

Los Instrumentos de Evaluación en la Investigación Educativa

Gloria Mousalli-Kayat

Los instrumentos de evaluación en la investigación educativa

Para referenciar este documento:

Mousalli-Kayat, G. (2017). *Los instrumentos de evaluación en la investigación educativa*. Mérida.

Este documento está bajo una licencia de Creative Commons Reconocimiento 4.0 Internacional



Tabla de contenido

Introducción	2
Los instrumentos de evaluación en la investigación educativa:	4
Validez de los instrumentos	5
Determinación de la Validez de contenido de la prueba	6
Aspectos asociados al proceso de validez de los instrumentos	7
Confiabilidad de los instrumentos	8
Métodos para estimar la confiabilidad de instrumentos	9
Resumen del proceso de construcción de pruebas objetivas como instrumentos en la investigación educativa	12
Referencias:	12
Anexos	13
Anexo 1: Formato de solicitud de evaluación de pruebas	13
Anexo 2: Formato de solicitud de evaluación de pruebas	14
Anexo 3: Formato Autorización del juez para el uso de la información suministrada.....	14

Introducción

En el enfoque cuantitativo de la ciencia, donde la generación de conocimiento es posible solo a través del método hipotético-deductivo, sostiene que el encuentro con la realidad solo es posible a través de la experimentación, es decir, mediante la medición de los elementos que

definen al objeto de estudio. Sin embargo, este proceso de medición resulta, muchas veces, problemático en la educación y las ciencias sociales ya que las medidas están relacionadas a constructos abstractos, intangibles y no observables.

Antes de hablar sobre cómo se atiende el problema de la medida en la investigación educativa, es vital acotar algunos aspectos previos:

- La mayoría de los conceptos en la educación y las ciencias sociales tienen un significado dentro del contexto de la teoría de la que forman parte.
- Cada concepto, por lo tanto, tiene una definición operacional que se rige por la teoría general.
- Si un concepto está involucrado en la prueba de la hipótesis para apoyar la teoría, debe ser medido (primacía del enfoque cuantitativo).

Entonces, la primera decisión a la que se enfrenta el investigador es ¿cómo se medirá el concepto?, en el enfoque cuantitativo este concepto se convierte en variable y se operacionaliza en diversas dimensiones a fin de lograr su “medición”.

Algunos podrían argumentar que en perspectiva amplia, la medición puede venir de la observación, autoevaluación, entrevista, etc. Sin embargo, en el método experimental, la objetividad forma parte de la rigurosidad científica exigida, es decir, la medición no puede venir de la interpretación *de alguien* sobre el concepto medido.

Todo enfoque o paradigma se sustenta sobre principios y supuestos, cuando se desarrolla una investigación es necesario conocer a fondo el enfoque donde se enmarca (cuantitativo o cualitativo) y desarrollar cada fase de tal manera que cumpla con los requisitos que exige. Uno de los errores que cometen los investigadores, es que no tener claridad sobre los criterios que hacen válido los resultados en el paradigma en el que se desarrolló la investigación, por lo que sus conclusiones pueden ser cuestionada.

En este apartados abordamos lo concerniente a la investigación en el enfoque cuantitativo, aquí la medición debe ser objetiva, ya que todos los resultados se sustentan en este principio de objetividad, por ello se debe garantizar que los instrumentos utilizados sean los adecuados para una medida objetiva, significativa, precisa y eficiente.

Otros aspectos que debe responder el investigador al momento de iniciar una investigación en el enfoque cuantitativo son:

- Las unidades de análisis, es decir, los sujetos de donde se obtendrá la medida. Esta decisión no depende por completo del paradigma teórico, sino más bien de la pregunta o hipótesis de investigación.
- El objetivo de la medición, este aspectos se determina respondiendo a las siguientes interrogantes: ¿Qué es lo que el investigador quiere hacer con la medida? ¿Se desarrolló para un estudio específico o pretende confirma experimentos previos en nuevas poblaciones?

Una vez que se toman algunas de estas decisiones y se logra obtener la medida (proceso cuidadoso y riguroso), en el marco de la actitud investigativa surgen nuevas preguntas **¿cómo sabemos que estamos midiendo lo que queremos medir?** Partiendo del hecho que se mide un aspecto abstracto.

La medición es propia de las ciencias exactas, los instrumentos utilizados son calibrados a fin de garantizar la consistencia de la medida se mantenga, es decir, que en la repetición de la medición se obtenga el mismo resultado, el otro cuestionamiento que tiene el investigador es **¿si repito la medición tendré el mismo resultado?**

La primera pregunta apunta a la característica de validez de los instrumentos y la segunda se relaciona con la confiabilidad de los mismos. La validez y la fiabilidad son dos características que deben garantizarse en la investigación educativa y en las ciencias sociales, aclarando que su importancia es vital en el marco del enfoque cuantitativo, aclarando que el instrumento de medición empleado realmente mide el concepto o el constructo que se desea medir.

A pesar de lo simple que parece, a menudo se omite o se menciona pasivamente en la propuesta o en informe de investigación poniendo en dudas las conclusiones generadas. Algunos autores indican que esto se debe la escasa destreza en materia de estadística de los investigadores en la educación y las ciencias sociales (Bolarinwa, 2015).

[Los instrumentos de evaluación en la investigación educativa:](#)

Ya hemos dejado claro que la medición objetiva es clave en la investigación cuantitativa, por tal motivo en las investigaciones exploratorias, descriptivas, correlacionales o explicativas (Mousallikayat, 2015), la evaluación es una pieza clave cuando uno de los constructos es rendimiento académico, aprendizaje o conocimiento. Aquí la evaluación no solo obedece a los objetivos, sino que también responde a las preguntas de investigación y a las hipótesis que el investigador haya planteado.

En esta sección debemos retomar las características deseables de un instrumento de evaluación:

- Validez: que mida con precisión el rasgo o características que se pretende medir
- Confiabilidad: que la medición que aporte sea consistente en el tiempo
- Objetividad: que la medición no vea afectada por variables externas, por ejemplo el criterio de la persona que corrige.
- Discriminación: que logre diferenciar entre los participantes que saben de los que no saben
- Integralidad: que la batería de preguntas cubran todos los objetivos que se esperan valorar con su aplicación
- Facilidad: que su elaboración y administración sea sencilla

Todas estas características están relacionadas con las pruebas objetivas de evaluación, por lo que en este documento estudiaremos la validez y confiabilidad de este tipo de pruebas a fin de ser parte de la investigación educativa.

Validez de los instrumentos

De manera sencilla, la validez de un instrumento se refiere a la cualidad del instrumento de medir lo que debe medir. Cuando una madre toma la temperatura de su hijo usando el termómetro, ella tiene la certeza que mide la temperatura corporal, no hay espacio para la duda, pues lo mismo debe pasar en la medición de constructos complejos como conocimiento, aprendizaje, motivación, entre otros.

En general, podemos decir que la validez se clasifica en dos componentes validez interna y externa (Bolarinwa, 2015). La **validez interna** se refiere a qué tan exactamente las medidas

obtenidas de la investigación estaban realmente cuantificando lo que se esperaba medir, mientras que la **validez externa** se refiere a la precisión con que las medidas obtenidas de la muestra del estudio describen la población de donde se sacó la muestra del estudio. En lo referido a los instrumentos trabajaremos con la validez interna.

Como vemos, la validez es una característica compleja y su definición depende también del tipo de investigación, la escala de medida, la población en estudio y la hipótesis que plantea el investigador.

En lo que respectan las pruebas objetivas de evaluación podemos hablar de validez de una prueba en la medida en que sea verdaderamente representativa de los contenidos de la asignatura y a profundidad del conocimiento adquirido (objetivos). Aquí hemos definido la **validez teórica** (Flores 2001) que viene dada por la validez de constructo y de contenido, la primera hace referencia al universo de conocimientos asociados al tema y la segunda al muestro realizado en la prueba, el muestreo de contenido sobre ese tema que se expone en la prueba. De manera alterna a la validez teórica, existe una validez empírica relacionada con el impacto de lo aprendido en el futuro (**validez predictiva**) y la relación de lo aprendido con el resto de contenidos del programa de la asignatura o de la carrera (**validez concurrente**).

Determinación de la Validez de contenido de la prueba

En la investigación educativa, es necesario confirmar la validez teórica de contenido, la cual inicia con el diseño de la prueba y la especificación de cada ítem de la prueba (conocimiento) y su nivel de dominio (profundidad). El método comúnmente utilizado se concentra en establecer la concordancia de un panel de expertos quienes valoran la prueba en su totalidad y en cada uno de los ítems, con base en ese resultado se ajusta la prueba garantizando que los ítems planteados son representativos de las dimensiones del constructo que se pretende medir con la prueba.

La concordancia de los expertos se determina a través de coeficientes de concordancia como el de validez de contenido (CVC), concordancia W de Kendall, Coeficiente Kappa, entre otros. Todos ellos representan una medida de asociación de variables, que en este caso determina el acuerdo inter-jueces, en otras palabras, una medida de coincidencia de un número determinado de jueces durante la validación de un instrumento.

El Coeficiente de Validez de Contenido (CVC) propuesto por Hernández-Nieto (2011), permite valorar el grado de acuerdo de los expertos (el autor recomienda la participación de entre tres y cinco expertos) respecto a cada uno de los diferentes ítems y al instrumento en general. El método requiere la aplicación de una escala tipo Likert de cinco alternativas, para luego calcular la media obtenida en cada uno de los ítems y, con base a esta, se calcula el CVC para cada ítem

$$CVC_i = M_x / V_{\max}$$

donde M_x representa la media del ítem en la puntuación dada por los expertos y V_{\max} la puntuación máxima que el ítem alcanzada por el ítem. Por otro lado, debe calcularse el error asignado a cada ítem (Pe_i), de este modo se reduce el posible sesgo introducido por alguno de los jueces, obtenido mediante $Pe_i = (1/j)^j$ siendo j el número de expertos participantes. De esta manera el CVC se calcula aplicando $CVC = CVC_i - Pe_i$.

Respecto a su interpretación el autor recomienda mantener únicamente aquellos ítems con un CVC superior a 0.80.

Para revisar otros coeficientes de Validez se puede revisar el artículo de Pedrosa, Suárez-Álvarez y García Cueto (2013), quienes hacen una comparación interesante de diferentes coeficientes.

Aspectos asociados al proceso de validez de los instrumentos

Hay diversos aspectos a tomar en cuenta al momento en que se someta el instrumento al juicio de expertos, estos son:

1. **Número de expertos:** se recomienda que sea un número impar que puede ser 3 o más.
2. **Sobre la cualidad de experto:** es importante indicar que la técnica indica que quien valora el instrumento o la prueba debe ser un experto. La experticia debe ser en el constructo que se está evaluando o que se pretende medir, no es un experto en metodología de la investigación, ni en estadística, sino en lo que pretendemos medir.
3. **Criterios de valoración por parte de los expertos:** no hay un estándar al respecto, pero Hernández-Nieto (2002) recomienda los siguientes:

- **Pertinencia:** El grado de correspondencia entre el enunciado del ítem y lo que se pretende medir.
 - **Claridad Conceptual:** Hasta qué punto el enunciado del ítem no genera confusión o contradicciones.
 - **Redacción:** Si la sintaxis, ortografía y la terminología utilizadas son apropiadas.
 - **Escala y Codificación:** Si la escala empleada en cada ítem es apropiada y la misma ha sido debidamente codificada.
4. **Escala para la valoración de cada ítem:** se recomienda una escala tipo Likert con 3 o 5 niveles, por ejemplo
- 1= Inaceptable 2= Deficiente 3= Regular 4= Bueno 5= Excelente

Es recomendable siempre dejar una columna para por ítem donde el juez pueda agregar comentarios y sugerencias.

5. **Formatos** a entregar a cada juez: el protocolo de valoración por juicio de expertos implica que a cada experto se le entregue los siguientes formatos:
- Carta de solicitud de evaluación del instrumento o la prueba (Ver Anexo 1).
 - Tabla de operacionalización de las variables y bosquejo general de idea de investigación y sus objetivos
 - La prueba con todos los elementos, tal como la va a recibir el estudiante
 - Formato de evaluación de la prueba, donde se incluyen los criterios y la escala (Ver Anexo 2)
 - Autorización del juez para el uso de la información suministrada (Ver Anexo 3)
6. **Valoración cualitativa de los ítems.** Una vez calculado el CVC, el investigador debe realizar una revisión cualitativa de los ítems en cada uno de los criterios evaluados, eliminando aquellos que no tengan un CVC adecuado, además de tomar en consideración algunas mejoras sugeridas por los jueces.

Confiabilidad de los instrumentos

La investigación requiere una medición confiable, esto se refiere a que las mediciones se repiten en cada ocasión que se apliquen los instrumentos. Volvamos al ejemplo de la madre que

mide la temperatura de su hijo, la madre puede realizar dos o tres medidas seguidas, con el termómetro, y espera que la medida siempre sea la misma o muy cercanas entre sí; si la mamá consigue diferencias entre las medidas repetidas entonces llegará a la conclusión que el termómetro no es confiable.

Aunque la confiabilidad contribuye de manera importante a la validez de una prueba, no es una condición suficiente para garantizar su validez. Eliminando la subjetividad de un observador, podemos decir que la falta de confiabilidad o consistencia de una medida proviene de los instrumentos de medición, dejando claro que esto afectará invariablemente a su validez.

Es importante tener claro, que la consistencia o la confiabilidad se refiere a los puntajes de la prueba no a las personas. Lo que se pretende indicar aquí es que en el grado en que las respuestas de un individuo (sus puntajes) en una prueba se mantengan a lo largo del tiempo esa prueba es confiable.

En la educación y las ciencias sociales, el tiempo es un factor que afecta directamente la estabilidad de las medidas, en el caso del aprendizaje ya que lo aprendido en el nivel de conocimiento podría olvidarse, o en los niveles superiores pueden cambiar la evaluación en función de nuevos aprendizajes. En tal sentido, por el mismo dinamismo del aprendizaje la confiabilidad puede determinarse por el método de test-retest, con períodos que no excedan los tres meses. Otro método que puede implementarse es el de equivalencia o similitud entre instrumentos, también llamado forma alternativa, consiste en comparar los puntajes del sujeto en dos instrumentos diferentes.

Métodos para estimar la confiabilidad de instrumentos

Test-Retest

La correlación test-retest proporciona un coeficiente de confiabilidad de la medida a lo largo del tiempo. Se dice que este aspecto de confiabilidad o consistencia ocurre cuando se obtienen puntajes iguales o similares con pruebas repetidas con el mismo grupo de encuestados. En otras palabras, los puntajes son consistentes de 1 vez al siguiente. La estabilidad se evalúa a través de un procedimiento de prueba + repetición que implica la administración del mismo instrumento de medición, como una prueba, a los mismos individuos en las mismas condiciones después de un período de tiempo.

Los cálculos estadísticos inmersos en este procedimiento son de correlación de los puntajes de cada individuo antes y después. Dos supuestos subyacen al uso del procedimiento de test-re-test:

- La característica que se mide no cambia durante el período de tiempo
- El período de tiempo sea suficientemente largo para que los recuerdos de los individuos, sobre la prueba, no influyan en los puntajes del re-test, asociada al efecto de la memoria.

Forma Alternativa

La forma alternativa estima la correlación entre los puntajes obtenidos por el mismo individuo en dos pruebas diferentes sobre el mismo constructo, las cuales se administran casi al mismo tiempo.

El procedimiento consiste en preparar dos pruebas donde las preguntas o respuestas se reformulan, se reordenan para producir dos elementos que son similares pero no idénticos. Luego se separan los resultados de cada individuo en las dos formas y se correlacionan las medidas, cuanto mayor es el grado de correlación entre las dos formas, más equivalentes son. El problema del uso de este método es la construcción de las formas paralelas, ya sabemos que es bastante difícil tener un instrumento o prueba bien desarrollada para medir un constructo ahora imagínense elaborar dos.

Consistencia interna

Un método muy utilizado hace referencia a la consistencia interna, que correlaciona el puntaje de cada estudiante con los puntajes en los ítems que integran la prueba. Flores (2001) propone que el cálculo de la consistencia interna, permite obtener automáticamente la validez teórica de la prueba siempre que se haya tenido cuidado en la especificación en el diseño de la prueba, como esta última condición no se puede garantizar, recomendamos siempre realizar la validez de contenido mediante juicio de expertos como técnica de confirmación de validez teórica de contenido.

En cualquiera de los métodos utilizados, la confiabilidad es una medida de correlación el coeficiente que se utilice dependerá del nivel de los puntajes, si son medidas en escalas de intervalo o razón (Mousalli-Kayat, 2015) se utiliza en Alpha de Cronbach, si son dicotómicos se puede calcular el KR-20, para mayor detalle revisar Hernández, Fernández y Batista (2014)

Alpha de Cronbach

El coeficiente α de Cronbach Alfa representa la consistencia interna del test, el grado que todos los ítems del test covarían entre sí.

$$r_{xy} = \frac{n}{n-1} \left(1 - \frac{\sum s_i^2}{s_T^2} \right)$$

Donde:

n = # de ítems

s_i^2 = varianza de cada ítem

s_T^2 = varianza de la suma de los ítems

Nota: Hoy día se utilizan software estadísticos (SPSS, R, minitab) para el cálculo de los coeficientes

El valor de alfa que se obtiene mientras más se acerque a 1 es mayor la consistencia interna de los ítems analizados, George y Mallery (2003) indican la siguiente escala de interpretación para el coeficiente

- Alfa > 0,9 Excelente
- Alfa > 0,8 Bueno
- Alfa > 0,7 Aceptable
- Alfa > 0,6 Cuestionable
- Alfa > 0,5 Pobre
- Alfa < 0,5 Inaceptable

Por lo general el alpha de Cronbach se utiliza para ítems con puntuaciones numéricas, sin embargo, para aquellas pruebas o instrumentos de medición que utilicen escalas tipo Likert o escalas aditivas, solo se requiere hacer una transformación de la escala a valores como se muestra a continuación:

Escala	Valor de codificación
Totalmente en desacuerdo	1
En desacuerdo	2
Ni de acuerdo ni en desacuerdo	3
De acuerdo	4
Totalmente de acuerdo	5

Resumen del proceso de construcción de pruebas objetivas como instrumentos en la investigación educativa

Ya hemos visto la importancia de la medición en el enfoque cuantitativo de la investigación educativa, es por ello, que debemos retomar los aspectos del diseño de la prueba incorporando nuevos elementos en el proceso:

Fase 1: Operacionalización de las variables de medición. en esta etapa se descomponen las variables a medir, partiendo del constructo, sus dimensiones e indicadores (Hernández, Fernández y Batista, 2014)

Fase 2: Elaboración de los ítems. En esta sección hacemos uso de todas las herramientas y formatos empleados en la etapa de diseño de pruebas objetivas y de desarrollo (Mousalli-Kayat y Luzardo, 2017), aquí se determinan los ítems de la prueba en función de los contenidos, los objetivos de aprendizaje y las preguntas de investigación. Tomando en consideración todos las recomendaciones dadas para tal fin.

Fase 3: Validez de la prueba. Una vez elaborado el instrumento se somete al juicio de expertos, revisando aspectos tales como pertinencia, claridad conceptual, redacción, codificación. El número de expertos debe ser 3, 5 o 7. Se determina el coeficiente de validez de contenido (CVC), que indica la concordancia y luego se realiza la revisión cualitativa, donde se verifica si la concordancia es para la eliminación, modificación o aprobación. Se realiza el ajuste de la prueba de acuerdo a los resultados.

Fase 4: Confiabilidad de la prueba. Se selecciona un grupo piloto de estudiantes, se aplica la prueba y se calcula el coeficiente correspondiente. Se valora la confiabilidad, en caso de no tener un nivel aceptable realizar los ajustes necesarios

Una vez cumplidos todos estos pasos, se cuenta con una prueba válida y confiable que puede ser utilizada en la investigación educativa de enfoque cuantitativa, sin riesgo a invalidar la investigación.

Referencias:

Bolarinwa O. A. (2015). Principles and methods of validity and reliability testing of questionnaires used in social and health science researches. *Nigerian Postgraduate Medical Journal*, 195-201. Obtenido de http://www.npmj.org/temp/NigerPostgradMedJ224195-3275178_090551.pdf

Flores O. R. (2001). Evaluacion Pedagogica y Cognicion. Mc Graw Hill. Colombia

George y Mallery (2003). SPSS for Windows step by step: A Simple Guide and Reference 11.0 Update (4ª ed.) Boston

Hernández, Fernández y Batista (2014). Metodología de la Investigación. (6ª ed.) Mc Graw Hill. México.

Hernández-Nieto, R. A. (2002), Contribuciones Al Análisis Estadístico de Datos. Mérida, Venezuela: Universidad de Los Andes.

Mousalli-Kayat, G. y Luzardo H. (2017). Técnicas e Instrumentos de la Evaluación Centrada en Resultados. Mérida.

Mousalli-Kayat, G. (2015). Métodos y Diseños de Investigación Cuantitativa. Mérida. Obtenido de https://www.researchgate.net/publication/303895876_Metodos_y_Disenos_de_Investigacion_Cuantitativa

Pedrosa, Suárez-Álvarez y García Cueto (2013). Evidencias sobre la Validez de Contenido: Avances Teóricos y Métodos para su Estimación. [Content Validity Evidences: Theoretical Advances and Estimation Methods]. Acción Psicológica, 10(2) <http://dx.doi.org/10.5944/ap.10.2.11820>

Anexos

Anexo 1: Formato de solicitud de evaluación de pruebas

Fecha

Estimado(a)

XXXXX

Reciba un cordial saludo, la presente misiva tiene por objetivo informarle que en este momento me encuentro desarrollando la investigación _____, en conocimiento de su trayectoria profesional y su vinculación a la temática en la que estamos trabajando, solicito su colaboración para la valoración del instrumento para medir _____

Agradeciendo de antemano su amable colaboración como experto, me suscribo de usted.

Atentamente

XXXXX

Anexo 2: Formato de solicitud de evaluación de pruebas

Dr./ Sr.. Especialista se le pide su colaboración para que luego de un riguroso análisis de los ítems del instrumento anexo, marque con una X en la casilla que considere conveniente de acuerdo a su criterio y experiencia profesional indicando si cuenta o no con los requisitos mínimos, en cuanto a pertinencia, claridad conceptual, redacción y codificación

- **Pertinencia:** El grado de correspondencia entre el enunciado del ítem y lo que se pretende medir.
- **Claridad Conceptual:** Hasta qué punto el enunciado del ítem no genera confusión o contradicciones.
- **Redacción:** Si la sintaxis, ortografía y la terminología utilizadas son apropiadas.
- **Escala y Codificación:** Si la escala empleada en cada ítem es apropiada y la misma ha sido debidamente codificada.

FORMATO DE EVALUACIÓN DE INSTRUMENTO. CRITERIOS GENERALES.

Evaluator:

Fecha: _____

Instrumento:

Escala Evaluativa

1= Inaceptable 2= Regular 3= Excelente

Contenido			Evaluación		
Ítem	Criterios Generales	Observaciones	1	2	3
1	Pertinencia				
	Claridad Conceptual				
	Redacción				
	Codificación				
2	Pertinencia				
	Claridad Conceptual				
	Redacción				
	Codificación				

Se repiten las filas hasta completar el número de ítems

Anexo 3: Formato Autorización del juez para el uso de la información suministrada

AUTORIZACIÓN

Mediante la firma de este documento, doy mi consentimiento para que _____, utilice la información por mí suministrada, en el marco de la validación del instrumento titulado _____, en las publicaciones de carácter académico y científico que se generen a partir de la investigación desarrollada.

Nombre del Experto

CI experto

Huella digital