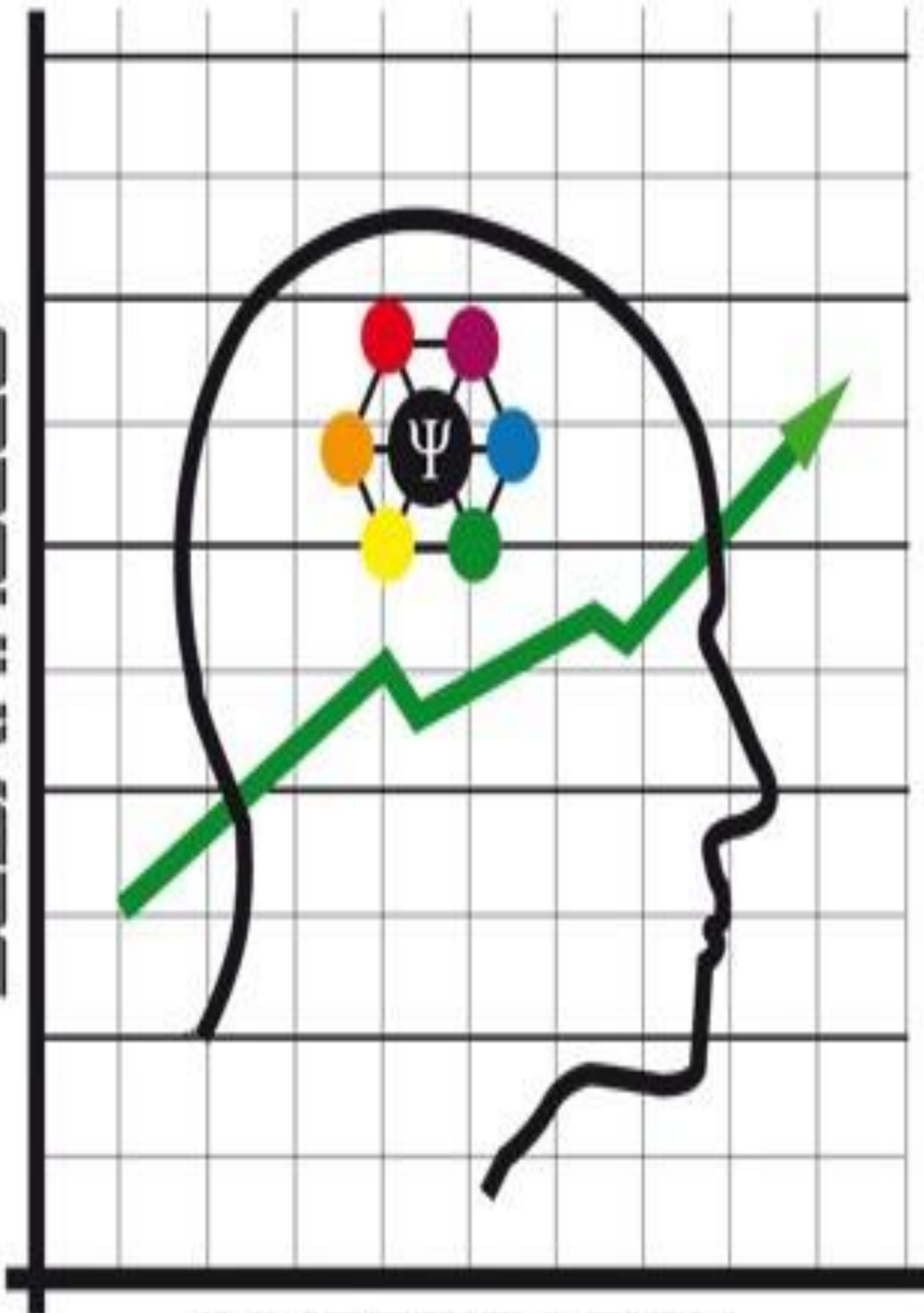


**DESARROLLO**

**INVESTIGACION**



**Nombre del alumno: cleopatra Edith maza gomez**

**Nombre del profesor: Lic. KENNETH ALEJANDRO REYES ESCALANTE**

**Nombre del trabajo: Trabajo de investigación**

**Materia: MODELOS DE INTERVENCION EN PSICOTERAPIA**

**Grado: 7**

**Grupo: "A"**

## Introducción

Los instrumentos de medición como las escalas, índices y cuestionarios, se utilizan para calificar o cuantificar atributos, cualidades, propiedades o constructos (conceptos completamente teóricos), que es imposible medir o cuantificar de otra forma. Preparación: Antes de proceder a medir algo, debemos definir de manera clara y precisa lo que se quiere medir, esto es definir el constructo. Debe explicitarse el objetivo de la medición del constructo, contestando la pregunta ¿para qué lo queremos medir? y la siguiente pregunta ¿en quién lo queremos medir?

Aunque en diferentes niveles, cuando se entrevista a un paciente y se decide que tiene o no una enfermedad, se está realizando un proceso de medición. Medir es asignar un valor a objetos o sucesos de acuerdo con ciertas reglas.<sup>1</sup> Si necesitamos determinar la tensión arterial de un paciente o su nivel de albúmina en sangre, no tenemos ninguna duda ya que contamos con instrumentos específicos para ello; cuando las medidas o cuantificaciones tienen alta probabilidad de ser engañosas, existe dificultad para separar los sujetos o manifestaciones normales de los considerados anormales como sucede con algunos síntomas clínicos aislados, el bienestar, la satisfacción del paciente, grado de funcionalidad, etc. es necesario contar con otro tipo de instrumentos.<sup>2</sup> Los instrumentos de medición como las escalas, índices y cuestionarios, se utilizan para calificar o cuantificar atributos, cualidades, propiedades o constructos (conceptos completamente teóricos), que es imposible medir o cuantificar de otra forma.<sup>3</sup> Las características deseables en un instrumento son: fácil aplicación, amenidad del formato, breve, claro, fácil interpretación de resultados, de fácil puntuación, registro y codificación. Las no deseables: que requieran cálculos complejos para obtener el resultado, que implique alto costo, largo tiempo para ser aplicado.

La evaluación es una de las acciones que constantemente se realiza en la institución escolar, los resultados obtenidos permiten tomar decisiones con base en dos elementos: 1. determinar los aspectos que hay que mejorar y 2. definir y orientar las acciones que permitan elevar la calidad de la propuesta educativa (Instituto Colombiano Para el Fomento de la Educación Superior ICFES, 2001). El evaluar por competencias supone la recolección de evidencia que permita determinar el saber hacer

de un sujeto en un contexto determinado, utilizando herramientas conceptuales y procedimentales adquiridas a través del proceso de formación.

Al evaluar por competencias se deben revisar los desempeños de los estudiantes, es decir, es una medición indirecta a partir de las actuaciones determinadas y esperadas de acuerdo al nivel manejado. Estos desempeños están directamente relacionados con el área a evaluar, se evidencian las características en el manejo de la información y su aplicación en situaciones problema definidas en el área de conocimiento

## ¿Qué es medición?

Una medición es el resultado de la acción de medir. Este verbo, con origen en el término latino metiri, se refiere a la comparación que se establece entre una cierta cantidad y su correspondiente unidad para determinar cuántas veces dicha unidad se encuentra contenida en la cantidad en cuestión.

Por ejemplo: «De acuerdo a la última medición, el pino que se encuentra junto al lago ya mide más de treinta metros», «Antes de comprar el sillón, tenemos que realizar la medición del espacio disponible», «La final del torneo fue seguida por TV por más de veinte millones de espectadores, según una medición de los organizadores».

La medición, en definitiva, consiste en determinar qué proporción existe entre una dimensión de algún objeto y una cierta unidad de medida. Para que esto sea posible, el tamaño de lo medido y la unidad escogida tienen que compartir una misma magnitud.

La unidad de medida, por otra parte, es el patrón que se emplea para concretar la medición. Es imprescindible que cumpla con tres condiciones: la inalterabilidad (la unidad no debe modificarse con el tiempo ni de acuerdo al sujeto que lleva a cabo la medición), la universalidad (tiene que poder usarse en cualquier país) y la facilidad de reproducción. Cabe destacar que es muy difícil realizar una medición exacta, ya que los instrumentos usados pueden tener falencias o se pueden cometer errores durante la tarea.

En la música, se suele repetir el término «La 440» o «tono de concierto». Esto representa la estandarización de la cantidad de vibraciones por segundo que tiene una determinada nota, el La central del piano, y permite una coherencia en la afinación de todos los instrumentos. La unidad utilizada es Hz y para la medición se suele utilizar un diapasón, compañero inseparable de cantantes de ópera y coreutas.

### Medición de sismos

Países como Japón, donde ocurren terremotos todos los años, invierten millones en mejorar las técnicas de detección de sismos. Siendo que no existe forma alguna de detenerlos, la única esperanza de quienes los sufren es adelantarse a los temblores con el tiempo suficiente para la inevitable evacuación. Los aspectos que se pueden medir de un terremoto son su magnitud y su intensidad, para lo cual se utilizan diversas escalas. Las más conocidas son la de Richter, que se enfoca en la causa, y la de Mercalli, que observa el efecto.

La escala de Richter o de magnitud local, permite conocer la energía que se libera en el foco del terremoto, el lugar en que se origina la fractura de las rocas y que luego se extiende a través de las ondas sísmicas. Su cálculo se basa en una expresión logarítmica que ofrece como resultado valores que van aumentando de manera tal que cada unidad represente una magnitud 10 veces más fuerte que la anterior. El grado más alto jamás registrado pertenece a un terremoto ocurrido en Chile en el año 1960, con un valor de 9.5 Richter.

Si bien la magnitud tiene un valor único, la intensidad varía a lo largo del diámetro que alcance el sismo. Para esta medición se utiliza la escala de Mercalli, y su aplicación varía de acuerdo

a la gravedad del terremoto. Básicamente, se observan los efectos en un sitio determinado de la superficie terrestre y se les asigna un grado que va desde el I hasta el XII, en números romanos. La percepción de la intensidad de un sismo es personal y varía de acuerdo al lugar en que se encuentre la persona a la hora de la apreciación, a la altura, al tipo de edificación y de suelo, entre otros aspectos. Gracias a esta medición, es posible mejorar las estrategias de protección de los ciudadanos ante este tipo de catástrofes.

La medición de las variables puede realizarse por medio de cuatro escalas de medición. Dos de las escalas miden variables categóricas y las otras dos miden variables numéricas (Therese L. Baker, 1997). Los niveles de medición son las escalas nominal, ordinal, de intervalo y de razón. Se utilizan para ayudar en la clasificación de las variables, el diseño de las preguntas para medir variables, e incluso indican el tipo de análisis estadístico apropiado para el tratamiento de los datos.

Una característica esencial de la medición es la dependencia que tiene de la posibilidad de variación. La validez y la confiabilidad de la medición de una variable depende de las decisiones que se tomen para operacionalizarla y lograr una adecuada comprensión del concepto evitando imprecisiones y ambigüedad, por en caso contrario, la variable corre el riesgo inherente de ser invalidada debido a que no produce información confiable.

#### a) Medición Nominal.

En este nivel de medición se establecen categorías distintivas que no implican un orden específico. Por ejemplo, si la unidad de análisis es un grupo de personas, para clasificarlas se puede establecer la categoría sexo con dos niveles, masculino (M) y femenino (F), los respondientes solo tienen que señalar su género, no se requiere de un orden real.

Así, si se asignan números a estos niveles solo sirven para identificación y puede ser indistinto: 1=M, 2=F o bien, se pueden invertir los números sin que afecte la medición: 1=F y 2=M. En resumen en la escala nominal se asignan números a eventos con el propósito de identificarlos. No existe ningún referente cuantitativo. Sirve para nombrar las unidades de análisis en una investigación y es utilizada en cárceles, escuelas, deportes, etc. La relación lógica que se expresa es:  $A \neq B$  (A es diferente de B).

#### b) Medición Ordinal.

Se establecen categorías con dos o mas niveles que implican un orden inherente entre si. La escala de medición ordinal es cuantitativa porque permite ordenar a los eventos en función de la mayor o menor posesión de un atributo o característica. Por ejemplo, en las instituciones escolares de nivel básico suelen formar por estatura a los estudiantes, se desarrolla un orden cuantitativo pero no suministra medidas de los sujetos. La relación lógica que expresa esta escala es  $A > B$  (A es mayor que B). Clasificar a un grupo de personas por la clase social a la que pertenecen implica un orden prescrito que va de lo mas alto a lo mas bajo. Estas escalas admiten la asignación de números en función de un orden prescrito.

Las formas mas comunes de variables ordinales son ítems (reactivos) actitudinales estableciendo una serie de niveles que expresan una actitud de acuerdo o desacuerdo con respecto a algún referente. Por ejemplo, ante el ítem: La economía mexicana debe dolarizarse, el respondiente puede marcar su respuesta de acuerdo a las siguientes alternativas:

\_\_\_ Totalmente de acuerdo

\_\_\_ De acuerdo

\_\_\_ Indiferente

\_\_\_ En desacuerdo

\_\_\_ Totalmente en desacuerdo

las anteriores alternativas de respuesta pueden codificarse con números que van del uno al cinco que sugieren un orden preestablecido pero no implican una distancia entre un número y otro. Las escalas de actitudes son ordinales pero son tratadas como variables continuas (Therese L. Baker, 1997).

#### c) Medición de Intervalo.

La medición de intervalo posee las características de la medición nominal y ordinal. Establece la distancia entre una medida y otra. La escala de intervalo se aplica a variables continuas pero carece de un punto cero absoluto. El ejemplo más representativo de este tipo de medición es un termómetro, cuando registra cero grados centígrados de temperatura indica el nivel de congelación del agua y cuando registra 100 grados centígrados indica el nivel de ebullición, el punto cero es arbitrario no real, lo que significa que en este punto no hay ausencia de temperatura.

Una persona que en un examen de matemáticas que obtiene una puntuación de cero no significa que carezca de conocimientos, el punto cero es arbitrario por que sigue existiendo la característica medida.

#### d) Medición de Razón.

Una escala de medición de razón incluye las características de los tres anteriores niveles de medición anteriores (nominal, ordinal e intervalo). Determina la distancia exacta entre los intervalos de una categoría. Adicionalmente tiene un punto cero absoluto, es decir, en el punto cero no existe la característica o atributo que se mide. Las variables de ingreso, edad, número de hijos, etc. son ejemplos de este tipo de escala. El nivel de medición de razón se aplica tanto a variables continuas como discretas.

## Construcción de instrumentos de medición

Preparación Antes de proceder a medir algo debemos definir de manera clara y precisa lo que se quiere medir, esto es definir el constructo. Es importante tener presente, que un mismo problema puede definirse desde diferentes perspectivas teóricas y por lo tanto puede haber diferentes definiciones de un mismo constructo. En ocasiones es necesario construir la definición, seleccionar una de varias disponibles o simplemente explicitarla. Esto se puede hacer mediante consulta a expertos, revisión bibliográfica o ambos.<sup>5</sup> Es fundamental establecer los límites y discriminar lo que pertenece o no al evento de estudio. Sabremos que nuestra definición es útil cuando podamos contestar ¿qué sí es? ¿qué no es? ¿qué partes lo forman? De tal manera que evitemos que se traslape con otros constructos o que se omitan partes importantes en su medición. Además, de la revisión bibliográfica se desprenderá la certeza de que no existe ya tal instrumento, lo cual entre otras cosas justifica su elaboración. Debe explicitarse el objetivo de la medición del constructo, contestando la pregunta ¿para qué lo queremos medir? y la siguiente pregunta ¿en quién lo queremos medir? La respuesta a estas preguntas nos dará el camino para definir la población de interés y con ello las características de formato y manera de administración.<sup>3</sup> Llegado este momento, ya tenemos los elementos básicos para hacer un modelo teórico con el cual podremos explicar la complejidad del constructo, ¿qué partes lo definen?, ¿cómo se expresa en términos operativos? ¿cuáles son los indicadores que nos permitirán medirlo? y anticipar ¿cómo se comportará la medición del atributo en la población de interés en circunstancias particulares?<sup>6</sup>

Elaboración del instrumento Del significado del constructo, propósito de la medición y población de estudio, dependerá en gran medida el contenido del instrumento. Respecto a la función a desempeñar por el instrumento, se requerirán consideraciones metodológicas específicas, por ejemplo, en el área de diagnóstico se puede tratar de un instrumento para detección y por lo tanto será relevante su sensibilidad, para identificar personas en estadios preclínicos o clínicos tempranos o para categorización de los pacientes, siendo de especial importancia su capacidad de discriminación. En el área de pronóstico, un instrumento tendrá que demostrar su capacidad para predecir eventos específicos a futuro. En el área de tratamiento, será trascendente la capacidad de detectar cambio en el estado de salud a través del tiempo.<sup>7</sup> Es esencial tomar en consideración estos aspectos, ya que van a definir si será un instrumento para ser llenado por el médico, por el paciente, a distancia con dispositivos electrónicos o inclusive mediante software. Para la elaboración del primer borrador, se identifican los indicadores que la evidencia científica avala como representantes del constructo. Por ejemplo, si se piensa en un instrumento para identificar personas con mayor susceptibilidad de desarrollar urticaria crónica, será factible si se cuenta con estudios originales, donde se determinen los factores de riesgo y más aún si contamos con sus riesgos relativos podría pensarse inclusive en un instrumento para cuantificar riesgo, los ítems serían

los factores de riesgo. En cambio, si lo que se necesita es medir el trato digno durante la atención médica y en la bibliografía no se ha determinado en forma operativa y válida cuáles son los componentes de trato digno, se tendrá que iniciar con la operacionalización del concepto. En este caso se tendría que empezar por revisiones bibliográficas, para determinar los dominios que componen el concepto de trato digno en la atención médica y luego proceder a validarlos mediante consensos formales con expertos en bioética. En un segundo paso, se tendrían que elaborar indicadores que derivarán de esos dominios para con ellos elaborar los ítems. También se puede dar el caso de que se necesite utilizar metodología cualitativa como grupos focales o entrevistas a profundidad, si lo que se necesita medir son percepciones, capacidades, desempeño de roles o responsabilidades

de los pacientes. Se pueden elaborar instrumentos para categorizar a los pacientes mediante criterios clínicos, que pueden ser elaborados por sociedades profesionales reconocidas internacionalmente.<sup>8,9</sup> Una vez que se han identificado los indicadores del constructo deberán ser transformados en ítems, se aconseja que al menos sean el doble de los que se piensa debe contener el instrumento, pero esto depende de qué tan subjetivo sea lo que mide cada ítem, si son signos o síntomas de enfermedad tan sólo variará la redacción, pero si se mide algo más subjetivo como calidad de la atención en etapas posteriores, seguro se eliminarán gran parte de ellos.<sup>10</sup> En la figura 1 se puede observar la secuencia para realizar el proceso antes descrito. Se debe decidir si se manejarán afirmaciones o preguntas. En el caso de preguntas decidir el número de alternativas de respuesta. Las decisiones para el formato derivarán del tipo de población que lo va a contestar: si el usuario será personal profesional o técnico se usa. Proceso para elaborar un instrumento de medición en salud.

¿Existe con qué medir?	Sí	No	Juzgar	Validez	Confiabilidad	Sí	No	Construcción	Seleccionar
el mejor	Definición conceptual del constructo	Definición operacional del contenido	Primer borrador del instrumento	Bibliografía	Expertos	Técnicas cualitativas	¿Qué sí es?	¿Qué partes lo constituyen?	Dominio
Indicaciones	Ítems	No falta ítems	Validez de contenido	¿Qué no es?	Eliminar ítems ajenos al constructo	No sobran ítems	1. Objetivo del instrumento	• Investigación, práctica clínica	• Predecir, clasificar, diagnosticar, detectar cambio
2. Ámbito de aplicación	• Hospitalización, comunidad, etc.	3. Usuario	• Médicos, enfermeras, pacientes, etc.	4. Población a medir	• Niños, ancianos	5. Forma de aplicación	• Autorreporte	• Entrevista, etc.	

minología habitual en el campo, si es para pacientes se redactará en función a la edad y escolaridad, piloteando para determinar las palabras de uso común. Revisar que los ítems se hayan redactado de forma correcta, evitando que puedan propiciar respuestas socialmente aceptadas, que los ítems o las alternativas de respuesta no sean mutuamente excluyentes, que no se formulen preguntas negativas, etc. De la misma forma se deberá revisar la redacción de las instrucciones, que sean claras y apropiadas a la población en quien se piensa aplicar. Elaborado el primer borrador se debe someter a prueba de usuario. No existe una regla para determinar el tamaño de muestra, dependiendo de la objetividad de los ítems, se podrán aplicar a grupos de 10 en 10, hasta que estemos seguros de que todos los ítems son entendidos por el usuario, con el significado que se les atribuyó cuando fueron elaborados, que no sean ambiguos, identificar y eliminar la carga afectiva. Para ver la potencial utilidad de un ítem, sus respuestas deben variar en los diferentes pacientes y tener un rango de respuesta sin patrones, lo cual también refleja que la muestra de personas participantes, cubre el espectro del concepto objetivo del instrumento.<sup>8</sup> Es útil la entrevista cognitiva mediante la cual documentaremos los ajustes que se van realizando.<sup>11</sup> Debe elaborarse a la par un manual donde se detalle contexto de aplicación, capacitación necesaria, tiempo de llenado y características clinimétricas del instrumento. Es indispensable asegurarse de que la última versión del instrumento que va a ser sometido a validación, contenga los ítems o preguntas



que representen totalmente el evento que se pretende medir, lo que le dará la validez de contenido.<sup>12</sup> Para que el instrumento de medición sea útil, se necesita someterlo a pruebas empíricas que aportarán evidencias de confiabilidad y validez. Procedimiento para determinar la confiabilidad La confiabilidad es la propiedad de mostrar resultados similares, libres de error, en mediciones repetidas.<sup>13</sup> Asumiendo que las condiciones en las cuales se realiza la aplicación de la escala y el estado del atributo, se mantienen estables en el tiempo<sup>14</sup> o en aplicaciones simultáneas por diferentes evaluadores, previa estandarización. El paso siguiente previa estandarización será determinar: a) Si el instrumento que generamos es capaz de ofrecer la misma información, cuando se aplica por un mismo observador dos veces consecutivas (se recomienda 15 días entre mediciones), en ausencia de cambios en el atributo que se está midiendo.<sup>15</sup> b) Si cuando se aplica por más de una persona al mismo tiempo pero en forma independiente a un mismo sujeto, se obtienen datos muy similares. La primera prueba se denomina test-retest y la finalidad es obtener evidencias de que el instrumento por sí mismo no ofrece variaciones ajenas a la cantidad del atributo que se pretende medir; diríamos que el instrumento es «estable». En este caso se elimina la variación que podría provenir de los aplicadores. Como se obtiene con dos mediciones a través del tiempo, se deben ofrecer pruebas de que el atributo no ha sufrido cambios entre una y otra medición La segunda prueba mide la reproducibilidad del instrumento.<sup>16</sup> Pretende verificar si la prueba puede ser aplicada por personas diferentes, por ejemplo dos enfermeras, o inclusive de diferente formación académica, una enfermera y una licenciada en nutrición. El objetivo de esto, es determinar qué tanto afecta a los resultados la variabilidad introducida por un segundo o tercer observador. Por otro lado, es lógico pensar que los ítems contenidos en un instrumento se interrelacionan, pues todos buscan medir lo mismo, esto es la fiabilidad o consistencia interna. Esta propiedad nos informa del grado de homogeneidad que tiene un instrumento; a mayor homogeneidad mayor consistencia. Cuando el instrumento pretende medir atributos con alto grado de subjetividad y que no existe estándar de oro para una validez de criterio, esta propiedad da una medida indirecta de validez (aunque estrictamente no es una medida de validez), cuando un instrumento no demuestra su homogeneidad es muy probable, que los ítems midan constructos diferentes y se pone en duda su validez.<sup>17</sup> Esta prueba se realiza aplicando el instrumento al menos en 100 personas.<sup>15</sup> Procedimiento para determinar la validez La validez es el grado en que un resultado obtenido refleja el fenómeno bajo estudio, o la capacidad del instrumento para medir la cualidad para la que fue construido.<sup>8</sup> La validez de un instrumento se genera ofreciendo evidencias que pueden ser de diferentes tipos: La validez de contenido es el grado en que un instrumento representa la totalidad de los contenidos del fenómeno que pretende medir. Demostrar que no omite ningún ítem ni incluye alguno extraño al atributo de interés. Esto se puede documentar haciendo explícito cómo se definió el constructo, los dominios, indicadores e ítems, ejemplo en cuadro I. La validez de constructo se establece mediante la elaboración de hipótesis, en relación a cómo se comportarán las puntuaciones obtenidas con el instrumento en diferentes circunstancias:

Cuadro I. Cuadro sugerido para transitar desde un constructo hasta la elaboración de los ítems.<sup>18</sup>

Constructo	Dimensión	Indicador	Ítems
Riesgo de cáncer de piel  Definición conceptual: Conjunto de características de una persona que hacen incrementar la probabilidad de desarrollar cáncer de piel	Características constitucionales	Fototipo de Fitzpatrick	¿Cuál es el color de tu piel?
	Antecedentes personales y familiares	Número de nevos melanocíticos	¿Aproximadamente cuántos lunares tiene en total en el cuerpo?
	Exposiciones ambientales	Antecedentes personales y/o familiares y personales de cáncer de piel	¿Su padre, madre, o hermanos tienen o han tenido cáncer de piel? ¿Usted tiene o ha tenido cáncer de piel?
		Exposiciones al sol Ingestión de agua contaminada con arsénico	¿Ha sufrido quemaduras de sol? ¿Cuántas veces? ¿Ha vivido en lugares con playa, en la montaña? ¿Ha consumido agua de pozo por 10 años o más?

a) Aplicando instrumentos que miden constructos diferentes, por ejemplo, si se aplica el cuestionario para medir trato digno por enfermería y el cuestionario de Maslach Burnout; si el cuestionario que construimos realmente mide trato digno, con los resultados se demostrará que a mayor Burnout (desgaste por el trabajo) menor trato digno (validez divergente). b) Comparando grupos extremos, por ejemplo, se construyó un cuestionario para medir riesgo de cáncer de piel, para generar evidencia de que mide lo que dice medir. Se aplicó en dos grupos similares en características sociodemográficas; uno de pacientes con cáncer de piel y el otro sin cáncer de piel y se demostró diferencia clínica y estadísticamente significativa en la puntuación.<sup>18</sup> c) Si se crea un instrumento para predecir muerte en niños que se ingresan a cuidados intensivos. Se aplica a una muestra de pacientes y se siguen a través del tiempo y se verifica si la puntuación obtenida fue capaz de predecir la mortalidad. La validez de constructo sólo se limita por la creatividad y conocimiento del investigador, ya que consiste en demostrar que el instrumento funciona para lo que fue creado. La metodología de la investigación y la estadística, son los instrumentos para generar y someter a prueba empírica estas hipótesis. La validez de criterio es la más robusta, pero requiere de que el atributo que pretende medir el instrumento ya tenga un estándar de oro y consiste en demostrar que el nuevo instrumento mide el atributo de manera comparable, como lo hace el estándar de oro y puede ser de dos tipos: a) Validez concurrente: grado en que se relaciona la nueva medida con la de referencia, siendo ambas administradas simultáneamente. b) Validez predictiva: cuando el criterio de referencia no esté disponible hasta un tiempo después, por ejemplo, desarrollo de una enfermedad y valora hasta qué punto la nueva medida es capaz de predecirlo correctamente. La validez de criterio es un estudio de proceso que se realiza con metodología para pruebas diagnósticas.<sup>19</sup> La validez de un instrumento es cuestión de grado, instrumentos como el Apgar nunca se sometieron a una validación formal, sin embargo, hasta ahora tienen validez de apariencia pues se ha usado durante mucho tiempo por considerarse útil. La validez es cuestión de grado, y las evidencias pueden ser desde las más débiles como la validez de apariencia y contenido hasta la más robusta que es la validez de criterio.<sup>20</sup> Por último cuando un instrumento es aceptado por los usuarios y lo consideran útil, aceptado y contextualizado hablamos de otra característica que es la viabilidad.

Según Herrera (2003), las pruebas de carácter objetivo son tradicionalmente instrumentos usados en la evaluación de conocimientos y habilidades. Generalmente están constituidas por preguntas cerradas y cumplen con las siguientes características:

1. La calificación no depende del evaluador, dado que dichas condiciones de asignación de puntajes están previamente estandarizadas,
2. Existe un control máximo de las condiciones de aplicación, que reduce la influencia de variables extrañas que pudiesen contaminar los resultados,
3. El proceso de diseño y construcción de la prueba contempla una serie de pasos organizados de manera sistemática y que implica trabajo interdisciplinario, tanto de expertos conceptuales como de expertos en técnicas de medición.

A este respecto cabe mencionar que en el contexto colombiano se han implementado evaluaciones de este tipo de forma masiva, ejemplo de ellos son las pruebas SABER aplicadas por la Secretaria de Educación Nacional y los exámenes de estado ICFES, en asocio con el Ministerio de Educación Nacional (MEN) y las instituciones de escolares del país.

Bajo este panorama se hace necesaria la proyección de este tipo de procedimientos de evaluación al interior del aula, buscando hacer un seguimiento constante de los estudiantes y potencializando de esta forma, las competencias pertinentes para su desempeño tanto en las pruebas masivas, como en situaciones de carácter evaluativo de su cotidianidad.

En el ambiente educativo, específicamente el nivel de educación básica formal, los docentes de forma un tanto empírica, construyen sus propias pruebas de evaluación, que son aplicadas, pero nunca son sometidas a procesos de validación y estandarización que permitan hacer inferencias de carácter sólido y científico; desde el mismo proceso de construcción se carece del apoyo de elementos de carácter teórico y técnico proporcionados por la teoría de la medición y evaluación, resultado de la carencia en formación específica en esta área, dentro de los programas de licenciatura y formación en pedagogía.

De acuerdo a la panorámica anteriormente presentada se hace necesario llevar a cabo procesos de construcción, validación y estandarización de las pruebas aplicadas en el aula de clase, siguiendo rigurosamente los lineamientos tradicionales de la medición y evaluación. El aporte del presente estudio es entonces llevar a cabo la validación de una prueba en el área de matemáticas, por dos razones fundamentales: primero por considerar esta área como fundamental acorde con lo expuesto anteriormente y segundo pretende servir de modelo metodológico en procesos similares desde diversas áreas de la educación básica formal.

Dentro del proceso de construcción de una prueba se contemplan las fases de: conceptualización, construcción de reactivos, validación, análisis de confiabilidad y por último la estandarización.

La fase de conceptualización inicia estableciendo la finalidad de la prueba en donde se especifica el constructo a medir la población y los objetivos del instrumento, por medio de la definición conceptual y operacional.

La fase de construcción de reactivos incluye el plan de prueba en donde se construye la tabla de especificaciones y se elige el formato de la prueba, es decir, tipo de ítems y estructura.

Al final de esta fase se lleva a cabo el análisis de ítems en donde se observan parámetros como índice de dificultad e índice de discriminación, retomando los parámetros establecidos para la aceptación de ítems planteados por Guilford (1975), donde el rango aceptable para el índice de dificultad oscila entre 0.2 y 0.85. Y en el caso de la discriminación retomando al mismo autor se asume un índice mínimo de 0.3

Y también se puede hacer un análisis de flujo de opciones con el fin de identificar la calidad de los distractores de dichos ítems. Retomando a Haladyna (1994) y Downingl (1993) (citados por Rodríguez, et al, 2005). Se tendrá en cuenta el criterio de porcentaje de respuesta:

Poco razonables: si son respondidos por menos del 5% de los sujetos.

No discriminan: si son elegidos en igual proporción por quienes tienen puntajes altos y por quienes tienen puntajes bajos.

No claros: si varían a lo largo de los distintos rangos de puntajes de la prueba.

Indeseables: si imitan el comportamiento de la clave, es decir aumentan su frecuencia de respuesta a medida que aumentan los puntajes de la prueba. La fase de validación contempla la recolección de diferentes tipos de evidencia acorde con la finalidad

## **1. Definir la finalidad de la prueba**

El primer paso a dar tras el proceso inicial en la elaboración de una prueba es el de definir específicamente la finalidad que tendrá la prueba.

Resulta importante realizar esta descripción, dado que de esta forma se tendrá claro cuál es el objetivo de la prueba, que variable psicológica busca medir, a que población esta dirigida y de que modo se desarrollará la creación del instrumento de modo que conlleve al objetivo plantado inicialmente.

Lo primero y fundamental que hay que hacer cuando se planifica una evaluación es definir de forma clara, concisa y operativa aquello que se desea evaluar. Esto parece obvio y de sentido común, pero a menudo los objetivos de la evaluación aparecen confusos y pobremente definidos, con lo cual la evaluación está condenada al fracaso, utilícese la metodología que se utilice (Muñiz y Fonseca-Pedrero, 2008, p. 14).

Como se hace mención anteriormente, cuando se desea evaluar determinada variable es importante reducir el objetivo de dicha medición para así mismo determinar precisamente la finalidad de la prueba que evaluará dicha variable. Un ejemplo para aclarar la primera etapa en el proceso de elaboración de pruebas sería: El instrumento a desarrollar tiene como fin medir y evaluar la ansiedad en mujeres que cursan un proceso de duelo por la separación conyugal.

## **2. Aspectos de diseño preliminares**

En la elaboración de pruebas se debe tomar decisiones en que debe tener en cuenta la finalidad de la prueba y para llegar a esto se debe tener presente los siguientes aspectos:

- 1.

- **Modo de aplicación:** Se busca saber si la prueba será individual o grupal. En una prueba colectiva es eficaz pero en la prueba individual permitirá mayor adaptabilidad en los formatos de los reactivos (Hogan, 2004, p.158).
- **Duración:** Aquí se determina si la prueba será breve o larga. Una prueba breve es más eficaz pero su confiabilidad es muy limitada. El tiempo no solo se da por la cantidad de reactivos que se presentan sino también en la sensibilidad de la prueba.
- **Formato de los reactivos:** Hay varios formatos que se utilizan en los reactivos: la opción múltiple, verdadero o falso, acuerdo o desacuerdo, construcción de respuestas (son de mayor flexibilidad pero la más difícil de calificar).
- **Cantidad de las puntuaciones:** Las puntuaciones permite interpretaciones adicionales pero esto genera más reactivos y a causa de esto más tiempo en la prueba.
- **Informes de las puntuaciones:** Se busca saber qué clase de puntuaciones se producen. Si hay registro simple, escrito a mano, de la puntuación o informes a computador en que se informara una puntuación total de la prueba o el conjunto de reactivos.
- **Capacidad del aplicador:** Se plantea cuanta capacidad tiene la prueba para su aplicación y calificación. También se debe tener en cuenta el perfil de quien implemente la prueba ya que debe tener un adiestramiento profesional para aplicar, calificar e interpretar el instrumento.
- **Investigación de antecedentes:** Esta se lleva presente si se lleva una investigación presente si la prueba está pensada para una aplicación práctica generada en que se debe incluir las conversaciones de los practicante (Hogan, 2004, p.158).

### **3. Preparación de reactivos:**

Incluye tanto la redacción de estos como su revisión. Esta redacción esta definida por la finalidad de la prueba y las consideraciones de diseño preliminares. Un REACTIVO de una prueba posee cuatro partes:

- 1) El estímulo al que responde el examinado denominado a menudo la base del reactivo, puede ser una pregunta, una imagen acompañada por una pregunta oral, puede ser un aparato, etc.
- 2) El formato (o método) de respuesta incluye factores, respecto de si éste es un reactivo de opción múltiple o de construcción de pruebas.
- 3) Las condiciones que rigen la respuesta se hallan variables como el límite de tiempo para responder o si quien aplica la prueba puede sondear las respuestas ambiguas y como se registra exactamente la respuesta.

### **TIPOS DE REACTIVOS DE LAS PRUEBAS**

Se clasifican en términos del formato de respuesta a los reactivos. Hay DOS tipos:

- Reactivos de selección de respuesta se le presenta al examinado al menos dos, pero no más de una razonable y pequeña cantidad de opciones, de las cuales tiene que elegir una respuesta. Se denominan reactivos de respuesta múltiple, opción múltiple o de elección forzada. Es el más popular. Por ejemplo, se utiliza una estructura de opción múltiple de 4 o 5 opciones, o la escala likert (cinco puntos que van de “completamente de acuerdo” a “completamente en desacuerdo” en cada respuesta.

Se califican como CORRECTOS O INCORRECTOS; se concede un punto o cero a una respuesta correcta o incorrecta respectivamente luego, se califica la prueba contabilizando la cantidad total de aciertos. Otra forma de calificar consiste en

conceder un crédito parcial por elegir una opción que no es la mejor respuesta posible, pero que no está del todo equivocada. Otra variante supone otorgar una ponderación adicional a los reactivos que son muy importantes. Los sistemas de clasificación más complicados suelen generar solo puntuaciones más confiables y más validas, pero hay quienes afirman que no vale la pena molestarse en utilizar sistemas complicados.

Las consideraciones de diseño preliminares a la elaboración de la prueba determinan en parte el método de calificación de estos reactivos.

- Reactivos de construcción de respuesta a se presenta un estímulo, pero no se restringe al examinado a elegir un conjunto fijo de respuestas; él debe crear o construir la respuesta. Es una respuesta libre en el sentido de que no está limitado a una serie fija de opciones, hay condiciones que rigen la respuesta. Una versión es el formato de LLENAR EL ESPACIO EN BLANCO, donde se omite una palabra o frase clave de un enunciado para que el examinado coloque las palabras ausentes. Otra es la PRUEBA DE ENSAYO, en la cual el reactivo de la prueba se presenta una situación o tema, y el examinado escribe una respuesta que puede ir de unos cuantos enunciados a varias páginas de texto.

La calificación plantea desafíos especiales, debido a que las respuestas pueden ser muy diversas. Hay DOS factores clave para crear puntuaciones útiles a los reactivos de construcción de respuesta: el primero consiste en asegurar la confiabilidad entre calificadores (es crucial); el SEGUNDO es concebir un ESQUEMA DE CLASIFICACIÓN, que varían mucho dependiendo el tipo de prueba (para calificar



los ensayos hay 4 métodos: el HOLÍSTICO, el ANALÍTICO, el RASGO PRIMARIO y el SISTEMA DE PUNTOS).

## **¡VENTAJAS Y DESVENTAJAS DE LOS REACTIVOS DE SELECCIÓN Y DE CONSTRUCCIÓN DE RESPUESTA!**

Aunque en muchas evaluaciones de aplicación individual se emplea el formato de construcción de respuesta, también es posible utilizar este formato en las pruebas de aplicación de grupos, y viceversa con el de selección de reactivos; es decir cualquier prueba que se aplique a un grupo también puede hacerse en forma individual.

### **LOS REACTIVOS DE SELECCIÓN DE RESPUESTA POSEEN 3 PRINCIPALES VENTAJAS:**

1. Dado que la calificación exige poco o ningún juicio, la confiabilidad entre los calificadores o entre jueces es en esencia perfecta (en cambio en la de CONSTRUCCIÓN DE RESPUESTA es un problema sustancial). El reactivo de opción múltiple se convirtió en el formato preferido. Ya que genera puntuaciones más confiables que los otros.
2. Su EFICACIA TEMPORAL al examinado por lo general responde a mas reactivos de selección de respuesta que de construcción de respuesta.
3. La EFICACIA DE LA CALIFICACIÓN, ya que un empleado o escáner electrónico pueden calificar estos reactivos de manera veloz. Los reactivos de CONSTRUCCIÓN DE RESPUESTA poseen también 3 ventajas:
4. Permiten una OBSERVACIÓN MAS SENCILLA DEL COMPORTAMIENTO Y LOS PROCESOS al realizar la prueba facilita observar en el examinado su motivación, ansiedad, forma de abordar los problemas y otros, en formas que los reactivos de selección de respuesta no lo permitirían.

5. PERMITE EXPLORAR ÁREAS INUSUALES que probablemente tal vez nunca aflorarían con uno de selección de respuesta.
6. Influye en los HÁBITOS DE ESTUDIO DE LOS ESTUDIANTES, ya que promueven una aproximación mas holista y significativa al estudio (en cambio los otros más memorización y aproximación atomista al aprendizaje de las materias).

### **CONSIDERACIONES PRÁCTICAS PARA REDACTAR LOS REACTIVOS**

¿Cuántos deben redactarse? No hay una respuesta definitiva a esta pregunta, pues depende en parte de que se tomen las decisiones adecuadas en la etapa de diseño preliminar.

Una regla práctica consiste en preparar entre 2 y 3 veces más reactivos de los necesarios para la prueba final. Así que, si la prueba final contará con 50 reactivos, uno debería preparar entre 100 y 150 para probarlos. Si la prueba final tendrá 50 reactivos y sólo se someten a prueba 55, es casi seguro que el análisis de los reactivos revele más de 5 reactivos con características indeseables.

### **EDICIÓN Y REVISIÓN DE REACTIVOS**

Una vez que se han redactado los reactivos, se someten a una revisión desde varias perspectivas. Primero se revisan en función de su claridad, corrección gramatical y conformidad con las reglas para la redacción de reactivos. En segundo lugar, en el caso de los reactivos de las pruebas de aprovechamiento, se hace una revisión de la corrección del contenido. En tercer lugar, se busca posibles sesgos de género, raciales o étnicos: la cuestión es si algo en el reactivo perjudica o es injusto con algún miembro del grupo, quizás estos respondan de manera inadecuada. Si se sospecha que un reactivo afecta de manera no equitativa el desempeño en la prueba, la tendencia común es eliminarlo de la reserva.

#### **4. Análisis de reactivos:**

Los ítems se han creado con el fin de medir variables cuantificables. Cada ítem cuenta con indicadores como el índice de dificultad, índice de homogeneidad y el índice de validez.

##### **INDICE DE DIFICULTAD**

Este indicador tiene como fin cuantificar el grado de dificultad de cada cuestión.

##### **INDICE DE HOMOGENEIDAD**

El índice de homogeneidad o de discriminación de acuerdo al ítem (H) busca determinar la correlación de Pearson de acuerdo a las puntuaciones de los sujetos (N) en el ítem J y las puntuaciones X, en el total del test.

##### **INDICE DE VALIDEZ**

Este indicador busca determinar si el test está evaluando lo que se pretende lo cual se hace por medio de una correlación.

#### **5. Programas de estandarización e investigación secundarios:**

Durante las investigaciones científicas se hace uso de metodologías, herramientas, técnicas e instrumentos los cuales ayudan con la finalidad de aumentar el conocimiento. La metodología cualitativa, en el modelo general de ciencia, ha propuesto diferentes tipos de investigación que buscan alguno de los tres siguientes objetivos; describir, predecir y controlar (Clark- Carter, 2002). En el programa de estandarización se genera las normas de la prueba, en dicho caso se trata de unos de los aspectos más importantes en lo que corresponde a la creación de una prueba, la cual inicia después de tener seleccionado los reactivos en las bases finales de la etapa de análisis de éstos.

La muestra de estandarización está conformada por el conjunto de examinadores quienes se le aplican el instrumento, con base a sus respuestas se determina los rangos de manifestación del atributo y con ello establecer los intervalos en los resultados que se consideran bajos, medio y/o altos, la elección de dicha muestra es fundamental para el proceso de la medición, puesto que debe ser representativa de las población a la cual se está dirigiendo el instrumento ( o de lo contrario la estimación sería incorrecta).

En lo que corresponde a la estandarización de calificación, los instrumentos de medición permiten obtener un resultado numérico, lo que en la mayoría de instrumentos consiste en una suma simple de los valores dados a las respuestas quien las otorga el examinador, algunos instrumentos son más sofisticados los cuales requieren operaciones matemáticas algo complejas. Para el proceso de calificación esta comprende dos etapas; la primera es la obtención de la puntuación cruda o directa, el segundo es la transformación del puntaje directo a escalas estándares de resultados.

Durante la obtención del puntaje directo es estipulada durante la construcción de los instrumentos, es ahí donde se establece los procesos que tendrán las diferentes señales recolectadas por el instrumento.

## **6. Preparación de los materiales finales y publicación:**

Esta es la etapa final de todo el largo y complejo proceso de la construcción de instrumento de evaluación, pero no la menos importante dentro del ciclo, puesto que de la publicación y éxito de la misma dependerá la excelencia de la prueba elaborada. “Pero la publicación de un instrumento psicométrico constituye algo más que eso, tiene que ver también con las instrucciones de aplicación e interpretación,

los manuales técnicos, los reportes de puntuaciones y otro materiales complementarios”

# Teoría de la medición

Medición es el procedimiento de asignación de numerales a objetos o acontecimientos de acuerdo con ciertas reglas.

- Variable es cualquier característica o propiedad de un objeto o acontecimiento que contenga dos o más categorías posibles en las que un objeto o acontecimiento puede ser clasificado potencialmente.
- Indicador son elementos cuantitativos que sirven para medir un concepto.

- Nominales
- Ordinales
- De intervalo

Ejemplo:

Medir en la población estudiantil de una primaria de 3° a 6° el número de alumnos que ha repetido un grado escolar por cada salón (A, B y C). Escala estimativa: de 1 a 5 alumnos por salón.

## Tipos de Medidas

- Medidas nominales

La propiedad estudiada en los objetos o acontecimiento solo puede agruparse en categorías lógicamente exhaustivas y mutuamente exclusivas, sin que se puedan ordenar.

- Medidas ordinales

Se incluyen las propiedades de la medida nominal y la propiedad de que las categorías pueden ser ordenadas en el sentido de menor que o mayor que. El nivel de medición ordinal no ofrece ningún tipo de información sobre la magnitud de las diferencias entre las categorías.

- Medidas de intervalo

Se asigna al objeto o acontecimientos estudiados números que, además de poseer las características de la medida ordinal, permiten la interpretación de la diferencia entre dos medidas. Lo que caracteriza a una escala de intervalo es la existencia de una unidad de medición común y, constante, que permite asignar un número real a todos los pares de objetos del conjunto ordenado

- Medidas de cociente o proporción

Cuando una medición tiene todas las características de una medida de intervalo y, además, se le puede asignar un punto de origen verdadero de valor 0, se tiene entonces una medida de cociente o proporción.



