

VALIDEZ Y CONFIABILIDAD DE LOS INSTRUMENTOS DE INVESTIGACIÓN PARA LA RECOLECCIÓN DE DATOS

RESUMEN

El presente ensayo pretende orientar a los noveles investigadores en el procedimiento para elaborar un instrumento adecuado para la recolección de datos, válido y confiable, que proporcione un basamento relevante para el logro de los objetivos planteados y sustente los hallazgos que realicen con sus investigaciones. Se describe la validez y la confiabilidad de un instrumento. Cómo se realiza el cálculo de la confiabilidad y cuáles son los instrumentos que requieren de este procedimiento, ya que no todos los instrumentos para recabar datos ameritan se realice una prueba piloto con esta finalidad.

Palabras Clave: Validez. Confiabilidad. Instrumento de Investigación.

.....
Autor (a):

Prof(a). Yadira Corral

yadi_jcm@hotmail.com

*Facultad de Ciencias Económicas Y
Sociales Universidad de Carabobo
Valencia Estado Carabobo
Venezuela.*

Recibido 28/10/08 Aprobado: 09/02/09

*Prof(a). Adscrita a la Dirección de
Estudios de Postgrado. Facultad
de Ciencias Económicas y Sociales
Maestría en Administración de
Empresas Mención Mercadeo. MSc.
Educación Superior. Profesora,
mención Matemática. UPEL-
Caracas.*

VALIDITY AND RELIABILITY OF THE INSTRUMENTS TO COLLECT DATA

ABSTRACT

This essay aims to provide guidance to junior researchers in the procedure for developing a suitable instrument for data collection, valid and reliable, providing a basement relevant to achieving the goals set and underpins the findings to conduct its investigations. It describes the validity and reliability of an instrument. How is the calculation of reliability and what are the instruments that require this procedure? Since not all instruments to collect data warrant a pilot test is conducted for this purpose.

Keywords: Validity. Reliability. Instrument Research.

INTRODUCCIÓN

La problemática al momento de la recolección de datos en la realización de los Trabajos de investigación se centra en la construcción de los instrumentos a emplear con esta finalidad, de manera que permitan recabar información válida y confiable. Porque el valor de un estudio depende de que esta información refleje lo más fidedignamente el evento investigado, dándole una base real para obtener un producto investigativo de calidad. Por lo anterior, el presente ensayo pretende orientar a los noveles investigadores en el procedimiento para elaborar instrumentos adecuados que proporcionen un basamento relevante para el logro de los objetivos planteados en la investigación abordada y sustenten los hallazgos que se realicen.

Siempre que se quiere recopilar información a fin de ayudarnos a tomar decisiones, nos enfrentamos al problema de qué tipo de instrumento usar o si realmente hay un cuestionario que nos ayude a tomar la decisión. En este sentido, un instrumento según Alvarado, Canales y Pineda (1994) "...es el mecanismo que utiliza el investigador para recolectar y registrar la información" (pp.125). Con este propósito, es imperativo saber si el instrumento a diseñar nos ayudará a contar con información útil y fidedigna sobre el problema de estudio, en cuyo caso ¿cuál será el mejor que podamos usar?, ¿cómo hacer la selección tanto del instrumento como de los reactivos que la formen?. Existen muchas consideraciones específicas a tomar en

cuenta en la evaluación de un cuestionario; las cuales las englobaremos bajo tres encabezados principales: validez, confiabilidad y utilidad práctica.

Ebel (1977, citado por Fuentes, 1989) establece que validez "...designa la coherencia con que un conjunto de puntajes de una prueba miden aquello que deben medir" (p. 103). La validez se refiere al grado en que una prueba proporciona información que es apropiada a la decisión que se toma. La confiabilidad tiene que ver con la exactitud y precisión del procedimiento de medición. Los coeficientes de confiabilidad proporcionan una indicación de la extensión, en que una medida es consistente y reproducible. La utilidad práctica está relacionada con factores, tales como: economía, conveniencia e interpretación para determinar si una prueba es práctica para usarla ampliamente.

Validez

La validez responde a la pregunta ¿con qué fidelidad corresponde el universo o población al atributo que se va a medir?. La validez de un instrumento consiste en que mida lo que tiene que medir (autenticidad), algunos procedimientos a emplear son: Know groups (preguntar a grupos conocidos), Predictive validity (comprobar comportamiento) y Cross-check-questions (contrastar datos previos). Al estimar la validez es necesario saber a ciencia cierta qué rasgos o características se desean estudiar. A este rasgo o característica se le denomina variable criterio. Al respecto, Ruiz Bolívar (2002) afirma que "...nos interesa saber qué tan bien corresponden las posiciones de los individuos en la distribución de los puntajes obtenidos con respecto a sus posiciones en el continuo que representa la variable criterio" (p. 74). Existen tres tipos de validez:

- **Validez de Contenido:** se refiere al grado en que un instrumento refleja un dominio específico del contenido de lo que se quiere medir, se trata de determinar hasta dónde los ítemes o reactivos de un instrumento son representativos del universo de contenido de la característica o rasgo que se quiere medir, responde a la pregunta cuán representativo es el comportamiento elegido como muestra del universo que intenta representar. Por ejemplo, un cuestionario sobre la actitud de los alumnos ante la investigación no tendrá validez de contenido si explora la opinión de los alumnos sobre las

características de los docentes dentro de la cátedra de estadística. También se le denomina *validez racional o lógica*.

El análisis del instrumento se hace en gran parte en términos de su contenido. Sin embargo, no se debe pensar en el contenido de manera estrecha, porque puede ser que estemos interesados en un proceso tanto como en el contenido simplemente. El problema de apreciar la validez de contenido está vinculado íntimamente con la planificación del cuestionario y después con la construcción de los ítems ajustados a esos planes y a los contenidos del marco teórico de la investigación. Un instrumento de medición debe tener representados a todos los ítems del dominio de contenido de las variables a medir.

Los investigadores deben elaborar una serie de ítems, acordes con las variables empleadas y sus respectivas dimensiones. Luego de la selección de los ítems más adecuados para el proyecto, se elabora el instrumento, para ser validado por un grupo impar de expertos, normalmente de tres o cinco, que certifiquen, efectivamente, que las preguntas, reactivos o afirmaciones seleccionadas son claras y tienen coherencia con el trabajo desarrollado.

Hay que considerar que, la validez de contenido no puede expresarse cuantitativamente es más bien una cuestión de juicio, se estima de manera subjetiva o intersubjetiva empleando, usualmente, el denominado **Juicio de Expertos**. Se recurre a ella para conocer la probabilidad de error probable en la configuración del instrumento. Mediante el juicio de expertos se pretende tener estimaciones razonablemente buenas, las «mejores conjeturas». Sin embargo, estas estimaciones pueden y deben ser confirmadas o modificadas a lo largo del tiempo, según se vaya recopilando información durante el funcionamiento del sistema. Los juicios de expertos se pueden obtener por métodos grupales o por métodos de experto único. Se pueden seguir, entre otros, el método de Agregados Individuales, el método Delphi, la técnica de Grupo Nominal y el método de Consenso Grupal.

- *Método de Agregados Individuales:* Se pide individualmente a cada experto que dé una estimación directa de los ítems del instrumento. Éste es un método económico porque, al igual que el método Delphi, no exige que se reúna a los expertos en un lugar determinado. Puede parecer un método limitado porque los expertos no pueden

intercambiar sus opiniones, puntos de vista y experiencia, ya que se les requiere individualmente; no obstante, esta limitación puede ser precisamente lo que se esté buscando para evitar los sesgos de los datos ocasionados por conflictos interpersonales, presiones entre los expertos, etc. Se procede de la siguiente manera:

- ✓ Se seleccionan al menos tres expertos o jueces, para juzgar de manera independiente la relevancia y congruencia de los reactivos con el contenido teórico, la claridad en la redacción y el sesgo o tendenciosidad en la formulación de los ítemes, es decir, si sugieren o no una respuesta.
- ✓ Cada experto debe recibir la información escrita suficiente sobre: el propósito de la prueba (objetivos), conceptualización del universo de contenido, tabla de especificaciones o de operacionalización de las variables del estudio.
- ✓ Cada experto debe recibir un instrumento de validación que contenga: congruencia ítem-dominio, claridad, tendenciosidad o sesgo y observaciones.
- ✓ Se recogen y analizan los instrumentos de validación y se decide:
 - 1) Los ítemes que tienen 100% de coincidencia favorable entre los jueces (congruentes, claros en su redacción y no tendenciosos) quedan incluidos en el instrumento
 - 2) Los ítemes que tengan 100% de coincidencia desfavorable entre los jueces quedan excluidos del instrumento
 - 3) Los ítemes que tengan una coincidencia parcial entre los jueces deben ser revisados, reformulados o sustituidos, si es necesario, y nuevamente validados.

Cuadro 1. Formato para validar instrumentos a incluir en el instrumento de validación.

ÍTEM	CRITERIOS A EVALUAR										Observaciones (si debe eliminarse o modificarse un ítem por favor indique)	
	Claridad en la redacción		Coherencia interna		Inducción a la respuesta (Sesgo)		Lenguaje adecuado con el nivel del informante		Mide lo que pretende			
	Sí	No	Sí	No	Sí	No	Sí	No	Sí	No		
1												
2												
3												
....												
n												
Aspectos Generales										Sí	No	*****
El instrumento contiene instrucciones claras y precisas para responder el cuestionario												
Los ítems permiten el logro del objetivo de la investigación												
Los ítems están distribuidos en forma lógica y secuencial												
El número de ítems es suficiente para recoger la información. En caso de ser negativa su respuesta, sugiera los ítems a añadir												
VALIDEZ												
APLICABLE					NO APLICABLE							
APLICABLE ATENDIENDO A LAS OBSERVACIONES												
Validado por:							C.I.:			Fecha:		
Firma:							Teléfono:			e-mail:		
Nota. Modificado de Formato de la Facultad de Odontología de la Universidad de Carabobo (2007).												

- *Método Delphi:* Este método fue creado en 1948 para obtener la opinión de expertos de una manera sistemática. En un primer momento, cada experto responde de manera individual y anónima a un cuestionario. Después se analizan las respuestas del conjunto de expertos, se remite a cada uno la respuesta mediana obtenida, así como el intervalo intercuartil para cada cuestión y se les pide que reconsideren su juicio anterior, teniendo en cuenta estos datos.

En cada una de las tres o cuatro «vueltas» siguientes, se informa a los expertos de cuál es la mediana del grupo y se les propone revisar su juicio anterior. Todo juicio individual que quede fuera del intervalo intercuartil en que se mueve el grupo de expertos tiene que estar debidamente justificado o argumentado. La mediana de las respuestas obtenidas en esta última vuelta es el valor que se estaba buscando. Se emplea la mediana, porque se presupone que las puntuaciones posibles de los expertos se distribuyen de forma asimétrica. Con este método los expertos comparten en cierto modo sus opiniones, sin que existan discusiones ni confrontaciones directas entre ellos.

- *Técnica de Grupo Nominal:* El primer paso es reunir a los expertos (entre ocho y diez personas) y pedirles que registren, individualmente y sin intercambiar opiniones, sus propias puntuaciones y consideraciones respecto a las probabilidades de error para cada una de las preguntas o tareas que se les detallan. Después, cada experto expone a los demás las puntuaciones y principales consideraciones registradas y al acabar esta ronda, se establecen las coincidencias del grupo. Luego se realiza un debate de cada uno de los apartados de ésta. Finalmente, cada experto, de manera individual y por escrito, puntúa y argumenta las probabilidades de error para cada tarea/pregunta considerada.

En general, se procede como con el método Delphi, sólo que en esta técnica se permite algún debate entre los expertos, para que aclaren y compartan la información que cada uno está considerando. Aunque hay intercambio de pareceres, los juicios se emiten de forma individual y la estimación final suele ser la media aritmética del conjunto de las estimaciones dadas por los expertos. El éxito de la técnica depende, por una parte, de la habilidad y la experiencia del moderador del grupo y de la buena voluntad de los expertos para trabajar juntos en un marco altamente estructurado.

- *Método del Consenso Grupal:* Se reúne a los expertos en un lugar determinado, se indica al grupo que su tarea consiste en lograr una estimación de la pertinencia y otros aspectos relacionados con la elaboración de los ítemes, que sea satisfactoria para todos los

expertos. Con estas instrucciones se maximizan los intercambios de información y opiniones dentro del grupo de expertos. Si el grupo no logra un consenso, puede intentarse un consenso artificial recogiendo las estimaciones individuales y sintetizándolas estadísticamente. Este método, como el anterior, también precisa que el grupo de expertos sea pequeño, se fomente la libre expresión y se eviten las discusiones tensas y los sistemas de votación.

- **Validez de Constructo:** intenta responder la pregunta ¿hasta dónde el instrumento mide realmente un rasgo determinado y con cuánta eficiencia lo hace?. Al respecto Gronlund (1976, citado por Ruiz Bolívar, op. cit.) señala que esta validez interesa cuando se quiere usar el desempeño de los sujetos con el instrumento para inferir la posesión de ciertos rasgos. Para estudiar este tipo de validez es necesario que exista una conceptualización clara del rasgo estudiado basado en una teoría determinada. La teoría sugiere las tareas pruebas que son apropiadas para observar el atributo o rasgo y las evidencias a considerarse en la evaluación. Cronbach (1960, citado por Ruiz Bolívar, op. cit.) sugiere los siguientes pasos:
 - 1) Identificar las construcciones que pudieran explicar la ejecución en el instrumento
 - 2) Formulación de hipótesis comprobables a partir de la teoría
 - 3) Recopilación de los datos para probar las hipótesis

El término *constructo* se usa en psicología para referirse a algo que no es observable, pero que literalmente es *construido* por el investigador para resumir o explicar las regularidades o relaciones que él observa en la conducta. Por tanto, la mayoría de los nombres de rasgos se refieren a constructos. Para las preguntas acerca de si el instrumento revela algo significativo respecto de las personas, se usa el término *validez de constructo*.

Muchas pruebas psicológicas, y en menor medida algunas pruebas educativas, intentan medir rasgos generales o cualidades de un individuo, tales como: razonamiento verbal, visualización especial, sociabilidad, introversión e interés mecánico son designaciones de constructos o de rasgos. Las pruebas de estas funciones son válidas en tanto se comporten de la manera que cabría esperar que se comportara el rasgo.

- **Validez Predictiva o de Criterio Externo o Empírica:** se asocia con la visión de futuro, determinar hasta dónde se puede anticipar el desempeño futuro de una persona en una actividad determinada. La validez predictiva se estudia comparando los puntajes de un instrumento con una o más variables externas denominadas variables de criterio. Se establece una correlación, la cual se interpreta como **índice de validez**.

Para este propósito, estamos interesados en el grado en que el instrumento se correlaciona con algunos de los criterios escogidos para medir, por ejemplo: el éxito en los estudios. Así, se toma alguna otra medida, como el criterio del “éxito”, y juzgamos el cuestionario elaborado en términos de su relación con esa medida de criterio. El procedimiento básico consiste en aplicar la prueba a un grupo de personas que ingresan a un trabajo o a un programa de entrenamiento y, posteriormente, seguirlas observando para obtener de cada una, una medida criterio de éxito específica y luego calcular la correlación entre la puntuación de prueba y la medida criterio de éxito. Cuanta más alta sea la correlación, mejor será el instrumento, es decir, la evaluación como predictor es principalmente una evaluación empírica y estadística.

Cualidades deseables en una medida criterio: Existen cuatro cualidades deseables en una medida criterio, en orden de importancia son:

- *Atingencia* – Se consideran que un criterio es atingente en la medida en que esta medida criterio corresponde con o ejemplifica el éxito en el trabajo.
- *Libre de sesgos* – La medida debería ser aquella en la que cada persona tiene las mismas oportunidades de obtener una buena puntuación.
- *Confiabilidad* – Debe ser estable o reproducible si ha de ser predicha por algún tipo de prueba.
- *Disponibilidad* – Debe tener en cuenta límites prácticos como: ¿Qué tanto se debe esperar para obtener una puntuación?, ¿Cuánto dinero costarán las interrupciones de las actividades?

Factores que Afectan la Validez

Existen varios factores que tienden a distorsionar los coeficientes de validez y a complicar su interpretación, entre ellos:

- **Construcción del Instrumento.** Algunos de los factores a evitar, para que el instrumento sea válido son:
 - Instrucciones imprecisas o vagas
 - Estructura sintáctica de las oraciones muy difíciles
 - Preguntas inadecuadas respecto a las especificaciones (sobre todo en pruebas escritas y orales)
 - Preguntas que sugieren la respuesta
 - Ambigüedad en la formulación de los reactivos, que lleven a diferentes interpretaciones
 - Cuestionarios demasiado cortos (no incluyen una muestra adecuada de los indicadores a medir), con pocos reactivos
 - Ítems incongruentes con el universo de contenido, sin relación con los rasgos o características a medir
 - Ordenamiento inadecuado de los ítems
 - Patrón identificable de respuestas (sobre todo en pruebas de selección)
- **Administración y Calificación del Instrumento.** Entre estos factores están: tiempo insuficiente para responder, ayuda adicional de otros sujetos, en caso de pruebas: las chuletas o copias, la subjetividad en los puntajes de las preguntas.
- **Respuestas de los Sujetos.** Algunos factores: bloqueo de los sujetos al responder debido a situaciones emocionales y las respuestas al azar.
- **Naturaleza del Grupo y del Criterio.** Debe ser aplicado el instrumento para el grupo al cual fue diseñado, puede ser inapropiado para otros grupos.

Confiabilidad

Antes de iniciar el trabajo de campo, es imprescindible probar el cuestionario sobre un pequeño grupo de población. Esta **prueba piloto** ha de garantizar las mismas condiciones de realización que el trabajo de campo real. Se recomienda un pequeño grupo de sujetos que no pertenezcan a la muestra seleccionada pero sí a la población o un grupo con características similares a la de la muestra del estudio, aproximadamente entre 14 y 30 personas. De esta manera se estimará la confiabilidad del cuestionario.

La confiabilidad responde a la pregunta ¿con cuánta exactitud los ítems, reactivos o tareas representan al universo de donde fueron seleccionados?. El término confiabilidad "...designa la exactitud con que un conjunto de puntajes de pruebas miden lo que tendrían que medir" (Ebel, 1977, citado por Fuentes, op. cit., p. 103). Entre los **métodos para estimar la confiabilidad**, se tienen:

- ✓ **Método Test-Retest:** una forma de estimar la confiabilidad de un test o cuestionario es administrarlo dos veces al mismo grupo y correlacionar las puntuaciones obtenidas. Este método tiene la desventaja de que los puntajes pueden verse afectados por el recuerdo, la práctica, etc. Este procedimiento no es adecuado para aplicarlo a pruebas de conocimientos sino para la medición de aptitudes físicas y atléticas, tests de personalidad y motores.

El coeficiente que se obtiene recibe el nombre de **coeficiente de estabilidad** porque denota la coherencia de las puntuaciones en el tiempo. Para un desarrollo adecuado y sean confiables deben variar entre 0,80 y 0,95 (Popham, 1980, citado por Fuentes, op. cit.).

Se usa la **correlación por el método de los puntajes directos (Correlación r de Pearson)**:

$$r_{xy} = \frac{N \sum XY - \sum X \sum Y}{\sqrt{[N \sum X^2 - (\sum X)^2] [N \sum Y^2 - (\sum Y)^2]}}$$

En donde:

r_{xy} : es el coeficiente de correlación

N: número de sujetos

X: valores de X (1ª aplicación)

Y: valores de Y (2ª aplicación)

XY: producto de cada valor X por su correspondiente valor en Y

Cuadro 2. Tabla de Datos para aplicar el Coeficiente de Correlación r de Pearson.

Sujetos	1ª Aplicación Test		2ª Aplicación Retest		XY
	X	X ²	Y	Y ²	
1					
2					
3					
ETC					
Σ					

- ✓ **Método común de división por mitades o Hemitest:** este método computa el coeficiente de correlación entre los puntajes de las dos mitades del test o cuestionario aplicado. Esto supone que las dos test mitades son paralelos, tienen igual longitud y varianza entre sí. Se estima a través del coeficiente de confiabilidad de Spearman-Brown:
- i) Se establece la correlación entre los dos puntajes de las dos mitades del test a través del **método de los puntajes directos, Correlación r de Pearson:**

$$r_{12} = \frac{N \sum X_1 X_2 - (\sum X_1)(\sum X_2)}{\sqrt{[N \sum X_1^2 - (\sum X_1)^2] [N \sum X_2^2 - (\sum X_2)^2]}}$$

ii) Estimación del test completo (Spearman-Brown) con la fórmula:

$$r_{tt} = \frac{2r_{12}}{1 + r_{12}}$$

Se interpreta la prueba de hemitest como **coeficiente de consistencia interna**, ya que una sola prueba contiene las dos formas equivalentes y su énfasis lo pone en las puntuaciones de los sujetos, no en los ítemes.

- ✓ **El método de división por mitades de Rulon:** utiliza la división del test en mitades, pero su método no supone necesariamente varianzas iguales en los sub-tests. **coeficiente de consistencia interna.**

$$r_{tt} = 1 - \frac{s_d^2}{s_t^2}$$

Donde:

r_{tt} : coeficiente de confiabilidad

s_d^2 : varianza de la diferencia entre las puntuaciones de las mitades

s_t^2 : varianza de las puntuaciones del test total

- ✓ **El método de división por mitades de Guttman:** también se denomina **coeficiente de consistencia interna**. Su fórmula es:

$$r_{tt} = 2 \left[1 - \frac{s_a^2 + s_b^2}{s_t^2} \right]$$

Donde:

r_{tt} : coeficiente de confiabilidad

s_a^2 : varianza de las puntuaciones de los ítemes pares

s_b^2 : varianza de las puntuaciones de los ítemes impares

s_t^2 : varianza de las puntuaciones del test total

- ✓ **Coficiente Alfa de Cronbach:** Para evaluar la confiabilidad o la homogeneidad de las preguntas o ítems es común emplear el coeficiente alfa de Cronbach cuando se trata de alternativas de respuestas policotómicas, como las escalas tipo Likert; la cual puede tomar valores entre 0 y 1, donde: 0 significa confiabilidad nula y 1 representa confiabilidad total. El coeficiente α de Cronbach puede ser calculado por medio de dos formas:

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

- a) Mediante la varianza de los ítems y la varianza del puntaje total (Hernández Sampieri et al, 2003)

Donde:

r_{tt} : coeficiente de confiabilidad de la prueba o cuestionario.

k: número de ítems del instrumento.

s_t^2 : Varianza total del instrumento.

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

Cuanto menor sea la variabilidad de respuesta, es decir, que haya homogeneidad en las respuestas dentro de cada ítem, mayor será el Alfa de Cronbach.

- b) Mediante la matriz de correlación de los ítems.

$$\alpha = \frac{np}{1 + p(n - 1)}$$

Donde:

n: Número de ítems **p:** Promedio de las correlaciones lineales entre cada uno de los ítems

Cuanto mayor sea la correlación lineal entre ítemes, mayor será el alfa de Cronbach.

- ✓ **Método de Kuder-Richarson 20:** permite obtener la confiabilidad a partir de los datos obtenidos en una sola aplicación del test. **Coefficiente de consistencia interna.** Puede ser usada en cuestionarios de ítemes dicotómicos y cuando existen alternativas dicotómicas con respuestas correctas e incorrectas.

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

Donde:

n: número total de ítemes

s_t^2 : varianza de las puntuaciones totales

p: proporción de sujetos que pasaron un ítem sobre el total de sujetos

q = 1- p

- ✓ **Método de Kuder-Richarson 21:** permite obtener la confiabilidad a partir de los datos obtenidos en una sola aplicación del test. La suposición básica es considerar que todos los ítemes presentan igual varianza. **Coefficiente de consistencia interna.**

$$KR_{21} = \frac{n}{n-1} \left[1 - \frac{M(n-M)}{ns_t^2} \right]$$

Donde:

n: número total de ítemes

M: media aritmética de las puntuaciones obtenidas por los individuos

s_t^2 : varianza de las puntuaciones totales

Los datos obtenidos de la prueba piloto deben organizarse según el cuadro 3.

Cuadro 3. Distribución de las respuestas dadas al instrumento o al cuestionario en la prueba piloto (Alfa de Cronbach y Kuder-Richarson).

Sujeto	Ítemes o Reactivos							
	1	2	3	4	5	6	...	k
1	Ptaje	Ptaje	Ptaje	Ptaje	Ptaje	Ptaje	...	Ptaje
2	Ptaje	Ptaje	Ptaje	Ptaje	Ptaje	Ptaje	...	Ptaje
3	Ptaje	Ptaje	Ptaje	Ptaje	Ptaje	Ptaje	...	Ptaje
4	Ptaje	Ptaje	Ptaje	Ptaje	Ptaje	Ptaje	...	Ptaje
5	Ptaje	Ptaje	Ptaje	Ptaje	Ptaje	Ptaje	...	Ptaje
6	Ptaje	Ptaje	Ptaje	Ptaje	Ptaje	Ptaje	...	Ptaje
7	Ptaje	Ptaje	Ptaje	Ptaje	Ptaje	Ptaje	...	Ptaje
m	Ptaje	Ptaje	Ptaje	Ptaje	Ptaje	Ptaje	...	Ptaje

Nota. m = número de sujetos a los cuales se aplicó la prueba piloto; k = número de ítemes del cuestionario o instrumento. Se usa para el Alfa de Cronbach y para Kuder-Richarson. En cada casilla se colocará el puntaje asignado a la opción seleccionada por el informante. Por ejemplo: Siempre (5), casi siempre (4), Indeciso (3), Casi nunca (2), nunca (1); Sí (1) – No (0).

Interpretación del Coeficiente de Confiabilidad

El coeficiente de confiabilidad es un coeficiente de correlación, teóricamente significa la correlación del test consigo mismo. Sus valores oscilan entre 0 y 1. Entre las escalas empleadas se tienen las mostradas en los cuadros siguientes.

Cuadro 4. Escala de interpretación de la magnitud del Coeficiente de Correlación o Asociación.

Coeficiente de Correlación	Magnitud
0,70 a 1,00	Muy fuerte
0,50 a 0,69	Sustancial
0,30 a 0,49	Moderada
0,10 a 0,29	Baja
0,01 a 0,09	Despreciable

Nota. Tomado de Sierra Bravo (2001).

Cuadro 5. Interpretación de la magnitud del Coeficiente de Confiabilidad de un instrumento.

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

Nota. Tomado de Ruiz Bolívar (2002) y Pallela y Martins (2003).

Cuadro 6. Interpretación de la magnitud del Coeficiente de Confiabilidad de un instrumento.

Correlación	Directa	Indirecta
Alta	0,70 a 1,00	-1,00 a -0,70
Moderada	0,40 a 0,70	-0,70 a -0,40
Baja	0,10 a 0,40	-0,10 a -0,10
Nula	-0,10 a 0,10	-0,10 a 0,10

Nota. Tomado de Orozco, Labrador y Palencia (2002). Esta caracterización es válida para $n > 20$, puesto que el tamaño puede afectar el valor del coeficiente de correlación.

En el Cuadro 7, se muestra un resumen de los métodos.

Cuadro 7. Resumen de los Métodos y Técnicas para estimar la confiabilidad de un instrumento.

Método	Técnica	Propósito
Test-Retest	Coefficiente de correlación r de Pearson	Consistencia en el tiempo de los puntajes
Formas equivalentes	Coefficiente de correlación r de Pearson	Coefficiente de equivalencia, variación en el tiempo de los puntajes
División por mitades	Correlación de Pearson Correlación de Spearman-Brown Coefficiente de Rulon Coefficiente de Guttman	Establecer la homogeneidad de los ítems al medir el constructo
Homogeneidad de las preguntas o ítems	Coefficiente Alfa de Cronbach Coefficiente de Kuder Richardson 20 Coefficiente de Kuder Richardson 20	Para escalas policotómicas como las tipo Likert Para ítems con escalas dicotómicas

Nota: Tomado de Pallela y Martins (2003), p. 155.

Como colofón, cabe destacar que, existen instrumentos para recabar datos que por su naturaleza no ameritan el cálculo de la confiabilidad, como son: entrevistas, escalas de estimación, listas de cotejo, guías de observación, hojas de registros, inventarios, rúbricas, otros. A este tipo de instrumentos, sin embargo, debe estimarse o comprobarse su validez, a través del juicio de expertos, para establecer si los reactivos que los configuran o integran se encuentran bien redactados y miden lo que se pretende medir.

Igualmente, instrumentos como la Historia Clínica no ameritan el cálculo de la confiabilidad; debido a que, usualmente, ya se encuentra validada por su uso frecuente y su confiabilidad se ha comprobado por sus aciertos, en otras palabras, es un instrumento ya estandarizado. Al respecto, la Historia Clínica es definida por Guzmán Mora (2006) como una de las formas de registro del acto médico y, en esencia, son los profesionales de la salud quienes están en capacidad de practicar una buena Historia Clínica; conceptualizada como el documento médico y clínico por excelencia.

REFERENCIAS BIBLIOGRÁFICAS

- Alvarado, E., Canales, F. y Pineda, E. (1994). *Metodología de la Investigación* (2ª ed.). Washington D.C: Organización Panamericana de la Salud
- Arias, F. (2006). *El Proyecto de Investigación. Introducción a la metodología científica*. (5ª ed.). Caracas: Episteme C. A.
- Arquer, M. (1994) *NTP 401: Fiabilidad humana: métodos de cuantificación, juicio de expertos* [Documento en línea]. Disponible: http://www.mtas.es/insht/ntp/ntp_401.htm [Consulta: 2007, febrero 10]
- Ary, D; Jacob, L. y Razavieh, A. (1992). *Introducción a la Investigación Pedagógica*. (2ª ed.). México: Mc Graw-Hill Interamericana.
- Ávila Baray, H. L. (2006). *Introducción a la metodología de la Investigación*. (Edición electrónica) Disponible: <http://www.eumed.net/libros/2006c/203/2i.htm> [Consulta: 2007, junio 2]
- Balestrini, M. (2002) *Como se Elabora el Proyecto de Investigación*. (6ª ed.). Caracas, Venezuela: BL Consultores Asociados Servicio Editorial.
- Busot, A. (1991). *Investigación Educativa*. (2ª ed). Maracaibo, Venezuela: LUZ.

- Chávez Alizo, N. (2001). *Introducción a la investigación educativa*. Maracaibo: sin editorial.
- Facultad de Odontología. Universidad de Carabobo. (2007). *Manual de Normas de Presentación del Informe de Investigación*. Bárbula: Departamento Formación Integral del Hombre.
- Fuentes, R. (1989, abril). Estudios sobre confiabilidad. *Paradigma*, 4(2), 101-126
- Guzmán Mora, F. (2006) *El acto médico: consideraciones Generales*. [Documento en línea]. Disponible: http://www.medspain.com/ant/n10_mar00/actomed.htm. [Consulta: 2007, junio 2]
- Hernández Sampieri, R. y Fernández Collado, C. (2000) *Metodología de la Investigación. Manual de apoyo para profesores*. (2ª ed.) México: Mc Graw-Hill Interamericana Editores S.A.
- Hernández Sampieri, R., Fernández Collado, C. y Baptista Lucio, P. (2003) *Metodología de la Investigación*. (4ª ed.) México: Mc Graw-Hill Interamericana Editores S.A.
- Kinpear, T. y Taylor, J. (1993). *Investigación de Mercados*. Colombia: McGraw-Hill Interamericana S. A.
- Labrador, M. E. (2006). *Formato para Validación de Instrumentos*. Bárbula, Venezuela: Facultad de Odontología de la Universidad de Carabobo.
- Malhotra, N. (1997). *Investigación de Mercados. Un enfoque práctico*. México: Prentice-Hall Hispanoamericana S. A.
- Méndez, C. (1999). *Metodología*. (2ª ed.). Colombia: McGraw-Hill Interamericana S. A.
- Menéndez, A. (2006, enero 24). *Taller CES (Consejo de Educación Superior de Puerto Rico). Validez, confiabilidad y utilidad*. [Documento en línea]. Disponible: <http://www.gobierno.pr/NR/rdonlyres/5CF112BB-5811-4A9A-8D1E-1BA213C5EEF7/0/14Validez.pdf>
- Orozco, C., Labrador, M. y Palencia, A. (2002). *Metodología. Manual teórico Práctico de Metodología para tesis, asesores, tutores y jurados de trabajos de investigación y ascenso*. Venezuela: Ofimax de Venezuela.

- Parella, S. y Martins, F. (2003). *Metodología de la Investigación cuantitativa*. Caracas: Fedupel.
- Pérez, M. (2006). *Metodologías de investigación. Modelo para el diseño de una investigación educativa*. [Documento en Línea]. Disponible: <http://dewey.uab.es/pmarques/edusoft.htm> [Consulta: 2007, Mayo 30]
- Rodríguez, Y. y Pineda, M. (2001). *La Experiencia de Investigar. Recomendaciones precisas para realizar una investigación y no morir en el intento*. Venezuela: Predios.
- Pita Fernández, S. y Pértegas Díaz, S. (2005). *Metodología de la investigación*. [Documento en línea]. Disponible: <http://www.fisterra.com/mbe/investiga/graficos/graficos.htm> [Consulta: 2007, Mayo 30]
- Ramírez, T. (1999) *Cómo Hacer un Proyecto de Investigación*. (2^{da} ed.) Caracas, Venezuela: Panapo de Venezuela, C.A.
- Ruiz Bolívar, C. (2002). *Instrumentos de Investigación Educativa*. Venezuela: Fedupel.
- Sabino, C. (2002). *El proceso de investigación*. Buenos Aires: Editorial El Cid.
- Sierra, C. (2004). *Estrategias para la Elaboración de un Proyecto de Investigación*. Maracay, Venezuela: Insertos Médicos de Venezuela C.A.
- Sierra Bravo, R. (2001). *Técnicas de Investigación Social. Teoría y Ejercicios*. Madrid: Paraninfo S.A.
- Tamayo y Tamayo, M. (2004) *El proceso de la investigación científica*. México: Limusa
- Universidad Nacional Abierta (1991). *Proyecto de Investigación I en Dificultades de Aprendizaje*. Caracas: Autor.
- Universidad Pedagógica Experimental Libertador-UPEL. (2006). *Manual de trabajos de grado de especialización y maestría y tesis doctorales*. (4^a ed.). Caracas: Fedupel.
- Wikipedia.com (2007) *Encuesta* [Documento en línea]. Disponible: <http://es.wikipedia.org/wiki/Encuesta> [Consulta: 2007, noviembre 5]