

VALIDACIÓN Y ESTANDARIZACIÓN DE INSTRUMENTOS¹

La investigación científica hace uso, entre otros, de teorías, conceptos, metodologías, técnicas e instrumentos, con el fin de aumentar el conocimiento existente respecto a algo (Bunge, 1989). La metodología cuantitativa, en el modelo general de ciencia, ha propuesto diferentes tipos de investigación que buscan alguno de los siguientes tres objetivos: describir, predecir y controlar (Clark-Carter, 2002), haciendo para ello uso de instrumentos de medición que permitan obtener la información ya sea con fines descriptivos, o la usada para predecir fenómenos o la que da cuenta del grado de control que se tiene sobre lo conocido.

Los instrumentos son por tanto aquellas herramientas que permiten obtener evidencia, y cuando ésta se provee en términos de unidades de medida entonces tal instrumento es catalogado como científico. En el presente documento se hará referencia de manera exclusiva a este tipo de instrumentos y el proceso que se debe seguir para asegurar su calidad y modo de uso. En este sentido, primero se expondrá brevemente que implica la medición, luego el papel que juegan los instrumentos en la misma; posteriormente, se presentarán las características técnicas que debe tener un instrumento para que sea considerado válido y se concluirá con algunos lineamientos para garantizar el adecuado uso de los instrumentos y sus resultados.

Instrumentos de Medición

En primera instancia y tal como lo indica Herrera (1998, p.16):

Un instrumento de medida es la técnica o conjunto de técnicas que permiten la asignación numérica a las magnitudes de la propiedad o atributo ya sea por comparación con las unidades de medida o para provocar y cuantificar las manifestaciones del atributo cuando éste es medible sólo de manera indirecta. Un instrumento debe satisfacer tres exigencias básicas:

¹ Documento elaborado por Juan Carlos Espinosa Méndez miembro del Equipo Interdisciplinario de la Universidad Nacional de Colombia, coordinado por José Ignacio Ruíz, para la Subdirección de Evaluación y Tratamiento del INPEC en el marco del Convenio Interadministrativo de Cooperación Académico-Científico No. 29 de 2008

1. Detectar 'la señal' sin interferencia y, en especial, sin intervención del operador. La operación de medida es la interacción objeto de medida-instrumento, por tanto el interés no es ya el objeto de medida sino el complejo objeto-instrumento.
2. No provocar reacción en el objeto de medida o, de ser así, tal reacción debe ser calculable.
3. Basarse en supuestos determinados sobre la relación entre la propiedad y el efecto observado.

Como se aprecia en el texto de Herrera (1998), los instrumentos no se reducen a cuestionarios, inventarios o test, aunque estas sean las formas más comunes y reconocidas de los instrumentos de medida, sino que este concepto se extiende a todas aquellas herramientas que cumplan con las condiciones y exigencias mencionadas. En el texto anterior también se evidencian otros aspectos tales como la medición directa e indirecta, sensibilidad para captar las señales, resultados y su interpretación, las cuales se desarrollan brevemente a continuación.

La medición se conceptualiza como el proceso de cuantificación, el cual consiste en atribuir números a las características observadas de acuerdo con reglas previamente definidas (Bunge, 1989, Kerlinger y Lee, 2002). En este sentido, Herrera (1998, p.11) indica que

...puede entenderse la medición como el procedimiento necesario para la adquisición de información sobre un atributo de un objeto, sujeto o evento en un momento determinado; tal adquisición de información obedece a un objetivo específico e implica la estimación, con ayuda de un instrumento de medición, de la magnitud de dicho atributo en términos de la unidad de medida; el resultado de dicho procedimiento es expresado mediante números.

El procedimiento que señala Herrera (1998) inicia con la definición conceptual del atributo que se desea medir, continúa con la elección del tipo de medición: directa o indirecta, en la que se evalúa el atributo o manifestaciones que se atribuyen a dicho atributo. Posteriormente, se elige o construye el instrumento de medición, se aplica, se obtienen los resultados numéricos, que son interpretados en un marco de referencia particular, para un momento y condiciones también particulares.

La definición conceptual del atributo que se desea medir, es un aspecto fundamental de la medición, pues de ello depende el tipo de medición que se realice, el instrumento que se use y la interpretación de los resultados obtenidos (Aiken,

2003, Coolican, 2005). En este sentido la definición de variables, dimensiones y componentes es el primer paso en el diseño e implementación de un proceso de medición.

La medición directa o indirecta depende en gran parte del atributo que se desea medir, en algunos casos no se tiene acceso al atributo sino de manera mediada por las manifestaciones que se consideran producto de tal atributo; como por ejemplo, cuando se determina la presencia del atributo por sus consecuencias. Por tal razón, cuando el proceso de medición es indirecta, se requiere tener evidencia robusta de la relación existente entre lo que realmente se está midiendo y el atributo que se supone se mide, cuanto más fuerte la evidencia, más precisa la medición, pues tal como lo expresan Kerlinger y Lee (2002), cuando se realiza una medición no se hace realmente ni sobre un objeto, sujeto o evento, ni siquiera sobre sus propiedades como lo indican Brown (1980) y Aiken (2003), sino que la medición se realiza sobre indicadores de las propiedades, por lo que podría afirmarse que la mayoría de los procesos de medición son indirectos.

Tal como se indicó el proceso de medición se realiza con apoyo en algunos instrumentos de medida, los cuales se pueden elegir dentro de la amplia gama existente en el mercado o se pueden construir de acuerdo con los requerimientos que se tengan (Hernández, Fernández y Baptista, 2007). La elección o construcción de un instrumento se basa en cuestiones tan variadas tales como las necesidades que se tengan, los objetivos que se persigan, de la accesibilidad al instrumento, la calidad psicométrica del mismo, la población a la cual va dirigida, la adaptación para las condiciones con las que se cuentan, etc. Sin embargo, los aspectos más importantes son el atributo que se desea medir, el objetivo de la medición (describir, comprobar modelos o evaluar intervenciones) y las propiedades psicométricas del instrumento (confiabilidad y validez), aspectos que se desarrollan posteriormente en el apartado de validación.

El siguiente paso en el proceso de medición es la aplicación y calificación de los instrumentos, el cual es de vital importancia en este proceso, ya que en algunos casos hay una adecuada conceptualización del atributo, los instrumentos cumplen con los criterios de calidad necesarios, pero si el usuario incurre en errores de aplicación

o calificación entonces no se obtendría la información real y verdadera que se persigue.

Por norma general, todo instrumento de medición cuenta con un manual de uso, en el que se especifican los parámetros para su correcta aplicación y calificación, pues la objetividad de los instrumentos de medición se basa fundamentalmente en que el resultado obtenido dependa del atributo medido y no del operador del instrumento, es decir, que sin importar quién aplique o califique el instrumento, el resultado debe ser el mismo para un atributo de la misma magnitud medido de la misma manera (Herrera, 1998). Por tal motivo, es fundamental que antes de usar un instrumento de medición se reciba el adecuado entrenamiento, el cual algunas veces es tan sencillo como leer previamente el manual y seguir las instrucciones que ahí se indican y algunas veces tan complejo que se requiere un aprendizaje asistido por expertos en el uso de tales instrumentos.

Autores como Cohen y Swerdlik (2001), Hogan (2004), Kaplan y Saccuzzo (2006), entre otros, hacen recomendaciones generales para la adecuada aplicación y calificación de los instrumentos de medición. En general, las recomendaciones son sobre el contar con una adecuada planeación de la evaluación, es decir, no hacer la evaluación por solo la oportunidad de hacerla, sino que ésta debe estar enmarcada en un plan general con objetivos claramente definidos y elección de instrumentos basados en consideraciones tales como la población meta, calidad técnica de los instrumentos y posibles usos de los resultados.

Otro aspecto en el que enfocan las recomendaciones es el desarrollo de la evaluación, desde contar con el espacio y tiempo necesario hasta la relación entre evaluado y evaluador, la cual entre otras, debe estar mediado por normas éticas de ejercicio profesional, adecuado entrenamiento en el uso de las herramientas, etc. Finalmente, hacen recomendaciones sobre la calificación, la cual debe seguir parámetros clara y previamente establecidos en el manual de uso, pues de otra manera no se puede garantizar aspectos tales como la objetividad, la replicabilidad y la transparencia.

El paso final del proceso de medición es la interpretación de resultados, momento en el cual autores como Brown (1980) señalan que se pasa de la medición a la

evaluación, pues la primera está interesada en responder la pregunta de ¿Cuánto?, es decir, el resultado expresado en unidades de medida, mientras que la segunda pretende emitir un juicio con base en tal resultado. Sin importar si la interpretación hace parte de la evaluación o de la medición, lo fundamental es que tal interpretación solo es posible dentro del marco conceptual que se usó como referencia para iniciar el proceso de medición. En otras palabras, el resultado de la medición sólo tiene sentido e interpretación dentro del marco conceptual del atributo que se está midiendo, por lo cual las conclusiones solo son sobre dicho atributo y no sobre otro, y solo tienen el alcance que ha sido demostrado y no el de especulaciones con pretensiones de verdad que se basan en el hecho de haber usado un instrumento de medición, aunque éste cuente con los criterios de calidad que se examinan a continuación.

Validación de Instrumentos de Medición

En este apartado se presentará de forma resumida el sustento conceptual en el que se basa el diseño y construcción de los instrumentos de medición, también se expondrán los principales aspectos teóricos sobre la calidad técnica o propiedades psicométricas de tales instrumentos, así como los procedimientos genéricos para su estimación.

Teoría Clásica de los Test

Actualmente existen tres teorías psicométricas denominadas *Teoría de Respuesta al Ítem* (ver Hambleton, Swaminathan y Rogers, 1991; Muñiz, 1997), *Teoría de la Generabilidad* (ver Cronbach, Gleser, Nanda y Rajaratnam, 1972; Shavelson y Webb, 1991), y *Teoría Clásica de los Test* (Brown, 1980; Nunally, 1987; Thorndike, 1989). La última de estas tres teorías es la que sustenta la mayor parte de los instrumentos de medición existentes, por tal motivo será la expuesta en este documento.

El supuesto básico del que parte la *Teoría Clásica de la Medición* es que el atributo medido tiene una magnitud que puede ser expresada mediante números y que los instrumentos de medida tienen la función de detectar dicha magnitud y

traducirla en tales números denominados *unidades de medida* (Herrera, 1998). Por lo tanto, existen dos valores de la magnitud, la que verdaderamente tiene el atributo objeto de medición y la observada en el instrumento de medida, lo deseable es que ambos valores sean el mismo, es decir, que lo verdadero y lo observado sean lo mismo, cualquier diferencia entre estos se denomina *error de medición* (Kelinger y Lee, 2002).

Sin embargo, el único que se conoce de los tres elementos anteriores es el valor observado, lo cual lleva a que se realicen un serie de supuestos y de procedimientos para determinar hasta qué punto los valores observados se asemejan a los verdaderos. Planteando, entonces, que el valor observado es la suma del valor verdadero y el error. Se parte del hecho que el valor verdadero es desconocido, pero si el error se puede cuantificar, entonces la ecuación de tres términos, implícita en los párrafos anteriores, dejaría de tener dos incógnitas y se plantearía que el valor verdadero es el valor observado menos el componente de error que está presente en la medición. Esto llevo a que durante mucho tiempo los especialistas en medición dedicarán grandes esfuerzos a identificar las fuentes de error, a diferenciar tipos de errores, a cuantificarlos y a analizar el impacto que tienen en la medición (Hogan, 2004).

En términos generales, se ha llegado a la conclusión de que existen dos tipos de errores de medición que tienen efectos importantes en este proceso: el error aleatorio y el error sistemático (Brown, 1980). El primero es fluctuante y afecta de manera diferente cada vez que se realice el proceso de medición; mientras que el segundo es constante y afecta siempre de la misma manera este proceso (Kaplan y Saccuzzo, 2006).

Estos errores se debe a fuentes o causas diferentes, las del primero están relacionadas con uso inadecuado del instrumento de medición y con las variaciones en las condiciones de aplicación que pueden generar inconsistencia en los resultados; mientras que el error sistemático generalmente se debe a una deficiencia en el proceso de diseño y construcción del instrumento (Aiken, 2003).

Teniendo en cuenta que lo deseable de un instrumento de medición es que permita realizar mediciones consistentes y precisas, los errores afectan las

propiedades las propiedades psicométricas (Confiabilidad y Validez). A continuación se exponen los aspectos conceptuales de cada una de ellas así como los principales procedimientos para estimarlas y por ende identificar la calidad del instrumento usado en el proceso de medición (Hogan, 2004).

Confiabilidad

La confiabilidad se define como la cantidad de la varianza presente en los resultados de la medición que se debe a diferencias reales en la magnitud de atributo medido, o en otras palabras, es la proporción de varianza observada que es varianza verdadera. El hecho de usar el termino estadístico *varianza* se debe a que se tiene como supuesto que el error aleatorio presente en la medición tiene una distribución normal y por ende el promedio de todos valores observados en las mediciones realizadas, teóricamente infinitas, sería la mejor estimación del valor verdadero, lo cual es una aplicación del *Teorema del Límite Central* procedente de la estadística. Una consecuencia del uso de este teorema es que se puede aplicar las teorías estadísticas de distribución de probabilidades y con ello establecer intervalos de confianza para los resultados de las mediciones realizadas.

Cuando se hace referencia a la confiabilidad, en términos más aplicados y menos conceptuales, se está haciendo referencia a la consistencia de la medición obtenida, es decir, hasta qué punto se obtendría el mismo resultado (valor observado) si se hace nuevamente la medición con ligeras variaciones en el tiempo, o en el espacio, o en el formato y extensión del instrumento, entre otros. Cuanto más consistente el resultado (cuanto menos cambie), la medición será más exacta, es decir, el resultado observado es más similar al valor verdadero de la magnitud del atributo medido.

Existen diferentes procedimientos para estimar la confiabilidad de un instrumento de medición, los más populares son estabilidad, equivalencia y consistencia interna.

Estabilidad: Este procedimiento consiste en aplicar el instrumento a las mismas personas en distinto momento, garantizando similares condiciones de aplicación, de ahí que también se le denomine *tests-retest* (aplicación-reaplicación). Si el atributo no es susceptible de variaciones por desarrollo o envejecimiento de los examinados o entre aplicación y aplicación no se han hecho intervenciones que modifiquen el

atributo, entonces los resultados serán similares en las dos mediciones si el instrumento realmente está midiendo un único atributo estable; en caso que los resultados no sean similares entonces el error de medición es lo suficientemente alto como para alterar el funcionamiento del instrumento.

El procedimiento técnico para estimar la estabilidad de los resultados es la aplicación de una medida de asociación entre los dos conjuntos de datos (valores observados de la medición), el procedimiento más utilizado es el *coeficiente de correlación producto-momento de Pearson*, el cual es una medida estandarizada de la covariación de dos variables. El coeficiente de confiabilidad obtenido varía entre 0,00 y 1,00 donde el límite inferior indica ausencia de confiabilidad y el límite superior indica confiabilidad total, es decir, que el valor verdadero es igual al valor observado.

Equivalencia: Este procedimiento se basa en el hecho que al momento de construir un instrumento realmente se está desarrollando una de sus múltiples versiones, es decir, que de la misma manera como se diseñó y construyó un instrumento *A*, se pudo desarrollar la versión *B* que sea su equivalente en extensión, forma, contenidos y demás; por esto a este procedimiento también se le denomina *Formas Paralelas*. El procedimiento es viable cuando se cuenta con dos versiones de un mismo instrumento (Forma *A* y Forma *B*), las cuales son aplicadas a las mismas personas con mínimo intervalo de tiempo entre aplicaciones. Si el instrumento realmente está midiendo el atributo meta, entonces sus dos versiones deberían hacerlo de manera similar.

El procedimiento técnico para estimar la equivalencia de las dos mediciones obtenidas con versiones diferentes del instrumento pero procedente de las mismas personas se realiza igual que en el caso anterior con una medida estadística de asociación, popularmente con el *coeficiente de correlación producto-momento de Pearson*, los parámetros de interpretación son los mismos.

Consistencia Interna: Este procedimiento parte del hecho de que todos los elementos que constituyen el instrumento están midiendo el mismo atributo y en consecuencia las respuestas de los examinados deberían ser consistentes a lo largo del instrumento y las posibles variaciones en tales patrones deberían ser menores que las diferencias presentes entre las magnitudes de atributo de los examinados, es

decir, que las respuestas particulares de los examinados crean patrones particulares. También implica que diferentes subconjuntos o partes del instrumento estiman de la misma manera el atributo medido, es decir, que cada una de las múltiples partes en las que se puede dividir el instrumento constituye una forma paralela del mismo.

Existen diferentes procedimientos técnicos para la estimación de la consistencia interna, el más general es el *Alpha de Cronbach* el cual se define como la intercorrelación promedio de un universo de reactivos o de todas las posibles partes en las que se puede dividir el instrumento.

El procedimiento técnico para la estimación del *Alpha de Cronbach* se basa en una sola aplicación del instrumento a un solo grupo de examinados, lo cual se debe a que la estimación de la confiabilidad se hace sobre el contenido representado en el conjunto de elementos que constituyen el instrumento y no en variaciones del tiempo de aplicación o en los conjuntos de elementos (ítems) aplicados.

La estimación del *Alpha de Cronbach* se ve afectada tanto por las intercorrelaciones de los elementos como por la cantidad de elementos, de tal manera que un mismo grado de intercorrelación se ve potenciado por diferencias en la cantidad de elementos, pues a mayor cantidad de elementos mayor representación del atributo medido en el instrumento usado, aunque existe un punto de inflexión en el que ésta cantidad de elementos deja de tener efectos significativos.

Teniendo en cuenta las dificultades prácticas de aplicarle el mismo instrumento a las mismas personas en diferente tiempo (estabilidad) o de construir y aplicar dos versiones del mismo instrumento a las mismas personas (Equivalencia), el procedimiento de consistencia interna es la manera más frecuente como se estima la confiabilidad del instrumento. Los parámetros de interpretación son los mismos, valores próximos a 1,00 indican mayor confiabilidad.

Tal como se ha indicado, la confiabilidad es una cuestión de grado, lo deseable es que este cercana al límite superior (1,00). Sin embargo, en ocasiones esto no se logra, por lo que debe estimarse el *Error de medición* que no es otra cosa que desviación media de la inconsistencia expresada en términos de unidades de medida. Con base en este error se pueden establecer intervalos de confianza para el valor observado, cuanto más estrecho sea el intervalo más información real esta

proporcionando el instrumento de medición. También se han desarrollado ecuaciones que permiten establecer el tamaño adecuado del instrumento para aumentar determinada cantidad de confiabilidad; estos aspectos exceden los propósitos de este documento, pero pueden consultarse los trabajos especializados de Aiken (2003) y Hogan (2004) para profundizar al respecto.

Validez

Hogan (2004) indica que los criterios de aplicación y calificación claros y exactos al igual que altos niveles confiabilidad son deseable en un instrumento, pero que lo más importante es la validez. La razón de esto es que la validez hace referencia al grado en que realmente el instrumento está midiendo el atributo que se supone mide.

En ese sentido, la validez es la proporción de varianza observada que es producida por diferencias individuales reales en el atributo que se pretende medir. Tal como ya se ha indicado, la validez se ve afectada por el error sistemático, el cual es un error en la construcción del instrumento que lleva a que cada vez que el instrumento se use el error este presente, debido principalmente a que no hubo una adecuada definición del atributo meta o que la operacionalización del atributo no fue completa o que simplemente se cree que se está midiendo algo cuando en realidad no hay correspondencia entre lo medido y lo que se desea medir.

En psicometría se han definido distintos aspectos relevantes de la validez, en este documento sólo se hará referencia a dos de ellos: validez de contenido y validez de constructo. El primero no cuenta con procedimientos matemáticos o estadísticos que permitan establecer un índice, por el contrario es una medida bastante simple pero útil para examinar si los elementos que constituyen el instrumento realmente son una representación adecuada de todos los posibles elementos que podrían constituir el instrumento para realizar la medición que se pretende. Como se puede inferir del texto, el análisis consiste en la emisión de un juicio por parte de alguien que tiene profundos conocimientos del atributo que se pretende medir, es decir, se realiza a través de una evaluación de expertos o de jueces.

El segundo aspecto de la validez, el constructo, es más sofisticado y se enfoca en la obtención de evidencia sobre el grado en que el instrumento realmente mide el

atributo (constructo) para el cual fue desarrollado. Las estrategias para la obtención de tales evidencias se basa en la diferenciación de grupos criterios, la convergencia con otros instrumentos que miden el mismo atributo, la congruencia con criterios externos, la correspondencia entre la estructura empírica del instrumento fruto de las intercorrelaciones de elementos con la estructura teórica fruto de las conceptualizaciones realizadas sobre el atributo medido.

La estimación de la validez de constructo por medio de grupos criterios consiste básicamente en tener grupos previamente constituidos por examinados con diferentes niveles de la magnitud de atributo; el supuesto subyacente es que el instrumento será válido si logra diferenciar correctamente los grupos que presentan magnitudes de atributo diferentes. Por lo regular, se usan pruebas estadísticas de inferencia sobre diferencias tales como la prueba *t de Student* para grupos independientes. Si grupos diferentes presentan puntuaciones medias significativamente diferentes entonces el instrumento logra capturar adecuadamente las distintas magnitudes del atributo medido. Otro procedimiento similar consiste en tener grupos criterios (conjunto de personas que presentan determinadas características particulares) aplicarles el instrumento y analizar si el instrumento logra detectar las personas que pertenecen a este grupo criterio de otras que no pertenecen a tal grupo.

De otra parte, la estimación de la validez de constructo por convergencia se basa en el hecho que existen diversos instrumentos que permiten hacer mediciones de un mismo atributo, en tal sentido si los instrumentos empleados que pretenden medir el mismo atributo llegan a conclusiones similares en un mismo grupo de examinados, entonces se arguye la existencia de la validez, es decir, que el instrumento que se está usando efectivamente mide el atributo para el que fue diseñado. En otras palabras, este procedimiento consiste en comparar las mediciones obtenidas por instrumentos de medida. El procedimiento para realizar la estimación de la convergencia, es la correlación entre los conjuntos de resultados obtenidos por cada instrumento usado.

Otra manera de estimar la validez de constructo es por procedimiento de análisis factorial, el cual es una técnica de análisis multivariado en el que con base en las

intercorrelaciones entre los ítems que conforman el instrumento se identifican variables latentes, es decir, configuraciones de variables que se denominan factores o componentes.

Para una exposición detallada de los diferentes tipos de validez, los procedimientos para su estimación y la manera de mejorarla, consultar obras especializadas clásicas tales como las de Brown (1980), Thorndike (1989), Nunally (1987), Anastasi y Urbina (1997) o más contemporáneas tales como las de Aiken (2003), Hogan (2004), Kaplan y Saccuzzo (2006).

Estandarización de Instrumentos de Medición

En este apartado se abordarán aspectos relacionados con la estandarización del proceso de medición, específicamente lo concerniente con la aplicación y calificación de los instrumentos de medida. También se revisaran recomendaciones sobre la muestra de estandarización y los diferentes procedimientos para la transformación de puntajes.

Estandarización de la aplicación

Tal como ya se mencionó la aplicación de los instrumentos es un punto sensible dentro del proceso de medición, pues se encuentran múltiples amenazas que de no ser prevenidas pueden quitarle muchas de las ventajas que tiene el uso de los instrumentos de medidas sobre otras estrategias de consecución de información.

Las amenazas más frecuentes tienen que ver con la falta de entrenamiento para el correcto uso de los instrumentos de medición. Este riesgo se controla de manera sencilla mediante el entrenamiento a los usuarios de los instrumentos; tal como también fue mencionado previamente, en algunos casos el correcto uso de los test solo requiere la lectura previa y el seguimiento de instrucciones puntuales consignados en los manuales de uso. Sin embargo, por lo regular una correcta interpretación de tales instrucciones así como el seguimiento de las mismas implica el dominio disciplinar por parte del evaluador. Por ello, en la literatura especializada se recomienda leer el manual antes de usar el instrumento además de entrenamiento dirigido por personas con experiencia en el uso de los instrumentos.

Una segunda amenaza al proceso de medición que está presente en la aplicación ocurre de manera previa a la misma. Este riesgo se hace evidente en la planeación del proceso evaluativo, durante ésta se deben contemplar no solo aspectos como la disponibilidad del material sino también sobre cuestiones logísticas de la aplicación: que hacen referencia no solo al cómo sino al cuándo y dónde se usa el instrumento.

El primer aspecto involucra un tiempo calendario y un tiempo dentro del proceso de intervención; en cuanto al tiempo calendario, se deben tener en cuenta el cronograma de actividades que realizan tanto el evaluador como el evaluado y la duración del proceso de medición; por su parte el momento dentro del proceso de intervención es aún más relevante en tanto que se requiere un parámetro claro de por qué realizar el proceso de medición en ese momento y no en otro.

Un tercer aspecto que debe controlarse dentro del proceso de aplicación del instrumento de medición está relacionado con la interacción entre evaluador y evaluado. Esto tiene que ver básicamente con el dar las instrucciones claras y precisas para que el evaluado puede abordar el proceso de medición de manera tranquila, sepa en que consiste tal proceso y que se busca con el mismo; igualmente, en las instrucciones debe estar claro cuales son los comportamientos que debe emitir el examinado respecto al instrumento.

El otro aspecto se refiere al trato personal que realiza el evaluador con el fin de tranquilizar y orientar el proceso de medición que está realizando, en general, los procesos de medición generan tensión, en especial si de ella se desprenden consecuencias importantes para la vida del evaluado, lo cual puede llevar a que éste se bloqué y la medición no se realice de manera adecuada, llevando a posibles resultados erróneos.

Estandarización de calificación

Tal como se ha indicado, los instrumentos de medición permiten obtener un resultado numérico, que en la mayoría de los instrumentos consiste en una suma simple de los valores dados a las respuestas dadas por el examinado. Algunos instrumentos son un poco más sofisticados y requieren de operaciones matemáticas algo complejas, pero son la excepción y no la regla. En la actualidad la mayoría de

instrumentos de medición vienen acompañados de un programa de computo que permite sistematizar este paso, es estos casos, la atención debe dirigirse en alimentar correctamente el programa de calificación.

El proceso de calificación comprende dos etapas, la primera es la obtención de la puntuación cruda o directa, mientras que el segundo es la transformación del puntaje directo a escalas estándares de resultados.

La obtención del puntaje directo se estipula durante la construcción de los instrumentos, ahí se establecen los pesos que tendrán las diferentes señales recolectadas por el instrumento, así como el parámetro mediante el cual se obtendrán el o los resultados del proceso de medición. Por lo tanto, el usuario debe estar atento a comprender y aplicar los procedimientos que se deben realizar para pasar de la aplicación del instrumento a la obtención del resultado.

Por su parte, el cálculo del puntaje típico presenta dos niveles, el primero que se realiza durante la construcción y evaluación del instrumento, donde no solo se estipula el tipo de transformación de puntajes que se realizará sino también se hace elección de la muestra de estandarización y la equivalencias entre puntajes directos y típicos denominados baremos.

En cuanto a las escalas utilizadas, en general lo que se hace es transformar una escala ordinal en una escala intervalar en la que existe un punto de referencia definido y se pueden establecer comparaciones entre los distintos valores de la escala a diferencia de la escala ordinal donde solo se indica que existe una mayor o menor magnitud de atributo, pero no cuánta en realidad. Las transformaciones más populares se basan en modificaciones de la *puntuación normal estándar* que tiene media cero con desviación 1, por lo que la magnitud de atributo se expresa en distancias respecto al comportamiento promedio de la muestra de estandarización utilizada en el proceso de medición.

La muestra de estandarización esta constituida por el conjunto de examinados a quienes se le aplica el instrumento y que con base en sus respuestas se determinan la los rangos de manifestación del atributo y con ello establecer los intervalos de resultados que se considerarán bajos, medios o altos. La elección de ésta muestra es fundamental en el proceso de medición, pues debe ser representativa de la

población a la cual se dirige el instrumento; de lo contrario la estimación será incorrecta, ya por sobreestimación o por subestimación de la magnitud del atributo del examinado en relación con su grupo de referencia.

Finalmente, los baremos son tablas que los constructores de instrumentos elaboran para establecer la equivalencia entre los puntajes directos y típicos. Estas herramientas descargan al usuario de la labor de realizar cálculos para la transformación de los puntajes. Así pues, la labor de éste consiste en saber a qué escala se transformaron las puntuaciones directas y como se interpretan las mismas, pues el significado de una magnitud particular solo es interpretable dentro del marco conceptual y disciplinar en el cual se circunscribe el atributo medido.

Consideraciones Finales

El diseño, construcción, evaluación de la calidad técnica y estandarización de instrumentos de medición es un área especializada en las distintas disciplinas y dominios del conocimiento, por tanto el presente texto no pretende ser un compendio de tal especialidad, sino que pretende únicamente hacer una breve presentación de algunos aspectos que los usuarios de instrumentos de medición deben conocer para hacer un mejor uso de herramientas y tecnologías valiosas que implican grandes esfuerzos académicos y profesionales propios de las labores de investigación y desarrollo.

Referencias

- Aiken, L. (2003) *Test Psicológicos y Evaluación (11va Edición)*. México: Prentice Hall.
- Anastasi, A. y Urbina, S. (1998) *Tests Psicológicos (7ma edición)*. México: Prentice-Hall.
- Brown, (1980) *Principios de la Medición en Psicología y Educación*. México: Manual Moderno.
- Bunge, M. (1989). *La Investigación Científica (2ª ed. Corregida)*. Barcelona: Ariel.
- Clark-Carter, D. (2002) *Investigación Cuantitativa en Psicología: Del diseño experimental al reporte de investigación*. México: Oxford University Press.

- Cohen, R. y Swerdlik, M. (2001) *Pruebas y Evaluación Psicológicas: Una introducción a las pruebas y a la medición (4ta Edición)*. México: McGraw-Hill
- Coolican, H. (2005) *Métodos de Investigación y Estadística en Psicología*. México: Manual Moderno.
- Cronbach, L.J., Gleser, G.C., Nanda, H., & Rajaratnam, N. (1972). The dependability of behavioral measurements: theory of generalizability for scores and profiles. New York: John Wiley and Sons.
- Hambleton, R., Swaminathan H. y Rogers H. (1991). *Fundamentals of item response theory*. Newbury Park/ London/ New Delhi: Sage publications.
- Hernández, R., Fernández, C. y Baptista, P. (2007) *Metodología de la Investigación (4ta edición)*. México: McGraw-Hill.
- Herrera, A.N. (1998) *Notas sobre psicometría*. Bogotá, D.C.: Universidad Nacional de Colombia.
- Hogan, T. (2004) *Pruebas Psicológicas: Una introducción práctica*. México: Manual Moderno.
- Kaplan, R. y Saccuzzo, D. (2006) *Pruebas Psicológicas: Principios, aplicaciones y temas (6ta edición)*. México: Thompson
- Kerlinger, F. y Lee, H. (2002) *Investigación del Comportamiento: Métodos de Investigación en Ciencias Sociales*. México: McGraw-Hill
- Muñiz, J. F. (1997). *Introducción a la teoría de respuestas a los ítems*. Editorial Pirámide.
- Nunally, J. (1987) *Teoría Psicométrica*. México: McGraw-Hill
- Thorndike, R. (1989) *Psicometría Aplicada*. México: Limusa.
- Shavelson, R. J. & N.M., Webb. (1991). *Generalizability Theory: A Primer*. Newbury Park: SAGE Publications.