CRITERIOS METODOLÓGICOS PARA EL DISEÑO DE INSTRUMENTOS DE MEDICIÓN PARA LA INVESTIGACIÓN CIENTÍFICA

DISEÑO DE INSTRUMENTOS DE MEDICIÓN PARA LA INVESTIGACIÓN CIENTÍFICA

AUTORES: Manuel Villarruel-Fuentes[1]

Elvira Monserrat Villarruel-López[2]

DIRECCIÓN PARA CORRESPONDENCIA: dr.villarruel.fuentes@gmail.com

Fecha de recepción: 19-07-2019

Fecha de aceptación: 24-09-2019

RESUMEN

La travesía científica del investigador dentro de las ciencias sociales, humanas y de la conducta se encuentra acotada por el paradigma de referencia, mismo que dicta el sentido y la naturaleza del método a emplear. Al respecto existen serias dificultades conceptuales y metodológicas asociadas al diseño de cuestionarios o test aplicados en la investigación, lo que determina una escasa validez y confiabilidad en los resultados. Bajo este contexto problemático se exponen una serie de recomendaciones dirigidas al logro de mayores niveles de rigor científico dentro de los test para la investigación, con el objetivo de coadyuvar a la preparación del novel investigador, así como para aquellos que incursionan en el paradigma empírico-analítico.

PALABRAS CLAVE: Validez; confiabilidad; investigación; instrumentos; medición.

METHODOLOGICAL CRITERIA FOR THE DESIGN OF MEASURING INSTRUMENTS FOR SCIENTIFIC RESEARCH

ABSTRACT

The scientific journey of the researcher within the social, human and behavioral sciences is bounded by the reference paradigm, which dictates the meaning and nature of the method to be used. In this regard there are serious conceptual and methodological difficulties associated with the design of questionnaires or tests applied in the research, which determines a low validity and reliability in the results. Under this problematic context, a series of recommendations aimed at achieving higher levels of scientific rigor within the tests for research are exposed, with the aim of contributing to the preparation of the new researcher, as well as for those who venture into the empirical-analytical paradigm.

KEYWORDS: Validity; reliability; research; instruments; measurement.

[1] Doctor en Educación. Miembro del Sistema Nacional de Investigadores en México (CONACYT) y del Consejo Mexicano de Investigación Educativa (COMIE). Líder del Cuerpo Académico "Cultura Académica y Desarrollo Social Sustentable". Profesor-Investigador del Tecnológico Nacional de México - Instituto Tecnológico de Úrsulo Galván. Ver. México.

[2] Licenciada en Negocios Internacionales por la Universidad Veracruzana, México. Especialista en Comercio Exterior por la Universidad Veracruzana. Estudiante de la Maestría Profesionalizante en Gestión del Desarrollo Social, Colegio de Postgraduados, Campus Puebla. México. E-mail: monsevillarruel@gmail.com

© Centro de Investigación y Desarrollo. Universidad Regional Autónoma de Los Andes - Extensión Santo Domingo. Ecuador.

INTRODUCCIÓN

Al margen de los debates sostenidos en torno a la validez del conocimiento, de su pertinencia con la realidad que dice reflejar, o del fundamento desde el cual se concibe como pensamiento científico, la ciencia centra el núcleo de sus planteamientos teóricos en la metodología que le acompaña.

Sobre la base de un ascetismo metodológico que valida y orienta el quehacer del investigador, la rectoría de la razón científica se ajusta a la máxima establecida por Bourdieu, Chamboredon y Passeron (1975), quienes siguiendo a Bachelard (2000) puntualizan en que el hecho científico se conquista, se construye y se comprueba. La tarea se completa cuando el investigador rompe con la ingenuidad que el dato supone, para de ahí reconstruirlo, al asociarlo con la teoría que lo explica, que lo interpreta. Mucho de la vigilancia epistemológica que proponen pasa por reconocer la jerarquía epistémica que el dato implica, esclareciendo la tautología que los estudios experimentales arropan; la idea es trascender el algoritmo para ir tras de la heurística, pasar del «saber inmediato» al «hecho».

Se trata de dos fases que el investigador debe reconocer y atender dentro de su quehacer indagatorio. Para ello es necesario pensar en las limitantes inherentes a la obtención del dato, como reflejo y referente de la realidad que se dice estudiar. No hay quehacer tan complejo como este. Pese a ello existe en el imaginario colectivo, sobre todo en el de muchos noveles investigadores, la idea de que la apropiación del dato resulta de un proceso lineal, por necesidad inmanente a la condición humana, denominador común de todo proceso indagatorio. Se pasa por alto que independientemente de la lógica del error que domina en las principales doctrinas de pensamiento científico (Pérez-López, 2008; Tello, 2011; Sánchez-Rubio, 2016), la relatividad suele dictar los límites del propio pensamiento, ya que es la construcción teórica de los objetos la que impone su sentido y naturaleza.

De esta manera la apropiación del dato termina por ser la evidencia empírica necesaria para respaldar cualquier saber que pretenda ser verdadero, elemento consustancial de la teoría que lo explica e interpreta. En atención a este nivel de incertidumbre, se han diseñado diversos esquemas metodológicos para asegurar que los instrumentos o test aplicados midan lo que dicen medir, y además, lo hagan cada vez que se apliquen en un grupo y contexto determinado (Prieto y Delgado 2010; Romero, 2011; Soriano 2014).

Sobre el empleo de los test, Gómez-Benito, Hidalgo y Guilera (2010) establecen que "un test se administra con un objetivo concreto, generalmente para tomar decisiones que en la mayoría de ocasiones son relevantes para la vida del individuo receptor" (p.75), esto es, son instrumentos valiosos para evaluar y tomar decisiones que en menor o mayor nivel puedan incidir en la intervención concreta de una realidad, una vez que esta es descrita, caracterizada e incluso comprendida.

Por su parte López-Mezquita (2007), establece que "el nacimiento de un test surge de la necesidad de cubrir un hueco en algún área del campo de la evaluación" (p. 271), lo que exige elaborar un riguroso plan para su diseño.

En este cauce de propósitos útiles que la propia investigación supone, existe en todos los investigadores una preocupación latente: la presencia de un posible sesgo en la medición, es decir, la existencia de un error proveniente del diseño del propio instrumento. Al respecto se tiene que:

> …la presencia de un posible sesgo en los ítems que componen el test es una preocupación central en la evaluación de la validez de los instrumentos de medida, entendiendo por validez el grado en que la evidencia empírica y el razonamiento teórico apoyan la adecuación e idoneidad de las interpretaciones basadas en las puntuaciones de acuerdo con los usos propuestos. (Gómez-Benito, Hidalgo y Guilera, 2010, p.75)

De manera análoga es posible afirmar que el test es el fino bisturí con el cual un investigador disecciona la realidad. De ahí la permanente vigilancia que debe asignársele a su diseño.

Al respecto, las tradicionales estimaciones de confiabilidad se hallan ligadas a los sujetos, los ítems y los evaluadores, en tanto la situación de aplicación, las interacciones entre estas fuentes y el error aleatorio «puro» habitualmente se reducen mediante una estricta estandarización de la misma (Cervantes, 2005).

No obstante, este problema puede ser atendido a partir de verificar estadísticamente los resultados de una prueba piloto, a la que se le aplique un análisis de correlación de ítems (Van Zyl, Neudecker, & Nel., 2000; Merino-Soto y Lautenschlager, 2003; Cervantes, 2005; Corral, 2009; Soler-Cárdenas y Soler-Pons, 2012), atendiéndose la consistencia interna de la prueba, lo que involucra esperar a que el test esté terminado y haya sido aplicado a un grupo previamente seleccionado.

Otra alternativa sería el empleo de la validez de contenido (Pedrosa, Suárez-Álvarez y García-Cueto, 2013), constructo (Pérez-Gil, Chacón-Moscoso y Moreno- Rodríguez, 2000) o criterio (García-Garro, Ramos-Ortega, Díaz de León-Ponce y Olvera-Chávez, 2007) donde se busca dar sentido y coherencia a la serie de ítems de que consta el instrumento, antecedente necesario para lograr confiabilidad en las pruebas. Diversos autores han respaldado el empleo de una u otra técnica (Mikulic, s/f), e incluso algunos han sostenido que ninguna de ellas es cien por ciento confiable (Cohen & Swerdlik, 2001; Gempp-Fuentealba, 2006), ya que todas están determinadas por el nivel de error con que se juzga su validez, y no tanto por el grado de confianza que proyectan.

En medio de estas incertidumbres, es natural que el joven investigador vea comprometido su trabajo de observación a partir del cuestionamiento que se haga a su instrumento de medición. La comunidad científica en el área social, humana y de la conducta suele externar severos juicios acerca del diseño de las preguntas incluidas, la escala de medición implicada en ellas –sobre todo si se trata de variables de atributo-, y la orientación seguida dentro del análisis estadístico, así como el análisis de resultados, que muchas veces se ve sesgado por la orientación personal del investigador.

Si se atiende este contexto problemático, se estará en condiciones de avanzar en todo proyecto de investigación que involucre el diseño de instrumentos propios, y no de pruebas estandarizadas, muchas de las cuales no corresponden con las realidades locales, ya que fueron pensadas y probadas en grupos sociales y humanos particulares y por tanto distintos al de interés; incluso su «adecuación» puede representar un reto mayor que comprometa el propio estudio. Esto no es nuevo, Arias-Beatón (1991) lo aclara al indicar que:

> La fuerte necesidad de utilizar los instrumentos de medición en la actualidad, puede provocar un uso indiscriminado de los que existen e incluso, realizar procesos de construcción y validación inadecuados que sí nos conducirían a errores imperdonables en el desarrollo de la ciencia. Es preciso enfrentar el momento con un sentido práctico, pero

> con la suficiente preparación técnica que el hecho requiere, con el objetivo de que la medición en psicología encuentre su justo lugar. (p.94)

Con base en estas consideraciones, y en atención a lo postulado por López-Mezquita (2007), quien establece que todo plan para el diseño de un test debe contemplar los criterios y principios que se van a seguir para construirlo, a continuación se desglosan una serie de recomendaciones dirigidas al diseño básico de un instrumento de medición, a partir de su estructura lógica, su contenido sustancial y su secuencia de entendimiento. Se espera que con este apoyo, el novel investigador pueda alcanzar altos niveles de construcción teórica y racionalidad operativa.

DESARROLLO

Recomendaciones generales para el diseño

A continuación se describen las recomendaciones que deben ser consideradas al momento de diseñar un test para la investigación científica:

1.1. Establecer con claridad el problema de investigación, los objetivos e hipótesis

Cuando se diseñen y redacten las preguntas que se incluirán en el instrumento es importante que estas guarden relación directa con el objeto de estudio y las hipótesis de trabajo, de tal manera que se puedan cumplir los objetivos de la misma.

1.2. Cuando redacte las preguntas tome en cuenta a quién irá dirigido el instrumento

Niños, jóvenes, adultos, personas en estado de vulnerabilidad social, psicológica o emocional, hombres y mujeres ubicados en medios rurales y urbanos, personas empoderadas, sujetos inmersos en conflictos, por citar algunas condiciones de contexto y situacionales, tienen que ser tomadas en cuenta al momento de redactar los ítems. Para ello es necesario una evaluación previa de las condiciones en las que se encuentra la población objeto de estudio. Ello implica saber acerca de aquello que se pretende estudiar, de sus idiosincrasias e improntas. En todo caso es necesario evitar preguntas que representen algún tipo de violencia simbólica. Al respecto "es necesario considerar también que los términos utilizados pueden adquirir diferentes significados según el contexto socio-cultural en que se empleen" (Fernández, Pérez, Alderete, Richaud y Fernández, 2010, p.65).

Adicionalmente es conveniente que los contestantes estén familiarizados con el formato del test y con la estrategia de aplicación. Esto asegura bridar las condiciones adecuadas para su administración, lo que evitará interpretaciones equivocadas que puedan invalidar los resultados.

1.3. Asegúrese que existe un continum en la variable a medir y que usted lo conoce

Se entiende que una variable se puede medir o no es variable. Esto es, la variable es una característica que puede tomar cualquier valor en un continuo de valores; se trata de una característica medible del fenómeno objeto de estudio. Es importante que estas sean identificadas e identificables por todos.

1.4. Verifique el tipo de variable a medir y sus relaciones teóricas

Esto es, el investigador debe diferenciar claramente entre una variable cuantitativa (continua o discreta) y una cualitativa (de atributo). Esto permitirá identificar el manejo que se le dará a la variable, sobre todo a los indicadores que supondrán su medición (variables de respuesta). Para

ello es relevante saber qué variable deberá ser manipulada y controlada (independiente) y cuál se medirá (dependiente). Sobra señalar que se trata de relaciones de causa y efecto (si "p" entonces "q"), de donde se derivarán posibles manejos estadísticos de los resultados.

1.5. Operacionalice las variables a medir

Mientras las variables cuantitativas se miden directamente -peso, edad, estatura, número de hijos, entre otras-, las variables de atributo o cualitativas imponen un reto conceptual y operativo al investigador. La complejidad de estos constructos exige un correcto desglose de sus propiedades, a fin de permitir una observación clara y precisa, libre de interpretaciones personales. Variables como motivación, eficiencia, liderazgo, empatía, entre otras, deben ser primero definidas conceptualmente, para de ahí precisarlas operativamente. De esto depende que sus «dimensiones» y «categorías» puedan ser clarificadas, es decir, se trata de las formas posibles en que pueden ser observadas en un contexto real. Si esto se realiza adecuadamente, el siguiente paso es identificar los «indicadores» bajo los cuales será medida dicha variable; se trata de las expresiones que una categoría puede tomar en un entorno particular, observables para ser medidas.

1.6. Ubique el nivel de medición del dato a obtener

Es necesario puntualizar que «contar» es la forma más simple de medir. A partir de aquí es posible alcanzar altos niveles de medición, los cuales deben ser identificados para cada indicador. Se trata de ubicar la medición que en cada caso será considerada en las futuras operaciones estadísticas. Un error muy común es querer manejar con una estadística paramétrica -por ejemplo, la correlación lineal-, un conjunto de datos obtenidos de variables cualitativas, en un esfuerzo por darles un valor más profundo a las decisiones y conclusiones derivadas de ellas. Es imprescindible considerar que «los datos deben resistir el análisis estadístico» al que serán sometidos. No es posible pensar en aplicar un análisis de varianza para todos los datos, solo porque se conoce muy bien este tipo de pruebas.

Las escalas o niveles conocidos son: nominal, ordinal, de intervalo y de razón. Los dos primeros asociados generalmente a las variables cualitativas, los dos últimos, a las cuantitativas, sin que ello sea una regla rigurosa, ya que, en estudios descriptivos de corte cuantitativo, es viable emplear escalas nominales y ordinales –conteos, relaciones, porcentajes-. A su vez, existen estudios cualitativos cuyos resultados son tratados en los niveles más altos –intervalo y de razón-, siempre y cuando los atributos medidos sean identificados con una expresión numérica, tal como ocurre cuando se emplean pruebas de x^2 (Ji-Cuadrada), con escalas tipo Likert definidas por asignaciones numéricas.

1.7. Reconozca la naturaleza de la variable a medir

No todas las variables pueden ser medidas de forma única, algunas permiten más de una expresión manipulable por el investigador. Ejemplo de ello es la variable de atributo «sexo», la cual es dicotómica estricta, es decir, solo tiene dos modos de expresarse en la realidad: macho y hembra –genéticamente-, a diferencia de la variable «género», que puede tomar distintas expresiones: masculino, femenino y las identidades transgénero.

En este mismo orden de ideas se puede mencionar la variable «preferencia», que al ser de atributo o cualitativa, puede tomar varios niveles en su expresión –siempre, casi siempre, de vez en cuando o nunca-; pero aún hay más: «preferencia» puede manejarse como dicotómica, al desglosarla en dos niveles dentro de la escala, «se prefiere» o «no se prefiere». Así pasa con otras

variables, tales como calidad, liderazgo, compromiso y actitud, las cuales pueden ser manejadas como dicotómicas o politómicas, según convenga al investigador.

Cuide la redacción de las instrucciones

Evite los circunloquios y el "cantinfleo" en la redacción. Considere que si la pregunta no es clara y precisa puede desalentar al contestante. Preguntas redundantes, carentes de redacción y sintaxis, obligan a una doble o triple lectura de las mismas. Es necesario redactarlas en oraciones o enunciados breves y directos, con el uso de la voz activa de los verbos. Como todo enunciado, debe expresar una idea en particular que se manifiesta por una cadena de palabras coherentemente ordenadas.

Al tratarse preferentemente de preguntas o cuestionamientos se debe identificar fácilmente la idea principal, la cuestión a la que se hace referencia, hecho que evita la posibilidad de obtener distintas respuestas a las mismas preguntas.

Por ello se deben evitar frases largas, las que generalmente provocan ambigüedad, o incluso que exista información innecesaria para el lector. Sobre este particular, una inadecuada organización de las ideas puede provocar desánimo y falta de interés por su lectura. Adicionalmente se debe considerar que: 1) las oraciones no deben pasar de las 16-17 palabras; 2) la puntuación debe ser correcta; 3) no debe haber ambigüedad sintáctica o semántica; 4) de preferencia no deben emplearse oraciones subordinadas; 5) prefiéranse ideas simples y organizadas.

1.8. Privilegie las preguntas cerradas sobre las abiertas, a fin de facilitar el manejo estadístico de las respuestas

Las preguntas cerradas brindan la posibilidad de orientar las respuestas bajo distintos niveles de intensidad dentro de una escala, lo que permite manipular y controlar mejor las variables en estudio. Adicionalmente, al tratarse de escalas diseñadas para graduar la magnitud de una variable, bajo asignaciones numéricas, es fácil identificar y definir el manejo estadístico al que serán sujetas.

Las respuestas a las preguntas abiertas comúnmente se valoran como «positivas» o «negativas», asignándoles valores nominales u ordinales –por ejemplo, cero y uno-. Al respecto es poco lo que se puede hacer estadísticamente con esta información, ya que no permite el empleo de operaciones aritméticas. Un inconveniente puede encontrarse en la necesidad de «triangular» al momento de valorar las respuestas, lo que hace necesario contar con personas capacitadas para ello.

1.9. Evite emplear términos como «cree usted», «supone» o «sospecha», ya que constituyen factores de confusión

Considerar claridad en la redacción exige buscar precisión en lo que se solicita, lo que supone evitar expresiones ambiguas que causen confusión en los sujetos. Preguntar por lo que se «cree» o «supone» lleva implícito un mensaje que se asienta más en la vaguedad que en la certeza. Creer es un acto de fe que no puede ser cuestionado por el investigador, y a menos que sea lo que se busque, es preferible evitar su uso dentro de los ítems.

1.10. No incluya más de una incógnita en cada pregunta

Sobre la base de lo indicado anteriormente, la claridad y pertinencia de los ítems descansa también en aquello que se pregunta. La idea es básica: una pregunta en cada ítem. En ocasiones

se cae en el error de redactar la pregunta con ambivalencias, donde es posible que la cuestión planteada pueda interpretarse de maneras contradictorias entre sí, o bien, que remita a dos asuntos relacionados pero distintos. Un ejemplo sería: ¿De qué manera influye el ambiente escolar en el desempeño y rendimiento del estudiante? Frente a este cuestionamiento, el contestante puede sentirse atraído hacia el concepto o idea que mejor conozca o esté familiarizado, para este caso «desempeño» o «rendimiento». El mismo investigador no podría saber si la respuesta otorgada se refiere a uno u otro concepto. Mención aparte merece el trato estadístico que se le brinde a la respuesta, que exigirá una identificación clara al respecto.

La recomendación es no tratar de conseguir demasiados datos con una sola pregunta. En vez de ello diversifique el número de estas.

1.11. *No sugiera la respuesta dentro de la pregunta*

Preguntas tendenciosas pueden llevar a orientar las respuestas, situación que debe evitarse. Cuestionamientos como: ¿Hasta dónde las políticas institucionales están condicionando el bajo desempeño laboral? Están sostenidas por premisas que el contestante debe aceptar antes de emitir un juicio al respecto; para este caso se asume que las políticas están condicionando un bajo desempeño laboral, lo cual implica contar con la evidencia para ello y, sobre todo, que el contestante la conozca.

1.12. Ordene las preguntas en una forma lógica (deductiva)

No «salte» arbitrariamente de un tema a otro. Mantenga la unidad de pensamiento, de tal forma que no se muestren cambios bruscos en las temáticas abordadas. El principio radica en estructurar un test que permita un interrogatorio fluido, sin tener que regresar a temas ya tratados ítems atrás. En este sentido es necesario no incluir dentro del test hechos o sucesos que no guarden relación lógica.

1.13. Considere el tiempo que llevará contestar el cuestionario

Este aspecto es particularmente subjetivo, ya que para tener una idea más precisa es necesario pilotear el instrumento, a fin de identificar tiempos mínimos y máximos para su completa solución. Una recomendación que puede servir de guía es asignar a los sujetos el doble de tiempo que le lleva a un investigador contestar el test.

1.14. No asuma como entendibles y válidas las preguntas formuladas

Lo apropiado es "pilotear" el instrumento; de no ser posible, al menos remítalo a un grupo de expertos para su revisión, la estrategia más empleada es el método Delphi, el cual consiste en una revisión secuenciada de un test, donde un grupo de expertos revisa el diseño del instrumento y aporta sus observaciones al investigador, quien acopia todas las sugerencias y rediseña el test, para luego reenviarlo a los mismos expertos, hasta completar de tres a cinco rondas. El propósito del diseño sucesivo es disminuir el espacio intercuartil, esto es, estimar cuánto se desvía la opinión del experto de la opinión del conjunto, calculando la mediana de las respuestas obtenidas. Para mayores detalles consúltese a Reguant-Álvarez y Torrado-Fonseca (2016) y López-Gómez (2018).

1.15. Asegúrese de aplicar el cuestionario en el mejor momento posible, a fin de evitar introducir más variación al estudio

No debe subestimarse el contexto de aplicación de los test. El espacio físico, el horario de aplicación, la condición del inmueble, los materiales empleados, entre otros aspectos, pueden modificar la respuesta que las personas otorguen al test. Espacios muy reducidos, horarios inapropiados –muy tarde o muy temprano-, mobiliario inadecuado, insuficiente o en mal estado, son condiciones que deben evitarse.

1.16. Tome el tiempo necesario para capacitar a los aplicadores

Toda evaluación puede sesgarse, e incluso perderse, por un mal manejo del proceso de aplicación. La capacitación de los aplicadores del test es clave en este sentido. Hay que tomarse el tiempo para entrenarlos, aclarar sus dudas, plantearles las posibles consultas que pueden surgir en el momento de la aplicación y establecerse los criterios que normarán el proceso.

1.17. Considere incluir preguntas «ambientadoras»

Lo que se busca es ganar la confianza del interrogado. Estas pueden no ser de interés para la investigación, pero facilitan la resolución del cuestionario. En muchas ocasiones sirve de enlace o de sustento para el entendimiento lógico de los ítems siguientes.

Mención aparte merecen las preguntas generales, tales como el sexo, edad, nivel educativo y procedencia del contestante, conocidas como «variables de agrupación», útiles para segmentar la información en el manejo estadístico de los datos.

CONCLUSIONES

Sin dejar de reconocer que el diseño de instrumentos y test de medición/evaluación es un campo ampliamente explorado por los psicológicos experimentales y clínicos, es necesario reconocer que para muchos profesionales de las ciencias humanas y de la conducta el ámbito de las escalas y sus métricas es poco conocido. Esto se observa al revisar los numerosos trabajos que actualmente se publican en revistas y se presentan dentro de los congresos de investigación, donde basta con señalar la construcción de un test para suponer de facto la validez del estudio presentado. Para este caso es indispensable considerar que incluso “si un test es válido para un determinado grupo o población, no necesariamente lo será para otros” (Covacevich, 2014, p.8).

Pero transformar una variable de atributo o cualitativa en una cuantitativa es un proceso que reclama la mayor seriedad posible. Conceptos o constructos como calidad, actitud, valoración, motivación, liderazgo y experiencia tienen que ser operacionalizados con el propósito de asignarles un referente numérico que permita su análisis estadístico y una interpretación que atienda a estándares previamente establecidos, que no permita el sesgo a partir de las apreciaciones del investigador.

Si bien existe la posibilidad de acudir a test estandarizados, probados en ambientes muy específicos -y por tanto situados-, se debe reconocer la necesidad de validarlos y adaptarlos a situaciones particulares, ya que la varianza del error tiende a ser distinta en condiciones diferenciadas, por ejemplo, si se piensa en «creencias religiosas», no es lo mismo pensarlas en contextos culturales anglosajones que latinos, por lo que el test empleado puede variar en sus resultados. De esta forma el margen de error incluso puede variar; “en este contexto ‘error’ se refiere a la suposición de que factores distintos al que pretende medir la prueba influirán en el desempeño de ésta” (Mikulic, s/f, p. 11).

Una forma de minimizar el impacto de estos errores –y lo sesgos que suponen- es complementar los resultados del test, es decir, no centrar la validez del estudio solo en aquello que derive del

test, ya que existen diversas técnicas e instrumentos para abordar el estudio de un constructo. Aplicar más de un test puede ser otra medida recomendable.

En todo caso, tal como lo expresa Mikulic, (s/f):

> Si se quieren usar pruebas adecuadas se deberá tener en cuenta: cómo se elaboró la prueba, las condiciones para su aplicación, cómo y a quién se debe administrar, cómo deberían interpretarse los resultados de la prueba y a quienes, y cuál es el significado de la puntuación. Ello implica conocer las limitaciones de las pruebas y compensarlas con datos de otras fuentes. (p. 11)

Al final, incluso aplicando todos los criterios aquí señalados, no existe un test o cuestionario -diagnóstico o de evaluación-, que mida al 100% lo que dice medir. La teoría de los test así lo señala, y la estadística, con sus niveles de significancia, se encarga de confirmarlo. De esta manera, el principio de validez y confiabilidad descansa en el logro de una objetividad que desde la ciencia se asume relativa, ya que los objetos de estudio se entienden a partir de un referente metafísico que los explica a través de sus cualidades –características-, identificables por el investigador, quien hace de estos atributos algo inteligible, entre otros medios mediante el diseño, empleo e interpretación de los test. Se trata ni más ni menos que de una «subjetividad objetivada», que remite a una especie de «correlato del objeto de estudio». Esto es ampliamente documentando en la teoría científica, pero ello no anula la utilidad de los test como instrumento de medición, siempre que mediante ellos se evite que lo estudiado sea enteramente dependiente del observador.

En conceptos de Tristán-López y Pedraza-Corpus (2017), la objetividad se alcanza mediante dos aspectos fundamentales: la especificidad y la interpretación. En este sentido, "la especificidad es la representación de la realidad, contenida en una definición completa, pertinente, precisa del objeto y que lo distingue de otros" (p.14). Mientras que la interpretación "se asocia con las justificaciones de los usos y juicios de valor que pueden postularse a nivel contextual, cultural, grupal, o de otra índole, a partir de datos obtenidos de la realidad" (p.15). Destaca además que las interpretaciones y justificaciones "responden a la necesidad de identificar, prevenir, medir y, de preferencia, eliminar o reducir al mínimo la presencia de sesgos en las apreciaciones de las personas que van a emitir juicios de valor sobre los objetos en estudio (p.15). De aquí la necesidad de estar atentos a los problemas de diseño ya indicados anteriormente.

Mención aparte merecen las consideraciones éticas que deben atenderse por quien diseña, aplica y valora un test. Si bien es cierto que "la conducta moral del investigador dependerá, en gran medida, de su propio desarrollo moral y de las circunstancias en que se dé su acción investigativa" (Perales 2010, p.439), es indispensable reconocer que existen principios éticos universales, consagrados en documentos rectores, como lo es Ethical Principles of Psychologists and Code of Conduct, de la American Psychological Association (2018), que orientan la conducta de los investigadores, psicólogos en particular, normas que deben seguirse en busca de no perder el sentido humano y social de toda investigación.

A manera de cierre, se debe enfatizar en que entender los postulados que gobiernan la ciencia significa reconocer sus alcances y limitaciones. No se trata de caer en posturas radicales, ajenas al verdadero sentido de búsqueda/descubrimiento/construcción del conocimiento. En todo caso la ciencia encuentra en el camino que toma para evitar el dogmatismo, la vía de acceso a él. Bunge (1983) lo deja claro al sentenciar que "si la sustancia (objeto) no puede ser lo distintivo de toda la ciencia, entonces tiene que serlo la forma (el procedimiento)" (p. 22). El diseño de los test

científicos se encuentra inmerso en este enfoque epistemológico, absolutismo científico heredado de la modernidad.

REFERENCIAS BIBLIOGRÁFICAS

American Psychological Association (APA). (2018). *Ethical Principles of Psychologists and Code of Conduct*. Recuperado de: http://www.apa.org/ethics/code/index.aspx

Arias-Beatón, G. (1991). La medición en psicología: una reflexión impostergable. *Revista Cubana de Psicología*, 8 (2-3), 93-100. Recuperado de: http://pepsic.bvsalud.org/pdf/rcp/v8n2-3/02.pdf

Bachelard, G. (2000). *La formación del espíritu científico*. Buenos Aires: Siglo XXI Editores, S.A. de C.V.

Bourdieu, P., Chamboredon, J. C. y Passeron, J. C. (1975). *El oficio de sociólogo*. Buenos Aires: Siglo XXI editores.

Bunge, M. (1983). *La Investigación Científica. Su Estrategia y su Filosofía.* Barcelona: Editorial Ariel, S.A.

Cervantes, V. H. (2005). Interpretaciones del coeficiente Alpha de Cronbach. *Avances en Medición*, 3, 9-28. Recuperado de: http://www.humanas.unal.edu.co/psicometria/files/1113/8574/8604/Articulo_1_Alfa_de_Cronbach_9-28_2.pdf

Cohen, R. J. & Swerdlik, M. E. (2001). *Pruebas y evaluación psicológica. Introducción a las pruebas y a la medición*. México: McGrawHill.

Corral, Y. (2009). Validez y confiabilidad de los instrumentos de investigación para la recolección de datos. *Revista ciencias de la educación*,19 (33), 228-247. Recuperado de: http://servicio.bc.uc.edu.ve/educacion/revista/n33/art12.pdf

Covacevich, C. (2014). *Cómo seleccionar un instrumento para evaluar aprendizajes estudiantiles*. Banco Interamericano de Desarrollo. Recuperado de: https://publications.iadb.org/bitstream/handle/11319/6758/C%C3%B3mo-seleccionar-un-instrumento-para-evaluar-aprendizajes-estudiantiles.pdf

Fernández, A., Pérez, E., Alderete, A. M., Richaud, M. C. & Fernández, L. M. (2010). ¿Construir o Adaptar Tests Psicológicos? Diferentes Respuestas a una Cuestión Controvertida. *Evaluar*, 10, 60-74. Recuperado de: file:///C:/Users/Manuel/Downloads/459-1500-2-PB%20(3).pdf

García-Garro, A. J., Ramos-Ortega, G., Díaz de León-Ponce, M. A. y Olvera-Chávez, A. (2007). Instrumentos de evaluación. *Revista Mexicana de Anestesiología*, 30 (3), 158-164. Recuperado de: http://www.medigraphic.com/pdfs/rma/cma-2007/cma073f.pdf

Gempp-Fuentealba, R. (2006). El error estándar de medida y la puntuación verdadera de los tests psicológicos: Algunas recomendaciones prácticas. *Terapia Psicológica*, 24 (2), 117-130. Recuperado de: http://www.redalyc.org/pdf/785/78524201.pdf

Gómez-Benito, J., Hidalgo, M. D. y Guilera, G. (2010). El sesgo de los instrumentos de medición. Tests justos. *Papeles del Psicólogo*, 31 (1), 75-84. Recuperado de: http://www.redalyc.org/pdf/778/77812441008.pdf

López-Gómez, E. (2018). El método Delphi en la investigación actual en educación: una revisión teórica y metodológica. *Educación XX1*, 21(1), 17-40. DOI: 10.5944/educXX1.15536

López-Mezquita, M. M.T. (2007). *Evaluación de la competencia léxica: Test de vocabulario. Su fiabilidad y validez*. Granada: Ministerio de Educación y Ciencia.

Merino-Soto, C. y Lautenschlager, G. J (2003). Comparación Estadística de la Confiabilidad Alfa de Cronbach: Aplicaciones en la Medición Educacional y Psicológica. *Revista de Psicología de la Universidad de Chile*, 12 (2), 127-136. Recuperado de: file:///C:/Users/Manuel/Downloads/17668-52156-1-PB.pdf

Mikulic, M. I. (s/f). *Construcción y adaptación de pruebas psicológicas.* Argentina: Universidad de Buenos Aires-Facultad de Psicología. Recuperado de: http://23118.psi.uba.ar/academica/carrerasdegrado/psicologia/informacion_adicional/obligatorias/059_psicometricas1/tecnicas_psicometricas/archivos/f2.pdf

Pedrosa, I., Suárez-Álvarez, J. y García-Cueto, E. (2013). Evidencias sobre la validez de contenido: avances teóricos y métodos para su estimación. *Acción Psicológica*, 10 (2), 3-18. Recuperado de: http://revistas.uned.es/index.php/accionpsicologica/article/view/11820/12588

Perales, A. (2010). Reflexiones sobre ética de investigación en seres humanos. *Rev Peru Med Exp Salud Pública*, 27 (3), 438-42. Recuperado de: http://www.scielo.org.pe/pdf/rins/v27n3/a18v27n3

Pérez-Gil, J. A., Chacón-Moscoso, S. y Moreno-Rodríguez, R. (2000). Validez de constructo: el uso de análisis factorial exploratorio-confirmatorio para obtener evidencias de validez. *Psicothema*, 12 (2), 442-446.

Pérez-López, C. G. (2008). La estadística como herramienta en la investigación psicológica: un estudio exploratorio. Tesis de Maestría, Universidad de Granada, España. 92 p. Recuperado de: http://www.ugr.es/~batanero/pages/ARTICULOS/Cuauhtemo.pdf

Prieto, G.; Delgado, A. R. (2010). Fiabilidad y Validez. *Papeles del Psicólogo*, 31 (1), 67-74. Recuperado de: http://www.redalyc.org/pdf/778/77812441007.pdf

Reguant-Álvarez, M. y Torrado-Fonseca, M. (2016). El método Delphi. Universitat de Barcelona. Institut de Ciències de l'Educació. *REIRE*, 9 (1). DOI:10.1344/reire2016.9.1916//

Romero, E. (2011). Confiabilidad y validez de los instrumentos de evaluación neuropsicológica. *Subjetividad y Procesos Cognitivos*, 15 (2), 83-92. Recuperado de: http://www.redalyc.org/pdf/3396/339630258004.pdf

Sánchez-Rubio, A. (2016). El sistema contradictorio como método de investigación importado de la ciencia: una novedosa configuración de la prueba científica. *Revista internacional de Estudios de Derecho Procesal y Arbitraje*, 2, 1-34. Recuperado de: http://www.riedpa.com/COMU/documentos/RIEDPA21603.pdf

Soler-Cárdenas S. F. y Soler-Pons, L. (2012). Usos del coeficiente alfa de Cronbach en el análisis de instrumentos escritos. *Rev Méd Electrón* [Internet]. Recuperado de: http://www.revmatanzas.sld.cu/revista%20medica/ano%202012/vol1%202012/te ma02.htm

Soriano, A. M. (2014). Diseño y validación de instrumentos de medición. Diálogos 14, 19-40. Recuperado de: http://www.redicces.org.sv/jspui/bitstream/10972/2105/1/2%20disenoyvalidacion_dialogos14.pdf

Tello, C. (2011). El objeto de estudio en ciencias sociales: entre la pregunta y la hipótesis. *Cinta Moebio*, 42, 225-242. Recuperado de: www.moebio.uchile.cl/42/tello.html

Tristán-López, A. y Yahibé-Pedraza, C. N. (2017). La Objetividad en las Pruebas Estandarizadas. *Revista Iberoamericana de Evaluación Educativa*, 10 (1), 11-31. DOI: https://doi.org/10.15366/riee2017.10.1.001

Van Zyl, J. M., Neudecker, H., & Nel, D. G. (2000). On the distribution of the maximum likelihood estimator of Cronbach's alpha. *Psychometrika*, 65, 271-280.