

*Melissa García Meraz*<sup>\*1</sup>

# MEDICIÓN EN PSICOLOGÍA: DEL INDIVIDUO A LA INTERACCIÓN

---

\*UNIVERSIDAD AUTÓNOMA DEL ESTADO DE HIDALGO

La evaluación es necesaria e indispensable para el psicólogo que desea conocer cierto constructo psicológico; más aún, es indispensable en la intervención psicológica. ¿Cómo sabemos que cierto tratamiento funciona?, ¿cómo sabemos que hemos logrado incrementar la autoestima de nuestros pacientes?, ¿son los resultados de nuestra intervención significativos y relevantes? El diagnóstico que realicemos acerca de una situación particular dependerá de los instrumentos o técnicas que utilicemos para su evaluación. Las técnicas utilizadas pueden consistir en entrevistas y cuestionarios cerrados hasta instrumentos válidos y confiables; la utilización de estos últimos implica la realización de mediciones y cuantificaciones.

La evaluación en psicología implica una gran variedad de técnicas que los profesionistas pueden utilizar con la finalidad de obtener un diagnóstico adecuado y certero. Johnston, Frech, Bonnetti y Johnston (2004) mencionan que en la evaluación en psicología pueden utilizarse diversas técnicas:

---

<sup>1</sup> La doctora García Meraz es profesora investigadora del área académica de Psicología de la Universidad Autónoma del Estado de Hidalgo, pertenece al Sistema Nacional de Investigadores, nivel candidato.

- a) **Entrevistas:** pueden ser utilizadas como una técnica para recoger información demográfica, autorreporte de conductas, creencias y sentimientos, entre otros.
- b) **Cuestionarios o instrumentos psicométricos:** este tipo de evaluación refiere a una medición realizada por medio de instrumentos válidos y confiables.
- c) **Observación:** implica la evaluación por medio del registro preciso de conductas.
- d) **Psicofisiología:** realizada por medio de la evaluación de reacciones fisiológicas como la actividad cardiovascular.
- e) **Índices clínicos:** la temperatura y el pulso.
- f) Finalmente, información obtenida a través de **registros médicos**.

Todas estas formas de evaluación nos permiten un diagnóstico adecuado. En particular, la evaluación a través de la utilización de instrumentos psicométricos es una opción muy frecuente en psicología. Los instrumentos son una opción práctica ya que en muchos casos, son de autoaplicación; su administración requiere de poco tiempo y permiten realizar una evaluación adecuada con más facilidad. Sin embargo, cuando utilizamos una prueba psicométrica debemos preguntarnos varias cosas: ¿es el instrumento válido?, ¿evalúa de manera adecuada aquello que pretende evaluar?, ¿la prueba es confiable?, es decir, ¿los resultados obtenidos por medio de la prueba son consistentes a través del tiempo? Estas preguntas son fundamentales cuando decidimos, o no, utilizar una prueba para realizar un diagnóstico psicológico. De igual manera, si al plantearnos una intervención decidimos evaluar los resultados a través de los puntajes obtenidos en un instrumento, antes y después de la intervención, necesitamos asegurarnos de que el instrumento que utilizamos sea válido y confiable. De esta manera, seremos capaces de contrastar la puntuación de ambas mediciones y saber si las diferencias entre ambas son significativas, finalmente, podremos conocer si nuestra intervención ha sido adecuada o no para el grupo y cultura estudiados.

## SOBRE EL PROCESO DE CONSTRUCCIÓN DE INSTRUMENTOS

Davis (2008) describe una serie de pasos que son necesarios para construir de manera adecuada un instrumento de medición:

- a) Es necesario realizar una serie de investigaciones previas o estudios preliminares que incluyan entrevistas a profundidad, grupos focales, etc., con el objetivo de que los reactivos que componen el instrumento correspondan a la definición del constructo, no sólo de forma teórica sino empírica.
- b) Determinar de qué manera serán recopilados los datos preliminares y del instrumento final, es decir, si mediante entrevistas cara a cara, cuestionarios, vía correo electrónico, entrevistas por teléfono o en grupos de trabajo.

- c) Determinar el tipo de preguntas que serán utilizadas. Pueden ser preguntas abiertas, cerradas, tipo escala Likert, o alternadas, tomando en cuenta 1) el formato de respuesta, 2) el número de reactivos, 3) la redacción de los reactivos; este punto es fundamental ya que los reactivos se deben redactar con un solo sentido, es decir, ningún reactivo debe evaluar más de dos cosas ya que confundirá a los participantes y no logrará evaluar de manera adecuada ninguna de las dos, 4) El orden en el cual se presentarán los reactivos. Si cubren varias dimensiones, entonces primero deberán presentarse los que evalúan hechos y conductas y después las perceptuales y actitudinales. Si se considera que puede haber reactividad, esto es, que los reactivos influyan sobre los participantes, entonces la presentación debe realizarse de forma contrabalaceada.
- d) Determinar el tipo de análisis al que serán sometidos los datos: análisis exploratorios, descriptivos o inferenciales.
- e) Se debe decidir que más se le preguntará a los participantes, por ejemplo: edad, sexo, escolaridad, estatus de pareja, estado civil, etcétera.

De acuerdo con Juve (2008), la utilización de pruebas con reactivos de opción múltiple plantean varias ventajas, entre las que se encuentran: mejores puntajes de confiabilidad, la posibilidad de realizar el estudio en una muestra amplia, mejores y más eficientes puntajes, así como la posibilidad de obtener mayor información diagnóstica. Las dificultades –de acuerdo con Juve– estarían asociadas con el problema de crear reactivos cuando el tema es muy complejo; se incrementa el tiempo de respuesta para las pruebas que evalúan algún tipo de maestría, existe algún grado de dificultad para proporcionar beneficios a los participantes, además, existe la posibilidad de que los resultados se vean influenciados por la capacidad de lectura de los participantes.

La construcción de instrumentos es tema fundamental para los psicólogos: conlleva toda una planeación que comienza con los estudios previos que anteceden a la construcción de los reactivos hasta el aseguramiento de que la prueba es válida y confiable. Es decir, que realmente mide lo que pretende medir y que lo hace de manera consistente.

## SOBRE LA VALIDEZ

Temas como la validez y la confiabilidad no son triviales para los psicólogos, son constructos que deben contemplarse en todas las áreas de la psicología, ya sea en el caso de un psicólogo experimental que trata de reconocer la significatividad de una variable manipulada de manera experimental, un psicólogo educativo que realiza una evaluación en un grupo escolar o un psicólogo social que busca evaluar el cambio social y cultural de un grupo específico. En los tres casos, el psicólogo deberá verificar que aquello que ha evaluado lo ha hecho de manera correcta, asegurando la exactitud, firmeza, solidez y credibilidad de

sus hallazgos (Barron, Brown, Egan, Gesualdi y Marchuk, 2008). La validez implica que el juicio que hemos realizado, la evaluación o diagnóstico, se refleja de manera adecuada y precisa a través de los datos empíricos que obtenemos (Rogelberg, 2004).

Cuando hablamos de un instrumento de medición, la validez implica que hemos medido de manera adecuada aquello que pretendemos medir. Decimos que la validez está relacionada con la certeza y precisión de nuestra medición, es decir, medimos de manera correcta el constructo que queremos o pretendemos medir (Bryman, 1988). Por ejemplo, en 2004, Flores y Díaz-Loving construyeron un instrumento para medir el grado de *asertividad* en la población mexicana. Una de las primeras observaciones de los autores fue que para desarrollar la escala hacía falta algo más que la traducción de la prueba original del inglés al español. Los autores señalan, de manera precisa, que en Estados Unidos, la asertividad se refiere a la defensa de los derechos individuales. Las personas asertivas son directas al pedir que se respeten sus derechos; y al solicitar aquello que desean. En cambio, en México no sucede así, por efecto de la cordialidad, la cultura mexicana se maneja de manera indirecta. En general, los mexicanos no somos tan directos como los estadounidenses porque en nuestra cultura sería una forma ruda e, incluso, grosera de comportarse. Esto no significa que en México las personas no sean asertivas, sino que el constructo mismo necesita replantearse y acomodarse a la cultura mexicana. Por ello, el simple proceso de traducción al español no es suficiente para poder evaluar la asertividad en México. En este caso, la validez implica que nuestro instrumento mide de manera adecuada y precisa la asertividad, y no otro constructo. Si en nuestro país aplicáramos la prueba de asertividad que se utiliza en los Estados Unidos ésta no tendría validez ya que en ese país este constructo tiene un significado diferente al que se le da en el nuestro. La prueba de asertividad desarrollada en México es válida porque refleja lo que significa realmente el constructo entre la población mexicana. Lo mismo sucede en el caso de que decidiésemos medir la inteligencia, necesitamos asegurarnos de que la prueba que utilizemos pueda medirla de manera adecuada. Una prueba válida le permite al psicólogo tener la certeza de que está evaluando con precisión aquello que desea conocer, además, obtiene la información adecuada para realizar un tratamiento, diagnóstico o intervención.

Lo fundamental de la validez radica en el hecho de que conocemos de manera amplia el constructo que deseamos evaluar. Por ejemplo, cuando un investigador decide construir y validar una prueba acerca del constructo de ansiedad generalizada debe, en primera instancia, conocer a profundidad el constructo evaluado, reconocer la sintomatología, las características esenciales que llevarán a una evaluación adecuada, su caracterización en cierto grupo social, cultural y del desarrollo, con el fin de crear una prueba que refleje realmente aquello que el constructo significa. Este primer paso es fundamental en la construcción de una prueba válida. Si la validez se refiere al hecho de que la prueba mide de manera adecuada y precisa cierto constructo, entonces el primer paso para conocer lo que significa el constructo es fundamental. Algunas veces el constructo

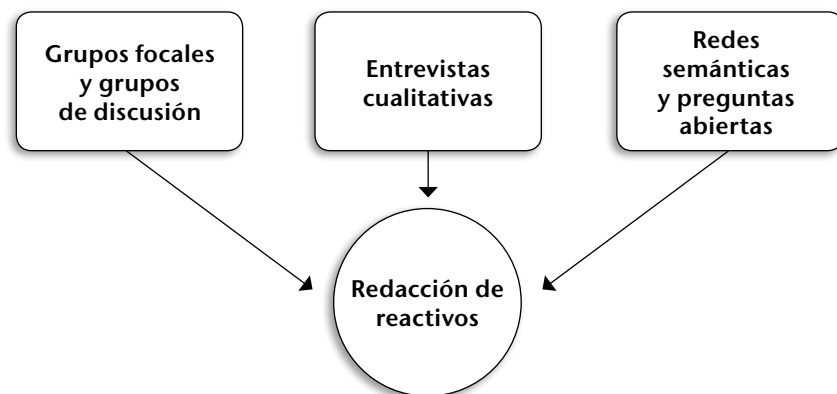
es claro y preciso, existe mucha literatura al respecto y es fácil reconocer qué preguntas debe contener el instrumento que lo evalúe. Sin embargo, en la mayoría de los casos, es deseable y aun necesario realizar estudios previos que nos lleven a reconocer de manera adecuada aquellos rasgos o características que definen el constructo. Antes de redactar las preguntas que constituirán el cuerpo del instrumento, el investigador puede realizar una serie de estudios cualitativos o preguntas abiertas que lo lleven a identificar qué significado tiene el constructo.

Los estudios previos le permiten al psicólogo que construye el instrumento conocer aquellos rasgos, características, conductas o aspectos que definen el constructo en la población estudiada. Por ejemplo, en el *capítulo 2* de este libro, Saucedo-Molina y Guzmán realizan la construcción y validación de un instrumento que evalúa los factores de riesgo asociados a trastornos alimentarios en púberes. De manera precisa, las autoras realizan una investigación teórica acerca de la definición de trastornos alimentarios, después, analizan los diversos instrumentos que se han utilizado en la psicología para evaluarlos. Sin embargo, y aún y cuando las autoras conocen de manera precisa el constructo, realizan un primer estudio piloto en el que simulan grupos de enfoque de ocho participantes (cuatro niñas y cuatro niños) que discuten los diversos factores biológicos, psicológicos y socioculturales que actúan como factores de riesgo en el desarrollo de trastornos de la conducta alimentaria. Esta primera fase exploratoria permite a los investigadores conocer de manera adecuada las características y dimensiones que constituyen el constructo y que necesitarían ser incluidas en el instrumento. Estas fases exploratorias son fundamentales ya que permiten conocer las dimensiones del constructo a evaluar y, además, porque en algunos casos se presentan como fases piloto en las que se presenta el instrumento a un pequeño grupo de participantes que serán evaluados, creando un diálogo en el que los participantes comentarán los aspectos más relevantes del instrumento y los que les parezcan irrelevantes.

En el *capítulo 5* Cruz y Díaz-Loving realizan la validación de la escala de deseo sexual en población mexicana, para ello, los autores proceden como en el caso anterior, primero revisan el constructo en la literatura, analizan de manera detallada los instrumentos más relevantes que evalúan el constructo y, finalmente, decidieron realizar una serie de estudios exploratorios que les permitirán conocer el constructo a fondo, así como redactar preguntas significativas. Como primer paso realizan un estudio con redes semánticas en el que participan 122 personas, el estímulo utilizado es la palabra “deseo sexual”. De este estudio, los autores recopilan una serie de términos definidores como amor, deseo, necesidad, gusto, placer, atracción y gusto. También presentan a los participantes una pregunta abierta: ¿Qué significa “deseo sexual”? Los autores sintetizan los resultados en ocho categorías que agrupan las motivaciones, sensaciones, emociones, objetos, consumación, finalidad y temporalidad del deseo sexual. Así, utilizan la información obtenida a través de estas técnicas para redactar una serie de reactivos para presentarla a una población más amplia.

De igual manera, en los *capítulos 4 y 7* presentamos la validación de un instrumento de elección de pareja y uno de disolución de la relación. Ambos tienen como antecedente la realización de 70 entrevistas cualitativas que fundamentaron las áreas que componen tanto la elección como la disolución y que permitieron redactar los reactivos de manera adecuada. Mientras que Rivera Aragón, Díaz-Loving, Flores Galaz y Montero realizan la validación del Inventario Multidimensional de Celos con base en una revisión de la literatura y la realización de grupos focales que permiten ahondar más sobre los celos y sus diferentes dimensiones. Con los datos obtenidos mediante ambas técnicas, realizan una lluvia de ideas para construir los reactivos del inventario.

Es así que, con el fin de obtener una mayor representatividad del constructo y de la redacción de las preguntas que constituirán los instrumentos, los constructores de la prueba pueden optar por estudios cualitativos o exploratorios, como los grupos focales, grupos de discusión, entrevistas cualitativas, redes semánticas y preguntas abiertas.



**Figura 1.** Estudios exploratorios que preceden a la construcción de los reactivos del instrumento psicométrico.

La validez de un instrumento de medición es tan importante en la práctica del psicólogo que su uso comienza con el investigador que la construye y termina con el psicólogo que la utiliza en la práctica diaria de su profesión. Suponga que usted se dedica a la evaluación de pacientes con problemas emocionales, una buena parte de su consulta se centra en el trabajo con pacientes que padecen depresión. Por tanto, debe conocer de manera adecuada y precisa el constructo de la depresión, debe reconocer la sintomatología, es decir, debe ser un experto en el tema. Si desea evaluar la depresión de manera adecuada y decide utilizar un instrumento que evalúe depresión, debe asegurarse de que la prueba sea válida, de que los psicólogos que la construyeron hayan reportado la forma en que obtuvieron o demostraron la validez del instrumento. Es así que un primer paso en la construcción y el uso de una prueba implica conocer de manera adecuada el cons-

tructo, para reflejarlo de manera adecuada en los reactivos que conformarán la prueba. Este tipo de validez comúnmente se denomina validez de contenido. A ella se suman la validez de criterio externo y de constructo.

## VALIDEZ DE CONTENIDO

Una vez que el investigador conoce de manera adecuada el constructo en cuestión, a través de una revisión exhaustiva de la investigación más relevante y la realización de estudios cualitativos o piloto que le ayuden a reconocer la autenticidad del constructo en su cultura y grupo social particular, es tiempo de redactar los diferentes reactivos que constituirán el instrumento de medición. La validez de contenido implica un primer paso en la validación del instrumento, podemos decir que si queremos obtener un instrumento válido, nuestro instrumento debe, de una o otra forma, tener validez de contenido. Por ello, Groth-Marnat (2003) enfatiza que esta validez debe ocupar el primer punto de interés de aquellos psicólogos que desean construir un instrumento de medición. Para el autor, esta validez refleja el significado del constructo en los reactivos que ha diseñado el psicólogo. Por lo cual, se encuentra muy cercana a una validez de facie, una validez que corresponde al juicio subjetivo de los usuarios del instrumento. Es simplemente, una primera evaluación que permite afirmar o no que el instrumento mide aquello que pretende medir. Sin embargo, y aún cuando este es el primer paso certero en la construcción de un instrumento de medición, no debe ser el único, ya que sólo refleja la opinión, en todo caso subjetiva, de aquel que construye el instrumento, o del usuario que a primera vista emite el juicio de que la prueba mide aparentemente aquello que pretende medir (Barron, Brown, Egan, Gesualdi y Marchuk, 2008). La validez de contenido es tan sólo el inicio de la construcción de una prueba. Fundamental pero no suficiente, a ella se le suman la validez de criterio externo y de constructo.

Fitzpatrick, Davey, Buxton y Jones (2001) mencionan que si bien la validez de facie y la de contenido están altamente relacionadas, no son lo mismo. La validez de facie es tan sólo un primer juicio que a simple vista nos indica si un instrumento mide de manera adecuada aquello que pretende medir, mientras que la validez de contenido examina de manera más detallada si el contenido del constructo se ve reflejado en los reactivos que constituyen el instrumento, reflejando de manera precisa las dimensiones de interés que teóricamente deben constituirlo.

## VALIDEZ DE CRITERIO EXTERNO

Una vez que hemos recabado la información más relevante acerca del constructo que deseamos evaluar, es momento de dar el siguiente paso y comprobar de forma más exigente

la validez del instrumento de medición. Si bien, la validez de contenido se mantiene como una mera valoración acerca de que el instrumento mide de manera adecuada aquello que pretende medir, cuando utilizamos la validación de criterio externo, como su nombre lo dice, utilizamos un criterio externo a nuestro instrumento. De acuerdo a Groth-Marnat (2003), la validez de criterio también es conocida como validez empírica o predictiva. Como criterio externo podemos utilizar los puntajes de algún otro instrumento o la ejecución de una conducta fuera de la medición que realizamos. Es empírica porque ya no depende de nuestra evaluación subjetiva sino de una contrastación directa con la realidad. La validez de criterio externo se divide en validez concurrente y predictiva.

Por ejemplo, García, Rivera Aragón, Díaz Loving y Reyes Lagunes (2004) realizaron un estudio cualitativo-cuantitativo con 60 participantes, 15 parejas que reportaron interacciones violentas y 15 parejas con una interacción sin violencia. A los participantes se les pidió que contestaran el instrumento de estrategias y estilos de poder, estilos de comunicación y enfrentamiento en la pareja, así como la realización de una entrevista cualitativa. Las estrategias de poder fueron evaluadas con un instrumento creado por Rivera Aragón (2000). Con base en estudios cualitativos, la autora construyó una prueba que evalúa las formas o estrategias que las personas utilizan para lograr que sus parejas realicen lo que ellas desean. Por ejemplo, algunas personas piden las cosas de manera cariñosa, utilizan argumentos razonables, lo piden en nombre del amor que los une, etc., todas estas estrategias son positivas. Sin embargo, también existen estrategias negativas, como el uso de la fuerza, la coerción y el chantaje. Dentro de estas estrategias, la estrategia sexual negativa implica que, ante la necesidad de convencer a la pareja, la persona se rehúsa a tener relaciones sexuales o intimar con ella. La hipótesis de nuestra investigación era que, aquellas parejas que reportaban violencia utilizarían en mayor medida esta estrategia. Como se esperaba, las parejas en interacciones violentas utilizaban esta estrategia en mayor medida que el otro grupo. Sin embargo, cuando se realizaron análisis estadísticos más a fondo, se encontró que son las mujeres en este tipo de interacción quienes utilizan más esta estrategia en comparación con cualquier otro grupo, incluso que sus compañeros masculinos. Dentro de los resultados de las entrevistas cualitativas, éste fue un tema recurrente en el caso de las mujeres en interacciones violentas. Las dos evaluaciones, tanto la entrevista como la aplicación de los instrumentos, se realizaron en tres sesiones consecutivas. Los resultados mostraron que ambas evaluaciones coincidieron en que las mujeres en interacciones violentas utilizan en mayor medida la estrategia sexual negativa. La entrevista cualitativa funciona como un criterio externo para validar el instrumento. Es decir, se utilizan dos técnicas que, finalmente, llegan a los mismos resultados. Este tipo de validez de criterio externo se denomina validez concurrente. Podemos, entonces, concluir que el instrumento de estrategias de poder, mide de manera adecuada la estrategia sexual negativa. Para cumplir con los requisitos de una validación concurrente, es necesario que las dos técnicas sean aplicadas al mismo tiempo o de manera muy cer-

cana en el tiempo. La validez concurrente implica que tengamos otro instrumento que mida lo mismo o estructuramos de manera adecuada otra técnica que evalúe el mismo constructo (Clark-Carter, 2004).

El segundo criterio externo, la validez predictiva, implica una medición que se realiza en el futuro, es decir, el criterio externo es una evaluación que se realiza tiempo después de la primera evaluación. Por ejemplo, supongamos que se desea realizar una prueba de aptitudes en alumnos que han ingresado recientemente al nivel de educación superior. El instrumento evalúa diversas aptitudes necesarias para la culminación exitosa de la carrera en lenguas. Al cabo de tres años, se regresa a evaluar a los participantes y se encuentra que aquellos con más aptitudes son los alumnos que han terminado en tiempo y forma su educación superior. En este caso, tenemos un criterio externo en el futuro que nos indica que la prueba midió de manera adecuada estas aptitudes. Este tipo de validación es fundamental para la selección de alumnos, por ejemplo, de posgrado. Todos los programas de posgrado tienen pocos lugares que ofrecer y muchos solicitantes. Por ello, los evaluadores de los candidatos tienen que elegir a aquellos que mejor cumplan con el perfil de ingreso y terminarán en tiempo y forma con el programa educativo.

Es así, que la validez de criterio externo echa mano de puntajes, pruebas o conductas que se dan fuera del instrumento y que nos permiten afirmar que el instrumento mide de manera exacta y precisa lo que pretendemos medir. Groth-Marnat (2003) afirma que el uso de la validez concurrente y predictiva depende del uso que se le dé a una prueba. Si se realiza selección, clasificación de personal o de población estudiantil es más adecuado utilizar la validez predictiva. En cambio, cuando se está ante diagnósticos de tipo clínico o de predicción de conducta es mejor utilizar la validez concurrente. Imagine que, como psicólogo clínico que se especializa en el tratamiento de pacientes con cuadros depresivos, debe realizar un diagnóstico y para ello utiliza tanto la versión estandarizada en México del inventario de depresión de Beck, como una entrevista clínica. En este caso, tanto el instrumento como la entrevista deben permitirle diagnosticar depresión o descartarla. La entrevista servirá como un criterio externo para validar el diagnóstico realizado por el instrumento y, por supuesto, le servirá al psicólogo clínico para realizar un diagnóstico válido sobre la depresión en la práctica clínica. En cambio, sería poco ético utilizar un criterio a futura, ya que nuestro paciente necesitará de la evaluación y diagnóstico lo más pronto posible. Clark-Carter (2004) menciona que, en ambos casos, el evaluador no debe conocer los resultados de la primera evaluación antes de realizar la segunda ya que puede contaminar los resultados y afirmar que existe validez de criterio cuando en realidad no existe. En el caso del estudio realizado con parejas en interacciones violentas y sin violencia, es necesario que la entrevista se realice antes del procesamiento estadístico de los instrumentos, ya que de otra forma, el investigador podría contaminar los datos de la entrevista, sesgando sus percepciones durante la entrevista, tratando de corroborar aquello que encontró en los datos estadísticos, afirmando que existe validez de criterio externo cuando en realidad podría no ser así.

Para Barron, Brown, Egan, Gesualdi y Marchuk (2008), el criterio externo puede ubicarse incluso en el pasado, comparando los resultados con puntajes o evaluaciones ya realizadas en los participantes. Este tipo de validez es llamada validez retrospectiva.

## VALIDEZ DE CONSTRUCTO

Se puede decir que el tercer tipo de validez corrige los errores o limitaciones de las aproximaciones anteriores. La validez de constructo implica que hemos hecho explícito el constructo (Clark-Carter, 2004). Groth-Marnat (2003) explica que la validez del contenido está basada en un juicio personal, mientras que el criterio implica reconocer qué tipo de medida es utilizada como criterio externo. En cambio, la validez de constructo no depende de la planeación de una medida externa ni se centra solamente en el juicio subjetivo del psicómetra. En cambio, el autor propone tres pasos fundamentales para asegurar la validez de constructo:

- 1) El investigador debe asegurarse de conocer teóricamente el constructo, realizar una búsqueda exhaustiva o representativa de la literatura más reciente, además, debe asegurarse de reconocer la orientación teórica del constructo, si responde a alguna teoría o posición teórica particular,
- 2) Debe reconocer las relaciones o posibles relaciones con otros constructos o variables psicológicas, y
- 3) Finalmente, debe realizar ciertas hipótesis acerca del constructo y probarlas de manera empírica.

Estos tres pasos implican que el instrumento es probado en una población significativa para el constructo, por ejemplo, si se desea validar una prueba de depresión, realizamos el estudio en pacientes que reportan sintomatología relacionada con la depresión y pacientes que no lo hacen, con ello podemos establecer diferencias que serán contrastadas en la realidad. Recordemos el estudio cualitativo-cuantitativo realizado en las parejas que reportaron conflicto y las que no. Junto con el instrumento de estrategias de poder, se aplicó la prueba de estilos de comunicación de Sánchez Aragón (2002). La hipótesis planteada fue que las parejas en interacciones violentas no se comunicaban y si lo hacían tenían estilos de comunicación negativos, como por ejemplo, ser agresivos, demandantes y sarcásticos al comunicar algo a sus parejas. Como se había planteado, los resultados demostraron que las parejas en interacciones violentas utilizan estilos de comunicación negativos hacia sus parejas en comparación con el grupo que no reportó interacciones violentas. De esta manera, el instrumento muestra validez de constructo ya que muestra evaluaciones diferenciales en grupos que se plantea diferirán en ciertos constructos. En otro caso, Groth-Marnat (2003) menciona que si se desea validar un instrumento de

ansiedad, se establece que el instrumento tendrá correlaciones positivas en individuos que son evaluados durante situaciones que les provocan ansiedad. En las situaciones anteriores, la contrastación de hipótesis en grupos específicos revelan que la prueba mide realmente lo que pretende medir y no sólo que el investigador opina que así lo hacen. Este tipo de validez se relaciona con la validez conocida como Known-groups, en este caso, el investigador realiza un estudio en grupos que sabrá con anterioridad deberán puntuar de manera diferencial (Barron, Brown, Egan, Gesualdi y Marchuk, 2008). Como en el caso de las parejas en interacciones violentas o los participantes en situaciones estresantes.

Otros métodos utilizados para verificar la validez de constructo son la validez convergente y divergente (Kenny, Alvarez, Donohue y Winick, 2008). La primera implica que si una persona obtiene puntajes positivos para una escala de depresión, obtendrá puntuaciones positivas en un instrumento de malestar emocional. Decimos que las dos medidas convergen. La validez divergente implica que los puntajes del instrumento se correlacionan de manera negativa con otro constructo. Por ejemplo, si una persona puntúa alto en la medición de depresión, obtendrá puntuaciones bajas en constructos como el bienestar subjetivo. Las correlaciones altas y significativas, ya sea positivas o negativas, sugieren una alta validez de constructo, aunque en última instancia, refiere Satsky (2008), la validez depende del propósito y de la disciplina.

En el capítulo 3 de este libro, Del Castillo y Reyes-Lagunes muestran la validación convergente y divergente del cuestionario “Áreas problema en diabetes” (PAID) que evalúa el malestar emocional. Para realizar la validación, los autores relacionan los puntajes del PAID, de 240 pacientes con diagnóstico confirmado de diabetes tipo 2, con el puntaje del inventario de Depresión de Beck, el Índice de Bienestar Psicológico OMS-5 y el Instrumento de Autoeficacia en Diabetes. Como esperaban los autores, los datos muestran que el cuestionario de Áreas Problema en Diabetes obtiene correlaciones significativas y positivas con los puntajes del Inventario de Depresión de Beck y correlaciones significativas y negativas con la prueba de Bienestar Psicológico y Autoeficacia, demostrando así validez convergente y divergente.

Groth-Marnat (2003) menciona que otra forma de demostrar la validez de constructo es a través de la realización de una intervención experimental. Por ejemplo, si realizamos un estudio experimental para modificar cierta variable y aplicamos el instrumento antes y después del tratamiento, encontrando diferencias significativas entre la primera y la segunda evaluación, entonces concluimos que el instrumento es válido. Aunque para corroborar este tipo de validez, debemos asegurarnos en estudios previos que la intervención que utilizemos es ha probado ser realmente significativa por otros medios, evitando con ello costos y posibles daños a los participantes.

La validez de constructo también es probada por medio de técnicas estadísticas como el análisis factorial. Esta técnica es utilizada como una forma de reducir un amplio número de reactivos en dimensiones más pequeñas que muestran las dimensiones que el investigador propuso en un principio.

Los instrumentos presentados en este libro constituyen un excelente ejemplo de construcción y validación de instrumentos. En todos los casos, los autores realizan la reconstrucción del constructo a evaluar, describiendo lo que éstos significan y anotando las diversas formas en las cuales se ha evaluado típicamente el constructo. Además, reportan estudios cualitativos o exploratorios previos a la construcción de los reactivos que les permitirán ahondar más en los constructos psicológicos y reconocer las variaciones culturales y sociales de los grupos estudiados. Finalmente, someten los instrumentos a validación factorial, con lo cual reducen el número de reactivos en dimensiones que explican el constructo estudiado. Todos ellos utilizan el programa estadístico SPSS para realizar la validación de los instrumentos.

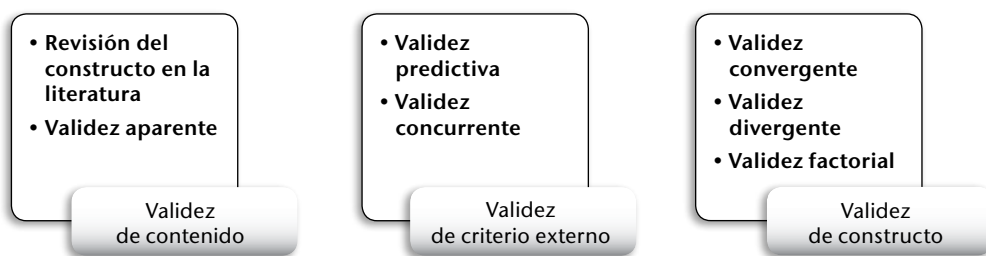


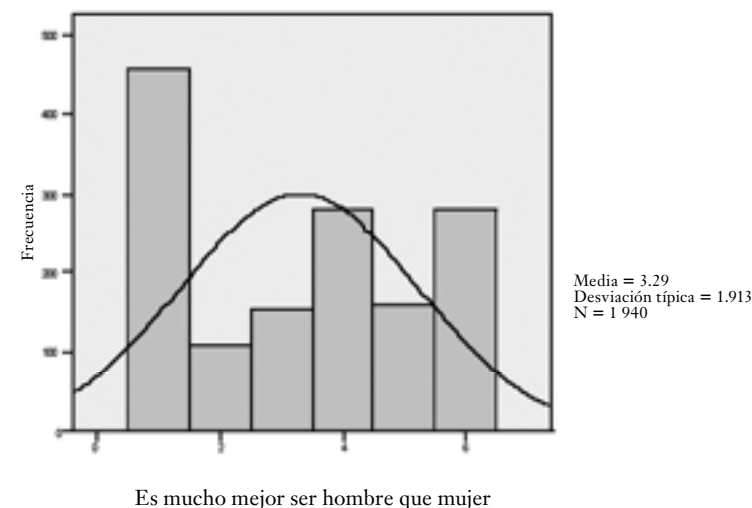
Figura 2. Tipos de validez contempladas en la validación de instrumentos psicométricos.

### LA VALIDACIÓN DE CONSTRUCTO A TRAVÉS DEL ANÁLISIS FACTORIAL

El análisis factorial es una técnica estadística utilizada para asegurar la validez de constructo. Esta técnica permite establecer relaciones cuantitativas en el constructo, como tal, los psicólogos evalúan aspectos de algo que no es directamente observable. A través del análisis factorial se pueden establecer puntajes cuantitativos que establecen relaciones entre los reactivos del instrumento (Fitzpatrick, Davey, Buxton, y Jones, 2001). El análisis factorial es una técnica multivariante que tiene como propósito principal encontrar la estructura subyacente en una matriz de datos, reduciendo un gran número de reactivos o variables en dimensiones, agrupando los reactivos en estas dimensiones conocidas comúnmente como factores (Hair, Anderson, Tatham y Black, 1999).

Sin embargo, antes de realizar el análisis factorial es conveniente realizar un estudio previo que permita identificar los reactivos que asegurarán un mayor índice de confiabilidad y varianza explicada. En el año 2008, Reyes-Lagunes y García sistematizaron un método que permite a los psicólogos realizar un procedimiento adecuado en la inclusión de reactivos al análisis factorial. El primer paso implica el reconocimiento de los datos, realizando un estudio descriptivo de las frecuencias de cada reactivo. De esta manera, ubicamos la distribución de los reactivos, el sesgo, mediana, media y curtosis. Para el

estudiante que apenas se inicia en el conocimiento de los análisis estadísticos le será beneficiosa la utilización del análisis de frecuencias para reconocer, de manera gráfica, la distribución de los reactivos, así como medidas de tendencia central y de dispersión o variabilidad. La exploración de los datos puede realizarse accediendo a la pestaña analizar el SPSS, y a la pestaña estadísticos descriptivos y frecuencias. Una vez incorporados los reactivos en la caja de análisis se accede a la sección de estadísticos para incorporar al análisis, las medidas de tendencia central, como la media, mediana y moda. Finalmente, en la sección de opciones, seleccionando las opciones histograma con curva normal se obtiene un gráfico de la distribución del reactivo con la curva sobre ella. Este primer análisis nos permitirá reconocer la distribución del reactivo. La *gráfica 1* muestra la distribución del reactivo Es mucho mejor ser hombre que ser mujer.



Gráfica 1. Histograma con curva normal para el reactivo “Es mucho mejor ser hombre que ser mujer”.

Reyes-Lagunes y García (2008) recomiendan una serie de análisis que implican la incorporación al análisis factorial de los reactivos que, o bien están sesgados o se encuentran en una distribución normal, pero nunca mezclar reactivos sesgados y normales, ya que el análisis factorial podría juntar distribuciones y no contenidos. Sin embargo, lo más relevante es que los autores proponen dividir los puntajes de cada reactivo en cuatro grupos (cuartiles), para después comparar el puntaje más alto con el más bajo, es decir, el primero y el último cuartil por medio de una prueba t. Los reactivos donde los puntajes considerados como altos no discriminan a los bajos son eliminados del análisis factorial. Este análisis permite demostrar que los participantes utilizan todas las opciones de respuesta, y que aquellos considerados como puntajes bajos y altos, realmente



lo son. El procedimiento de validación planteado por los autores es descrito en detalle en los *capítulos 4*, y *7* de este libro, con referencia a la validación de los instrumentos de elección y disolución de la relación de pareja. Sin embargo, es necesario mencionar que todos los instrumentos incluidos en este libro realizan este primer paso para reconocer la distribución de los reactivos que se someterán al análisis factorial.

Una vez que se ha establecido la distribución de los datos, el análisis factorial puede llevarse a cabo. Con ayuda del análisis se realiza una serie de interacciones con la finalidad de agrupar los diferentes reactivos en dimensiones. De acuerdo a Pardo y Ruiz (2002), el análisis factorial es una técnica de reducción de datos que permite encontrar grupos homogéneos de variables a partir de un número muy amplio. En nuestro caso, permite agrupar los diversos reactivos del instrumento en grupo o factores. Los factores se unen porque se correlacionan en un nivel alto entre sí, incluso cuando inicialmente se espera que sean independientes unos de otros. Por ejemplo, en el *capítulo 6*, Rivera Aragón y colaboradores realizan la validación de constructo del Inventario Multidimensional de Celos, como ya hemos mencionado, los autores construyeron la prueba basados en una extensa revisión de la literatura así como la realización de grupos focales que permitieron obtener 64 definidoras del constructo celos (por ejemplo: traición, amenaza de pérdida, celos de trabajo, orgullo herido, venganza entre otras). A partir de esta amplitud de definidoras construyen 162 reactivos con respuesta en escala Likert que reflejan cada una de las áreas obtenidas en los estudios cualitativos. Posteriormente, aplican la escala a un grupo de 813 participantes y someten los resultados al análisis factorial obteniendo 12 factores que explican 69.8% de la varianza. Cada factor agrupa un número concreto de reactivos que permiten realizar una medición adecuada. El inventario contempla dos áreas: emociones y sentimientos y cogniciones y estilos. En el primer caso, el primer factor, emociones y sentimientos, agrupa reactivos como “cuando pienso que mi pareja está con otra persona siento coraje”, “me enoja que mi pareja tenga amigos del sexo opuesto” y “muero silenciosamente por una cuestión de celos”. Mientras que en el segundo, el primer factor, Obsesión por la pareja, agrupa reactivos como “quiero que mi pareja piense sólo en mí”, “sospecho que mi pareja me engaña” y “siempre deseo saber con quién está mi pareja”. De esta manera, las 64 definidoras y los 162 reactivos planteados por los autores se agrupan en 12 factores divididos en dos dimensiones.

Para realizar este análisis es necesario conocer la relación entre los diferentes reactivos. Si la correlación es muy alta, entonces será necesario realizar un análisis factorial con rotación oblicua, si la relación es baja entonces se realizará un análisis factorial con rotación ortogonal. El primero implica que los reactivos serán forzados a separarse, por ejemplo, en instrumentos que evalúan la violencia en la relación es frecuente utilizar este tipo de rotación ya que los constructos se mantendrán muy cercanos entre sí y, probablemente, se encuentren en dos o más factores. El primer paso implica realizar una correlación bivariada con el coeficiente Pearson (datos intervalares) con el objetivo

de reconocer si los reactivos están muy relacionados y si es más adecuado realizar una rotación oblicua.

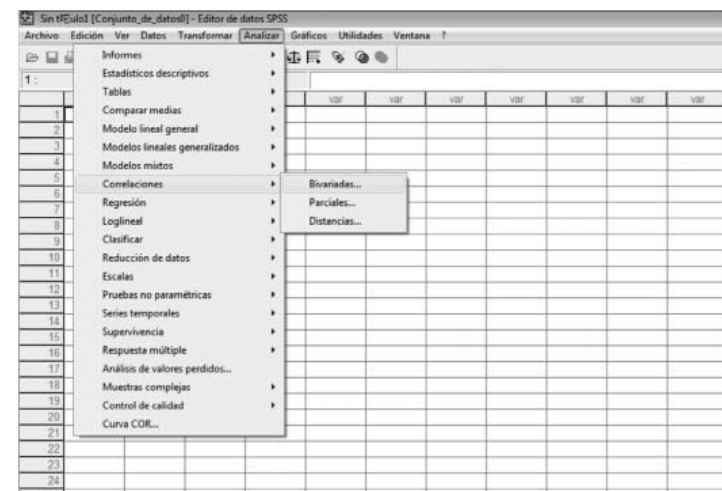


Figura 3. Cuadro de diálogo para realizar el análisis de correlaciones bivariadas.

El segundo paso es ingresar los reactivos en la opción Analizar del SPSS en el menú principal del SPSS, ingresando a la opción Reducción de datos y Análisis factorial.

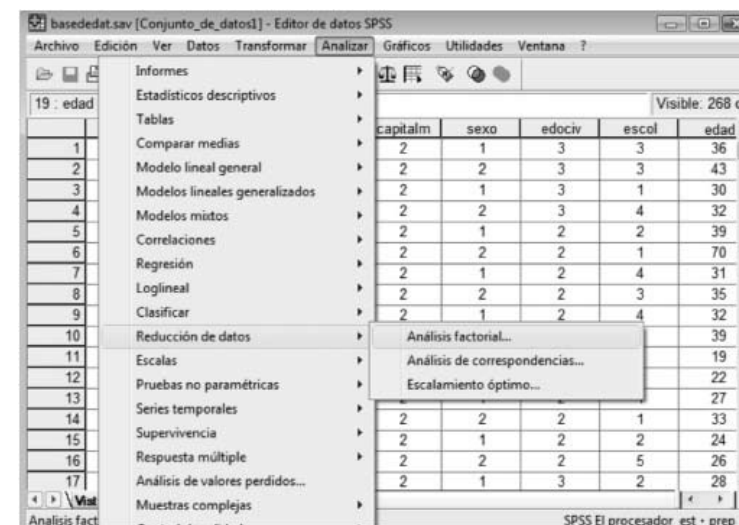


Figura 4. Pasos para realizar el análisis factorial.



Posteriormente, se agregan los datos que permitirán realizar el análisis estadístico. Los botones habilitados en la parte inferior del cuadro de análisis permiten elegir diversas características del análisis. En el primer botón denominado “Descriptivos” permite realizar análisis descriptivos y correlaciones. La segunda opción “Extracción” permite ubicar qué método de extracción se realizará en el análisis: el más utilizado en psicología es el análisis de componentes principales, también es necesario ordenar dentro de esta opción que el programa que extraiga los factores que tengan un autovalor o eigen value mayor a 1, finalmente, esta opción permite establecer un número determinado de interacciones para que converjan los factores. La opción “rotación” permite utilizar el método varimax, oblimin directo, quartimax, equamax y promax. En psicología utilizamos la rotación ortogonal (denominada varimax por el programa SPSS) y la oblicua (denominada Oblimin directa). El botón “Puntuaciones” permite guardar los datos como variables a través de los métodos: regresión, Bartlett y Anderson-Rubin. Finalmente, el último botón, “Opciones”, permite excluir datos con valores perdidos excluyendo casos según lista y según pareja o reemplazarlos por la media de los datos existentes. También permite modificar el formato de presentación de los datos ordenados por tamaño o suprimiendo los valores menores a .10 en adelante. Cada reactivo tiene un peso o puntaje que va de cero a 1. Este valor expresa el peso o cantidad con que cada reactivo contribuye al factor donde se agrupó con otros reactivos.

Una vez que se han ingresado los datos del análisis éste puede ejecutarse activando el botón “Aceptar”. En el visor de resultados aparecen diversos cuadros, que muestran la reducción de los factores. De acuerdo con Pardo y Ruiz (2002), el análisis factorial consta de cuatro fases:

- 1) Realización del cálculo de una matriz que expresa la variabilidad de cada uno de los reactivos del estudio.
- 2) Extracción de los reactivos en un número reducido de factores.
- 3) Rotación de la matriz.
- 4) Estimación de puntajes realizados por los participantes.

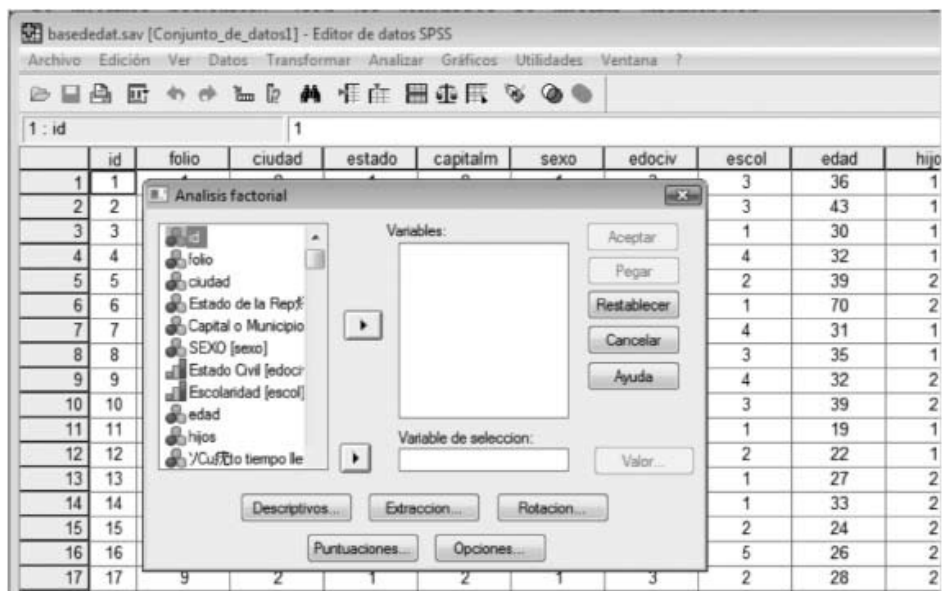


Figura 5. Cuadro de diálogo para elegir la rotación del análisis factorial.

Comunalidades obtenidas en el análisis factorial por componentes principales		
	Inicial	Extracción
t15 Nos llevaríamos mal	1.000	.342
t16 Conviviéramos menos con nuestras familias	1.000	.480
t17 Dejáramos de ser personas complementarias	1.000	.653
t18 Dejáramos de tener gustos e intereses diferentes pero complementarios	1.000	.642
tp1 Dejara de ser una persona simpática	1.000	.466
tp2 Dejara de ser una persona sincera	1.000	.466
tp3 Fuera una persona inequitativa	1.000	.615

Método de extracción: Análisis de componentes principales.

El primer cuadro muestra el método por el cual se llevó a cabo el análisis factorial. El método de extracción de componentes principales asume que se puede explicar el 100% de la varianza. Por ello, la extracción inicial comienza en 1.000, aunque cada reactivo contribuye con un puntaje en particular. Así, el factor “nos llevaríamos mal”, es el que menos explica, ya que de una unidad logra explicar sólo 34.2% (Pardo y Ruiz, 2002).

La figura 6 muestra el número de factores obtenidos, cantidad de varianza explicada por cada factor y en total. En nuestro ejemplo aparecen seis factores con valores propios mayores a 1.

**TABLA 2**  
**Tabla de resultados del análisis factorial, varianza total explicada**

Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción			Suma de las saturaciones al cuadrado de la rotación		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	8.531	47.395	47.395	8.531	47.395	47.395	4.354	24.192	24.192
2	1.711	9.508	56.903	1.711	9.508	56.903	3.658	20.322	44.514
3	1.125	6.249	63.152	1.125	6.249	63.152	3.355	18.638	63.152
4	.896	4.980	68.132						
5	.659	3.660	71.792						
6	.647	3.594	75.386						
7	.572	3.177	78.563						
8	.462	2.565	81.128						
9	.441	2.451	83.579						
10	.423	2.347	85.926						
11	.404	2.242	88.168						
12	.361	2.004	90.172						
13	.329	1.826	91.998						
14	.315	1.752	93.750						
15	.307	1.705	95.455						
16	.301	1.673	97.128						
17	.265	1.475	98.603						
18	.251	1.397	100.000						

En la *tabla 3* se muestra la matriz de componentes rotados donde se incluyen el número de factores y los reactivos que conforman cada uno de ellos. Es relevante mostrar que sólo se muestran aquellos reactivos que pesan más de .40 en cada una de las dimensiones ya que se considera un punto de corte adecuado, un peso inferior demostraría que el factor contribuye poco a este factor.

**TABLA 3**

**Matriz de componentes rotados del análisis factorial**

	Componente		
	1	2	3
t1 Faltaran acuerdos en nuestra relación			.705
t2 Ya no nos apoyamos en nuestras decisiones			.759
t3 Pasáramos poco tiempo juntos			.685
t4 Se terminara nuestra amistad			.699
t5 Dejáramos de platicarnos lo que nos pasa			.689
t6 Se terminara el cariño entre nosotros		.672	
t7 Tuviéramos un nivel educativo desigual	.754		
t8 Dejara de ser mi media naranja	.487		
t9 Faltara la química entre nosotros		.618	
t10 Dejáramos de ver la vida de la misma manera	.643		
t11 Tuviésemos diferentes ideales	.659		
t12 Ya no compartieramos amistades	.792		
t13 Dejáramos de demostrarnos amor		.740	
t14 Fuésemos muy diferentes en nuestra forma de ser	.563		
t15 Nos lleváramos mal		.801	
t16 Convivieramos menos con nuestras familias	.779		
t17 Dejáramos de ser personas complementarias		.569	
t18 Dejáramos de tener gustos e intereses diferentes pero complementarios	.619		

Método de extracción: Análisis de componentes principales.

Método de rotación: Normalización Varimax con Kaiser.

\* La rotación ha convergido en 6 iteraciones.

### OTROS TIPOS DE VALIDEZ

Tanto la validez de contenido, criterio externo y constructo son fundamentales para la construcción y uso de instrumentos psicométricos. En el caso de estudios de corte

experimental, es frecuente el uso de términos como validez interna y externa. La validez externa implica que los resultados encontrados son aplicables y generalizables a una población más grande. Para asegurar la validez externa deben incluirse un mayor número de participantes de diferentes edades y escolaridades que permitan extender los resultados a una población más amplia, de diferentes ambientes y tiempos (Barron, Brown, Egan, Gesualdi y Marchuk, 2008; Clark-Carter, 2004). Cuando los resultados se extrapolan a una población más amplia se le conoce como validez poblacional, cuando se generaliza a diferentes ambientes: validez ecológica, a través del tiempo: validez temporal, en distintos ambientes laborales: validez intraorganizacional (Barron, Brown, Egan, Gesualdi y Marchuk, 2008). De acuerdo con Clark-Carter (2004), dos aspectos son fundamentales para asegurar la validez externa: las condiciones del lugar, por ejemplo, el lugar donde se llevará a cabo el estudio, condiciones, tiempo y los aspectos relevantes de los participantes, como edad y escolaridad, y si ellos representan a una población más amplia.

La validez interna implica que en el estudio se ha evaluado el constructo o variable de manera precisa. De acuerdo con Barron, Brown, Egan, Gesualdi y Marchuk (2008) esto ocurre cuando se tiene un control preciso de la situación. Por ejemplo, en un estudio experimental, donde se han controlado todas las variables extrañas que puedan afectar los resultados. El investigador sabe que los cambios se deben a la manipulación de la variable o constructo que ha sido evaluado de manera precisa. Los estudios experimentales evalúan la relación causa efecto entre dos o más variables, la manipulación de una variable y la observación de si existe o no una relación permite conocer la validez del estudio. Se ha establecido una relación causal o, mejor dicho, en el ámbito experimental se ha demostrado que los cambios en una variable dependiente se deben a la manipulación de la variable experimental (Clark-Carter, 2004). Sin embargo, entre más control haya de las variables dentro del estudio, menor será la posibilidad de tener validez externa y extrapolar los datos a otras situaciones, ambientes y temporalidades. De acuerdo a Sanderson, McKee, Britton, Black, McPherson y Bain (2008), la validez interna está relacionada con la precisión de los resultados en el grupo particular que se ha estudiado, en caso de realizar generalizaciones serán únicamente a personas o pacientes muy similares al grupo experimental. De acuerdo con Clark-Carter (2004) varios aspectos pueden incidir en la validez interna: la selección de participantes, la maduración (los cambios en la variable dependiente se deben a cambios en el desarrollo), la historia (los cambios se deben a eventos de la vida diaria fuera del control del investigador), la instrumentación (cuando se evalúa a los participantes en dos ocasiones, los cambios pueden atribuirse al instrumento más que a lo que se está evaluando), al aprendizaje (se evalúa a los participantes con el mismo instrumento, pero los cambios pueden deberse a que los participantes han adquirido mayor destreza en contestarlo), mortalidad experimental (los participantes deciden abandonar el estudio), imitación (los participantes de un grupo copian o imitan la conducta de otro grupo), desmoralización y compensa-

ción (algunos participantes se pueden sentir más motivados que otros para realizar el estudio), y, finalmente, regresión a la media (los participantes tienden a contestar en un punto neutral, además, la estadística por sí misma tiene una tendencia de regresión a la media).

Groth-Marnat (2003) también menciona que, en el ámbito clínico, puede hablarse de validez conceptual como la capacidad o habilidad del clínico de crear descripciones coherentes de los pacientes, formulando hipótesis que pueden someterse a prueba en la realidad. De acuerdo con el autor no podemos afirmar en abstracto o de manera absoluta que una prueba es válida, lo que si podemos hacer es decir que es válida en un contexto específico para un grupo de personas en particular. Las pruebas son construidas con base en lo que el constructo psicológico en particular significa para cada población social y cultural. Es por ello que muchas veces la simple traducción de una prueba no resulta una medida adecuada. Recordemos el ejemplo de la asertividad, la prueba construida en Estados Unidos es diferente a la construida en México, porque el constructo significa dos cosas diferentes en ambos países.

## SOBRE LA CONFIABILIDAD

Barron, Brown, Egan, Gesualdi y Marchuk (2008) definen la confiabilidad como el grado de estabilidad, previsibilidad, exactitud y consistencia con la que se realiza una evaluación p. 12). Esto implica que las evaluaciones se mantendrán de manera constante a través del tiempo. La confiabilidad tiene como objetivo evitar, o al menos prevenir, los errores que son naturales dentro de toda medición psicológica. Juve (2008), la identifica como el hecho de que las mediciones son constantes o repetibles a través del tiempo, es un índice o, mejor dicho, una proporción que indica el radio entre la variación de los puntajes verdaderos de un instrumento y los puntajes verdaderos. Groth-Marnat (2003) plantea que los errores son naturales dado que son consecuencia de la ejecución de actos humanos. La consistencia se presenta en mayor medida en evaluaciones de la personalidad en comparación con evaluaciones de la ejecución. De hecho, menciona el autor, la confiabilidad tiene como fin estimar el grado de varianza que se presenta en la evaluación generada de un instrumento y que puede ser atribuible al error, recordemos que en psicología no se obtiene la evaluación de manera directa sino tan sólo como una medición indirecta.

Barron, Brown, Egan, Gesualdi y Marchuk (2008) mencionan que la consistencia interna o confiabilidad interna implica conocer como los reactivos de una prueba están relacionados entre sí. Imagine una escala compuesta por diversos reactivos que juntos evalúan cierto constructo, la consistencia implica en qué medida los reactivos evalúan el mismo constructo. De no ser así, diríamos que los participantes no responden de manera consistente el instrumento, o que la serie de reactivos fallan al evaluar el mismo

constructo. Este tipo de consistencia está relacionada con las mediciones que utilizan formato de instrumentos o índices (Bryman, 1988). Todos los instrumentos de medición deben obtener altos puntajes de confiabilidad interna, ya que de esta manera demostrarán que están altamente relacionados entre sí y evalúan el mismo constructo, y que los reactivos son coherentes entre sí.

En cambio, la confiabilidad a través del tiempo implica que al aplicar el instrumento en diferentes periodos en el tiempo se obtienen resultados consistentes en todas las aplicaciones (Bryman, 1988). El método para obtener la confiabilidad a través del tiempo o temporal de un instrumento es llamado: Test-retest. Esta técnica implica que el coeficiente de correlación es calculado correlacionando los puntajes obtenidos por la misma persona en dos administraciones diferentes. La prueba típicamente utilizada es la correlación producto-momento de Pearson. Es decir, se aplica la prueba en un primer momento y después en el futuro a la misma persona. Si la prueba es consistente, entonces la persona obtendrá puntajes muy similares, obtendrá correlaciones altas y significativas en la primera y segunda evaluación. Es importante destacar que el tiempo es fundamental, ya que si se deja transcurrir muy poco tiempo entre la primera y la segunda aplicación puede influir en el recuerdo de la prueba; mientras que demasiado tiempo puede permitir que cambios en el desarrollo y la maduración de los participantes muestren diferencias entre ambos puntajes. Fitzpatrick, Davey, Buxton, y Jones (2001) mencionan que el tiempo adecuado entre ambas aplicaciones deberá encontrarse en un periodo de 2 a 14 días, mientras que el coeficiente de correlación adecuado para mostrar consistencia debería estar en un 0.7, aunque lo deseable son puntajes superiores.

## CONFIABILIDAD INTERNA

La confiabilidad interna de un instrumento de medición puede ser evaluada a través de la aplicación de formas alternas de un mismo instrumento. Esta técnica implica que tenemos dos versiones de un mismo instrumento, ambos deben, por supuesto, medir el mismo constructo. La alta correlación entre los puntajes de ambos instrumentos muestra un alto coeficiente de confiabilidad.

La segunda técnica es llamada split half o dividir en mitades, implica que en lugar de aplicar dos instrumentos paralelos o el mismo instrumento en dos momentos diferentes, se utiliza la misma prueba y la misma evaluación pero dividiéndola a la mitad, correlacionando ambas mitades. Una correlación alta evidencia un alto índice de confiabilidad. El investigador puede dividir la prueba exactamente a la mitad y correlacionar ambos puntajes. Groth-Marnat (2003) plantea que las personas pueden estar cansadas de contestar el instrumento y contestar de manera inadecuada la segunda mitad. Podemos, entonces, optar por dividir de manera aleatoria los reactivos. Sin embargo, entre menor número de reactivos tenga una prueba menor será la confiabilidad (Fitzpatrick, Davey, Buxton y Jones, 2001).

En tercer lugar, la confiabilidad puede ser evaluada con base en análisis estadísticos, como la prueba Kuder-Richardson y el Alfa de Cronbach. Estas técnicas estadísticas evalúan la consistencia interna del instrumento, el método utilizado es la correlación. Si encontramos correlaciones altas entonces decimos que los reactivos evalúan aspectos cualitativamente homogéneos (Reeves, MacLehose, Harvey, Sheldon, Russell y Black, 2001). El coeficiente Alfa de Cronbach se calcula por cada factor, la suma de todos los puntajes da como resultado la confiabilidad total de la prueba. Es necesario anotar que, entre mayor número de reactivos mayor será la confiabilidad de la prueba.

De acuerdo con Satsky (2008), lo ideal es evaluar la confiabilidad de los instrumentos a través del tiempo con diferentes grupos de diversas edades o contextos, para así poder generalizar los resultados. En el caso de los instrumentos de medición, Barron, Brown, Egan, Gesualdi y Marchuk (2008) mencionan que el tipo de confiabilidad utilizada depende de la naturaleza de la medida que utiliza el investigador: confiabilidad Test-retest, confiabilidad interna y confiabilidad interobservadores. Por ejemplo, en el capítulo 8, Díaz Loving y cols., realizan la validación del instrumento de premisas socioculturales planteadas por Díaz Guerrero, todos los reactivos se encuentran ampliamente relacionados entre sí, es decir, todos los reactivos miden el mismo constructo y, por tanto, están muy relacionados. Los autores reportan altos índices de confiabilidad interna obtenido a través del análisis Alfa de Cronbach.

La prueba o técnica utilizada para probar la confiabilidad dependerá de los objetivos del investigador. Satsky (2008) comenta que en el caso de la evaluación de rasgos de personalidad, podemos utilizar una prueba Test-retest ya que la personalidad se presupone como un rasgo relativamente estable a través del tiempo. En cambio, cuando evaluamos constructos como la motivación, tendríamos que utilizar otra técnica, ya que la motivación es variable y a menudo es afectada por diversas condiciones.

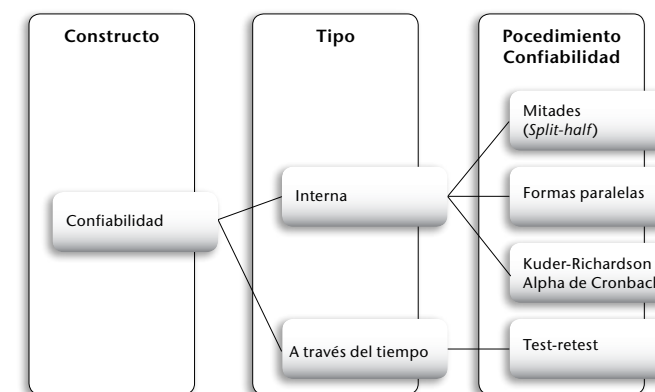


Figura 7. Tipos de confiabilidad utilizadas para demostrar el índice de confiabilidad de un instrumento psicométrico.

## LA CONFIABILIDAD INTERNA A TRAVÉS DEL COEFICIENTE ALFA DE CRONBACH

Para realizar el análisis de confiabilidad por medio del coeficiente Alfa de Cronbach se debe entrar a los menús, “Analizar”, “Escalas” y “Análisis de fiabilidad”.

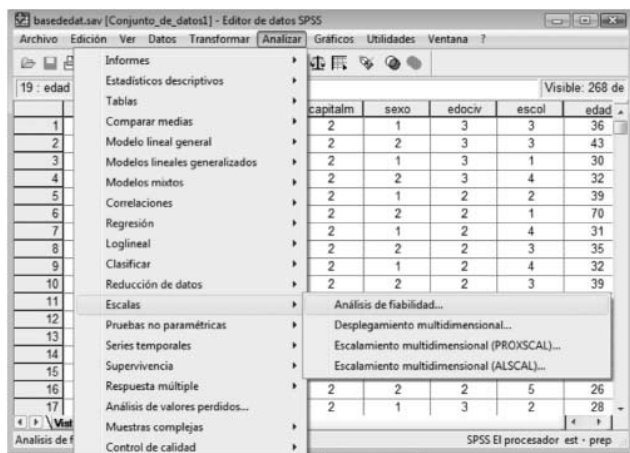


Figura 8. Pasos para realizar el análisis Alfa de Cronbach.

Dentro de las opciones estadísticas se puede solicitar la inclusión de estadísticos descriptivos para el elemento, la escala y la escala si se elimina el elemento.

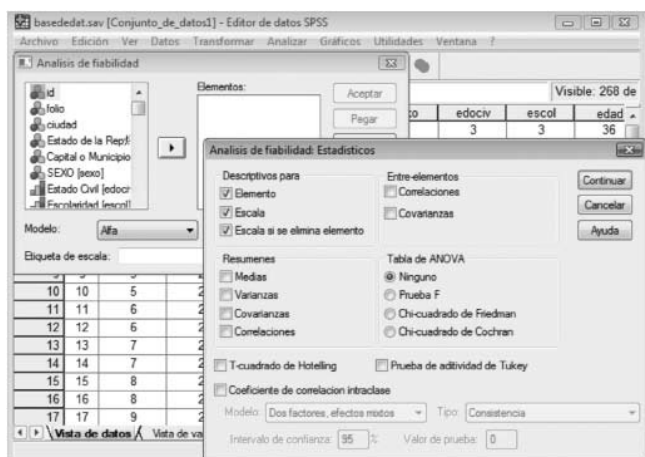


Figura 9. Subcuadro de diálogo para las opciones estadísticas del Alfa de Cronbach.

En la ventana de resultados aparece un primer cuadro donde se especifica el resumen del procedimiento, el número de casos válidos y el de los excluidos, así como un cuadro que especifica el estadístico del Alfa de Cronbach, en nuestro ejemplo .934, así como el número de reactivos del instrumento: en este caso 18.

ESCALA: TODAS LAS VARIABLES			
Resumen del procesamiento de los casos			
		No.	%
Casos	Válidos	1428	99.2
	Excluidos(a)	12	.8
Total		1440	100.0

a Eliminación por lista basada en todas las variables del procedimiento.

Estadísticos de fiabilidad	
Alfa de Cronbach	No. de elementos
.900	18

Figura 10. Visor de resultados del análisis de Alfa de Cronbach.

Un investigador puede aumentar la confiabilidad de un instrumento aumentando el número de reactivos, sin embargo Fitzpatrick, Davey, Buxton y Jones (2001) mencionan que una confiabilidad apropiada tendría que estar entre 0.70 y 0.90. Una confiabilidad más alta podría indicar que los reactivos son redundantes y que se está dejando de lado reactivos que reflejan una mayor complejidad y diversidad. Los autores mencionan que buscar a toda costa incrementar la confiabilidad de la escala podría traducirse en el absurdo de utilizar los mismos reactivos una y otra vez.



TABLA 4

**Estadísticos total-elemento del análisis de confiabilidad  
Alfa de Cronbach**

	Media de la escala si se elimina el elemento	Varianza de la escala si se elimina el elemento	Correlación elemento-total corregida	Alfa de Cronbach si se elimina el elemento
<b>7p</b> Nunca se debe dudar de la palabra del padre	24.69	49.090	.466	.644
<b>8p</b> Un hijo(a) debe siempre obedecer a sus padres	24.07	50.870	.429	.654
<b>9p</b> Las mujeres tienen que ser protegidas	23.75	52.237	.409	.659
<b>10p</b> La vida es más dura para una mujer que para un hombre	24.39	52.908	.300	.682
<b>11p</b> Es mucho mejor ser hombre que mujer	25.21	48.591	.427	.653
<b>12p</b> Uno debería respetar más a un hombre importante que a uno común y corriente	26.18	50.807	.401	.660
<b>13p</b> Es común, que los hombres casados tengan amantes	25.32	53.054	.235	.700
<b>14p</b> Las mujeres que llegan vírgenes al matrimonio son mejores esposas	25.87	49.233	.426	.653

Finalmente, en la *tabla 4* titulada Estadísticos total-elemento se muestra la media de la escala si cada reactivo es eliminado; la varianza, así como la correlación de cada reactivo con el total de reactivos y el Alfa de Cronbach si el reactivo es eliminado. Si el reactivo se correlaciona de manera baja con la escala total, la varianza total incrementa. Los reactivos que menos correlacionan con la escala total contribuyen menos a la confiabilidad. Sin embargo, no siempre es deseable eliminarlos. Como mencionamos anteriormente, un intervalo adecuado de confiabilidad se encuentra entre .70 y .90, cuando se excede este puntaje es posible que los reactivos estén demasiado relacionados y se corre el riesgo de estar evaluando lo mismo con los mismos reactivos. Eliminarlos sería un error ya que le restaría variabilidad al instrumento.

## OTROS TIPOS DE CONFIABILIDAD

Finalmente, en otros casos, la confiabilidad puede darse por medidas que son observables. Operacionalmente se denomina confiabilidad interobservadores (*interobserver*). Davis (2008) menciona que la confiabilidad puede obtenerse en estudios experimentales cuando se observa una conducta entre dos o más observadores. En este caso la confiabilidad consistirá en que los evaluadores observen y registren las mismas conductas o acciones. Es claro que por ser diferentes observadores habrá diferencias pero la confiabilidad implica encontrar consistencias. Si encontramos un mayor número de coincidencias entonces tendré una confiabilidad elevada, en cambio, si los resultados son muy diversos, entonces tendremos que concluir que los resultados se deben en mayor medida al azar. La forma de medir la confiabilidad en estos casos es dividir el número de veces que los observadores llegaron a un acuerdo entre el número de oportunidades que tuvieron de llegar a un acuerdo dividido por cien. Davis (2008) también menciona que puede medirse calculando el cuadrado del coeficiente de correlación entre los dos observadores multiplicado por cien ( $r^2 \times 100$ ).

El proceso de construcción de instrumentos válidos y confiables es fundamental en psicología. Es por ello, que el presente trabajo recopila siete trabajos que muestran la construcción y validación de instrumentos psicométricos culturalmente relevantes y sensibles a la cultura mexicana. El lector encontrará en estos capítulos, no sólo la oportunidad de tomar estos instrumentos y utilizarlos en su práctica profesional, sino que comprenderá, de mejor forma, el proceso de construcción y validación de una prueba.

En los capítulos que siguen a continuación se describe la construcción de instrumentos válidos y confiables que el profesional en psicología puede utilizar para realizar una evaluación adecuada. Los instrumentos recorren temas de evaluación individual –pasando de los trastornos alimentarios a las problemáticas asociadas a pacientes diabéticos– de evaluación en la interacción –involucrando temas que incluyen el inicio de la pareja, la sexualidad, los celos y la disolución de la relación de pareja– para finalmente llegar a la evaluación de normas y premisas que caracterizan a la sociedad mexicana.

Finalmente, no hay que olvidar que la evaluación es un término que implica más que el uso de instrumentos válidos y confiables (Rogelberg, 2004). La evaluación implica la medición pero también implica evaluaciones cualitativas como la entrevista cualitativa y clínica, así como las mediciones fisiológicas. Sin embargo, la evaluación en psicología no puede comprenderse sin la utilización de la medición.

## REFERENCIAS

- BRYMAN, A. (1988). *Quantity and quality in social research*. London and New York: Routledge.
- CLARK-CARTER, D. (2004). *Quantitative Psychological Research. A student's handbook*. USA & Canada: Psychology Press.
- COSTA, N. K. M. (1996). *Manual de pruebas de inteligencia y aptitudes*. México: Universidad Iberoamericana.
- DAVIS, F. S. (2008). Nonexperimental Research Methods. In S. F. Davis & W. Buskist (2008). *21st Century. Psychology a reference handbook*. USA: Sage publications.
- FITZPATRICK, R.; DAVEY, C.; BUXTON, M. J. & JONES, D. R. (2001). Criteria for Assessing Patient Based Outcome Measures for Use in Clinical Trials. In A. Stevens, K. Abrams, Brazier, J. Fitzpatrick, R. & Lilford, R. (eds.). *The advances Handbook of Methods in Evidence Based Health Care*. London, California, New Delhi: SAGE Publications.
- Flores, G. M. y Díaz-Loving, R. (2004). EMA: *Escala Multidimensional de Asertividad*. México: Manual Moderno.
- GARCÍA, M. M., RIVERA, A. S., DÍAZ, L. R. Y REYES, L. I. (2004). Poder y violencia: La pareja contemporánea. *Revista Mexicana de Psicología Social y Personalidad*, XX (1).
- Groth-Marnat, G. (2003). *Handbook of psychological assessment*. USA: Wiley & Sons, Inc.
- HAIR, J. F; ANDERSON, R. E; TATHAM, R. L. & BLACK, W. C. (1999). *Análisis multivariante 5ª ed*. Madrid: Prentice Hall.
- JOHNSTON, M; D. P. FRENCH, D. BONNETTI, & D. JOHNSTON, Assessment and Measurement in Health Psychology. In S. Sutton; A. Baum & M. Johnston (2004). *The SAGE Handbook of Health Psychology*. USA: SAGE Publications.
- JUVE, J. (2008). Testing and Assessment. In S. F. Davis & W. Buskist (2008). *21st Century. Psychology a reference handbook*. USA: Sage publications.
- KENNY, M. C; K. ALVAREZ, B. C. DONOHUE, & C. B. WINICK (2008). Overview of Behavioral Assessment with Adults. In M. Hersen & J. Rosqvist (eds.), *Handbook of Psychological Assessment, Case Conceptualization, and Treatment*. USA & Canada: John Wiley & Sons.
- PARDO, M. A. Y D. M. A. RUIZ (2002). SPSS 11. *Guía para el análisis de datos*. España: McGraw-Hill.
- REEVES, B. C; R. R. MACLEHOSE; I. M. HARVEY, T. A. SHELDON; I. T. RUSSELL & A. M. S. BLACK (2001). A Review of Observational, Quasi-Experimental and Randomised Study Designs for the Evaluation of the Effectiveness of Healthcare Interventions. In A. Stevens, K. Abrams, Brazier, J. Fitzpatrick, R. & Lilford, R. (eds.). *The Advances Handbook of Methods in Evidence Based Health Care*. London, California, New Delhi: SAGE Publications.
- REYES-LAGUNES, I. Y GARCÍA, I. L. F. (2008). Procedimiento de validación psicométrica culturalmente relevante: un ejemplo. *La psicología social en México*, XII (2), 625-630.
- RIVERA ARAGÓN, S. (2000). *Conceptualización, medición y correlatos de poder y pareja: una aproximación etnopsicológica*. Tesis de doctorado no publicada, México, UNAM.
- ROGELBERG, S. G. (2004). *Handbook of Research Methods in Industrial and Organizational*. USA: Blackwell Publishing.

- SÁNCHEZ, A. R. (2000). *Validación Empírica de la Teoría Bio-Psico-Socio-Cultural de la Relación de Pareja*. Tesis Doctoral. Facultad de Psicología, UNAM.
- SANDERSON, C. MCKEE, M. BRITTON, A. BLACK, N. MCPHERSON, K & BAIN, C. (2008). Randomised and Non-Randomised Studies: Threats to Internal and External Validity. In A. Stevens, K. Abrams, Brazier, J. Fitzpatrick, R. & Lilford, R. (eds.). *The Advances Handbook of Methods in Evidence Based Health Care*. London, California, New Delhi: SAGE Publications.
- SATSKY, M. K. (2008). Psychometrics. In S. F. Davis & W. Buskist (2008). *21st Century. Psychology a Reference Handbook*. USA: Sage Publications.

## RECOMENDACIÓN DE LECTURAS EN ESPAÑOL PARA ALUMNOS DEL PREGRADO

### En el caso de la construcción de instrumentos social y culturalmente relevantes

- REYES-LAGUNES, I. Y GARCÍA, I. L. F. (2008). Procedimiento de validación psicométrica culturalmente relevante: un ejemplo. *La psicología social en México*, XII (2), 625-630.

## PARA LA REALIZACIÓN DE LOS ANÁLISIS ESTADÍSTICOS

- HAIR, J. F; R.E. ANDERSON; R.L. TATHAM, & W.C. BLACK (1999). *Análisis multivariante, 5ª ed*. Madrid: Prentice Hall. Capítulos referentes al Análisis factorial y análisis de fiabilidad.
- PARDO, M. A. Y D. M. A. RUIZ (2002). SPSS 11. *Guía para el análisis de datos*. España: McGraw-Hill. Capítulos referentes al Análisis factorial y el análisis Alpha de Cronbach.

## CON REFERENCIA A LOS CONSTRUCTOS DE VALIDEZ Y CONFIABILIDAD

- ANASTASI, A. Y S. URBINA (1998). *Test psicológicos, 7ª ed*. México: Pearson Prentice Hall. Capítulos de validez y confiabilidad.