

# CAPÍTULO

## CONSTRUCCIÓN DE PRUEBAS PSICOMÉTRICAS

**Elizabeth Zanatta Colín**

### **Resumen**

El capítulo tiene como propósito de mostrar el proceso metodológico del diseño y validación de los instrumentos útiles en la recolección de datos de una investigación cuantitativa. Se inicia con un apartado dedicado a la naturaleza y uso de las pruebas psicológicas en el cual se revisa la definición de prueba psicológica y posteriormente se expone nociones y usos, los fundamentos e importancia de la medición en psicología, la definición y características de un instrumento de medición y la clasificación y, finalmente, se muestra el proceso de construcción de pruebas psicométricas. El diseño y validación de instrumentos de recolección de datos permite contar con herramientas que otorgan precisión de los datos recabados y aumentan la confiabilidad de los resultados de la investigación.

### **1 Introducción**

Las pruebas, tests o instrumentos de medición psicológicos son herramientas, y para obtener los beneficios que proporcionan es necesario tener presente este hecho esencial. Cualquier herramienta puede hacer el bien o el mal, dependiendo de cómo se emplee (Anastasi y Urbina, 1998).

Desde Brown (1999) una prueba es definida como un procedimiento sistemático para medir una muestra de conducta, entendiendo como “procesos sistemático” que una prueba se construye, se administra y se califica, según reglas preestablecidas. Un segundo término es la conducta. En un 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. Así no medimos directamente a la persona sino inferimos sus características (rasgos) a partir de sus respuestas dadas a los reactivos de la prueba.

Los instrumentos de medición en psicología tanto su elaboración como su aplicación implican una responsabilidad por parte del profesional que se encuentra al frente, como profesionales de la salud mental es relevante conocer los antecedentes fundamentales y los conceptos básicos al momento de aplicar, calificar y dar a conocer los resultados ya que las puntuaciones de dichos instrumentos son parte de la información que se proporciona al individuo para ayudarlo a la toma de decisiones.

Habitualmente, la función de las pruebas psicológicas ha sido medir las diferencias entre individuos o entre las reacciones de la misma persona en circunstancias distintas. El diagnóstico del retardo mental fue uno de los primeros problemas que estimularon su desarrollo y, hasta el momento, la detección de las deficiencias intelectuales sigue siendo una aplicación importante de ciertos instrumentos (Anastasi y Urbina, 1998).

Los usos clínicos incluyen el examen de personas con trastornos emocionales graves y otros problemas de conducta. La evolución de las necesidades educativas dio un fuerte impulso al desarrollo inicial de las pruebas, como fue el caso de los famosos tests de Binet que comenzaron el movimiento de las mediciones de la inteligencia (Anastasi y Urbina, 1998).

Siguiendo a Anastasi y Urbina (1998) la selección y clasificación del personal industrial es otra aplicación fundamental de las pruebas psicológicas. De igual manera las pruebas cumplen funciones importantes en la investigación básica. Por ejemplo, en casi todos los problemas de psicología diferencial se aplican test para obtener datos, como ocurre con los estudios sobre la naturaleza y el grado de las diferencias individuales, la organización de los rasgos psicológicos, la medición de las diferencias grupales y la identificación de los factores biológicos y culturales asociados con las variaciones conductuales. En todas estas áreas de investigación (como en muchas otras) es fundamental la medición precisa de las diferencias individuales que las pruebas bien formuladas hacen posible.

En esencia una prueba psicológica representa una muestra de conducta, es una medida objetiva y estandarizada de una muestra de conducta. Con las pruebas psicológicas como con las de cualquier otra ciencia, se hacen observaciones sobre una muestra pequeña, pero cuidadosamente elegida, de la conducta del individuo. El valor de diagnóstico o predictivo de un test psicológico depende de qué tanto funcione como indicador de un área de conducta relativamente amplia y significativa (Anastasi y Urbina, 1998).

Brown (1999) hace énfasis que una prueba va a contener sólo una muestra de todos los reactivos que sean posibles, a consecuencia de ello se plantean dos problemas: en primera debemos de asegurarnos de que los reactivos incluidos en la prueba sean una muestra representativa de todos los posibles existentes. El segundo problema incluye el saber si una persona obtendría la misma calificación al responder a una muestra diferente de reactivos extraída del mismo "dominio". En otras palabras, ¿obtendrá la misma calificación en una forma equivalente de la prueba? El primer problema hace referencia a la validez y el último al de la confiabilidad, ambos términos se abordarán más a detalle a lo largo de éste capítulo.

## **2 Noción y usos de la psicometría**

Para Morales (1990) la psicometría tiene como finalidad llevar a cabo la medición de la conducta. En el desenvolvimiento de la psicología como ciencia, ha sido necesario realizar estudios a fin de poder determinar si difieren los seres humanos entre sí y en qué grado se presentan estas diferencias, lo que ha originado, desde hace aproximadamente un siglo, el desarrollo de la instrumentación adecuada para evaluar cuantitativamente las diferencias y las semejanzas existentes entre los individuos. Aun cuando se ha expresado que los psicólogos deben estar más interesados en las diferencias individuales que en las constantes de las personas, es necesario, a veces, tener en cuenta cuales atributos aparecen en mayor o menor grado, consistentemente, en sujetos o grupos diferentes.

Desde Tovar (2007) la psicometría es entendida como una disciplina de la psicología cuya finalidad intrínseca es la de aportar soluciones al problema de la

medida en cualquier proceso de investigación psicológica. También es un campo metodológico que incluye teorías, métodos y usos de la medición psicológica, en que se circunscriben aspectos meramente teóricos y otros de carácter más práctico. La psicometría se distingue por el uso del lenguaje formal y estructurado de las matemáticas.

Existen muchas definiciones para la psicometría pero su esencia es siempre la misma, su importancia reside en el conocimiento e investigación de fenómenos conductuales, cognitivos, afectivos y actitudinales.

Desde las primeras décadas del pasado siglo los tests psicométricos han sido contruidos siguiéndose el enfoque de la Teoría Clásica de los Tests, creada por el psicólogo inglés Charles Spearman en la segunda década del siglo XX. En las tres últimas décadas del mismo siglo apareció otro enfoque para la construcción de tests que ha sido llamado Teoría de Respuesta al Ítem (TRI), pero todavía hay pocos tests basados en esta teoría. Los tests psicométricos han tenido un gran avance relacionado con los avances de la psicometría (Tovar, 2007).

Una de las bondades de la elaboración de instrumentos radica en la generación de validez de lo que se investiga, ayuda a la generación de conocimientos y crea propuestas de intervención, que equivale a la evolución y pleno funcionamiento de la psicometría, teniendo como base la teoría de la medida.

Finalmente una prueba es un instrumento de medición por lo que vale la pena detallar el proceso de la medición en psicología.

### **3 Importancia de la Medición en Psicología**

Una de las funciones de toda teoría de la medición consiste en emplear construcciones hipotéticas, y las desviaciones de la medida de esos valores hipotéticos constituyen la explicación de “eventos paradójicos”. Desde Morales (1990) deriva un concepto amplio y generalizado de lo que entendemos por medición: “es el procedimiento mediante el cual asignamos números (calificaciones, medidas) a las propiedades, atributos o características de los objetos (o de unidades

experimentales), estableciendo las reglas específicas sobre las que se fundamentan tales asignaciones”. El procedimiento de medición especifica una regla de correspondencia, un patrón a partir del cual a estos eventos les corresponderá un subconjunto de números reales.

Mientras que para Nunnally (1991) la medición consiste en reglas para la asignación de números a objetos en tal forma que representen cantidades de atributos. La palabra “objeto” se usa en un sentido amplio e incluye personas. En psicología, medir es dar la magnitud de cierta propiedad o atributo, por ejemplo, la inteligencia, la extraversión, el razonamiento verbal, de una o más personas, con ayuda del sistema numérico.

En congruencia con la definición anterior, Brown (1999) hace énfasis que la medición de cualquier característica implica la utilización de determinados procedimientos (operaciones), según reglas especificadas, para obtener como resultado la asignación de valores numéricos a la ejecución de una persona.

Las construcciones de la teoría psicológica se expresa por medio de las palabras; no obstante pueden traducirse en significados connotativos. Es pertinente aclarar que al momento de llevar a cabo una medición ya sea en el área de psicología, de ciencias sociales o en educación, no se pretende medir al hombre, lo que se mide son sus atributos, descritos como características de los seres humanos.

Los psicólogos, como investigadores científicos, han dirigido su atención al estudio de aquellos eventos que pueden ser evidentes si se siguen las reglas fundamentales que la ciencia ha adoptado y que se han demostrado a lo largo de su desarrollo desde hace varios siglos (Morales, 1990).

Como en todo estudio científico, en el campo de la psicología debe haber diversas formas de medición; así, con excepción de la precisión que hay en las matemáticas, resulta difícil hacer afirmaciones absolutas, ya que toda ciencia se tiene que auxiliar de ellas para probar sus hipótesis y darlas por verdaderas (Morales, 1990).

La asignación de números a los atributos de los objetos se deriva de lo que denominamos *niveles de medición*, mismos que se diferencian por su grado de complejidad de los sistemas matemáticos que utilizan y las transformaciones matemáticas que admiten los datos para ser manejados en la obtención de resultados comprensibles (Morales, 1990).

### **Niveles de Medición:**

Los niveles de medición son llamados con más propiedad *escalas de medición*, son elementos que se emplean frecuentemente, los cuales se clasifican desde la medida más elemental hasta la más compleja. Refiriendo a Brown (1999) éstos niveles de escalas son jerárquicos, ya que las de nivel superior satisfacen todos los requisitos de las de nivel inferior más otros adicionales que caracterizan a ese nivel dado. Del orden inferior a superior, de lo más simple a lo más complejo, de la siguiente manera:

- Escalas Nominales (o de categorización).- La medición en una escala nominal consiste simplemente en asignar las personas a categorías cualitativamente distintas. La operación fundamental es la de determinar si hay dos personas que sean miembros de la misma categoría o clase; o sea, si poseen características comunes. Por ejemplo, las personas se pueden clasificar como hombres o mujeres, o votantes republicanos, demócratas independientes. Si se desea, se podrán asignar números a cada uno de esos grupos. Así los hombres pueden denominarse 1, y las mujeres 2, o bien, los republicanos 1, los demócratas 2 y los independientes 3.

En una escala nominal, las cifras se utilizan sólo con fines de identificación, para denotar la pertenencia a cierta clase; no implican magnitud.

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

Así, en una escala ordinal, tenemos clasificación y magnitud; pero no tenemos ningún conocimiento sobre el tamaño de las unidades de la escala de medición.

- Escalas Intervalares.- En una escala de intervalos; una diferencia de cierta magnitud significa lo mismo en todos los puntos de la escala. En otras palabras, las unidades de clasificación se puede demostrar que son equivalentes en todos los puntos de la escala. Así, una escala de intervalos implica clasificación, magnitud y unidades de tamaños iguales. En consecuencia, la operación crucial en el desarrollo de una escala de intervalos es la de establecer el que las unidades sean de tamaños iguales. También se debe de poder demostrar que un intervalo dado en la escala, por ejemplo, una diferencia dada de cinco puntos, es equivalente a un intervalo de cinco puntos en cualquier otro lugar de dicha escala.
- Escalas de Razón.- El cuarto nivel de medición es el de las escalas de proporciones. Además de intervalos iguales, las escalas de proporciones tienen también un punto cero significativo. Puesto que es virtualmente imposible definir un punto cero absoluto significativo para la mayoría de las características psicológicas (por ejemplo, tomemos en consideración cómo podríamos definir “inteligencia cero”) y puesto que las escalas de intervalo son suficientes para la mayoría de los fines, no estamos interesados en desarrollar pruebas que midan sobre escalas de razón.

De estas cuatro escalas, Morales (1990) coincide que en el campo de la medición psicológica sólo se pueden utilizar las tres primeras, ya que la última (la escala de razón), por partir de cero absoluto, no es posible emplearla, puesto que sabemos que el universo que medimos en psicología no parte de la nada, sino que siempre se inicia en “algo”.

Desde Anastasi y Urbina (1998) es importante describir los siguientes aspectos:

*Validez.* Se trata del grado con el que verdaderamente mide lo que pretende medir. La validez proporciona una comprobación directa de qué tan bien cumple una prueba su función. Por lo general para determinarla se requiere de *criterios* independientes y externos de lo que la prueba intenta medir. Por ejemplo, una correlación, o *coeficiente de validez*, elevada significa que los individuos que en la prueba obtuvieron una calificación relativamente alta, han sido más o menos exitosos en la escuela de medicina, mientras que los que obtuvieron bajas calificaciones en la prueba han tenido un pobre desempeño académico (Anastasi y Urbina, 1998).

Una correlación baja indicaría que existe poca correspondencia entre la puntuación obtenida en la prueba y la medida considerada como criterio y, por lo tanto, que la validez de la prueba es poca. El coeficiente de validez nos permite determinar qué tan bien se predice el desempeño que se toma como criterio a partir de las puntuaciones de la prueba (Anastasi y Urbina, 1998).

Para Nunnally (1991) un instrumento de medición es válido si cumple satisfactoriamente el propósito con el que se diseñó. Para dicho autor la validación siempre requiere de investigaciones empíricas, y la naturaleza de la evidencia requerida depende del tipo de validez. La validez es una cuestión de grados, y la validación un proceso permanente. En sentido estricto, no se valida un instrumento de medición, sino el uso que se le da.

*Validez de Contenido.* Desde Santisteban (2009) la validez de contenido hace referencia a la relevancia y al alcance del contenido. Los procedimientos para la validación del contenido consisten fundamentalmente:

- a) En el examen sistemático del contenido de la prueba para determinar si cubre todos esos contenidos, es decir, que es una muestra representativa del dominio de la conducta que se pretende medir.
- b) En el análisis de la relevancia de esos contenidos, pues deben incluir todo lo que es esencial o relevante, sin prejuicio de que también se incluyan cuestiones colaterales o irrelevantes, pero sabiendo que lo son y tratándolas



como tales. Por lo tanto, en la validación de contenido se trata de analizar tanto la representación como la relevancia del contenido.

Para establecer la validez de contenido se requiere de un análisis sistemático del dominio o área de conducta. Éste análisis deben realizarlo expertos que constaten que los ítems del test cubren todos los aspectos importantes y que de todos y cada uno de esos aspectos debidamente representados, incluyendo la proporción correcta de ítems (Santisteban, 2009).

Establecer la validez de contenido exige conocer muy bien, en amplitud y en profundidad, el constructo del que se quiera realizar la medición, así como las conexiones con sus manifestaciones observables. De este modo, un experto, y sólo un experto, podrá concluir si la prueba tiene validez de contenido porque incluye todos los aspectos relevantes (Santisteban, 2009).

La validez de contenido se introduce desde el inicio del proceso de construcción de la prueba mediante la elección de los ítems adecuados. Es el grado en que los contenidos de una medida reflejan fielmente los del dominio del constructo que se está evaluando, y quienes deciden sobre ello no son inexpertos o aficionados, sino verdaderos expertos en el campo, tanto en la vertiente teórica como en la empírica (Santisteban, 2009).

*Validez de Constructo.* Siguiendo a Santisteban (2009) la validez de constructo está fuertemente ligada a un contexto teórico en el que se considera la fundamentación conceptual de un constructo debe incluir las conexiones entre ese constructo y otros constructos psicológicos. La validez de constructo concierne más a lo que es capaz de especificar acerca de la naturaleza del constructo psicológico que subyace a las mediciones, que ha demostrar que un test mide aquello que se supone que mide.

Los aspectos de la validez de constructo que se consideran relevantes para la medición en Psicología y Educación (Messick, 1995, citado en Santisteban, 2009) son los relativos a:

- 1) Contenido: relevancia del contenido, representatividad del dominio de los contenidos y calidad técnica del instrumento de medición.

- 2) Sustantivo: justificación teórica y evidencia empírica de que los procesos que los sujetos ponen en juego al cumplimentar las pruebas están bien relacionados con las tareas que se proponen en esas pruebas.
- 3) Estructural: evaluación de la estructura de las puntuaciones en relación con la estructura del constructo.
- 4) Generalización: examen de en qué grado las propiedades e interpretaciones que se hacen a partir de las pruebas se pueden generalizar a otros grupos, a otras situaciones o a otras tareas.
- 5) Establecimiento de relaciones convergentes y discriminantes.
- 6) Consecuencias: valoración de las implicaciones e interpretaciones de las puntuaciones, así como las posibles fuentes de no-validez relacionadas con la equidad, el sesgo, etc.

*Validez de Criterio.* Se puede dividir en *validación concurrente* y *predictiva*: Los procedimientos de validación de criterio-predicción indican la efectividad de la prueba para predecir el desempeño del individuo e actividades específicas. El término “predicción” se utiliza en sentido amplio para referirse a la suposición que puede hacer la prueba sobre cualquier situación de criterio, o bien en el sentido más restringido de anticipación sobre un intervalo (Anastasi y Urbina, 1998).

Para ciertos usos de los tests la validación concurrente resulta más conveniente y puede justificarse por derecho propio. La distinción lógica entre la validación predictiva y concurrente no se basa en el tiempo, sino en los objetivos del examen.

La validación concurrente es adecuada para las pruebas que se emplean para *diagnosticar* el estado actual más que para predecir los resultados futuros. La diferencia puede ilustrarse al preguntar “¿califica Ortega como un buen estudiante?” o “¿posee Ortega los requisitos para convertirse en un buen profesional?” La primera pregunta requiere de validación concurrente; la segunda de validación predictiva (Anastasi y Urbina, 1998).

La validez es crucial tanto en los ámbitos de la investigación como en la práctica psicológica. Concretamente en el ámbito de los tests, sino tuviesen validez, serían inútiles en la contrastación de hipótesis en relación con las teorías, en la medición

de atributos psicológicos y en las predicciones y toma de decisiones. En la práctica, muchas decisiones se toman basándose en el total o parcialmente en la información que proporcionan los tests, por lo que deben proveerse de los procedimientos que aseguren que los tests poseen la validez suficiente para conducir a conclusiones acertadas (Santisteban, 2009).

*Confiabilidad.* Responde a que tanta funcionalidad tiene la prueba y la consistencia que hay en la misma. La única forma de obtenerlo es realizar una comprobación empírica. La evaluación objetiva de las pruebas psicológicas consiste principalmente en determinar su confiabilidad y validez en situaciones específicas. Como se emplea en la psicometría, el término “confiabilidad” significa consistencia (Anastasi y Urbina, 1998).

La confiabilidad de una prueba es la consistencia de las puntuaciones obtenidas por las mismas personas cuando se les aplica la misma prueba o una forma equivalente. Para comprobar la confiabilidad se comparan las puntuaciones obtenidas por las mismas personas en diversos momentos con diferentes conjuntos de reactivos examinadores o calificadores, o en cualquier otra condición de examinación pertinente (Anastasi y Urbina, 1998).

Desde Nunnally (1991) la fiabilidad de la consistencia interna de un instrumento se puede estimar a partir del *Alpha* de Cronbach. Esta herramienta es una medida de fiabilidad asumiendo que los ítems (escala tipo Likert) miden un mismo constructo y están altamente correlacionados. Si el valor del *Alpha* de Cronbach es más cercano a 1, es mayor la consistencia interna de los ítems. George (2003) sugiere lo siguiente para evaluar los coeficientes de alfa de Cronbach:

Coeficiente alfa > 0.9 es excelente

Coeficiente alfa > 0.8 es bueno

Coeficiente alfa > 0.7 es aceptable

*Estandarización.* La estandarización supone la uniformidad de los procedimientos en la aplicación y calificación de la prueba. Es evidente que si los resultados que

obtienen distintas personas han de ser comparables, las condiciones del examen tienen que ser las mismas para todos. Tal requisito es sólo una manifestación de la necesidad de tener condiciones controladas en todas las observaciones científicas. La formulación de las instrucciones es una parte importante de la estandarización de la nueva prueba, y se extiende a los materiales exactos que deben emplearse, los límites de tiempo, las instrucciones orales, las demostraciones previas, las formas de manejar las dudas de los examinados y cualquier otro detalle de la situación de examinación (Anastasi y Urbina, 1998).

Para Anastasi y Urbina (1998) otro paso importante en la estandarización de la prueba es el establecimiento de las normas, a fin de interpretar el resultado que obtienen el individuo en una prueba, éste se compara con los resultados de otros en la misma prueba, la norma es el desempeño normal o promedio.

*Datos Normativos.* Los resultados de las pruebas tienen por sí mismos poco significado. Sólo alcanzan un significado cuando se compara una calificación individual con los resultados obtenidos por otras personas de la misma prueba. Al comparar las calificaciones de un individuo con las de otras personas en una población especificada (que se llama grupo normativo), obtenemos una indicación de su desempeño relativo en comparación a la de otros de la misma población (Brown, 1999).

Así en el caso de un constructor de una prueba psicológica puede reunir datos que muestran el desempeño de un grupo pertinente de personas. Puesto que, en la mayoría de los casos, deseamos comparar las calificaciones de los individuos con las de varios grupos diferentes, el que desarrolla la prueba. Debe proporcionar datos útiles para varios grupos normativos (Brown, 1999).

#### **4 Instrumento de Medición en Psicología: Definición y Características**

Es necesario explicar qué se entiende por instrumento de medición en psicología. Desde Morales (1990) un instrumento de medición en psicología es aquella técnica metodológica producida artificialmente, que obedece a reglas explícitas y coloca al o a los individuos en condiciones experimentales, con el fin de extraer el segmento

del comportamiento a estudiar y que permita la comparación estadística conductora a la clasificación cualitativa, tipológica o cuantitativa de la característica o características que se están evaluando.

Simplificando esta definición, podríamos utilizar para fines prácticos la establecida por Pierre Pichot: “una prueba psicológica es una situación experimental y estandarizada que sirve de estímulo a un comportamiento y que, manipulada estadísticamente, permite clasificar a los sujetos tipológica o cuantitativamente”. La definición por Lee J. Cronbach: una prueba psicológica es una técnica sistemática que compara la conducta de dos o más personas (Morales, 1990).

El test psicométrico es un procedimiento estandarizado compuesto por ítems seleccionados y organizados, concebidos para provocar en el individuo ciertas reacciones registrables; reacciones de toda naturaleza en cuanto a su complejidad, duración, forma, expresión y significado (Rey, 1973, citado en Tovar, 2007).

La mayoría de las pruebas psicológicas están constituidas por una serie de elementos llamados estímulos, reactivos (ítems), reunidos en una sola cédula, a la cual le llamamos *prueba* (test) si se utiliza como único instrumento de medida, o *subprueba* (subtest) si forma parte de lo que en medición psicológica denominamos batería o escalas de medición, porque están constituidos por más de dos pruebas que evalúan diversos aspectos o características del comportamiento en una misma cédula (Morales, 1990).

Cada reactivo, prueba o *subprueba* produce una medida que, dependiendo de cualquiera de los casos, se adiciona o sintetiza para obtener una medida total, *compuesta* de tantas partes como las mediciones que la *componen*; es decir, las características de una prueba compuesta están en función de las características de sus componentes, y el fin primordial de la teoría de una prueba es determinar el tipo y número de componentes que deben combinarse para construir “un compuesto de propiedades específicas” (Morales, 1990).

Tovar (2007) menciona varios requisitos importantes para que un test sea llamado test psicométrico:

- a) El contenido y la dificultad de los ítems están sistemáticamente controlados (construcción del test).
- b) La situación de aplicación del test: el ambiente en el cual se le administra, el material del test, la administración, debe estar bien definida y debe ser reproducida idénticamente para todos los sujetos examinados con el test.
- c) El registro del comportamiento provocado en el sujeto examinado debe ser preciso y objetivo. Las condiciones de cómo hacer este registro deben estar bien definidas y deben ser cumplidas rigurosamente.
- d) El comportamiento registrado debe ser evaluado estadísticamente con respecto al de un grupo de individuos llamado grupo de referencia o normativo.
- e) Los sujetos examinados son clasificados en función de normas resultantes del examen previo del grupo de referencia o normativo (baremo), lo que permite situar cada una de las respuestas, totales o parciales, en una distribución estadística (contraste).
- f) Las respuestas a las cuestiones planteadas dan una medida correcta del comportamiento al que el test apunta (validez).

## **5 Clasificación de las Pruebas Psicológicas**

Morales (1990) hace una distinción de los términos empleados en el lenguaje común, por ejemplo cuando se habla de un instrumento psicológico, nos referimos a todo tipo de implemento que, dentro de las diferentes áreas y formas de investigación psicológica, utilizamos, los cuales se rigen por los fundamentos metodológicos de la medición científica y tienen características tanto formales como de contenido, que nos permiten emplearlos como instrumento de medición.

La diferencia entre prueba y test no es más que distintiva del “para que” se está utilizando el instrumento en un área específica, que es la medición de ciertos aspectos del individuo y que, nos permite una comunicación exacta de lo que estamos haciendo y con qué fin, mediante éste tipo de instrumento (Morales 1990).

Tovar (2007) presenta la siguiente clasificación de las pruebas psicométricas:

<b>Criterio</b>	<b>Clasificación</b>
Por su forma de dar las instrucciones	<ul style="list-style-type: none"> <li>* Orales</li> <li>* Escritos (aunque en algunos casos hay que ejemplificar lo que se quiere que el sujeto realice en la tarea, como en el caso de personas con problemas auditivo)</li> </ul>
Por su administración	<ul style="list-style-type: none"> <li>* Individual</li> <li>* Colectiva</li> <li>* Autoadministrada</li> </ul>
Por la forma o tipo de respuesta que exigen	<ul style="list-style-type: none"> <li>* Objetivas</li> <li>* Subjetivas</li> </ul>
Por el material de la prueba	<ul style="list-style-type: none"> <li>* Lápiz y papel</li> <li>* Verbal completamente</li> <li>* De ejecución (material, manual, visomotor)</li> <li>* De aparatos especiales</li> <li>* Combinación de los tres primeros (Ej.: WAIS)</li> </ul>
Por su forma de calificación	<ul style="list-style-type: none"> <li>* Manual</li> <li>* Electrónica</li> </ul>
Por sus edades límites de aplicación	<ul style="list-style-type: none"> <li>* Pruebas para infantes (baby test)</li> <li>* Pruebas para preescolares</li> <li>* Pruebas para escolares</li> </ul>

	* Pruebas para adolescentes * Pruebas para adultos
Por su ejecución	* Pruebas de poder (power test) * Pruebas de rapidez (speedy test)

Fig.1 Clasificación de Pruebas Psicológicas (Tovar, 2007)

Los test psicológicos pueden clasificarse desde diversos puntos de vista que le son característicos. Morales (1990) los clasifica desde el punto de vista de su *objetivo* y pueden ser: *de potencia o de ejecución máxima, o de rasgos o peculiaridades (ejecución típica o cualitativa) llamados más comúnmente pruebas proyectivas.*

Pruebas de Potencia o de Ejecución Máxima: Exigen del examinado su máximo rendimiento en la tarea o tareas que se le pide que ejecute, tanto a lo que se refiere que “puede” hacer, como a lo que “rinde” en el momento del examen. Bajo éste rubro se incluyen las pruebas de inteligencia, de habilidades o aptitudes múltiples y específicas, de rendimiento (intelectual o de cualquier otro tipo que se esté investigando)

Pruebas de Ejecución de Rasgos (ejecución típica): Como su nombre entre paréntesis lo indica, sería más conveniente llamarlas “pruebas de clasificación tipológica o cualitativa”. En éste tipo de pruebas tenemos los siguientes test: de ajuste, temperamento, intereses, valores, actitudes, preferencias, escalas de personalidad, inventarios de personalidad, índices, técnicas proyectivas, pruebas situacionales, etc.

## 6 Proceso para la Construcción de las Pruebas Psicológicas

Si queremos obtener instrumentos con equivalencia conceptual, lingüística y métrica; debemos realizar un trabajo fundado en habilidades y conocimientos de especialistas que tengan en cuenta la cultura, el idioma, las metodologías de investigación y el correcto uso de los procedimientos estadísticos O sea que la calidad de la adaptación de una prueba es el resultado de la calidad de los diversos



pasos y de la cadena sucesiva de decisiones interdependientes que se deben tomar (Mikurlic, 2004).

Los procedimientos utilizados para la construcción de pruebas psicológicas tienen como fin asegurar que éstas alcancen sus finalidades y metas deseadas. Aun cuando el proceso exacto varía, dependiendo del tipo de prueba psicológica, se puede establecer una secuencia general en los pasos de construcción de las pruebas psicológicas, que incluyen la especificación de su finalidad, la construcción y comprobación de reactivos, el ensamblaje de la forma de la prueba, el análisis y estandarización de las calificaciones de la prueba psicológica (Brown, 1999). (Ver Fig. 1).

### **6.1 Las Pruebas Psicológicas: representación y predicción**

Las funciones de las pruebas psicológicas se pueden ver a partir de diversas perspectivas. Una clasificación que nos permite comprender el proceso de construcción establece una distinción entre lo que representan y lo que predicen las pruebas psicológicas. Una prueba psicológica es una representación cuando sus reactivos son similares a las conductas que nos interesan medir y es predictor, cuando se enfoca en alguna conducta, no considerada en la prueba que deseamos predecir (Brown, 1999).

Desde un punto de vista pragmático, el constructor de pruebas tiene que tomar dos decisiones importantes: determinar el contenido de la prueba y su formato. O sea se debe determinar las conductas, los conocimientos o las habilidades que cubrirán ésta y cómo se presentarán los reactivos. (Brown, 1999).

Brown (1999) afirma que el constructor de una prueba debe tomar en cuenta variables importantes que de igual manera delimitan y hacen que sea uniforme la muestra, debe tomar en consideración variables como la edad, nivel intelectual, educación, orígenes socioeconómicos y culturales y nivel de lectura. Las variables son pertinentes e circunstancias específicas, dependen del tipo y finalidades de la prueba.

## Construcción de Pruebas Psicológicas

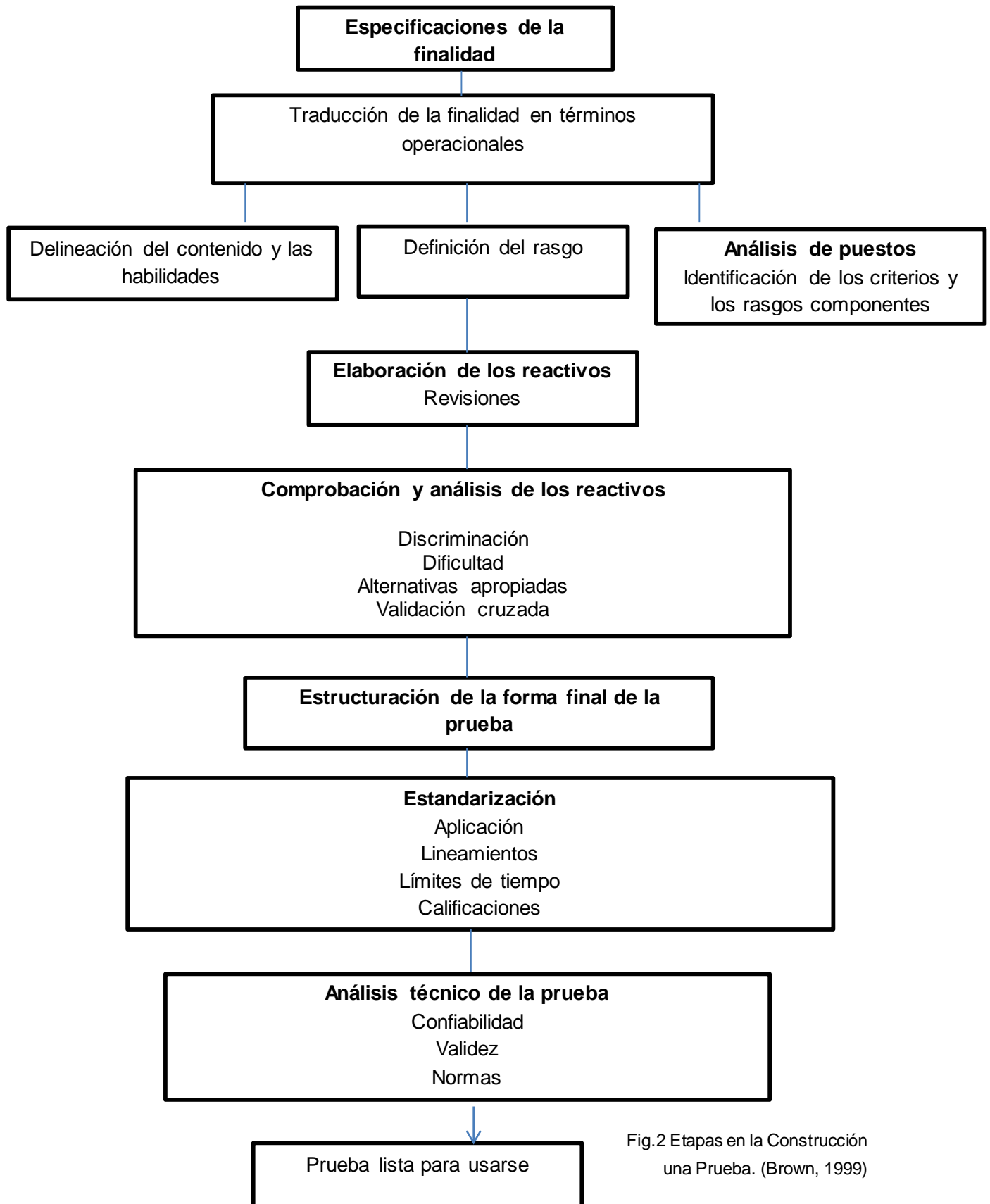


Fig.2 Etapas en la Construcción una Prueba. (Brown, 1999)

## **6.2 Contenido de las Pruebas**

Para proceder con la elaboración de la prueba psicológica, el constructor debe traducir sus propósitos en términos operacionales, lo que nos permite llevar a cabo la medición de la conducta para los fines de lo que se pretende explorar, describir o clasificar.

Para las pruebas representativas, en el caso de las pruebas de rendimiento, la primera etapa consiste en bosquejar el contenido y las habilidades que cubre la prueba. Si la prueba psicológica se desarrolla para medir algún rasgo o alguna característica psicológica dada, la tarea del constructor será distinta. En primer lugar, deberá definir, de modo explícito, el rasgo que se medirá. A continuación, tendrá que indicar las conductas a través de las cuales se manifiesta el rasgo. Es decir, describirá las conductas observables, las habilidades o las aptitudes que sean indicadoras del rasgo que se medirá (Brown, 1999).

En las pruebas de predicción la primera etapa consiste en efectuar un análisis sistemático de la ejecución que se trata de predecir. Por ejemplo los análisis de puestos que implican la predicción de las conductas apropiadas. Después de establecer una hipótesis respecto a las aptitudes, las habilidades o los rasgos que se necesitan para realizar el trabajo, el constructor de la prueba podrá escoger los reactivos que midan esas aptitudes y esas habilidades (Brown, 1999).

El vehículo utilizado para especificar la amplitud de la prueba, es el plan de la misma. Ésta no es más que una tabla que muestra los tópicos que se cubren y las habilidades que se medirán en la prueba, junto con la importancia relativa que se atribuye a cada categoría de contenido de las habilidades. El plan de una prueba psicológica sirve para dos fines. En la etapa de la elaboración de los reactivos, indica la cantidad y el tipo de ellos que es preciso redactar (Brown, 1999).

### 6.3 Formato de la Prueba

El constructor de una prueba debe determinar el modo en que serán presentados los reactivos, para ello, Brown (1999) especifica los formatos más comunes:

1.- Respuesta alternativa vs libre. El examinado escoge la respuesta apropiada de entre varias alternativas, como en los casos de reactivos de comparación, elección múltiple o verdadera y falsa. En los de respuesta libre, el sujeto proporciona una respuesta, como en el completamiento de frases, respuestas cortas o preguntas de ensayo.

2.- Pruebas de velocidad vs pruebas de poder. En una prueba de velocidad los reactivos suelen ser muy sencillos; pero existe un tiempo límite estricto; 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. En esa forma, la calificación refleja 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. En las pruebas de ejecución típica, nos interesa conocer su comportamiento habitual o normal. En general, las pruebas de rendimiento y de aptitudes son medidas de ejecución máxima en tanto que las de personalidad, son medidas de ejecución típica.

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 repuestas. Las pruebas de ejecución implican, con frecuencia, la manipulación de algún aparato u objeto.

5.- Aplicación colectiva vs individual. Las pruebas colectivas se aplican a más de un 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.

#### **6.4 Construcción de la Prueba**

Elaboración de los Reactivos: Después de tomar las decisiones preliminares, quien desarrolla la prueba, está listo para comenzar a redactar los reactivos. Una vez más, tendrá varias opciones para actuar. El proceso de desarrollo de buenos reactivos es la redacción, corrección, prueba y revisión. A continuación, se repiten estas etapas hasta desarrollar reactivos satisfactorios (Brown, 1999).

Al principio, se escriben muchos más reactivos de los que se necesitarán, debido a que serán numerosos los que se eliminen mediante los análisis sucesivos. A continuación, se revisan y corrigen las primeras redacciones de las preguntas, tanto por parte de los escritores originales como por otras personas. La corrección implica la eliminación de la redacción ambigua, el fortalecimiento de las alternativas débiles y la exclusión de reactivos duplicados y, por ende, inútiles (Brown, 1999).

Análisis y Comprobación de Reactivos: Los reactivos que sobreviven a esta selección inicial se combinan, a continuación, en una o más formas de prueba previa. Estas pruebas previas se aplican a una muestra de personas similares a las que se examinará con la prueba terminada. La meta de la comprobación previa es obtener información sobre cómo reaccionan los sujetos ante los reactivos. Esta prueba puede incluir comentarios cualitativos, tales como las ambigüedades percibidas en los reactivos, e índices cuantitativos de la dificultad y el poder de discriminación de los reactivos (Brown, 1999).

Ensamblaje de la Prueba: La etapa que sigue es la de preparar la o las formas finales de la prueba. Los resultados de los análisis de reactivos se utilizan para seleccionar los reactivos que proporcionen la mejor discriminación, tengan la dificultad apropiada y no posean ambigüedad ni alternativas deficientes. Después de efectuar los cambios editoriales finales que se consideren oportunos, se imprime

la prueba. Ahora, por primera vez, la prueba existirá como entidad clara (Brown, 1999).

#### **6.4.1 Procedimiento para la Construcción de Instrumento**

- Investigación teórica respecto a referentes del constructo que se desea medir.
- Cómo se caracteriza, qué elementos o dimensiones comprende el rasgo o cualidad a medir.
- Contextualizar los datos a través de entrevistas semiestructuradas
- Aplicación de Redes Semánticas Naturales
- Detección de indicadores y sub-indicadores que permiten explorar y caracterizar el rasgo o cualidad.
- Elaboración de reactivos, clasificándolos en las dimensiones del constructo a medir.
- Validación inter-jueces para determinar validez de contenido y comprensión de reactivos.
- Piloteo de la encuesta en un grupo, modificación con base en resultados.
- Aplicación de la encuesta.
- Determinación de validez de constructo y de consistencia interna para corroborar dimensiones que explora el instrumento.
- Depuración del cuestionario.

#### **6.5 Análisis de la Prueba**

Siguiendo a Brown (1999) se establecen los conceptos siguientes:

Estandarización: En la medición al administrar una prueba, la meta es obtener una estimación tan precisa como sea posible de la ejecución de los sujetos. La estimación precisa en las pruebas psicológicas, igual que en otros procedimientos científicos, depende del control de los errores, es decir, de la minimización de la influencia de los factores irrelevantes para los fines de la medición. 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 son la base de la estandarización e implica que a cada sujeto le sean presentados los mismos reactivos (o equivalentes) y que existan reglas específicas para la administración y calificación de la prueba.

Los elementos de la estandarización son:

**Contenido:** El primer elemento esencial es el conjunto de reactivos comunes que se aplican a todos los sujetos que se someten a la prueba.

**Aplicación:** Se trata de cuidar las condiciones en las que se aplica la prueba, se trata de considerar a las variables extrañas, utilizando además instrucciones prescritas para la aplicación de las pruebas, límites de tiempo estándar y métodos objetivos de calificación. Se pueden establecer dos conjuntos de instrucciones: uno para quien se somete a la prueba y otro para quien la aplica. La mayoría de las pruebas de ejecución típica se administran sin límite de tiempo; sin embargo, cuando se mide rendimiento o aptitudes, la velocidad suele ser un factor importante, además de colocar los reactivos en orden de dificultad creciente.

**Calificaciones:** éste tercer elemento hace énfasis a las calificaciones objetivas. La objetividad implica un acuerdo entre dos o más calificadores competentes y se reducen a tres etapas básicas: la primera es el registro inmediato y carente de ambigüedades de la respuesta, el segundo requisito es una lista de respuestas estándar o correctas o sea, una clave de calificación, mientras que el tercer requisito es un procedimiento para comparar las respuestas dadas por el sujeto con las que figuran en la clave esto es, un procedimiento para clasificar objetivamente las respuestas. En los reactivos de elección alternativa, tales como los de opciones múltiples, éste procedimiento es directo y evidente.

## **6.6 Análisis Psicométrico**

La aplicación de los procedimientos citados, sólo asegurará que la calificación dada de un individuo refleje sus capacidades o sus características de personalidad, y no

las peculiaridades o las vicisitudes de la situación de la prueba. Se trata de retomar hasta qué punto son consistentes las calificaciones de la prueba, qué es lo que mide, qué inferencias se pueden sacar de las calificaciones de la prueba, etc. Para ello es necesario el análisis que mencionábamos con anterioridad retomando los conceptos de confiabilidad, validez y normalización.

Por otra parte en la actualidad, las demandas propias en investigación exigen la obtención de valores que proporcionan mayor validez y sustento a los fenómenos estudiados en psicología: Estadístico Kaiser-Meyer-Olkin (KMO) busca un valor muy próximo a 1, un requisito importante que nos indica que la cantidad de sujetos es pertinente y se espera un buen ajuste factorial.

## **Referencias**

Anastasi, A; Urbina, S. (1998). *Test Psicológicos*. 7ma Ed. México: Prentice Hall.

Brown, F. (1999). *Principios de la Medición en Psicología y Educación*. México: Manual Moderno.

George, D; Mallery, P. (2003). *SPSS for Windows step by step: A Simple Guide and Reference*. 11.0 Update (4.<sup>a</sup> ed.). Boston: Allyn & Bacon.

Mikurlic, I; Muiños, R. (2004). *La Construcción y Uso de Instrumentos de Evaluación en la Investigación e Intervención Psicológica: El Inventario de Calidad de Vida Percibida (ICV)*. Facultad de Psicología-UBA, Secretaría de Investigaciones, XII Anuario de Investigaciones, de la pág. 193-202.

Morales, M. (1990). *Psicometría Aplicada*. México: Trillas.

Nunnally, J. (1991). *Teoría Psicométrica*. México: Trillas.



Santisteban, C. (2009). *Principios de Psicometría*. Madrid: Síntesis.

Tovar, J. (2007). *Psicometría: Test Psicométricos, Confiabilidad y Validez*. En: A. Quintana & W. Montgomery (Eds.), *Psicología: Tópicos de Actualidad*. (pp. 85-108).