que vives) ----- ()

6. ¿En tu familia eres?

Hijo único ----- Primer hijo ---- Segundo hijo ----- Tercer hijo ----- Cuarto hijo -----

Quinto hijo ----- Sexto o más ¿cual? ----- Ultimo hijo ----- ()

7. ¿Cuál es tu promedio escolar? ----- ()

8. ¿Tienes beca?

Sí ----- No ----- ()

INSTRUCCIONES.

En esta parte deseamos saber como te has sentido en los **últimos tres meses.**

Pon una palomita en la respuesta que mejor expresa cómo te has sentido.

Por favor, recuerda que esta información será tratada con la más estricta confidencialidad.

1. Mareado o débil.

Nunca ----- Algunas veces ----- A menudo ----- ()

2. Cansado o sin energía.

Nunca ----- Algunas veces ----- A menudo ----- ()

3. Nervioso.

Nunca ----- Algunas veces ----- A menudo ----- ()

4. Sensación de presión o tensión en la cabeza.

Nunca ----- Algunas veces ----- A menudo ----- ()

5. Con temores o miedos.

Nunca ----- Algunas veces ----- A menudo ----- ()

6. Con poco apetito.

Nunca ----- Algunas veces ----- A menudo ----- ()

7. Palpitaciones rápidas o fuertes sin razón aparente.

Nunca ----- Algunas veces ----- A menudo ----- ()

8. Sentimientos de que no hay esperanza.

Nunca ----- Algunas veces ----- A menudo ----- ()

9. Inquieto o asustadizo.

Nunca ----- Algunas veces ----- A menudo ----- ()

10. Con poca memoria.

Nunca ----- Algunas veces ----- A menudo ----- ()

11. Dolores en el pecho o dificultad al respirar o sensación de no tener suficiente aire.

Nunca ----- Algunas veces ----- A menudo ----- ()

12. Sentimientos de culpa

Nunca ----- Algunas veces ----- A menudo ----- ()

13. Preocupaciones.

Nunca ----- Algunas veces ----- A menudo ----- ()

14. Dolores musculares o reumatismo.

Nunca ----- Algunas veces ----- A menudo ----- ()

15. Sentimientos de que la gente me considera menos o piensa mal de mi.

Nunca ----- Algunas veces ----- A menudo ----- ()

16. Temblores o sacudimientos.

Nunca ----- Algunas veces ----- A menudo ----- ()

17. Dificultad para pensar claramente o para tomar decisiones.

Nunca ----- Algunas veces ----- A menudo ----- ()

18. Sentimientos de poca valía o fracaso.

Nunca ----- Algunas veces ----- A menudo ----- ()

19. Sentimientos de tensión o como si me hubiesen dado cuerda (como un reloj).

Nunca ----- Algunas veces ----- A menudo ----- ()

20. Sentimientos de ser inferior a otras personas.

Nunca ----- Algunas veces ----- A menudo ----- ()

21. Sensaciones de cosquilleo o adormecimiento en partes de mi cuerpo.

Nunca ----- Algunas veces ----- A menudo ----- ()

22. Irritabilidad.

Nunca ----- Algunas veces ----- A menudo ----- ()

23. Con pensamientos que no se pueden quitar de la cabeza.

Nunca ----- Algunas veces ----- A menudo ----- ()

24. Pérdida de interés en casi todas las cosas.

Nunca ----- Algunas veces ----- A menudo ----- ()

25. Descontento o deprimido.

Nunca ----- Algunas veces ----- A menudo ----- ()

26. Ataques de pánico.

Nunca ----- Algunas veces ----- A menudo ----- ()

27. Sensaciones de debilidad en partes del cuerpo.

Nunca ----- Algunas veces ----- A menudo ----- ()

28. Sin capacidad de concentración.

Nunca ----- Algunas veces ----- A menudo ----- ()

29. Dificultad para conciliar el sueño, o sueño desasosegado, o pesadillas.

Nunca ----- Algunas veces ----- A menudo ----- ()

30. Me despierto demasiado temprano y no puedo conciliar el sueño otra vez.

Nunca ----- Algunas veces ----- A menudo ----- ()

Muchas gracias.