

Validez y confiabilidad en instrumentos de investigación: una mirada teórica***Validity and reliability in research instruments: a theoretical look*****Yadira Corral de Franco**<https://orcid.org/0000-0003-2236-1328>

Facultad de Ciencias de la Educación.

Universidad de Carabobo. Valencia, Venezuela

yjcorral@gmail.com**Resumen**

En esta disertación se pretende retomar un tema ya expuesto por la autora en otros de sus escritos, en tal sentido, la presente revisión documental se propone orientar a los nuevos investigadores en los procedimientos y toma de decisiones para estimar validez y confiabilidad en instrumentos de recolección de información que proporcione datos relevantes en pro del logro de objetivos de investigación planteados. Se abordan algunos tipos de técnicas con este propósito, diversos procesos de validación, se describen ciertos métodos para estimar la confiabilidad de estos instrumentos y cuáles de ellos necesitan solamente la validación de expertos y aquellos que requieren ambas estimaciones. Se considera importante garantizar la validez, confiabilidad y objetividad de los instrumentos, pues cuando no se establecen éstas siempre habrá dudas sobre la veracidad de las evidencias y la calidad de los resultados y conclusiones derivados del análisis e interpretación surgidos a partir de la data recabada en la investigación.

Palabras clave: Validez, Confiabilidad, Instrumentos de Investigación.

Abstract

In this dissertation it is intended to return to a topic already exposed by the author in other of her writings, in this sense, the present documentary review intends to guide the new researchers in the procedures and decision making to estimate validity and reliability instruments of information that provides relevant data in favor of achieving the proposed research objectives. Some types of techniques will be addressed for this purpose, various validation processes, certain methods will be known to estimate the reliability of these instruments and their suitability requires only the validation of experts and those that require both estimates. It is considered important to guarantee the validity, reliability and objectivity of the instruments, because when these are not determined there will always be doubts about the veracity of the evidence and the quality of the results and conclusions derived from the analysis and interpretation arising from the data collected in the investigation.

Keywords: Validity, Reliability, Instrument Research.

Recibido: 27/06/2021**Enviado a árbitros:** 28/06/2021**Aprobado:** 12/04/2022

Introducción

La estabilidad y confianza en los resultados es el norte de la investigación científica y académica; he allí la importancia de contar con instrumentos bien contruidos que garanticen la pertinencia, calidad y veracidad de la información recolectada (sin sesgos) y midiendo con propiedad los aspectos a indagar. Esto es cierto, sea cual fuere el paradigma seleccionado.

En investigaciones cualitativas y sociocríticas, el énfasis se centra en el grado de credibilidad (validez interna), transferibilidad (validez externa) y consistencia (fiabilidad) de los hallazgos; donde se describa y analice la situación lo más ajustado a la realidad y los protagonistas se vean reflejados en esos hallazgos. Pero, en investigación cuantitativa o positivista es vital que los instrumentos registren los datos de manera tal que puedan ser procesados, organizados y analizados con procedimientos estadísticos apropiados al tipo específico del dato adquirido. Por tanto, la validez y confiabilidad de los mismos descansa en aspectos específicamente técnicos; al cumplirse cabalmente la construcción de los mismos. Sin embargo, no se descarta la utilización de algunos ya estandarizados, cuando cubren nuestras necesidades indagatorias.

Vale destacar, la problemática presentada usualmente en la recolección de datos (Corral, 2009) en investigaciones científicas y académicas se centra con cierta frecuencia en la construcción de instrumentos que permitan recabar información pertinente, veraz, válida, confiable y objetiva. De ello dependerá el valor científico de un estudio, dotando a la indagación de una base real, verídica y de calidad al producto investigativo.

Con este propósito, se amerita saber si el o los instrumentos diseñados o adoptados proporcionarán información realmente útil y fidedigna sobre el evento, situación o problemática abordada. Para tal fin, se deben realizar consideraciones específicas principalmente relacionadas con la validez, objetividad, confiabilidad y utilidad práctica, de estas herramientas.

En tal sentido, la *validez* de una investigación descansa en el grado en que refleja la realidad, en lo fidedigno de la información aportada por los datos recabados. La *objetividad* está

asociada a la presencia de sesgos, tendencias o influencias de los investigadores. La *confiabilidad* o *fiabilidad* corresponde a la exactitud y precisión de las técnicas, instrumentos y procedimientos empleados en la recogida de los mismos, su consistencia y su reproductividad o replicación. Y, la *utilidad práctica* se relaciona con la conveniencia del uso de un determinado instrumento, si es muy oneroso o no, adecuación, aceptación, entre otros aspectos a considerar.

Partiendo de lo anterior, el presente artículo se propone orientar a los noveles investigadores en los procedimientos para estimar validez y confiabilidad en instrumentos de recolección de información que proporcione datos relevantes en pro del logro de objetivos de investigación planteados. Para ello, se realizó la revisión de una diversidad de textos sobre la temática, dirigidos a la toma de decisiones adecuadas al momento de seleccionar las técnicas apropiadas y elaborar instrumentos ajustados al propósito indagatorio.

Sobre el Registro de la Información

Al momento de la recolección de información nos enfrentamos con la disyuntiva de qué técnica e instrumentos son pertinentes, según el tipo de investigación a emprender y conforme al diagnóstico realizado de la situación, fenómeno o evento problemático de interés. “Este diagnóstico inicial es el que va a arrojar las características del método a emplear y los procedimientos subsecuentes” (Corral, 2017, p. 197). Las características propias de cada uno de los tipos y diseños de investigación, influirán en la manera de recabar evidencias. Por tanto, es imprescindible conocer las posibles técnicas adecuadas al objetivo de un estudio; conforme a la procedencia (grupo, escenarios, tamaño de la población, etcétera) y naturaleza de los datos.

Ahora bien, ¿qué es un dato?, “la información, inherente al campo real, deviene en dato. Entendemos el dato como representación –teórica- de aquella realidad que fuera interpelada. La fiabilidad del dato es teórica y metodológica” (Cohen & Gómez, 2019, p. 32). Pues, una evidencia u

observación es información que se convierte en dato cuando el hecho, característica, rasgo, indicador o evento es registrado e interpretado. En efecto, producir datos cualitativos o cuantitativos implica “la sinergia entre la interpelación de la realidad en la que trabajamos y la representación de la información en dato” (ibídem). Para Abero (2015), el dato es un aspecto presente en la realidad.

Por tanto, las técnicas de recolección de información son procedimientos estandarizados para la recopilación y registro de evidencias pertinentes que se convierten en datos a analizar. Estas técnicas se constituyen en directrices metodológicas, permitiendo sistematizar y ordenar la recopilación de información pertinente a la investigación; integradas por las diversas formas a emplear en la obtención de la data, con miras al logro de los objetivos de investigación.

En cuanto a los instrumentos usados, éstos se derivan de las variables, incertidumbres, preguntas e hipótesis en estudio; deben ser reflejo de los objetivos de investigación y las bases teóricas que sustentan la indagación, como producto resultante de la intersección entre el problema, la contextualización y los referentes teórico/conceptuales.

El instrumento de registro, es el recurso que tenemos para intervenir en la realidad a partir de nuestros conceptos. Es el intermediario entre nuestras hipótesis, nuestras referencias teóricas y el campo empírico. Según sea nuestro abordaje, cuantitativo o cualitativo, así será: más o menos estructurado, más cerrado o más flexible. (Cohen & Gómez, 2019, p. 251)

Conforme a ello, investigaciones que requieren del registro de información cualitativa necesitan usar técnicas flexibles y abiertas; pero, los estudios cuantitativos ameritan técnicas para cuantificar los indicadores a observar. Datos cualitativos dan importancia a la descripción de contextos (categorías descriptivas) para comprender el comportamiento humano a partir de perspectivas internas y los datos cuantitativos son producto de mediciones (escalas numéricas continuas o discretas, de razón o intervalo), jerarquizaciones (escalas ordinales) y categorías descriptivas mensurables (escalas nominales y/o escalas categoriales).

Tipos de Técnicas e Instrumentos Asociados al Registro de Información

Previo a la recolección de datos, se recomienda precisar cuáles son las técnicas y procedimientos a usar y las condiciones como se recabarán los datos. En estudios con diseño documental, entre las técnicas de registro (Corral, Corral & Franco, 2019; Hernández Sampieri, Fernández & Baptista, 2014; Montero & Hochman, 2005; Van Dijk, 2005) se tienen:

- a) *Análisis documental*, discrimina fuentes inadecuadas;
 - b) *Presentación resumida*, resumen de ideas principales contenidas en los escritos;
 - c) *Resumen analítico*, devela la infraestructura del texto desglosando los elementos claves y su esqueleto básico;
 - d) *Análisis de contenido o de información documental*, evalúa mensajes subyacentes y sus significados, se vale de otras técnicas como la presentación resumida y el resumen analítico. Esta técnica incluye el análisis crítico de discurso (ACD) cuando se quiere establecer ideologías subyacentes en documentos escritos, orales, visuales o audiovisuales.
 - e) *Extracción de datos numéricos y/o estadísticos*, para ser procesados estadísticamente.
- Técnicas operativas e instrumentos apropiados: fichaje (fichas), subrayado (fotocopias, hojas impresas o resaltado electrónico), mapeo (elaboración de mapas), hojas de registro, matrices, ampliación de textos, otros.

En trabajos teórico-prácticos y de campo, la información teórica se obtendrá a través de la revisión documental y el registro de la información de campo empleará técnicas adecuadas al paradigma o enfoque, el tipo y diseño de investigación (cualitativo, cuantitativo o mixto) en los que se circunscriban.

Para investigaciones de naturaleza cualitativa naturalista, donde prevalece la descripción de los contextos, las técnicas más empleadas (Abero, 2015; Aravena, Kimelman, Micheli,

Torrealba & Zúñiga, 2006; Cabezas, Andrade & Torres, 2018; Corral, 2014, 2017; Corral et al., 2019; Hernández Sampieri et al., 2014) son las siguientes:

- *Observación directa no sistematizada.* Realizada a través de la observación libre no estructurada. *Instrumentos:* mapas, croquis, notas crudas, diarios de campo, bitácoras, registros anecdóticos, registros descriptivos, fichas de campo, listas de frecuencia, etc.
- *Entrevista cualitativa.* Incluye la entrevista individual (semiestructurada y en profundidad) y grupal (juegos de roles, grupos de discusión, focus groups, otros), y análisis de lenguaje. *Instrumentos:* soportes mecánicos para grabar y tomar imágenes (videos, películas, audios y fotografías), notas de campo (bitácoras, diarios de campo, son algunos de ellos) para plasmar gestos y/o posturas (lenguaje gestual) que complementen la comprensión del texto oral en grabaciones de audio. Se realiza cara a cara o por videoconferencia.
- *Historia de vida.* Relatos sobre la memoria personal y/o colectiva, testimonios y autobiografías sobre una historia colectiva o fenómeno de trascendencia social. Se emplean los mismos soportes usados en la entrevista.
- *Observación indirecta.* Revisión de documentos, entre ellos: escritos (autobiografías, diarios personales, cartas, mensajes, otros), narraciones, fotografías, videos y audios.
- *Encuesta etnográfica.* Se emplean cuestionarios de respuestas abiertas, para ser aplicados a un grupo y conocer sus percepciones sobre la problemática en estudio.

Al realizar investigación cuantitativa (Abero, 2015; Aravena et al., 2006; Cabezas et al., 2018; Corral et al., 2019; Hernández Sampieri et al., 2014), existe una variedad de técnicas de investigación a utilizarse, como son:

- *Observación directa sistematizada.* A través de la observación estructurada. Los instrumentos pueden ser individuales y/o grupales y se emplean formatos de registro estructurados: listas de cotejo, escalas de estimación (numéricas, gráficas y

descriptivas), rúbricas (holísticas o globales y analíticas), inventarios, cuadernos de protocolo, historias clínicas y guías de observación. Algunos de ellos pueden contener ítemes abiertos, cerrados o ambos.

- *Observación indirecta.* Asistida técnicamente por dispositivos mecánicos para captar eventos, fenómenos o situaciones de interés. Se sistematiza esta observación empleando instrumentos estructurados al igual que la observación directa.
- *Lista de frecuencia.* Permite registrar la cantidad de veces que se evidencia un indicador (característica, rasgo, conducta, actitud, situación, otros). También es usada en investigación cualitativa.
- *Entrevista estructurada.* Se realiza empleando como instrumentos de registro una guía (guion) de entrevista o un protocolo, con preguntas preestablecidas de respuestas usualmente cerradas o mixtas, donde se registran las respuestas. Sin embargo, se recomienda hacer la grabación de la misma sea en audio o video. Este tipo de entrevista puede realizarse cara a cara, por teléfono, videollamada o videoconferencia.
- *Encuesta.* Posibilita consultar simultáneamente a un grupo de personas sobre una temática particular. *Instrumentos:* cuestionarios y escalas de clasificación o medición (actitudinales y opináticas). Los cuestionarios pueden ser cerrados (respuestas dicotómicas y policotómicas) y semicerrados; autoadministrados (presenciales, por correo, e-mail o mensajería de texto) y verbales (por teléfono, videollamada o videoconferencia); también es posible formular preguntas con alternativas o proposiciones con opciones fijas de respuesta. Las escalas miden actitudes u opiniones sobre un aspecto determinado, son de naturaleza cerrada por lo que se deberá escoger una respuesta entre una serie de opciones preestablecidas; traducidas en proposiciones y alternativas de respuesta a las mismas. Las escalas más usadas son: Likert, Diferencial semántico, Stapel, entre otras reconocidas.

- *Prueba de rendimiento.* Posibilita medir el nivel de conocimiento sobre contenidos educativos. Se emplean sobre todo en el diagnóstico del rendimiento estudiantil. Según la forma de responderlas, se clasifican en pruebas: escritas (de papel y lápiz), orales (de base no estructurada, semiestructurada y estructurada) y prácticas de ejecución o funcionales (de proceso y de producto). Entre las escritas se tienen, pruebas: de ensayo, de respuesta guiada, de respuesta breve, objetivas (con preguntas de opciones múltiples, ordenamiento, verdadero y falso, pareo, otras).

Existen otros instrumentos para la recolección de información cuantitativa como los tests, datos numéricos y estadísticos extraídos de documentos, hojas de registros numéricos (planillas de notas, exámenes de laboratorio, etc.), son algunos de ellos.

Validez

En estudios de corte cualitativo y sociocrítico, el enfoque del tratamiento de la información es más abierto que en los cuantitativos, se le da un mayor protagonismo a los sujetos investigados y los instrumentos a usar deben tener una mayor flexibilidad. Para la investigación, en cualquiera de los paradigmas, los constructos de validez y confiabilidad son vitales, constituyen estándares indispensables de rigor científico. Así, Corral (2017) diserta:

Esto supone que el valor de una investigación, dependerá del grado de fidelidad de la descripción y análisis del evento, fenómeno o situación problemática abordada; implica que cobra mayor relevancia cuando los hallazgos se apegan a lo que acontece y a los contextos estudiados. (p. 197)

En *investigación cualitativa*, este rigor, se plasma en resultados y hallazgos que describan y analicen la realidad estudiada lo más fidedignamente posible. Alude a su credibilidad (validez interna), asociada al valor de verdad referida a la forma cómo se realizan las observaciones y conversaciones con los informantes, se realiza su registro y al reconocimiento de

los hallazgos por el grupo de participantes (Castillo & Vásquez, 2003). Otro aspecto se relaciona con la transferibilidad (validez externa) de los resultados. Así:

- **Validez interna** (valor de verdad, credibilidad). Requiere lograr la *validez de la observación*, categorizando al máximo y definiendo claramente las categorías a observar y la *validez de los informantes* se obtiene al presentar los datos aportados a éstos para que corroboren su veracidad y corregir posibles errores en el registro e interpretación de las respuestas. Para asegurar la credibilidad, el investigador cualitativo necesita (Rojas de Escalona, 2014): (a) estadía prolongada en el escenario (contexto societal); (b) tomar perspectiva separándose periódicamente del escenario; (c) observación sistemática persistente, detallada y amplia (saturación); (d) triangular la información cualitativa obtenida (cruzar datos); (e) analizar los casos negativos; (f) rechequear la información con los informantes para corregir posibles malentendidos, confirmar datos o añadir información; (g) registrar datos de las acciones e interacciones; entre otras posibles actividades.
- **Validez externa** (transferibilidad). Corresponde a la aplicabilidad en otros grupos. Moreira (citado en Corral et al., 2019) distingue los siguientes elementos a incluir en el informe para garantizar ésta: aserciones empíricas, viñetas narrativas, citas de las notas tomadas en el campo, citas extraídas de las entrevistas, sinópticos de datos (tablas comparativas, mapas, tablas de frecuencia, otros), comentarios interpretativos en las descripciones particulares y general (patrones), discusión teórica e informe sobre la historia natural de la indagación realizada. El grado de esa aplicabilidad está sujeto a la similitud (Castillo & Vásquez, 2013; Hernández Sampieri et al., 2014) entre contextos distintos y referentes teóricos establecidos en la investigación.

En relación con la *validez en investigaciones cuantitativas*, la validez se refiere al grado en que un instrumento proporciona información con fidelidad de la población, relacionado con el atributo a

medir, su autenticidad (fidedigna y veraz). Al establecer ésta, se precisa conceptualizar la variable criterio, saber cuáles son los indicadores (rasgos, características, categorías, factores) a estudiar. Destacan tres tipos de validez (Ruiz Bolívar, 2013): de contenido, de constructo y predictiva.

Validez de Contenido, racional o lógica

Es el procedimiento más empleado en investigación académica. Está referido al grado en que los ítemes o reactivos de un instrumento reflejan el universo de contenidos representativos de los indicadores (características, rasgos, factores, categorías) a ser ponderados; es decir, deben ser pertinentes según la descripción de la variable en cuestión. Ejemplo: si se mide la actitud hacia la matemática del estudiante, los ítemes deben referirse únicamente a esa variable y no a otra. Se analizará el contenido de acuerdo con los referentes teóricos (marco teórico) sustentadores de la investigación, acordes con la interpretación particular del investigador. El instrumento debe ser un compendio de ítemes o tareas representativos del dominio de contenido y de las variables en estudio, sus respectivas dimensiones e indicadores (extraídas de la teoría).

Elaborado el instrumento, esta validez se establece a través del denominado **Juicio de Expertos**. Se escoge un grupo impar de expertos (normalmente de tres o cinco) para verificar y certificar la claridad y coherencia de las preguntas, tareas, reactivos y/o proposiciones seleccionadas, sin sesgos que afecten las respuestas de los sujetos informantes. Esta validación no se expresa usualmente de manera cuantitativa, es más bien una valoración subjetiva (o intersubjetiva) de su fidelidad; sin embargo, algunos métodos sí cuantifican esa valuación (Tabla 1). Entre esos métodos usados (Corral, 2009, 2014; Palella & Martins, 2012; Ramírez, 2010; Ruiz Bolívar, 2013), grupales o de experto único, se encuentran:

- *Método de Agregados Individuales*. Es el más popular y económico, se solicita a cada experto una estimación directa sobre el contenido de los ítemes del instrumento, de

manera individual; no necesita la reunión de los expertos. Busca evitar sesgos en los reactivos ocasionados por posibles conflictos interpersonales, presiones u otros factores que desvirtúen el propósito del instrumento. *Procedimiento:*

- Se seleccionan al menos tres expertos o jueces para valorar de forma independiente - regularmente de manera cualitativa- la congruencia, pertinencia y coherencia de los reactivos con las bases teóricas, así como la claridad y ajuste al lenguaje de la población estudiada en la redacción de cada ítem, pulcritud del instrumento y si existe o no sesgo o tendenciosidad en su formulación (si sugieren o no una respuesta).
- A cada experto se le suministrará información suficiente sobre: propósito de la investigación (objetivos) y/o del instrumento, conceptualización del universo de contenido teórico, tabla de operacionalización de variables (de especificaciones o matriz de operacionalización de constructos del instrumento) y un formato de validación que contenga: congruencia ítem-dominio de contenidos teóricos, relación con los objetivos y las variables, claridad en la redacción, lenguaje adecuado al nivel de la población en estudio, tendenciosidad o sesgo, entre otros aspectos. (Tabla 1)
- Se analizan los formatos de validación y se toma la decisión de la inclusión, modificación o exclusión de los ítemes en la versión final del instrumento, conforme a los siguientes criterios, los ítemes con: (a) 100% de coincidencia favorable entre los expertos se incluyen; (b) 100% de coincidencia desfavorable se excluyen o sustituyen, y (c) coincidencia parcial u observaciones (sobre congruencia, tendenciosidad, claridad, presentación, etc.) se revisan, modifican o reformulan.
- En caso de modificaciones, se somete el instrumento a una nueva validación.
- *Método de Consenso Grupal.* Reunidos los expertos (3 a 5), se les instruye para estimar cuantitativamente la pertinencia de los ítemes del instrumento y otros aspectos

relevantes contenidos en el formato de validación. Se realizan intercambios de opiniones e información entre ellos en la búsqueda de consensos. Si no se logra consensuar, se recogen las estimaciones individuales y se calculan las medias de los puntajes por ítem. En este método se precisa la libre expresión, la evasión de discusiones tensas y los sistemas de votación.

Tabla 1

Ejemplo de formato para validar instrumentos

Criterios de Valoración: Coloque en la columna correspondiente la valoración de cada ítem, según los siguientes criterios: 5: Muy adecuado; 4 Adecuado; 3: debe mejorarse; 2: Inadecuado; 1: No pertinente							
Ítem	Indicadores					Total	Observaciones (Si debe eliminarse o modificarse un ítem por favor indique como mejorarlo)
	Redacción (claridad y concreción)	Coherencia interna (articulación lógica)	Relación con objetivos de investigación y variables	Lenguaje adecuado (ajustado al nivel del informante)	Mide lo que pretende (congruencia con el dominio de contenidos)		
1							
2							
....							
N							
TOTAL							-----
Aspectos Generales					Sí	No	Observaciones
El instrumento contiene instrucciones claras y precisas para responder							
Contiene ítemes que inducen las respuestas (sesgo)							
Se evidencian en el instrumento las variables del estudio							
Los reactivos se distribuyen de manera secuencial							
Pulcritud en la presentación							
El número de ítemes o reactivos es suficiente para recoger la información necesaria. En caso de ser negativa su respuesta, sugiera los ítemes a añadir							
VALIDEZ							
APLICABLE				NO APLICABLE			
APLICABLE ATENDIENDO A LAS OBSERVACIONES							
Validado por:			C.I.:		Fecha:		
Firma:			Teléfono:		e-mail:		

Nota. Elaborado por la autora

Validez de Constructo

Se refiere a la eficiencia y precisión del instrumento para medir los indicadores de las variables. Un constructo es una construcción del investigador para poder explicar patrones, regularidades o relaciones observables con el propósito de explicar un indicador (rasgo, conducta,

característica, actitud, factor). Es de interés establecer ésta para valorar desempeños, requiere una conceptualización clara de los rasgos a estudiar con sustento teórico definido (Corral, 2009, 2014; Corral et al., 2019; Ruiz Bolívar, 2013). Se usa en tests psicológicos y algunas pruebas educativas, como las pruebas de admisión, cuando miden rasgos generales o cualidades como: razonamiento verbal, sociabilidad, visualización espacial, entre otros; requiere puntuar los resultados. Interpretando a Hernández Sampieri et al. (2014), el procedimiento consta de tres etapas:

- Basándose en la literatura, establecer y especificar la relación entre las variables y conceptos teóricos, contenidos en los constructos.
- Estimar las correlaciones entre los constructos y analizarlas cuidadosamente.
- Interpretar las evidencias empíricas de acuerdo al nivel para clarificar la validez de constructo de una medición particular.

Validez Predictiva o de Criterio Externo o Empírica

Pertinente cuando el propósito es visualizar el desempeño futuro de personas ante tareas o actividades determinadas. Se establece comparando los puntajes arrojados por el instrumento en una prueba piloto, con una o más variables de criterio (variables externas); luego se calcula el coeficiente de correlación y se interpretará como **índice de validez**. Puede valorarse en instrumentos como pruebas escritas de respuesta cerrada estandarizadas, tests y escalas actitudinales estandarizadas. (Ruiz Bolívar, 2013)

El procedimiento básico consiste en aplicar una prueba a un grupo de personas de interés (grupo piloto) y, posteriormente, se obtiene una medida-criterio cuantitativa de éxito de forma individual (esos puntajes necesarios para la correlación se obtienen con otro instrumento que mida la variable criterio); finalmente, se calcula la correlación entre los datos obtenidos en ambas mediciones, de todos y cada uno de los sujetos que conforman el grupo. Cuanta más alta sea la correlación, el valor predictor es mayor. (Ruiz Bolívar, 2013)

Para un solo instrumento y criterio único, lo adecuado es: (a) Coeficiente r de Pearson cuando los datos son numéricos (variables continuas o discretas, escalas de intervalo o de razón) y (b) Coeficiente de correlación simple por rangos ρ de Spearman (variables y escalas ordinales).

Conforme a lo anterior, es de interés para los investigadores dejar establecido que los instrumentos aplicados en su estudio, proporcionan datos fidedignos, valiosos y de calidad. Por ello, ***deberán validarse aquellos instrumentos no estandarizados***, elaborados concretamente para la investigación, entre ellos: guías (guiones) de entrevistas, formatos de registro, guías de observación, listas de cotejo, escalas de estimación, rúbricas, pruebas de rendimiento, cuestionarios, escalas actitudinales u otros, de carácter cualitativo o cuantitativo. Sin embargo, algunos no precisan ser validados como son los usados para entrevistas en profundidad (no emplean guías preestablecidas), las historias clínicas (validadas por el uso), conversaciones libres y observaciones no sistematizadas (notas de campo de observaciones, bitácoras, entre otros).

Objetividad

La **objetividad** se evidencia en la neutralidad o confirmación de los datos y sus resultados (Corral et al., 2019; Ruiz Bolívar, 2013), se relaciona con el sesgo (si existe o no). Corresponde a la neutralidad para confirmar la información y los resultados, la coincidencia de apreciaciones sobre una realidad (evento, atributo, fenómeno, evento, situación). Para reforzar la objetividad es necesario estandarizar los instrumentos, su aplicación y la valoración de resultados.

Confiabilidad

En *investigación cualitativa*, algunos investigadores prefieren referirse a ella como *fiabilidad*, o emplean términos como auditabilidad o fidelidad. Se relaciona (Latorre, del Rincón & Arnal, 2014) con la consistencia (grado en que se repetirían los resultados al replicar el estudio) y conviene verificar que los procesos son aceptables (auditoría). Algunos posibles procedimientos: (a) control doble ciego, por dos o más investigadores sin participación en la recolección de los datos;

(b) reproducción (réplica paso a paso) independiente (comparación de resultados); (c) triangulación teórica; (d) triangulación de datos; (e) triangulación temporal (en diferentes momentos); (f) triangulación espacial; (g) triangulación de personas, al menos en dos de los niveles de relación: agregado (individuos), interactivo (parejas y familias) y colectivo (grupos, organizaciones, comunidades); (h) triangulación metodológica; (i) triangulación de fuentes de información; (j) triangulación múltiple; entre otros.

En *investigación cuantitativa*, la confiabilidad se asocia a la exactitud de la información aportada por los instrumentos; es la capacidad del instrumento de arrojar datos (respuestas o resultados) muy similares en ocasiones repetidas con una misma muestra y condiciones similares. Para instrumentos estandarizados, sus creadores regularmente han comprobado la confianza y estabilidad de los mismos. Sin embargo, como acción preventiva, podemos constatar la confiabilidad en una muestra piloto y corroborar su idoneidad en el nuevo contexto a estudiar.

Si son instrumentos elaborados propiamente para la investigación en curso, resulta imprescindible medir la capacidad de éste para arrojar datos que reflejen lo más fidedignamente posible la realidad de las variables (factores, descriptores o características) asociadas al tema de interés, evento o fenómeno investigado y a los objetivos planteados.

En la estimación de la confiabilidad, se recomienda (Corral, 2009, 2014; Hernández Sampieri et al., 2014; Ruiz Bolívar, 2013) seleccionar una porción de sujetos representativa (pequeña) del universo en estudio (muestra piloto) o un grupo con características similares o equivalente (sobre todo cuando el grupo es muy pequeño, menor a 10 sujetos); *la muestra piloto no formará parte de la muestra definitiva investigada*. Se aplicará el instrumento (prueba piloto) con condiciones lo más aproximadas posibles a las del campo real. Algunos investigadores sugieren: (a) en poblaciones pequeñas (10 a 50 personas) entre 5 a 10 individuos; (b) de 50 a 100 entre 10 a 20 sujetos y (c) en

grupos mayores de 100 aplicarlo a 20 o más. Puede tomarse 10% de la población, si el número es manejable y es no menor de 5.

En cuanto a los *métodos usados en la medición o estimación de la confiabilidad*, a continuación, se reseñan (Corral, 2009, 2014; Hernández Sampieri et al., 2014; Manterola, Grande, Otzen, García, Salazar & Quiroz, 2018) cuatro de los más usados:

- **Método Test-Retest.** Usado en instrumentos de respuestas cerradas sin escala fija, se administra el instrumento dos veces al mismo grupo y se correlacionan los puntajes obtenidos. Inadecuado para pruebas de conocimiento, es apropiado en instrumentos empleados para medir aptitudes físicas, atléticas y motoras. Se obtiene un *coeficiente de estabilidad* (denota coherencia de las puntuaciones en el tiempo).

Tabla 2

Tabla de datos en la correlación de datos numéricos (Coeficiente r de Pearson)

Sujetos	1ª Aplicación-Test		2ª Aplicación- Retest		$X_i Y_i$
	X_i	X_i^2	Y_i	Y_i^2	
1					
2					
...					
ETC.					
Σ					

Nota. Elaborada por la autora

Para escalas numéricas o cardinales (de intervalo o razón), el estadístico a usar es la

Correlación r de Pearson (por puntajes directos), se organizan los datos según la Tabla 2:

$$r_{xy} = \frac{N \sum X_i Y_i - \sum X_i \sum Y_i}{\sqrt{[N \sum X_i^2 - (\sum X_i)^2][N \sum Y_i^2 - (\sum Y_i)^2]}}$$

En donde:

r_{xy} : Coeficiente de correlación de Pearson N : Número de sujetos

X_i : Valores de X (1ª aplicación) Y_i : Valores de Y (2ª aplicación)

$X_i Y_i$: Producto de cada valor X por su correspondiente valor en Y

Al usar escalas ordinales o de rangos, se emplea el ***Coefficiente de correlación de rangos simple o ρ (ro) de Spearman***, la tabla 3 muestra cómo se organizan los datos obtenidos.

$$\rho_{xy} = 1 - \frac{6 \sum d^2}{N(N^2 - 1)}$$

Con:

ρ_{xy} : Coeficiente de correlación ρ de Spearman

N : Número de sujetos

d : Resta de los valores X_i (1ª aplicación) menos los valores Y_i (2ª aplicación)
 $d = (X_i - Y_i)$

Tabla 3

Tabla de datos en una correlación ordinal (Coeficiente ρ de Spearman)

Sujetos	1ª Aplicación-Test	2ª Aplicación- Retest	D ($X_i - Y_i$)	d^2
	X_i	Y_i		
1				
2				
3				
...				
ETC.				
Σ				

Nota. Elaborada por la autora

- **Confiabilidad del Corrector.** Se denomina *coeficiente de estabilidad*, relacionado a la coherencia en la corrección de los formatos. Usado para instrumentos de respuestas abiertas, como cuestionarios sin opciones y pruebas escritas de composición o ensayo.
Procedimiento: (a) se administra el instrumento una sola vez al grupo piloto; (b) dos expertos corrigen los formularios del instrumento, con criterios preestablecidos para su valoración; (c) se tabulan los puntajes de ambas correcciones (Tabla 4) y (d) correlacionan los puntajes obtenidos (o el porcentaje de coincidencias entre los observadores, ítem por ítem cuando se hagan jerarquizaciones) usando el Coeficiente r de Pearson, ya descrito.

Tabla 4

Tabla de datos en confiabilidad de corrector (Coeficiente r de Pearson)

Sujetos	Puntajes otorgados por 1 ^{er} experto		Puntajes otorgados por 2 ^{do} experto		$X_i Y_i$
	X_i	X_i^2	Y_i	Y_i^2	
1					
2					
3					
...					
ETC.					
Σ					

Nota. Elaborada por la autora

- **Hemitest (Split-halves) o Método común de división por mitades.** Correlaciona los puntajes de las dos mitades del instrumento (éste debe contener un número par de ítems y ser de respuestas cerradas). Se divide en dos sub-test, supone la igual de longitud y varianza, uno con los ítems impares y el otro con los pares. *Procedimiento:* (a) se estima la correlación entre los dos puntajes de cada sub-test (medio-test), utilizando el Coeficiente de correlación de puntajes directos r de Pearson (escalas numéricas) o el Coeficiente de correlación de rangos ρ de Spearman (escalas ordinales), ya descritos y (b) se mide la correlación del test completo usando el *Coeficiente de consistencia interna de Spearman-Brown*, con la siguiente fórmula:

$$r_{tt} = \frac{2 r_{xy}}{1 + r_{xy}}$$

r_{tt} : confiabilidad del instrumento en su totalidad

r_{xy} : correlación entre los sub-tests, mitades del test

- **Métodos de análisis de homogeneidad de los ítems.** Son *coeficientes de consistencia interna*, requieren sólo una aplicación del instrumento a la muestra piloto (una sola vez). Miden el constructo de los ítems y se seleccionan, a través de una escala con

puntuaciones, las respuestas optando por alguna de las alternativas fijadas en la escala. Los coeficientes toman valores en el intervalo [0, 1], 0 (cero) significa confiabilidad nula y 1 (uno) confiabilidad total o máxima.

Administrada la prueba piloto, se organizan los datos construyéndose una tabla de distribución de los puntajes (ver Tabla 5). En cada casilla se coloca el puntaje obtenido por la selección del informante según una escala fija que puede ser dicotómica (dupla) o policotómica (3 o más alternativas). Ejemplo de una escala policotómica: Definitivamente Sí (5), Probablemente Sí (4), Indeciso (3), Probablemente No (2), Definitivamente No (1).

Tabla 5

Organización de datos en la prueba piloto para estimar Alfa de Cronbach y Kuder-Richardson)

Sujeto	Ítemes, Propositiones o Reactivos							
	1	2	3	4	5	6	...	K
1	Ptaje	Ptaje	Ptaje	Ptaje	Ptaje	Ptaje	...	Ptaje
2	Ptaje	Ptaje	Ptaje	Ptaje	Ptaje	Ptaje	...	Ptaje
3	Ptaje	Ptaje	Ptaje	Ptaje	Ptaje	Ptaje	...	Ptaje
4	Ptaje	Ptaje	Ptaje	Ptaje	Ptaje	Ptaje	...	Ptaje
....	Ptaje	Ptaje	Ptaje	Ptaje	Ptaje	Ptaje	...	Ptaje
M	Ptaje	Ptaje	Ptaje	Ptaje	Ptaje	Ptaje	...	Ptaje

Nota. Elaborada por la autora

Las escalas dicotómicas son duplas como Sí-No, Verdadero-Falso, Acuerdo-Desacuerdo, Presente-Ausente, entre otras. También se consideran como dicotómicas las respuestas dadas en pruebas objetivas debido a que solamente puede seleccionarse una opción de respuesta correcta y las otras restantes son incorrectas, se asignan como puntajes 2-1 ó 1-0, considerando la opción positiva como la más alta; con estas puntuaciones se puede estimar el coeficiente de confiabilidad. Los métodos más aplicados o populares son:

- **Coeficiente Alfa de Cronbach.** Se usa cuando las alternativas de respuesta son policotómicas (3 o más alternativas de respuesta), como las escalas ordinales de

Likert, Staple, Diferencial semántico y otras escalas. El coeficiente α de Cronbach puede ser calculado usando la fórmula

$$\alpha = \frac{k}{(k-1) \left[\frac{1 - \sum s_i^2}{s_t^2} \right]}$$

Así: α : Coeficiente de confiabilidad alfa (α) de Cronbach del instrumento

k : Número de ítems, tareas o reactivos

s_t^2 : Varianza total del instrumento

$\sum s_i^2$: Sumatoria de las varianzas de los ítems

Cuanto sea menor la variabilidad en las respuestas (homogeneidad en las respuestas dentro de cada ítem), mayor será el Coeficiente Alfa de Cronbach.

- **Método o Coeficiente de Kuder-Richardson 20 (KR_{20})**. Puede ser usado en instrumentos con respuestas cerradas dicotómicas y cuando hay alternativas dicotómicas (correcto-incorrecto) como las pruebas objetivas de opciones múltiples y selección simple citadas. La fórmula a usar:

$$KR_{20} = \frac{n}{n-1} \left(1 - \frac{\sum pq}{s_t^2} \right)$$

En donde: KR_{20} : Coeficiente de Kuder-Richardson

n : número total de ítems

s_t^2 : varianza de las puntuaciones totales

p : proporción de sujetos respondientes de un ítem sobre el total de sujetos

$q = 1 - p$

Hoy día, los cálculos de los coeficientes normalmente se realizan utilizando herramientas tecnológicas como hojas de cálculo y paquetes estadísticos; sin embargo, es importante conocer cuál es el posible procedimiento estadístico a usar y la fórmula a emplear, de manera detallada.

Interpretación del Coeficiente de Confiabilidad

El coeficiente de confiabilidad es una medida estadística, corresponde a un coeficiente de correlación; teóricamente significa la asociación interna del instrumento consigo mismo. Es decir, la correlación entre los ítems que lo conforman; estos valores de correlación oscilan entre 0 y 1 (cero y uno).

Conforme a la Tabla 6 y su interpretación, se considera al 0 (cero) como correlación nula (no hay correlación, es inexistente) y 1 (uno) como correlación perfecta o de máxima confiabilidad (Corral, 2009, 2014); mientras más se acerque a 1 el instrumento será más confiable. Asimismo, en investigación académica y educativa, instrumentos con confiabilidad menor a 0,61 requieren ser reestructurados y sometidos de nuevo a los procesos de validación y confiabilidad (Palella & Martins, 2012), no deberán ser utilizados aquellos instrumentos que arrojen resultados menores a 0,61. Por otra parte, Kaplan & Sacuzzo (citados en Barraza, 2007) arguyen que un valor de 0,70 es suficientemente bueno para cualquier investigación.

Tabla 6

Escala de interpretación de la magnitud del Coeficiente de Confiabilidad de un instrumento

Rangos	Magnitud
0,81 a 1,00	Muy Alta
0,61 a 0,80	Alta
0,41 a 0,60	Moderada
0,21 a 0,40	Baja
0,01 a 0,20	Muy Baja

Nota. Tomada de Ramírez (2010), Ruiz Bolívar (2013) y Palella & Martins (2012)

Ahora bien, para que un instrumento tenga una confiabilidad alta, los ítems deben garantizar la posibilidad de respuestas diferentes (discriminación) y relacionarse entre sí (coherencia entre las respuestas). La confiabilidad disminuye cuando todos los sujetos de la muestra seleccionada responden igual; es decir, el ítem no discrimina. Los coeficientes de confiabilidad tienden a aumentar

cuando se ofrecen más de dos opciones de respuesta; sin embargo, no siempre es el caso. También, se incrementa ésta cuando los rasgos medidos son más específicos y bien delimitados.

Consideraciones Finales

Los instrumentos bien elaborados, no sólo deben ser válidos, confiables, objetivos y útiles, también deben ser pulcros, elegantes, bien presentados y organizados secuencial y lógicamente. Por ende, debe cuidarse el aspecto estético de un instrumento, como son: organización, secuenciación y presentación, lenguaje empleado, redacción y ortografía de los ítemes (reactivos, preguntas, tareas, proposiciones, etc.).

Además, existen instrumentos para recabar datos que, por su naturaleza, no ameritan el cálculo de la confiabilidad, tales como: guiones y protocolos de entrevistas, formatos de evaluación, pruebas orales, son algunos a mencionar; sólo requiere comprobarse, en investigación académica, su validez por juicio de expertos. Igualmente, la Historia Clínica es un instrumento estandarizado, está validada por su uso frecuente y su confiabilidad se comprueba por sus aciertos. Entre los instrumentos que sí demandan estimar el coeficiente de confiabilidad, cuando no están estandarizados, se encuentran: cuestionarios cerrados, escalas actitudinales y opináticas, pruebas académicas de respuesta cerrada y formatos de registro de observaciones sistematizadas como las rúbricas, escalas de estimación y listas de cotejo, si se asignan puntajes a las valoraciones.

Asimismo, vale recalcar la necesidad de construir una tabla o matriz de operacionalización de variables (de especificaciones o de constructos) del instrumento a construir, tomando como base el universo de contenidos teóricos que sustenta la indagación, conforme a los objetivos o propósitos de la investigación y la información precisada; ello permite discriminar las dimensiones y los indicadores (características, rasgos, conductas, actitudes, situaciones, factores, indicios, otros) necesarios para una selección y elaboración-redacción de ítemes (preguntas, proposiciones, tareas, reactivos, descriptores) adecuados y pertinentes que reflejen la realidad de lo estudiado.

Como colofón, es importante garantizar la validez, confiabilidad y objetividad de los instrumentos empleados en el transcurso de la investigación, pues cuando no se establecen claramente éstas siempre habrá dudas sobre la veracidad de las evidencias y la calidad de los resultados y conclusiones derivados del análisis e interpretación, surgidos a partir de la data recabada durante el proceso indagatorio.

Referencias

- Abero, L. (2015). Técnicas de recogida de datos. En L. Abero, L. Berardi, A. Capocasale, S. García & R. Rojas, *Investigación Educativa. Abriendo puertas al conocimiento* (pp. 147-158). CLACSO. <http://biblioteca.clacso.edu.ar/clacso/se/20150610045455/InvestigacionEducativa.pdf>.
- Aravena, M., Kimelman, E., Micheli, B., Torrealba, R. & Zúñiga, J. (Comps.). (2006). *Investigación Educativa I*. AFEFCE/Ecuador y Universidad Arcis/Chile. <https://www.redalyc.org/pdf/4759/475947763003.pdf>.
- Barraza, A. (2007). *¿Cómo valorar un coeficiente de Confiabilidad?* Investigación Educativa Duranguense, (6), 6-10. https://dialnet.unirioja.es/buscar/documentos?query=Dismax.DOCUMENTAL_TODO=c%C3%B3mo+valorar+un+coeficiente+.
- Cabezas, E., Andrade, D. & Torres, J. (2018). *Introducción a la metodología de la investigación científica*. Universidad de las Fuerzas Armadas ESPE. <http://repositorio.espe.edu.ec/handle/21000/15424>.
- Castillo, E. & Vásquez, M. L. (2003). *El rigor metodológico en la investigación Colombia Médica, cualitativa*. 134 (3), 164-167. <https://www.redalyc.org/pdf/283/28334309.pdf>.
- Cohen, N. & Gómez, G. (2019). *Metodología de la Investigación, ¿para qué?*. Teseo. http://biblioteca.clacso.edu.ar/clacso/se/20190823024606/Metodologia_para_que.pdf.

- 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. <http://servicio.bc.uc.edu.ve/educacion/revista/n33/art12.pdf>.
- Corral, Y. (2014). *Instrumentos de recolección de datos: validez y confiabilidad*. Editorial Académica Española-EAE.
- Corral, Y. (2017). *Validez y fiabilidad en investigaciones cualitativas*. Arjé. Revista de Postgrado FaCE-UC, 11 (20), 196-209. <http://arje.bc.uc.edu.ve/arj20/art19.pdf>.
- Corral, Y., Corral, I. & Franco, A. (2019). *La Investigación: tipos, normas, acopio de datos e informe final*. OPSU. <http://mriuc.bc.uc.edu.ve/handle/123456789/8374>.
- Hernández Sampieri, R., Fernández, C. & Baptista, P. (2014). *Metodología de la Investigación*. (6ª ed.). McGraw-Hill.
- Latorre, A., del Rincón, D. y Arnal, J. (2014). *Bases metodológicas de la investigación educativa*. Hurtado Editores.
- Manterola, C., Grande, L., Otzen, T., García, N., Salazar, P. & Quiroz, G. (2018). *Confiabilidad, precisión o reproducibilidad de las mediciones. Métodos de valoración, utilidad y aplicaciones en la práctica clínica*. Revista Chilena de Infectología, 35 (6), 680-688. <https://scielo.conicyt.cl/pdf/rci/v35n6/0716-1018-rci-35-06-0680.pdf>.
- Montero, M. & Hochman, E. (2005). *Investigación Documental: Técnicas y Procedimientos*. Panapo.
- Palella, S. & Martins, F. (2012). *Metodología de la Investigación cuantitativa*. Fedupel.
- Ramírez, T. (2010). *Cómo Hacer un Proyecto de Investigación*. Panapo.
- Rojas de Escalona, B. (2014). *Investigación cualitativa. Fundamentos y praxis*. (3ª ed.). Fedupel. <https://gsosa61.files.wordpress.com/2015/11/investigacion-cualitativa-rojas-2014-comprim-1.pdf>.

- Ruiz Bolívar, C. (2013). *Instrumentos y técnicas de Investigación Educativa. Un enfoque cuantitativo y cualitativo para la recolección y análisis de datos.* (3ª ed.). Danaga.
https://www.academia.edu/37886948/Instrumentos_y_Tecnicas_de_Investigaci%C3%B3n_Educativa_Carlos_Ruiz_Bolivar_pdf.
- Van Dijk, T. (2005). *Ideología y análisis del discurso.* Utopía y Praxis Latinoamericana, 10 (29).
http://ve.scielo.org/scielo.php?script=sci_arttext&pid=S1315-52162005000200002.

Yadira Corral de Franco:

Profesora mención Matemática, Instituto Universitario Pedagógico de Caracas- Universidad Pedagógica Experimental Libertador UPEL (1977-2004). Magister en Educación Superior, UPEL (1993). Investigador PEII B-2015. Profesor Asociado, Facultad de Ciencias de la Educación, Universidad de Carabobo (2020).