

# La rigurosidad científica: validez y confiabilidad en los paradigmas cuantitativo y cualitativo

Andrea Gaete Moreno\*

## Introducción

Comprender la realidad o el mundo que nos rodea sin sesgos es una tarea compleja, medirla mucho más, ya que es común que se presenten errores debido a la subjetividad propia del ser humano asociada a su particular comprensión e interpretación de la realidad (valores, percepciones, prejuicios) favoreciendo lo que Schutt denomina “observación selectiva”<sup>1</sup> donde se elige investigar aquello que es favorable o está en línea con nuestras preferencias o creencias. Ello se refleja en problemas que comúnmente ocurren en todo el proceso de la investigación científica, tales como la dificultad en la definición del problema, la sesgada formulación de hipótesis, un incorrecto diseño metodológico, así como errores en la medición y, por tanto, en la generalización y elaboración de las conclusiones, afectando finalmente la “calidad” de la información.

Es entonces la rigurosidad propia del método científico<sup>2</sup> la que permite controlar o reducir esos sesgos de percepción o interpretación, incluso de razonamiento, durante todo el proceso de investigación, facilitando la obtención de evidencia de calidad con el grado de precisión que la investigación requiera y generando conocimiento que se aproxime a la que podríamos denominar como “verdad científica”, tal como lo estableció el filósofo y físico francés René Descartes en el primero de sus cuatro preceptos del método, al no admitir como verdadera cosa alguna como no supiere con evidencia que lo es,<sup>3</sup> centrándose en la búsqueda de reglas para establecer verdades a través de un procedimiento claro y eficaz (criterio de verdad).

\* Socióloga, Universidad La República. Magíster en Ciencias Políticas, Universidad de Chile. Docente Facultad de Ciencias Sociales y Económicas, Universidad del Pacífico. Investigadora y Analista del Centro de Estudios Estratégicos de la Academia de Guerra del Ejército de Chile. [agaetem@acague.cl](mailto:agaetem@acague.cl)

<sup>1</sup> Russel Schutt, *Investigating the Social World. The process and practice of research* (Canada: Sage Publications, 2012), 6.

<sup>2</sup> Entendido como un procedimiento de la investigación científica para tratar un conjunto de problemas que se aplica al ciclo entero de investigación y cuya finalidad es el conocimiento objetivo del mundo. En Mario Bunge, *La investigación científica* (Barcelona: Ariel, 1989), 24.

<sup>3</sup> René Descartes, *El Discurso del Método* (Madrid: FGS, 2010), 47.

Esta aproximación objetiva del mundo, en su dimensión epistemológica (noción de verdad), metodológica (control) y axiológica (neutralidad valórica), se entiende como una actitud y un modo de proceder de los científicos, refrenando los elementos de valor puramente personales o subjetivos, para lograr un conocimiento objetivo que puede ser comprendido de la misma manera por todo sujeto competente (matemático, físico, sociólogo, etc.)<sup>4</sup>. Sin embargo, lograr la objetividad en el ámbito de las ciencias sociales (y militares) sigue siendo una tarea compleja, sobre todo en lo que respecta al diseño en una investigación científica.

Y es en este contexto donde cobran importancia la validez y confiabilidad, constituyéndose en la primera línea de defensa contra conclusiones espurias e incorrectas<sup>5</sup>, ya que comparativamente dos estudios pueden obtener una buena evidencia científica, pero diferir en la calidad y rigor de la misma debido a sesgos del propio investigador o bien debido a errores en el diseño utilizado<sup>6</sup>.

En este sentido, es importante aclarar que el concepto de precisión en la confiabilidad se entiende de manera distinta que en la validez. En la confiabilidad lo que importa es que el instrumento elaborado para la investigación (o bien adaptado de otro estudio) mida aquello que se pretende medir, es decir, que las preguntas midan de manera correcta el fenómeno con el menor error o sesgo posible; en cambio, la validez se concentra básicamente en que los conceptos o contenidos de interés para el estudio estén bien definidos y representados en el instrumento de medición.

Un clásico ejemplo que permite entender esta diferencia es el que ofrece Fred Kerlinger: Si alguien se sube a una pesa 50 veces durante el periodo de una hora, es probable que los resultados no difieran mucho entre cada medición, en este sentido, la pesa es precisa ya que indica consistentemente el mismo peso, sin embargo, si la pesa tiene una sobreestimación de 2 kilos, el valor obtenido en cada medición será errado, por tanto, la pesa puede ser considerada confiable pero no válida.

El ejemplo anterior indica dos cosas. La primera es que cualquier investigador que asuma el reto de elaborar un cuestionario o una escala manteniendo la rigurosidad como ente rector, debe distinguir claramente entre los conceptos de validez y confiabilidad<sup>7</sup>. La segunda es que, en cuanto a la confiabilidad, medir lo que se pretende medir no es nada fácil, ya que los sesgos propios de las ciencias sociales generan “errores” en la medición que obligatoriamente deben ser previstos y controlados por el investigador.

Considerando entonces la importancia de la rigurosidad en toda investigación que se defina como científica y, a la vez, lo complejo de esta tarea, es que este artículo se centra en distinguir conceptualmente la validez y confiabilidad como constructos elementales

<sup>4</sup> Alberto Cupani, “Acercas de la vigencia del ideal de objetividad científica”. *Revista Scientlei. Studia*. 9 3 (2011): 501.

<sup>5</sup> Neil J. Salkind, *Métodos de Investigación* (México: Prentice Hall, 1998), 72.

<sup>6</sup> Es pertinente aclarar que en los textos de metodología se entiende por diseño tanto al proceso completo de prueba de hipótesis, como a los “tipos de diseño” experimentales y no experimentales. Para profundizar en este aspecto, se sugiere revisar el cap. 7 del libro de Sierra Bravo.

<sup>7</sup> Recordar, como lo señala Alvarez-Gayou, que la confiabilidad se considera externa cuando otros investigadores llegan a los mismos resultados en condiciones iguales, e interna cuando varios observadores concuerdan en los hallazgos al estudiar la misma realidad (ver p. 31).

para la consecución de información de calidad, precisa y clara, así como las técnicas que comúnmente se utilizan para su obtención tanto en el enfoque cuantitativo como en el cualitativo, entendiéndose que, como bien lo señala la destacada metodóloga cualitativa norteamericana Janice Morse, toda investigación sin rigor, no tiene valor, se transforma en ficción y pierde toda su utilidad<sup>8</sup>.

### ***Validez y confiabilidad en el paradigma positivista (cuantitativo)***

En un sentido general, la *validez* en el paradigma positivista se entiende como la congruencia cercana y directa entre la naturaleza del objeto que se mide y el instrumento de medición<sup>9</sup>, o bien como el grado en que una medición empírica refleja adecuadamente el significado real del concepto<sup>10</sup>. Por ejemplo, medir el peso de un kilo de manzanas en una pesa correctamente calibrada resulta fácil, sin embargo, cuando lo que se pretende medir son constructos como la democracia, la seguridad o la defensa, establecer la validez se torna una tarea más compleja, ya que el paso a seguir es crear los instrumentos más adecuados de acuerdo al objeto o fenómeno a medir. Y es justamente en este proceso de creación donde es necesario establecer con claridad y precisión las condiciones de validez con el fin de evitar sesgos que afecten la medición y finalmente la obtención de evidencia que nos acerque a la verdad científica.

A continuación se expondrán brevemente los tres tipos de validez que todo investigador debiera conocer y utilizar en sus investigaciones: contenido, criterio y constructo.

La *validez de contenido* (también conocida como racional o lógica), se refiere a la adecuación o representatividad del contenido teórico del fenómeno a estudiar en el universo de reactivos (preguntas o ítems) del instrumento de medición.

Por ejemplo, si el objetivo del estudio es evaluar el impacto del Hospital Modular de Campaña (HMCE) en la Región metropolitana, se esperaría que las preguntas o reactivos del instrumento contengan aspectos tales como: satisfacción usuaria, calidad de vida, contribución al desarrollo sostenible nacional (definidos en marco teórico). En cambio, si el instrumento contiene preguntas sobre la industria de defensa, entonces no se estaría cumpliendo con este tipo de validez, ya que como se observa, el contenido cambia y no es compatible con el objetivo del estudio.

Con el fin de asegurar esta validez y evitar sesgos, se puede recurrir al juicio de otros (pares expertos) para establecer la representatividad conceptual de los reactivos y evaluar su supuesta relevancia respecto a la propiedad que se mide. Esta forma de evaluación por jueces es más bien subjetiva,<sup>11</sup> sin embargo, se puede cuantificar y comparar a través del índice Kappa de Cohen que estima hasta qué punto dos observadores coinciden en su medi-

<sup>8</sup> Janice Morse, Michael Barrett, Maria Mayan, Karin Olson and Jude Spiers, "Verification Strategies for Establishing Reliability and Validity in Qualitative Research". *International Journal of Qualitative Methods* 1 2 (2002), 14.

<sup>9</sup> Fred Kerlinger y Howard B. Lee, *Investigación del Comportamiento. Métodos de Investigación en Ciencias Sociales* (México: McGraw-Hill, 2001), 603.

<sup>10</sup> Earl Babbie, *Fundamentos de la Investigación Social* (Buenos Aires: International Thompson, 2000), 112.

<sup>11</sup> Los juicios de expertos se pueden obtener también usando métodos de agregados individuales (un solo experto a la vez y con evaluación independiente), grupales (varios expertos al mismo tiempo y con evaluación grupal),

ción, es decir, el nivel de concordancia entre ambos. Si existe variabilidad en las respuestas entre pares (inter observador), la posibilidad de error en la validez es mayor. En caso de no evaluar adecuadamente el contenido de los reactivos, solo se puede establecer una validez de tipo aparente, que en la práctica no asegura que los reactivos definidos representen verdaderamente el concepto que se pretende medir.

La *validez de criterio* compara las puntuaciones de una prueba o escala con una o más variables externas o criterios. Esta se divide en predictiva (criterios futuros) y concurrente (medición de ambas puntuaciones en el mismo tiempo). En este caso, el único criterio que diferencia ambos tipos de validez es “el tiempo” en el que se aplican los instrumentos.

Un ejemplo concreto de validez predictiva es utilizar pruebas como la Prueba de Selección Universitaria (PSU), que sirven para predecir el rendimiento académico futuro de los alumnos una vez que ingresen a la universidad (en este caso, las notas del alumno en el desarrollo de su carrera universitaria serían el “criterio” de comparación). En este caso, se realizan dos mediciones en tiempos distintos.

En el caso de la validez de criterio concurrente, se aplican dos instrumentos al mismo tiempo, permitiendo de esta manera probar hipótesis de tipo correlacional<sup>12</sup>. Por ejemplo, si la hipótesis del estudio es “a mayor modernidad individual, mayor ejercicio del liderazgo democrático en las jefaturas de la empresa XXX”, entonces se podría elaborar y posteriormente aplicar un cuestionario para medir el ejercicio del liderazgo democrático en la empresa (variable de interés) y, paralelamente y en los mismos sujetos, aplicar una escala de modernidad individual (criterio de comparación).

Quizás lo más difícil de este tipo de validación es definir el criterio más adecuado o pertinente para la comparación. Es por eso que se recomienda revisar el o los objetivos del estudio antes de decidir si este tipo de validez es pertinente para su investigación.

El último tipo de validez que un investigador puede utilizar es la de constructo. Esta presenta un grado mayor de dificultad ya que lo que pretende es validar la teoría que sustenta el instrumento. Veamos un ejemplo:

El Ejército de Chile necesita medir la cohesión institucional de su personal, para tal efecto, se hipotetiza que la cohesión está estrechamente ligada a la identificación con las normas y valores institucionales, a un alto sentido de pertenencia, de compromiso, compañerismo y solidaridad. Una vez definidos los conceptos y la relación entre ellos, se elabora un instrumento para medir la cohesión institucional (constructo).

Para validar el instrumento, se debe probar factorialmente<sup>13</sup> que las relaciones definidas y establecidas en el marco teórico efectivamente se cumplen, es decir, que la cohesión correlaciona positivamente con todas las variables antes mencionadas, esto debido principalmente a que los constructos son elaboraciones conceptuales que requieren del acuerdo o consenso

---

experto único o método Delphi. Recordar que la selección del tipo de evaluación depende de las características y necesidades de la investigación.

<sup>12</sup> Para comprender la lógica detrás de las hipótesis de correlación, se sugiere revisar el artículo sobre formulación de hipótesis en el proceso de investigación, escrito por Marjorie Gallardo.

<sup>13</sup> Se utiliza el análisis factorial (AF) buscando establecer de manera individual qué variable o propiedad aporta más o menos a la varianza total del instrumento. Este análisis tiene dos modalidades: el Análisis Factorial Exploratorio (AFE) y el Análisis Factorial Confirmatorio (AFC).

del mundo académico sobre sus significados para que sean considerados útiles y válidos<sup>14</sup>. Si se logra probar estadísticamente que existe esa correlación, entonces el instrumento es válido a nivel de constructo, de lo contrario, como lo señala Babbie, se pondría en tela de juicio la validez de la medición<sup>15</sup>.

Para algunos especialistas, este tipo de validez incorpora las otras dos, es decir, al lograr la validez de constructo se obtiene también la validez de contenido y criterio<sup>16</sup>. Cabe aclarar que no es necesario que todos los estudios desarrollen esta visión unificada de la validez, ya que, por ejemplo, para un estudio exploratorio-descriptivo, la validez de contenido puede ser suficiente, mientras que para un estudio de tipo correlacional o causal se necesitaría una validez de criterio o de constructo. Lo importante, es tener claro que la elección del tipo de validez de un estudio depende, en gran medida, del objetivo de la investigación.

La *confiabilidad*<sup>17</sup>, por su parte, tiene que ver con la *precisión* en el proceso de medición, donde los coeficientes de confiabilidad proporcionan una indicación en que una medida es *consistente, estable y reproducible* en el tiempo, logrando con ello, una mayor objetividad y rigurosidad. En otras palabras, la confiabilidad consiste en que repetidas aplicaciones de un instrumento, realizadas por distintos observadores en momentos de tiempo distintos, arrojen resultados iguales o similares con el menor error posible. Por ello es importante identificar los tipos de errores que comúnmente se presentan en el proceso de medición, entendiendo, como lo señala Shutt, que no se puede realmente medir un fenómeno, si la medición que se usa da resultados inconsistentes<sup>18</sup>.

Por ejemplo, si se aplica un único instrumento para evaluar el nivel de estrés de los alumnos de la ACAGUE antes y después de un Juego de Guerra (dos tiempos distintos), cabe la posibilidad que ambos puntajes difieran o varíen entre sí debido al instrumento utilizado (error de método) o a las condiciones del sujeto al momento de contestar (error de rasgo)<sup>19</sup>.

En cuanto al error de método, Salkind presenta un ejemplo muy claro: un sujeto se ha preparado para una prueba, ha estudiado y entiende bien la materia, sin embargo, al momento de rendir la prueba se da cuenta que las preguntas son de respuesta múltiple en vez de desarrollo, y que las instrucciones no son tan claras como esperaba. Tales condiciones pueden afectar su puntaje verdadero (lo que sabe y sería capaz de contestar), y su puntaje observado (lo que finalmente contestó)<sup>20</sup>. En este caso, el error o diferencia entre ambos puntajes (verdadero y observado) se debe a la estructura del instrumento.

<sup>14</sup> José Antonio Pérez-Gil, Salvador Chacón Moscoso y Rafael Moreno Rodríguez, "Validez de constructo: El Uso de Análisis Factorial Exploratorio-Confirmatorio para Obtener Evidencias de Validez", *Revista Psicothema* 12 2 (2000), 442.

<sup>15</sup> Babbie, *Fundamentos de la Investigación Social*, 114.

<sup>16</sup> Para entender mejor esta visión unificada, se sugiere revisar los trabajos del psicólogo americano Samuel Messick, así como también se recomienda el artículo escrito por Paula Elosua "Sobre la Validez de los Tests", 2003.

<sup>17</sup> Para una mayor comprensión del concepto de confiabilidad, se sugiere revisar la "teoría clásica de los tests", en autores como Charles Spearman, Mary J. Allen y Lee Cronbach.

<sup>18</sup> Shutt, *Investigating the Social World*, 124.

<sup>19</sup> Recordar que se pueden utilizar instrumentos ya validados y probados como confiables por otros investigadores. Lo importante es revisar que sean útiles para los objetivos del estudio, así como adaptarlos al contexto o realidad particular de la investigación en caso de ser necesario.

<sup>20</sup> Salkind, *Métodos de Investigación*, 73.

En cambio, el segundo tipo de error, considera las condiciones del sujeto al momento de contestar el instrumento (no maneja el contenido, alto nivel de estrés asociado a la carga laboral y académica, ansiedad, enfermedad, problemas familiares, ambiente de la prueba, u otros). A diferencia del anterior, aquí el error en la medición se debe a condicionantes externas e internas del sujeto, afectando o distorsionando su puntaje observado.

Al respecto, es importante tener en consideración lo siguiente: a) el puntaje verdadero se concibe como la puntuación que un individuo obtendría al responder realmente lo que sabe, piensa o cree, sin ningún tipo de interferencia; b) que el error aquí se refiere en concreto a la diferencia que existe entre el puntaje observado y el puntaje verdadero, por tanto, mientras menor sea esa distancia, menor será el error en la medición; y c) ambos puntajes nunca realmente se conocen, por tanto, lo que se hace es calcular la varianza del error. Esto significa que, cuanto menor sea la varianza del error, mayor será la *precisión* de los puntajes y la confiabilidad de la medición, derivando en conclusiones claras y de calidad.

Ahora bien, para controlar o minimizar el error, se utilizan técnicas específicas que permiten calcular el coeficiente de confiabilidad<sup>21</sup>, las que se encuentran, aunque con algunas diferencias, en prácticamente todos los libros o manuales de metodología (Kerlinger, Babbie, Salkind, Shutt, Sampieri, Bernal, Corbetta, entre otros). Estas se dividen en aquellas que buscan la *coherencia* de la medición (test-retest y formas paralelas), y las que buscan la *consistencia interna*. Estas son:

- Confiabilidad test-retest: Se aplica el mismo instrumento al mismo grupo de personas en dos ocasiones diferentes (*1 instrumento aplicado a los mismos sujetos en tiempo 1 y tiempo 2*), obteniendo de esta manera dos mediciones por persona, cuyos valores se utilizan para calcular la correlación entre ambas<sup>22</sup>. Este test mide la *estabilidad* a través del tiempo, ya que si se obtienen resultados similares en la segunda aplicación, se podría establecer que la prueba o instrumento es confiable.
- Confiabilidad de formas paralelas: En este caso, se elaboran dos formatos *equivalentes* del instrumento pero que no son idénticos (se mide lo mismo pero con preguntas distintas), los que se aplican a los mismos sujetos, obteniendo al igual que en el caso anterior, dos puntuaciones que se deben correlacionar (*2 instrumentos distintos aplicados a los mismos sujetos en tiempo 1 y tiempo 2*). En este caso, es importante considerar que, como son dos pruebas paralelas, cada sub muestra debe tener medias y desviaciones estándar estadísticamente equivalentes para poder concretar la comparación. Esta prueba es útil, por ejemplo, para medir un constructo en dos tiempos distintos.

Cabe aclarar que, tanto para el test-retest o las formas paralelas, se sugiere: definir muy claramente el intervalo de tiempo entre cada medición, ello con el fin de evitar la repetición de respuestas aprendidas o el aburrimiento de los participantes de la investigación; considerar la reactividad propia del ser humano, que puede afectar la propiedad que queremos medir debido a los cambios que se pueden producir en él

<sup>21</sup> Este coeficiente varía de -1,00 a 1,00, dónde este último significa confiabilidad perfecta. Esta es difícil de obtener, por tanto, valores de 0,8 o 0,9 son buenos, incluso hasta 0,7 es considerado aceptable.

<sup>22</sup> La literatura actualizada no sugiere el uso del r de Pearson para este cálculo de correlación, ya que este genera información interclase, mientras que lo necesita el test-retest es calcular el coef. intraclase (ICC). Este se puede obtener en programas estadísticos como SPSS, STATA o R, y su valor varía de 0 a 1.

entre mediciones; y, en la medida de lo posible, que sea el mismo investigador el que tome ambas mediciones.

- Confiabilidad por mitades partidas. Aquí el investigador debe elaborar un solo instrumento para, posteriormente, dividirlo en dos mitades iguales. Los reactivos o preguntas homogéneas se suman de manera independiente, obteniendo nuevamente dos puntajes que se deben correlacionar<sup>23</sup>. Para la suma de las mitades, se recomienda sumar los reactivos impares (total 1) con los reactivos pares (total 2), para luego correlacionarlos. A diferencia de las otras dos técnicas, esta requiere solo una aplicación.
- Confiabilidad por consistencia interna. Aquí se evalúa qué tan *consistentes* y unificados entre sí son los reactivos de una prueba. Por ejemplo, si un cuestionario o escala contiene un set de 50 preguntas distintas que pretenden medir la capacidad informática de un país, se esperaría que cada pregunta de manera independiente mida el constructo señalado, que todas las preguntas de manera conjunta estén relacionadas entre sí, y que también exista una relación reactivo-escala.<sup>24</sup> Esta técnica es útil cuando el objetivo es medir un solo concepto o constructo a través de varios ítems o preguntas. Esta técnica se divide en:
  - a. Coeficiente Alpha de Cronbach. Asume por defecto la unidimensionalidad de la escala, es decir, cuando se mide un solo factor o constructo en todo el instrumento. Se basa en la correlación interelementos, por ello, su uso es adecuado para preguntas con escala de respuesta tipo Likert, o bien, el mismo nivel de medición. Si ese no es el caso, se sugiere calcular el coeficiente por variables similares en nivel de medición o sub escalas.<sup>25</sup> Para el cálculo de este coeficiente, se puede calcular a) la varianza de los ítems, b) elaborar una matriz de correlación (pearson), o c) ir directamente al SPSS para el cálculo de fiabilidad. Todas las opciones anteriormente mencionadas son válidas.
  - b. KR-20 y KR-21 (Kuder-Richardson). Ambas fórmulas asumen que cada reactivo tiene la misma media y la misma varianza, y son aplicables solo en instrumentos con escala de respuesta dicotómica (si-no, correcto-incorrecto, verdadero-falso).

Otra técnica, aunque no todos los autores la mencionan, es la confiabilidad interobservador, en la que se usa más de un observador para medir personas, eventos o lugares. Si los distintos observadores al usar un mismo instrumento obtienen resultados similares, estos serán más confiables.

<sup>23</sup> Para el cálculo de la correlación, se puede usar la fórmula de Spearman-Brown, pero "solo" si tiene certeza que ambas mitades son efectivamente iguales. De lo contrario, el resultado del coeficiente será errado. Si existen dudas respecto a la homogeneidad de las mitades, se recomienda utilizar la fórmula de Rulon o de Guttman. Ambas estiman la confiabilidad de la prueba completa sin considerar la confiabilidad por mitades.

<sup>24</sup> En caso de querer utilizar este coeficiente en un cuestionario o escala no unidimensional, lo que se puede hacer es dividir o desagregar las dimensiones o variables diferentes en subescalas. Por ejemplo, un set de 100 preguntas que mide cinco dimensiones distintas, se puede fraccionar en 5 subsets de 20 preguntas cada una, para posteriormente calcular el Alpha para cada subset. No se puede aplicar el Alpha en niveles de medición diferentes (nominal o intervalo) o en variables que miden aspectos diferentes. \*Recordar que mientras más preguntas tenga un instrumento, mayor será el valor del coeficiente de confiabilidad.

<sup>25</sup> Para entender el sentido detrás de la fórmula del Alpha de Cronbach, se recomienda revisar el cap. 27 del libro de Fred Kerlinger y Howard B. Lee.

Para elegir el tipo de confiabilidad más adecuado, se recomienda considerar el objetivo, tipo de hipótesis que se quiera probar y los recursos económicos y humanos con los que cuenta el investigador, incluyendo el tiempo estimado para el desarrollo de la investigación y la capacitación de encuestadores y codificadores en caso de ser necesario.

Finalmente, cabe señalar respecto a la validez y la confiabilidad en este enfoque, que el error en su medición está directamente relacionado con la elaboración de datos empíricos imprecisos, una incorrecta prueba de hipótesis y conclusiones sesgadas.

La validez y confiabilidad en el paradigma cualitativo

Como bien lo señala Ruiz Olabuénaga, la metodología cualitativa tiene como fin captar el significado de las cosas (procesos, comportamientos, actos) tal como la viven y construyen los individuos a través de su propio lenguaje. Esta forma de comprender la realidad, si bien es rica en profundidad, también es compleja en su aproximación debido a ciertos factores que pueden llegar a distorsionar la calidad de las interpretaciones y conclusiones, tales como la dificultad de algunos analistas en mantener una postura neutral en todo el proceso investigativo (expresada en valores, prejuicios o preferencias del investigador), el tiempo que requiere realizar este tipo de análisis, y la complejidad propia de los fenómenos sociales.

Para controlar los sesgos y autenticar o legitimar los resultados, Guba y Lincoln proponen en 1985 cuatro criterios análogos a la validez y confiabilidad, estos son: la credibilidad (validez), transferibilidad (generalización), dependencia (fiabilidad) y confirmabilidad (objetividad)<sup>26</sup>.

La *credibilidad* (valor de verdad), alude a la confianza o veracidad de los descubrimientos realizados en una investigación, teniendo como objetivo el isomorfismo entre los datos recogidos por el investigador y la realidad<sup>27</sup>. Se obtiene a través de estrategias tales como la *observación persistente*, la *presencia prolongada en el campo*, la *búsqueda de evidencia negativa*, el *chequeo de la información con los informantes* y la *triangulación*.

De estas estrategias, la más utilizada en la metodología cualitativa es la triangulación, la que se entiende como la convergencia de distintas perspectivas y métodos en el estudio de un mismo objeto, asegurando así una aproximación más comprensiva del problema investigado (mayor validación). Norman K. Denzin, sociólogo experto en metodología cualitativa, identifica cuatro tipos básicos de triangulación: de datos, de investigadores, teórica y metodológica. Valerie Janesick agrega una quinta triangulación, a la que llama "triangulación interdisciplinaria", en la que participan profesionales de distintas disciplinas, enriqueciendo así la interpretación.

Es posible usar más de un tipo de triangulación en un estudio cualitativo, por cuanto la decisión de cuál usar o combinar en un estudio, depende principalmente de los objetivos de la investigación.

La *transferibilidad* apunta, según Briones, a la posibilidad de hacer inferencias lógicas a situaciones similares, centrándose en el grado de aplicabilidad de los descubrimientos. Las estrategias que se usan para lograr la transferibilidad son: el *muestreo intencional* de

<sup>26</sup> Para mayor detalle de estas técnicas, se sugiere revisar la primera unidad del libro de Guillermo Briones y el capítulo 3 del libro de José Ignacio Ruiz Olabuénaga.

<sup>27</sup> Gregorio Rodríguez, Javier Gil y Eduardo García, *Metodología de la Investigación Cualitativa* (Málaga: Aljibe, 1999), 286.



los casos y la *descripción densa* del fenómeno, referida a que mientras más información se pueda recabar sobre un tema, mayor será su comprensión y posibilidad de transferir las conclusiones a poblaciones similares.

La *Dependencia o Consistencia de los datos* se relaciona con la posibilidad de obtener los mismos resultados al replicar el estudio con los mismos o similares sujetos y contextos. Se obtiene a través de: la *auditoria de dependencia*, en el cual las decisiones metodológicas y el trabajo de interpretación de datos son examinados por un *investigador externo* (juicio crítico de un par auditor).

La *Confirmabilidad*, entendida también como neutralidad u objetividad. Según Rodríguez, Gil y García, se usa para aumentar la congruencia entre la realidad observada y la realidad que se pretendió representar, estableciendo la independencia de los descubrimientos frente inclinaciones, motivaciones, intereses o concepciones teóricas del investigador.

Para ello, se debe: *chequear la información obtenida en la investigación con un agente externo experto en la temática, verificar o contrastar la coherencia interna de los datos, la relación entre estos y las deducciones e interpretaciones obtenidas.*

Cabe aclarar que estos cuatro criterios no son independientes entre sí, y que a partir de su uso conjunto, como bien lo señala Olabuénaga, se enriquece la investigación (validez) y aumenta la fiabilidad (replicabilidad), sobre todo en este tipo de estudios que abordan fenómenos sociales que en su mayoría son inestables e irrepetibles.

Otra propuesta es la que presenta el profesor de metodología Burke Johnson, quien plantea que para asegurar el rigor científico en estudios cualitativos se deben considerar tres criterios de validez: descriptiva, interpretativa y teórica<sup>28</sup>.

Ahora bien, los criterios presentados por Guba y Lincoln, y Burke Johnson, son solo una forma de abordar el problema de la rigurosidad en el enfoque cualitativo,<sup>29</sup> más cercano al naturalismo y alejado del positivismo, y que buscan, como lo plantea Alvarez-Gayou, lograr y asegurar la obtención de la situación real y verdadera de las personas a las que se investiga buscando la “autenticidad”, es decir, lograr que las personas expresen realmente su sentir<sup>30</sup>. Sin embargo, autores como Margarite LeCompte y Judith Goetz (1982) abordan la calidad de la investigación manteniendo los conceptos clásicos de validez y confiabilidad.

Si bien las investigadoras se centran en la búsqueda de precisión y claridad científica en el trabajo etnográfico, su aporte se puede extrapolar a los distintos métodos y aproximaciones teóricas de la investigación cualitativa (fenomenología, estudio de casos, teoría fundamentada, investigación acción, biografía, entre otros).

A continuación se presenta un cuadro resumen con los criterios propuestos:

<sup>28</sup> Burke Johnson, “Examining the validity structure of qualitative research”, *Education* 118 2 (1997), 284-287.

<sup>29</sup> En algunos textos se distingue el “rigor” como concepto propio del enfoque positivista, mientras que en el enfoque cualitativo, se utiliza más bien el concepto de “trustworthiness” (credibilidad, confianza, veracidad, fidelidad, entre otros).

<sup>30</sup> Juan Luis Alvarez-Gayou, *Cómo hacer Investigación Cualitativa. Fundamentos y Metodología* (México: Paidós, 2009), 31.

Cuadro 1  
Validez y fiabilidad en estudios cualitativos, según LeCompte y Goetz

Tipo	Definición	Técnica
Validez interna	Correspondencia entre el significado que atribuye el investigador a las categorías conceptuales utilizadas en su estudio y el significado que atribuyen a esas mismas categorías los participantes	<ul style="list-style-type: none"> <li>-Revisión y contraste continuo del trabajo del investigador con informantes u observadores</li> <li>-Seleccionar criteriosamente a los informantes y el escenario para la recogida de datos</li> </ul>
Validez externa	Comparabilidad de los resultados con los de otros estudios	Diseños de escenarios múltiples
Fiabilidad Interna (nivel de concordancia)	Grado en que un segundo o tercer investigador llegaría a los mismos resultados partiendo de idénticos conceptos o constructos	<ul style="list-style-type: none"> <li>-Presencia de más de un investigador en el campo</li> <li>-Usar categorías descriptivas concretas cercanas a la realidad.</li> <li>-Confirmación con los informantes de la correspondencia entre los datos recogidos por el investigador y lo que ellos realmente perciben</li> <li>-Revisión del trabajo por parte de otros investigadores</li> <li>-Utilización de cámaras o grabadoras para el registro y conservación de los datos.</li> </ul>
Fiabilidad externa (replicabilidad)	Responde a la interrogante de si un investigador independiente llegaría a descubrir los mismos fenómenos o construiría los mismos constructos al estudiar el mismo escenario u otro similar	<ul style="list-style-type: none"> <li>-Contar con información clara del rol del investigador en el contextos estudiado</li> <li>-Descripción precisa de los informantes y de los criterios con los cuales fueron seleccionados</li> <li>-Descripción del contexto y circunstancias en que se lleva a cabo la recogida de datos</li> <li>-Presentar las premisas teóricas de las que parte el investigador</li> <li>-Definición clara de los conceptos y unidades de análisis, así como de las técnicas de recogida y análisis de los datos</li> </ul>

Fuente: Elaboración propia a partir del texto de Margaret LeCompte y Judith Goetz. *Problems of Reliability and Validity in Ethnographic Research*, 1982.

Una posición similar es la que adoptan Janice Morse<sup>31</sup>, quienes sugieren mantener los conceptos de validez y confiabilidad en el enfoque cualitativo, ya que el introducir términos paralelos o alternativos (como los de Guba y Lincoln) más que clarificar, debilitan la rigurosidad. La crítica se centra en que estos criterios evalúan el rigor durante una investigación,

<sup>31</sup> Janice Morse, *Verification Strategies*, 17.

pero no lo aseguran. Como lo señalan los autores, estas estrategias usadas apropiadamente fuerzan al investigador a corregir posibles errores durante todo el proceso investigativo.

A estas posturas, se agregan las de autores como el sociólogo holandés Peter Swanborn y la de los metodólogos Margaret Eisenhart y Keneth Howe. Estos últimos proponen el uso de cinco criterios generales de calidad de los datos (únicos) que puedan ser usados tanto en el paradigma positivista como interpretativo, centrándose básicamente en la obtención de argumentos válidos<sup>32</sup>. Un resumen de lo anterior se expone en el siguiente cuadro:

Cuadro 2  
Resumen con estrategias de validez y confiabilidad según Morse, Eisenhart y Howe

Autores	Criterios o Estrategias
Janice Morse	Responsabilidad del investigador, coherencia metodológica, un muestreo adecuado, una activa postura analítica y, finalmente, la saturación.
Margaret Eisenhart y Keneth Howe	a) la coherencia entre las preguntas de investigación, técnicas de recolección y análisis de datos, b) la aplicación eficaz de la recogida y análisis de datos, lo que implica manejar correctamente las entrevistas, diseñar adecuadamente los instrumentos, seleccionar criteriosamente los individuos y analizar sustantivamente los datos, c) la coherencia teórica con el conocimiento previo, es decir, que los supuestos y teorías sean expuestos con claridad de modo que permitan su comparación con otros estudios realizados desde la misma perspectiva teórica, d) consideraciones éticas de toda investigación, tanto internas como externas, y e) la exhaustividad del trabajo, que busca la claridad, coherencia y competencia en los procedimientos de recogida y análisis de datos, y la calidad técnica de la investigación.

Fuente: Elaboración propia.

En definitiva, este debate conceptual y de criterios en el enfoque cualitativo es pertinente en este artículo para que el lector comprenda la amplia gama de opciones a las que puede recurrir al momento de establecer la rigurosidad en este tipo de estudios, aunque teniendo siempre en consideración que todas las alternativas presentadas son válidas.

Por esta razón, solo es posible recomendar que para elegir la opción más pertinente para una investigación cualitativa, se deben considerar los objetivos del estudio, el método de investigación utilizado, los recursos y la factibilidad de aplicar los criterios propuestos por los distintos autores, manteniendo siempre, como lo señala Johnson, una actitud reflexiva y crítica que permita controlar los potenciales sesgos e inclinaciones siempre presentes en el investigador.

<sup>32</sup> Margaret Eisenhart and Kenneth Howe, "Validity in Educational Research", ed. M. LeCompte, W. Millroy, and J. Preissle, *The Handbook of Qualitative Research in Education* (San Diego: Academic Press, 1992), 655-663.

## **Conclusiones**

Como lo plantea Morse, el rigor es una meta deseable en toda investigación científica, independiente de la selección del método (cualitativo o cuantitativo) utilizado, ya que un estudio que no es válido, no es creíble ni tampoco útil. Los datos de dudosa calidad pueden conducir a errores importantes que no solo debilitan la comprensión del tema o fenómeno en estudio, sino que también conducen a la transmisión de errores, a deducciones limitadas o erradas, a planificaciones inexactas y a una inadecuada toma de decisiones.

Por tanto, es siempre responsabilidad del investigador crear un diseño apropiado para responder a los objetivos de su investigación y elegir las mejores técnicas para validar y establecer la confiabilidad de los instrumentos de recolección de información para, posteriormente, obtener resultados sustanciales con la claridad y precisión necesaria y suficiente propia del método científico, sin perder de vista los aspectos asociados a la objetividad de su investigación, ya que sin ella, sus datos no tendrán asidero en la realidad y no serán útiles para los propósitos en los que fueron pensados.

Específicamente, en el ámbito cuantitativo, es muy importante evitar todo tipo de sesgos (incluyendo los de método y de rasgo) en la elaboración de instrumentos. Esta etapa es muy compleja y a la vez sensible a las inclinaciones personales del investigador y al contexto en que está inmerso, pero sobre todo al conocimiento, comprensión y manejo de los aspectos técnicos asociados a la construcción de encuestas o escalas, entre ellos, la validez y la confiabilidad.

En el caso cualitativo, manejar la neutralidad valórica del investigador (quien determina qué, cómo y dónde va a investigar) es aún más difícil, por tanto, el riesgo de sesgar el levantamiento, posterior análisis de datos y conclusiones es permanente. Es por ello que el lector debe tener siempre presente los constructos de validez y confiabilidad para no adoptar ninguna posición personal que afecte los resultados. Tampoco debe olvidar que las realidades que se analizan cualitativamente son subjetivas, únicas e irrepetibles, por tanto, no necesariamente representan la visión de otros, ni tampoco se pueden presumir como verdades únicas, ya que son solo interpretaciones. Sin embargo, esta flexibilidad no impide reforzar constantemente la importancia y trascendencia de la rigurosidad para que justamente esa aproximación subjetiva de la realidad sea válida, útil y complementaria para la toma de decisiones.

Para finalizar, si bien este artículo aborda la rigurosidad de manera diferenciada para el enfoque cuantitativo y cualitativo, no hay que olvidar que ambas metodologías dialogan constantemente, ofreciendo una visión integral y complementaria que puede ser muy útil para la comprensión y resolución de problemas en el ámbito de las Ciencias Militares.

## **Bibliografía**

- Alvarez-Gayou, Juan Luis. *Cómo hacer Investigación Cualitativa. Fundamentos y Metodología*. México: Paidós, 2009.
- Babbie, Earl. *Fundamentos de la Investigación Social*. Buenos Aires: International Thompson, 2000.
- Briones, Guillermo. *Metodología de Investigación Cualitativa*. Santiago: Universidad de Chile, 2001.
- Burke, Johnson, "examining the validity structure of qualitative research", *Education* 118 2 (1997): 282-292 [en línea], [04.08.2017], [https://www.researchgate.net/publication/246126534\\_Examining\\_the\\_Validity\\_Structure\\_of\\_Qualitative\\_Research](https://www.researchgate.net/publication/246126534_Examining_the_Validity_Structure_of_Qualitative_Research).

- Bunge, Mario. *La investigación científica*, Barcelona: Ariel, 1989.
- Cupani, Alberto. "Acerca de la vigencia del ideal de objetividad científica". *Revista Scientle Studia* 9 3 (2011): 501-525 [en línea], [27.07.2017], <http://www.scielo.br/pdf/ss/v9n3/v9n3a04.pdf>.
- Del Rincón, Delio; Antonio Latorre, Justo Arnal y Antoni Sans. *Técnicas de Investigación en Ciencias Sociales*. Madrid: Dykinson, 1995.
- Denzin, Norman K. *The Research Act: A theoretical introductions of sociological methods*. Englewood Cliffs, N.J: Prentice Hall, 1989.
- Descartes René, *Discurso del Método*. Madrid: FGS, 2010.
- Eisenhart, M. and Kenneth Howe. "Validity in Educational Research". Eds. M. LeCompte, W. Millroy, and J. Preissle, *The Handbook of Qualitative Research in Education*. San Diego: Academic Press, 1992, [en línea], [20.07.2017], [http://nepc.colorado.edu/files/Eisenhart\\_Howe\\_Validity\\_in\\_Ed\\_Research.pdf](http://nepc.colorado.edu/files/Eisenhart_Howe_Validity_in_Ed_Research.pdf).
- Hernández, Roberto; Carlos Fernández y Pilar Baptista. *Metodología de la Investigación*. México: McGraw-Hill, 2014.
- Kerlinger, Fred y Howard, B. Lee. *Investigación del comportamiento. Métodos de Investigación en Ciencias Sociales*. México: McGraw-Hill, 2001.
- LeCompte, Margaret and Judith Goetz. "Problems of Reliability and Validity in Ethnographic Research". *Review of Educational Research* 52 1 (March 1982): 31-60 [en línea], [19.07.2017], <https://pdfs.semanticscholar.org/7253/c6cd672281576a96db1037f135ce3e78fe41.pdf>.
- Morse, Janice; Michael Barrett, Maria Mayan, Karin Olson and Jude Spiers. "Verification Strategies for Establishing Reliability and Validity in Qualitative Research". *International Journal of Qualitative Methods* 1 2 (2002): 13-22, [en línea], [01.08.2017], <http://journals.sagepub.com/doi/pdf/10.1177/160940690200100202>.
- Pérez-Gil, José Antonio; Salvador Chacón y Rafael Moreno. "Validez de constructo: el uso de análisis factorial exploratorio-confirmatorio para obtener evidencias de validez", *Revista Psicothema* 12 2 (2000): 442-446, [en línea], [05.07.2017], <http://www.psicothema.com/pdf/601.pdf>.
- Rodríguez, Gregorio; Javier Gil Flores y Eduardo García. *Metodología de la Investigación Cualitativa*. Málaga: Aljibe, 1996.
- Ruiz Olabuénaga, José Ignacio. *Metodología de la Investigación Cualitativa*. Bilbao: Universidad de Deusto, 1996.
- Salkind, Neil J. *Métodos de Investigación*. México: Prentice Hall, 1998.
- Sierra Bravo R. *Técnicas de Investigación Social. Teorías y Ejercicios*. Madrid: Thomson, 2007.
- Shutt, Russel K. *Investigating the Social World. The Process and Practice of Research*. Canada: Sage Publications, 2012.
- Swanborn, Peter. "A common base for quality control criteria in quantitative and qualitative research". *Quality and Quantity* 30, Issue 1 (February 1996): 19-35, [en línea], [06.07.2017], <https://link.springer.com/article/10.1007/BF0013983>