

2. La Medición en Psicología

La psicometría tiene como finalidad llevar a cabo la medición de la conducta en los seres humanos, y esto constituye uno de los pilares fundamentales de la ciencia psicológica, ya que el hombre se interesa cada día más por comprender su propia naturaleza y la de los demás (Morales, 1991).

Para comprender los procedimientos de medición en la Psicología, es importante entender el término "La medición consiste en reglas para asignar números a los objetos con el propósito de representar cantidades de atributos" (Nunally, 1991, p. 15), en donde el término reglas implica que el procedimiento para asignar números debe formularse clara y explícitamente.

Dentro del desarrollo de la Psicología como ciencia, se han realizado estudios a fin de poder determinar si los seres humanos difieren entre sí y en que grado se presentan estas diferencias, y esto a su vez ha originado el desarrollo de instrumentos para evaluar de forma cuantitativa las diferencias y las semejanzas existentes.

"Un instrumento de medición psicológica es aquella técnica metodológica producida artificialmente, que obedece a reglas explícitas y coloca a el individuo en condiciones experimentales con el fin de extraer el segmento del comportamiento a estudiar y que permita la comparación estadística conductora o la clasificación cualitativa, tipológica y cuantitativa de la característica o características que se están evaluando" (Morales, 1991 p. 15). Por lo tanto cuando se habla de los instrumentos de medición en Psicología se refiere a todos aquellos procedimientos u operaciones que permitan obtener objetivamente y con mayor certeza, información acerca de la expresión de los fenómenos que se suceden en una unidad biológica, social y psicológica, que es la conducta humana. Estos instrumentos contruidos para medir aspectos específicos de la conducta humana se

conocen con el nombre de pruebas psicológicas. (Morales, 1991 y Nunally, 1991).

En la actualidad existe una gran diversidad de pruebas psicológicas que se emplean en diferentes áreas. Para el uso de las mismas se requiere tomar en consideración los aspectos de confiabilidad y validez los cuales hacen alusión a cuestiones tales como si el instrumento realmente mide lo que queremos medir y a la consistencia de las puntuaciones obtenidas por los individuos.

La necesidad de obtener datos confiables y válidos, hace necesario contar con instrumentos adecuados a la población bajo estudio. Las pruebas ya elaboradas y estandarizadas tienen su propio procedimiento de aplicación, codificación e interpretación de los datos obtenidos.

El problema en el uso de dichas pruebas es que la mayoría han sido desarrolladas en contextos muy diferentes al latinoamericano y en ocasiones su uso puede conducir a la obtención de datos poco válidos y poco confiables. Por lo tanto, es recomendable seleccionar pruebas estandarizadas o adaptadas al contexto de estudio que hayan probado ser válidas y confiables. De aquí la importancia de tomar en consideración diversos criterios que se han de tener presentes para juzgar y determinar si el instrumento que se va a utilizar es el apropiado.

2.1. Criterios para la selección de una prueba

Frecuentemente se seleccionan instrumentos inapropiados ya sea para la investigación o solución de un problema, por lo que una de las principales consideraciones de quien utiliza los instrumentos psicológicos es que debe llevar en mente la elección de aquel o aquellos que reúnan los requisitos para el propósito que se persigue; es decir, si la función de una prueba psicológica es proporcionar la mayor y mejor información acerca de su problema, el

primer paso a seguir será la especificación cuidadosa de las interrogantes que deberán ser contestadas y el tipo de personas que deberán ser examinadas. También se debe considerar el hecho de contar con otro tipo de medición o instrumento que sean alternativos y que puedan ser utilizados en caso necesario para establecer comparaciones que permitan rectificar o ratificar los resultados obtenidos, a fin de asegurar que las decisiones alcanzadas son válidas.

Otra consideración importante consiste en determinar si la capacidad o potencialidad de la prueba es la adecuada para medir con el rigor necesario aquella característica que se está investigando.

La longitud de la prueba, en donde desempeña un papel relevante el tiempo, constituye otro de los criterios que deben tenerse en cuenta, tanto para la selección como en la construcción de una prueba, ya que esto influye en aspectos de esfuerzo y económicos.

La facilidad de administración y calificación de un instrumento, la disponibilidad de formas paralelas y la forma de adaptación del instrumento, son aspectos que también deberán considerarse para su elección.

Para poder hacer una buena elección del instrumento, el investigador debe poseer conocimientos sobre el manejo de las pruebas, a fin de establecer si existen los requisitos mínimos para considerar que se ha seleccionado el mejor instrumento. Estos conocimientos se refieren a aspectos tales como el grado de confiabilidad y validez, objetividad y sensibilidad del instrumento, así como al hecho de si ha sido estandarizado o cuando menos adaptado y en qué grupo o grupos de sujetos se han realizado tales estudios.

En el caso en que se elija una prueba diseñada en otro contexto, es necesario adaptarla a través de una muestra similar a aquella con la que se empleará, así como ajustarla a las condiciones de la investigación. El

instrumento o prueba debe demostrar ser válido y confiable para el contexto en el cual se va aplicar.

A continuación se describirán los conceptos de confiabilidad y validez así como los métodos para obtener la confiabilidad y los métodos para probar validez que se emplean en la medición en Psicología.

2.2. Confiabilidad

El error de medición se produce en dos formas: como un sesgo sistemático (error constante) o como un error aleatorio (error variable).

Existen factores que producen variación en las puntuaciones de un reactivo a otro dentro de una prueba y estas variaciones tienden a disminuir la confiabilidad; uno de ellos es la adivinación, que es la suposición del sujeto acerca de la respuesta que considera sea la correcta, también es posible que el sujeto presente trastornos nerviosos a mitad del desarrollo de la prueba, lo cual haría que su puntuación bajara en los subsecuentes reactivos. Es posible también que el sujeto se equivoque al responder ante un reactivo, debido a algún error, por ejemplo, que en lugar de contestar A, el sujeto contesta B. Alguien más puede saltarse inadvertidamente un reactivo que hubiera contestado en forma correcta. Si a la mitad de la prueba una persona se da cuenta de que malinterpretó las instrucciones para responder a la prueba, y no le queda tiempo para regresarse a los primeros reactivos, es seguro que tratará de hacerlo mejor en los siguientes (Nunally, 1991), estos factores se encuentran dentro del error aleatorio de medición, el cual varía de sujeto a sujeto.

Los errores aleatorios de medición nunca se pueden eliminar por completo, pero se hacen esfuerzos para reducirlos al mínimo. Se dice que una medición es confiable cuando el error de medición debido al azar es pequeño (Nunally, 1991). Es decir, la confiabilidad de una prueba indica hasta

qué punto puede atribuirse a errores de medida las diferencias individuales de la prueba y hasta qué punto cabe atribuir las diferencias verdaderas de las características que se están sometiendo a consideración. Por lo tanto se puede decir, que toda medida de confiabilidad de una prueba denota qué proporción de la varianza total de las puntuaciones es varianza de error, donde todas las fuentes de error muestran la tendencia a disminuir las correlaciones promedio que se dan ante los reactivos de la prueba.

Cualquier condición que no sea afín al propósito de la prueba, representa una varianza de error. Así, cuando el examinador trata de mantener uniforme las condiciones de aplicación de la prueba, controlando ambiente, instrucciones, límite de tiempo, rapport y otros factores similares, está reduciendo la varianza de error y haciendo más confiables las puntuaciones (Nunally, 1991). En conclusión las fuentes de error afectan a la confiabilidad.

La confiabilidad se define como la consistencia de las puntuaciones obtenidas en los mismos individuos cuando son examinados por la misma prueba en diferentes ocasiones, con conjuntos distintos de elementos equivalentes o bajo otras condiciones variables de examen (Anastasi, 1968).

El coeficiente de confiabilidad es un índice de la eficacia del instrumento; la confiabilidad es necesaria, sin embargo, no es condición suficiente para ningún tipo de validez (Woodworth, 1983).

Puesto que todos los tipos de confiabilidad se refieren al grado de consistencia interna o concordancia entre dos conjuntos de puntuaciones derivadas independientemente, pueden expresarse todos los conjuntos en función de un coeficiente de correlación. Los coeficientes que se encuentran en la práctica figuran entre dos extremos y tienen un valor superior a cero, pero inferior a uno (Nunally, 1991).

El primer tipo de confiabilidad es la denominada consistencia interna.

En este caso pueden utilizarse la prueba de mitades, la prueba de Kuder-Richarson y el alpha de Cronbach.

El segundo tipo es la consistencia temporal donde se puede utilizar la prueba de test-retest y el de muestras paralelas o de forma equivalente, en donde la correlación de las puntuaciones es dada en base a dos o más evaluaciones de la misma persona ya sea con el mismo instrumento o con una forma equivalente.

A continuación se describirán cada uno de los procedimientos de confiabilidad empleados en la determinación de la consistencia interna y temporal del instrumento.

2.2.1. Consistencia Interna

La consistencia interna es un requisito indispensable de los instrumentos. Todo cálculo de confiabilidad basado en la correlación promedio entre reactivos concierne a la consistencia interna de la prueba.

Al calcular la confiabilidad con base en la consistencia interna, se consideran fuentes de error que, estrictamente hablando, no dependen del muestreo de reactivos per se, sino del muestreo de factores situacionales que acompañan a los reactivos (Nunally, 1991).

2.2.1.1. División en mitades

Partiendo de una sola aplicación de un formato de prueba es importante llegar a la medida de confiabilidad siguiendo varios procedimientos de división en mitades. Se obtienen dos puntuaciones para cada individuo dividiendo la prueba en mitades comparables. Es evidente que la confiabilidad de la subdivisión nos proporciona una medida de equivalencia o adecuación de las muestras de elementos. En esta medida no entra la estabilidad temporal de las puntuaciones puesto que sólo requiere una sesión (Anastasi,

1968).

Para encontrar la confiabilidad por el método de las mitades, el primer problema reside en cómo dividir la prueba a fin de obtener las mitades equivalentes.

Un procedimiento adecuado es el de encontrar las puntuaciones de los elementos pares e impares de la prueba. Si los elementos estaban originalmente dispuestos en orden de dificultad aproximado, esta división proporciona puntuaciones equivalentes de las mitades.

Una vez obtenidas las puntuaciones por mitades para cada sujeto se puede correlacionar siguiendo los procedimientos para el análisis de correlación.

Cualquier diferencia entre las puntuaciones de una persona en las dos mitades de la prueba representa el error casual. La varianza de estas diferencias dividida por la varianza de las puntuaciones totales, es la proporción de la varianza error en las puntuaciones. Cuando esta varianza se resta al valor uno, da la proporción de la verdadera varianza, que es igual al coeficiente de confiabilidad. (Anastasi, 1968).

2.2.1.2. Kuder-Richardson

Es un procedimiento para determinar la confiabilidad de una prueba, a través de una sola aplicación con una sola versión del instrumento. Se basa en la consistencia de las respuesta de los sujetos a todos los elementos de la prueba. Esta consistencia entre los elementos se encuentra influenciada por dos fuentes de error: 1) El muestreo del contenido (como en la forma equivalente y en la confiabilidad de la división en mitades) y, 2) la heterogeneidad del área examinada. Cuanto más homogénea sea la prueba, más alta será la consistencia entre los elementos.

Cronbach, (citado por Anastasi, 1968) argumenta que se puede

demostrar matemáticamente que el coeficiente de confiabilidad de Kuder-Richardson, es en realidad la media de los coeficientes de división en mitades que resulta de las diferentes divisiones de una prueba.

2.2.1.3. Alpha de Cronbach

El coeficiente alpha es la fórmula básica para determinar la confiabilidad basada en la consistencia interna.

En la mayoría de las situaciones, con el coeficiente alpha se puede lograr un buen cálculo de la confiabilidad, pues la principal fuente de error de medición se debe al muestreo de contenido.

El coeficiente alpha establece el límite superior de la confiabilidad de las pruebas construidas según el modelo de dominio-muestra. Si el coeficiente es muy bajo, entonces la prueba es demasiado corta o los reactivos tienen muy poco en común.

2.2.2. Consistencia temporal

Este tipo de confiabilidad consiste en la correlación entre puntuaciones de dos o más aplicaciones de las pruebas, los procedimientos para evaluar este tipo de consistencia son los siguientes.

2.2.2.1. Test-retest

El método utilizado para probar la confiabilidad test-retest consiste en repetir la prueba en una segunda ocasión.

El coeficiente de confiabilidad, en este caso es simplemente la correlación entre las puntuaciones obtenidas con los mismos sujetos en dos aplicaciones de la prueba.

La varianza de error corresponde a las fluctuaciones en la puntuación debidas al azar de una sesión de prueba a otra.

La confiabilidad indica hasta que punto las puntuaciones de una prueba pueden generalizarse en distintas ocasiones; cuando más alta sea la confiabilidad, tanto menos susceptibles son las puntuaciones a los cambios diarios casuales que se producen en la condición del sujeto o en el ambiente en el que se aplica la prueba.

El tiempo que media entre una prueba y su repetición es muy importante en la determinación del valor del coeficiente de confiabilidad. En general, cuanto mayor sea dicho tiempo, menor es el coeficiente de confiabilidad, ya que el concepto de confiabilidad se limita generalmente a cambios casuales, de corto alcance, que caracterizan la ejecución de la prueba, más que al dominio total de la conducta que está siendo examinada (Anastasi, 1968).

2.2.2.2. Muestras paralelas

La confiabilidad de muestras paralelas también llamada confiabilidad de la forma equivalente, consiste en que puede darse a los sujetos una forma de prueba en la primera ocasión y otra forma paralela en la segunda. La correlación entre las puntuaciones obtenidas en las dos formas representa el coeficiente de confiabilidad de la prueba. Se advertirá que éste mide tanto la estabilidad temporal como la consistencia de las respuestas a las diferentes muestras de elementos.

En el desarrollo de formas equivalentes hay que asegurarse de que verdaderamente son paralelas. En realidad las formas paralelas de una prueba deberían ser pruebas construidas independientemente y destinadas a satisfacer las mismas especificaciones. Las pruebas han de contener el mismo número de elementos, y éstos tienen que expresarse de la misma forma y abarcar el mismo tipo de contenido. Así mismo, la amplitud total y el nivel de dificultad de los elementos deben ser iguales. Hay que controlar para

su comparabilidad las instrucciones, los límites de tiempo, los ejemplos ilustrativos, el formato y todos los demás aspectos de la prueba.

Una dificultad que presenta el empleo de este tipo de confiabilidad es que para muchas pruebas no se dispone de formas equivalentes a causa de las dificultades prácticas que existen en su elaboración (Anastasi, 1968).

2.3. Validez

La validez de la prueba se refiere a lo que ésta mide y a cómo lo mide (Anastasi, 1968).

En un sentido general, un instrumento de medición es válido si cumple satisfactoriamente el propósito con el que se diseñó. Solamente es posible definir el rasgo que mide una determinada prueba mediante el examen de los criterios específicos y otras fuentes objetivas de información utilizada para establecer su validez (Anastasi, citada por Anastasi, 1968). Además no puede decirse de ninguna manera que el instrumento tiene una validez alta o baja en abstracto. Su validez debe determinarse respecto al uso particular para la que se considera. Por lo tanto, se puede decir que el instrumento de medición no es válido per se, sino de acuerdo al uso que se le dará.

Existen tres tipos de validez: la validez predictiva, la validez de contenido y la validez de constructo (Bradley, McClelland, 1972).

2.3.1. Validez Predictiva

La validez predictiva se considera cuando se utiliza un instrumento para estimar alguna forma importante de conducta; a esto se le denomina criterio. (Nunally, 1991).

Este tipo de validez indica la eficacia de una prueba en la predicción de la conducta del individuo en situaciones específicas. Para ello se compara la actuación de la prueba con el criterio, es decir, una medida directa e

independiente de lo que está destinado a medir la prueba.

El empleo del término predicción se refiere en un sentido muy general a las relaciones entre un instrumento y los eventos que ocurrieron antes, durante y después de la aplicación.

La validez predictiva está determinada por el grado de correspondencia entre las dos medidas implicadas, la prueba y el criterio, la cual se calcula a través del análisis correlacional.

Este tipo de validez es de interés sólo en ciertos tipos de problemas aplicados en Psicología y Educación, por ejemplo, al usar pruebas para seleccionar oficinistas, empleados públicos, así como cuando se toman decisiones sobre la hospitalización y el tratamiento de enfermos mentales.

2.3.2. Validez de Contenido

Con ciertos instrumentos, la validez depende principalmente de la correcta adecuación con la que se haya muestreado un dominio específico de contenido.

Existen ocasiones en que la prueba no puede validarse en términos de su capacidad predictiva, porque su propósito no es predecir algo, sino medir directamente el rendimiento en una unidad de instrucción. Por lo tanto, la prueba debe considerarse como una medida adecuada de lo que se pretende medir. La validez no puede determinarse al correlacionar la prueba con un criterio, ya que la prueba misma es el criterio (Nunally, 1991).

Antes de probar la validez de las medidas es necesario iniciar adecuadamente el proceso para garantizar que la prueba cuente con un conjunto de reactivos representativos.

Para asegurar que los reactivos realmente representan la unidad de instrucción es necesario tener un perfil detallado, o anteproyecto de los tipos de preguntas y problemas que se incluirán. En tales casos, es necesario

Debido a que los constructos implican dominios de observaciones, lógicamente una buena medida de cualquier constructo debería obtenerse al combinar los resultados de un cierto número de tales observaciones y no al considerar individualmente cualquiera de ellos. Por lo tanto, podría pensarse que las puntuaciones combinadas de una cierta cantidad de medidas de observaciones propias del dominio tienen un grado de validez de construcción para el mismo, considerado éste como un todo.

Dentro de la medición de constructos que constituyen una parte vital de la actividad científica existen tres aspectos principales que son el procedimiento para elaborar o validar tales medidas:

- 1) Especificación del dominio de observaciones.
- 2) La determinación del grado en que todas o algunas de estas variables se correlacionan entre sí o son igualmente afectadas por los tratamientos experimentales.
- 3) La determinación de cuando alguna o todas las medidas de tales variables actúan como si midiesen el constructo.

El producto final de este complejo proceso es un constructo con las características siguientes:

- a) Está bien definido en términos de una variedad de observaciones, para las cuales hay una o más variables que representan adecuadamente su dominio.
- b) Eventualmente se correlaciona con otros constructos de interés.

Según Hernández, Fernández y Baptista (1991), la validez de constructo es la más importante y, se refiere al grado en que una medición se relaciona consistentemente con otras mediciones de acuerdo con hipótesis derivadas teóricamente y que conciernen a los conceptos (o constructos) que están siendo medidos.

La validación de constructo incluye tres etapas:

- 1) Se establece y especifica la relación teórica entre los conceptos.

2) Se correlacionan ambos conceptos y se analiza cuidadosamente la correlación.

3) Se interpreta la evidencia empírica de acuerdo a qué tanto clarifica la validez de constructo de una medición en particular.

No es posible llevar a cabo la validación de constructo, a menos que exista un marco teórico que soporte a la variable en relación con otras variables (Hernández et al. 1991).

Entre más elaborado y comprobado se encuentre el marco teórico que apoya la hipótesis, la validación de constructo puede arrojar mayor luz sobre la validez de un instrumento de medición. Se puede tener una mayor confianza en la validez de constructo de una medición, cuando sus resultados se correlacionan significativamente con un mayor número de mediciones de variables que teóricamente y de acuerdo con estudios antecedentes están relacionadas (Hernández et al. 1991).

Así pues, es la validez de constructo el tipo de validez que debe buscarse cuando la pretensión del investigador es contar con evidencia empírica que apoye el modelo teórico que pretenda poner a prueba para explicar un fenómeno dado.

2.4. Proceso de estandarización de un instrumento

De acuerdo con I. Reyes (octubre, 1996), si el instrumento elegido se construyó con una población diferente a la que se desea evaluar, es importante llevar a cabo el proceso de adaptación y estandarización.

Dentro del proceso de adaptación y estandarización de un instrumento se encuentran aspectos tales como la traducción literal del instrumento, el ajuste del mismo para lograr la equivalencia de lenguaje y equivalencia cultural, la validación externa por jueces, el piloteo, la selección y aplicación del instrumento a una muestra, el análisis de confiabilidad y validez del

instrumento.

En el caso de que el instrumento se presente en otro idioma, la traducción debe de realizarse literalmente, esto con el propósito de presentar todos y cada uno de los elementos de la prueba sin que se vea afectado su sentido o significado. Posteriormente debe de realizarse la equivalencia de lenguaje que tiene como finalidad la búsqueda de modismos propios de la cultura a la cual se va adaptar, permitiendo así la comprensión de los elementos.

Así mismo debe buscarse la equivalencia cultural, ésta implica el realizar cambios en algunos aspectos de la prueba, dadas las diferencias culturales entre la muestra en la que se estandarizó originalmente el instrumento y en la que se va emplear. Los cambios pueden realizarse abarcando desde un pequeño aspecto como podría ser alguna fecha, un dicho o refrán, hasta el realizar cambios radicales en el contenido del reactivo por no ser culturalmente aceptable.

El siguiente aspecto a considerar es la validación externa de la prueba, la cual se puede llevar a cabo por jueces en la materia para juzgar si los reactivos corresponden a las categorías originales. Donde los jueces agrupan los reactivos de acuerdo con su contenido, en categorías o dimensiones, las cuales posteriormente son comparadas con las categorías formuladas para el instrumento original.

El siguiente paso es el piloteo y consiste en la aplicación de la prueba a un número pequeño de personas con el objeto de observar si el formato y el lenguaje empleado son adecuados para la muestra que se va a emplear, si las instrucciones y los ejemplos ilustrativos son claros y entendibles, así mismo es posible estimar el tiempo requerido por las personas para contestar la prueba.

2.4.1. Selección de la muestra

Una vez realizada la traducción del instrumento, el ajuste para lograr la equivalencia de lenguaje y cultural, la validación por jueces y el piloteo, se puede decir que se ha completado el proceso de adaptación de un instrumento a la población o muestra en la que se va a emplear. El siguiente paso dentro del proceso de estandarización es la aplicación del instrumento a una muestra. La selección de la muestra depende de los objetivos del estudio, del esquema de investigación y de la contribución que se piensa hacer con dicho estudio.

Básicamente a las muestras se les categoriza en dos grandes ramas: las muestras no probabilísticas y las muestras probabilísticas. En éstas últimas todos los elementos de la población tienen la misma posibilidad de ser escogidos. En cambio, en las muestras no probabilísticas la elección de los elementos no depende de la probabilidad, sino de que estos cumplan con una serie de características de acuerdo con los objetivos que persiga el investigador y de la población a la que está dirigida dicha investigación (Fernández et al. 1991).

Las muestras probabilísticas tiene muchas ventajas, quizá la principal es que puede medirse el tamaño del error de las predicciones. Puede decirse que el principal objetivo en el diseño de una muestra probabilística es el de reducir al mínimo este error al que se le llama error estándar. Estas muestras pueden seleccionarse de varias formas: 1) por estratos, que consiste en dividir a la población en subpoblaciones o estratos y se selecciona una muestra para cada estrato, con la estratificación aumenta la precisión de la muestra; 2) por racimos el cual permite reducir costos, tiempo y energía al considerar que en ocasiones las unidades de análisis se encuentran encapsuladas o encerradas en determinados lugares físicos o geográficos a los que se les denomina racimos, este tipo de muestreo supone dos etapas:

se seleccionan los racimos y dentro de éstos se seleccionan a los sujetos u objetos que van a ser medidos (Fernández, et al. 1991).

Dentro de las muestras no probabilísticas, las cuales también se llaman muestras dirigidas suponen un procedimiento de selección informal. La desventaja de este muestreo es que los sujetos de la muestra no tienen la misma probabilidad de ser escogidos y la ventaja de este tipo de muestra es su gran utilidad para estudios que no requieren tanto de una representatividad de elementos de una población sino de una cuidadosa y controlada elección de sujetos con ciertas características especificadas previamente en el planteamiento del problema. Hay varias clases de muestras no probabilísticas; 1) la muestra de sujetos voluntarios o muestra fortuita, donde el investigador elabora conclusiones sobre personas que llegan a sus manos de manera casual. 2) muestra de expertos, frecuentemente utilizadas en estudios cualitativos y exploratorios que permiten generar hipótesis más precisas o para generar materia prima para diseño de cuestionarios; 3) muestra tipo, también 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 y estandarización, en este muestreo el investigador selecciona a los sujetos de la muestra que serán evaluados según ciertas características tales como la edad, el sexo, el estado civil; 4) la muestra por cuotas que se utiliza mucho en estudios de opinión y de mercadotecnia en la que los encuestadores reciben un número o cuota de cuestionarios que deben de aplicar.(Fernández, et al. 1991).

Para determinar el número de sujetos que se requieren para construir la muestra en la que se aplicará el instrumento y llevar a cabo los procedimientos para verificar confiabilidad y validez I. Reyes (octubre 1996), recomienda que sean 10 sujetos por cada uno de los reactivos, por el número de opciones de respuesta, por ejemplo, en un instrumento que cuente con

100 reactivos, y cada uno de ellos tiene dos opciones de respuesta, la muestra debería estar constituida por 2,000 sujetos.

Una vez establecido el diseño de muestra a emplear y el número de sujetos a evaluar se prosigue a la aplicación de la prueba, para tal actividad es necesario la especificación de ciertos criterios para la aplicación entre los que se encuentran las características de los escenarios físicos, el tiempo de aplicación y algunos otros.

Debe considerarse también las características del personal que va aplicar la prueba, tales como la edad y el sexo del entrevistador, además dicho personal debe de estar entrenado para realizar la aplicación, es decir, debe conocer todos y cada uno de los elementos de la prueba, desde las instrucciones, ejemplos y reactivos, debe tener la capacidad para crear un ambiente de empatía con la persona y poder persuadirla para obtener su colaboración.

2.4.2. Análisis de reactivos

Para llevar a cabo el análisis de reactivos de un instrumento es necesario evaluar la distribución de puntajes y estimar si los reactivos discriminan, es decir si contribuyen a la variación de los puntajes obtenidos por el instrumento (I. Reyes, octubre 1996).

Existen dos tipos de análisis que pueden realizarse: paramétricos y no paramétricos. Cada tipo posee sus características y supuestos. La elección del investigador sobre qué clase de análisis va a efectuar depende que esas suposiciones se ajusten a las características de los datos (Fernández et al. 1991).

Para realizar análisis paramétricos debe partirse de los siguientes supuestos:

- 1) La distribución poblacional de la variable dependiente es normal.

2) El nivel de medición de la variable dependiente es por intervalos o razón.

3) Cuando dos o más poblaciones son estudiadas, éstas tienen una varianza homogénea y la dispersión es similar en sus distribuciones (Weisma citado por Fernández et al. 1991).

En cambio para realizar análisis no paramétrico debe partirse de las siguientes consideraciones:

a) La mayoría de estos análisis puede basarse en distribuciones que no son normales.

b) Las variables pueden estar medidas en un nivel nominal u ordinal. De hecho, si se quiere aplicar análisis no paramétrico a datos por intervalos o razón estos deben de ser resumidos a categorías discretas. En el análisis no paramétrico se emplea la prueba chi cuadrada y en el análisis paramétrico la prueba t de Student (Fernández et al, 1991).

La prueba "t" de Student es una prueba estadística que permite evaluar si dos grupos difieren entre sí de manera significativa respecto a sus medias. La comparación se realiza sobre una variable. Los niveles de medición de la variable deben de ser de intervalos o razón (Fernández et al. 1991).

Para saber si el valor "t" es significativo, se aplica la fórmula y se calculan los grados de libertad. La prueba "t" se basa en una distribución muestral o poblacional de diferencia de medias conocida como la distribución "t" de Student. Esta distribución es identificada por los grados de libertad, los cuales constituyen el número de maneras como los datos pueden variar libremente.

Una vez calculado el valor "t" y los grados de libertad, se elige el nivel de significancia y se compara el valor obtenido contra el valor que le correspondería en la tabla de distribución de "t" de Student.

La chi cuadrada es una prueba estadística para evaluar hipótesis

acerca de la relación entre dos variables categóricas.

El nivel de medición de las variables es nominal u ordinal. La chi cuadrada se calcula a través de una tabla de contingencia o tabulación cruzada, que es una tabla de dos dimensiones y cada dimensión contiene una variable. A su vez, cada variable se subdivide en dos o más categorías. La chi cuadrada proporciona un índice de asociación entre dos variables (Fernández, et al. 1991).

Para el análisis de reactivos se hace una comparación entre 2 grupos: un grupo de sujetos con los puntajes totales altos y otro con puntajes totales bajos, de tal manera que si la forma en que respondieran ambos grupos al reactivo permite hacer una diferencia entre ellos y esta es significativa, se dice entonces que dicho reactivo discrimina.

Para llevar a cabo esta comparación se constituyen los grupos por el 25% de los sujetos con puntajes totales más altos en el instrumento (grupo alto) y por el 25% de los sujetos con puntajes totales más bajos (grupo bajo). Entonces aplicada la prueba correspondiente, chi cuadrada o t de Student para los datos correspondientes al reactivo bajo análisis y así establecer las diferencias y su nivel de significación.

2.4.3. Análisis de Confiabilidad

El análisis de confiabilidad se llevará a cabo en base a la elección del tipo de confiabilidad el cual estará en función de los objetivos de estudio. Los diversos tipos de confiabilidad y sus procedimientos de cálculo ya fueron descritos anteriormente.

2.4.4. Validez

La elección del tipo de validez a estimar está en función de los objetivos de la investigación.

La herramienta de análisis de datos que permite realizar dicha estimación es el análisis factorial.

El análisis factorial es un término que hace referencia a numerosos métodos de análisis que se pueden utilizar para determinar las variables que están asociadas y constituyen un factor (Nunally, 1991).

El análisis factorial trabaja con base en una matriz rectangular de datos. La matriz de datos o matriz de puntuaciones se simboliza con la letra s y contiene las puntuaciones de n sujetos en k medidas.

Cualquier combinación lineal de las variables en la matriz de datos es un factor de esa matriz.

A cada uno de los grupos o factores se les denota mediante un grupo de variables cuyos miembros se correlacionan más significativamente entre sí que con respecto a las variables no incluidas en el grupo. Se considera que cada factor es un atributo unitario que se mide, en mayor o menor grado, por medio de instrumentos particulares, después de obtener un factor pueden correlacionarse las puntuaciones del mismo con las puntuaciones de cada una de las variables en la matriz de datos. Como el factor es una columna de números, se puede correlacionar de manera directa con cualquier otra columna de la matriz de datos. Generalmente tales correlaciones se le denomina cargas factoriales lo que se refiere a las correlaciones factor-variable (Nunally, 1991).

La composición factorial de las medidas desempeña un papel específico en los tres tipos de validez que se han explicado.

El análisis factorial es muy importante cuando se trata de seleccionar instrumentos que se utilizarán como predictores. Así en lugar de elaborar una nueva prueba cada vez que surge un problema práctico, se selecciona un instrumento predictor a partir de un "depósito" de instrumentos predictores.

El análisis factorial proporciona útil evidencia circunstancial en relación

con las medidas que pretenden tener una alta validez de contenido. Por ejemplo, el análisis factorial de un conjunto de pruebas de rendimiento podría mostrar que una prueba cuyo objeto es medir habilidades aritméticas, se correlaciona en forma bastante significativa con un factor de comprensión verbal. Esto sugiere que las palabras y oraciones usadas para expresar los problemas fueron tan difíciles que introdujeron un factor no deseado en la prueba, lo cual produciría la necesidad de revisar varias veces la prueba de matemáticas.

El análisis factorial es el procedimiento estadístico principal para la medición de los constructos psicológicos. Donde explicar los constructos es determinar principalmente: la estructura estadística interna de un conjunto de variables que supuestamente miden un constructo y, las estructuras estadísticas entre diferentes medidas de un constructo con las de otro. Para considerar el caso más simple, si todos los elementos tanto del conjunto A como el B se correlacionan en forma altamente significativa los miembros de cada conjunto tienen altas correlaciones con un factor definido por el conjunto.

En relación con la validez predictiva, el análisis factorial adquiere especial importancia al sugerir predictores en la práctica; con la validez de contenido, su importancia estriba en sugerir la manera de revisar los instrumentos para mejorarlos y, con la validez de constructo, al proveer algunas herramientas útiles para determinar las estructuras interna y cruzadas de un conjunto de variables.

Con base en los procedimientos hasta aquí definidos para la adaptación y estandarización de un instrumento, a continuación se presenta la propuesta de análisis para el caso de la Escala de Ajuste Psicosocial a la Enfermedad de Derogatis (EAPE-AR).