



## EVIDENCIA DE VALIDEZ BASADA EN EL CONTENIDO: UNA APROXIMACIÓN CUALITATIVA

**Reyes Piñuelas Erika Paola**

Universidad Autónoma de Baja California  
erikareyes@uabc.edu.mx

**Área temática:** Evaluación educativa

**Línea temática:** Diseño y validación de instrumentos

**Tipo de ponencia:** Reporte final de investigación



### Resumen

El objetivo es describir una ruta metodológica cualitativa para la obtención de evidencias de validez basadas en el contenido a partir de un comité de expertos. La validez es un criterio imprescindible en la calidad de las interpretaciones de las puntuaciones derivadas de la aplicación de los instrumentos de medición, y en particular las evidencias de validez de contenido se refiere a, si los ítems que contiene el instrumento representan el dominio de contenido que se desea medir. Las metodologías de mayor desarrollo para su valoración son de enfoque cuantitativo, sin embargo, se recomienda contar con una metodología combinada para brindar solidez a los resultados, la presentación desde la perspectiva cualitativa es de interés en este trabajo. Cinco jueces especialistas analizaron la claridad, coherencia, suficiencia y relevancia de los componentes de un instrumento que busca medir las Habilidades para la vida en jóvenes, a partir de la técnica de análisis de contenido, se tomaron decisiones sobre la modificación de 75 de las 80 afirmaciones que componen el instrumento. Se concluye que el análisis detallado y exhaustivo de los elementos y características del instrumento, así como en la revisión de su pertinencia y representatividad con respecto al constructo que se desea medir, es uno de los aspectos que se privilegiaron en esta propuesta.

**Palabras clave:** Instrumentos de medición, validez, cuestionarios, análisis de contenido.

### Introducción

La validez en los instrumentos es un tema en debate constante, y en especial el de mayor impacto social es el asociado a las consecuencias de los usos de los test, sin embargo, los

usos potenciales se establecen cuando se determina el *marco general*; Muñiz y Fonseca-Pedrero (2019) lo enuncian como el primero de los 10 pasos que se requieren en la elaboración de instrumentos e incluye aspectos relevantes como la delimitación de la variable a medir, tipo de aplicación, el formato de administración, y tipo de decisiones a tomar a partir de las puntuaciones. Los siguientes pasos son la *definición de la variable medida, especificaciones, construcción de los ítems, la edición* y en el paso seis, en el *estudio piloto* es el momento que se obtienen evidencias de validez basadas en el contenido, posteriormente le siguen la *selección de otros instrumentos de medida, aplicación del test, propiedades psicométricas*, para concluir con la *versión final del test*.

En la cronología antes planteada en el diseño de los instrumentos, se ubica el momento en el cual, usualmente se realizan procedimientos para obtener evidencias de validez y confiabilidad, entre estas, la de contenido de la prueba. Sin embargo, cabe señalar que es en los pasos previos que se desarrollan la definición del dominio y la representación del dominio, fuentes principales de evidencias de validez de contenido. Según Pedrosa et al (2014), la definición del dominio se refiere a la definición operacional de la variable a medir, la matriz operacional es la muestra clara de esta definición, mientras que, la representación del dominio, se compone a su vez de dos elementos, la representatividad como una muestra del contenido esencial del dominio, y la relevancia, que alude al aspecto que cada ítem mide del fenómeno definido.

En un planteamiento tradicional la validez se refiere a la medida en que un instrumento mide de manera adecuada el constructo que se pretende evaluar, empero el concepto ha evolucionado, hacia un concepto unitario y se aleja de una tipología tripartita de contenido, criterio y constructo que por varias décadas prevaleció; en los estándares para pruebas educativas y psicológicas, validez “se refiere al grado en que la evidencia y la teoría respaldan las interpretaciones de los puntajes de una prueba para usos propuestos de las pruebas” (American Educational Research Association, American Psychological Association y National Council on Measurement in Education [AERA, APA y NCME], 2018, p. 11), busca validar las interpretaciones derivadas del instrumento para un uso particular, es decir no se valida el instrumento sino las interpretaciones; y se plantean cinco fuentes de evidencia de validez que representan diversos aspectos de esta:

1. Evidencia basada en el contenido de la prueba
2. Evidencia basada en los procesos de respuesta
3. Evidencia basada en la estructura interna
4. Evidencia basada en relaciones con otras variables
5. Evidencia de validación y consecuencias de las pruebas

En el caso de la obtención de evidencias basadas en el contenido de la prueba, existen diversas metodologías entre las más utilizadas se encuentran, la entrevista cognitiva y el juicio de expertos (Urrutia et al., 2014), la primera implica que los participantes piensan en voz alta al tiempo que resuelven el instrumento; mientras que, el juicio de expertos consiste en contar con

un número determinado de expertos que evalúan los diferentes ítems en función de diferentes atributos, entre los que destacan la relevancia y representatividad, con frecuencia emiten un valor a partir de una escala tipo Likert.

En cuanto a los criterios de selección de los jueces expertos la AERA, APA y NCME (2018) señala en uno de los estándares respecto al uso de información que proviene de opiniones o decisiones de jueces, “deben describir completamente los procedimientos para seleccionar a dichos expertos y para obtener los juicios o calificaciones. Deben presentarse las calificaciones y la experiencia de los jueces. La descripción de procedimientos debe incluir cualquier capacitación e instrucciones proporcionadas” (AERA, APA y NCME, 2018, pp. 27-28)

Por otro lado, en relación a los métodos basados en un comité de expertos, Pedrosa et al (2013) los analiza a partir de una revisión de la literatura, esta data de la década de los 60, y argumentan que la propuesta de Lawshe presenta mayores ventajas en comparación con el resto, por la facilidad de cálculo e interpretación; y el método propuesto por Rubio et al como el más completo y exhaustivo. Concluye que el análisis de la validez de contenido debe contener de manera imprescindible una estrategia combinada con los enfoques cualitativo y cuantitativo (ver tabla 1).

**Tabla 1. Síntesis de los métodos basados en el juicio de expertos para el análisis de la validez de contenido.**

Año	Autores	Método
1961	Tucker	Basado en Análisis Factorial
1975	Lawshe	CVR
1977	Rovinelli y Hambleton	Índice de congruencia ítem-objetivo
1980	Aiken	V
1980, 1984	Hambleton	Índice de congruencia
1986	Lynn	Índice de Validez de Contenido (IVC)
1992	Sireci y Geisienger	Escalamiento multidimensional y análisis de clusters
1997	Levine, et al.	Capacidades mínimas (MQ en inglés)
2001	Fitch, et al.	Rango Interpercentil Ajustado a la Simetría (IPRAS)
2002	Hernández-Nieto	Coeficiente de Validez de Contenido (CVC)
2003	Rubio, et al.	Índice de Validez Factorial (FVI en inglés)
2012	Claeys, et al.	Índice Promediado de la Desviación Media (ADm en inglés)

*Nota.* Elaborado por Pedrosa et al (2013).

En este mismo tema, se encuentran metodologías que combinan los enfoques cualitativo y cuantitativo, tal es el caso del Método Delphi en el que los jueces expertos se expresan de forma

libre al ser confidencial, bajo una serie de fases en la que participan los jueces, o la metodología Q que identifica de manera cualitativa y cuantitativa elementos comunes de opinión del comité de expertos (Urrutia et al., 2014).

Juárez-Hernández y Tobón (2018) consideran que la mirada cualitativa para analizar la validez de contenido destaca un análisis detallado de las observaciones y sugerencias que son expuestas por los expertos, dentro de las decisiones que pueden derivar son la eliminación o permanencia de ítems, modificación en las indicaciones, inclusión de ítems, y la incorporación de otras dimensiones, además permite ajustar el contenido y redacción de los ítems, tanto a nivel morfológico como sintáctico. Además, se asegura la representatividad del constructo a medir, dado que, permite examinar si los ítems del instrumento abarcan adecuadamente el dominio o ámbito del constructo que se pretende evaluar. Esto implica asegurarse de que las preguntas o afirmaciones reflejen de manera precisa y completa las dimensiones o aspectos clave del fenómeno en cuestión.

Además, en el caso de instrumentos que serán utilizados en contextos culturales distintos al de su origen, la aproximación cualitativa es especialmente relevante. Permite evaluar la pertinencia y relevancia cultural de los ítems, así como identificar posibles diferencias en la interpretación o comprensión de los conceptos evaluados. Esto es esencial para garantizar que el instrumento sea válido y aplicable en diferentes contextos culturales.

Por lo anteriormente planteado, el objetivo es describir una ruta metodológica cualitativa para la obtención de evidencias de validez de contenido a partir de un juicio de expertos.

## Desarrollo

La metodología propuesta por Escobar y Cuervo (2008) para evaluar la validez de contenido, contiene elementos de tipo cuantitativo y cualitativo, a partir del juicio de expertos, estos valoran la suficiencia, claridad, coherencia y relevancia de los ítems a partir de un escalamiento tipo Likert. A continuación, se expone el proceso realizado en cuanto a la información recabada de tipo cualitativo.

**Participantes.** Se conformó un grupo de cinco expertos como lo sugieren la American Educational Research Association, American Psychological Association y National Council on Measurement in Education (2014). El grupo de jueces se integra por dos especialistas en evaluación con 12 y siete años de experiencia, respectivamente; una especialista en jóvenes e intervenciones socioeducativas con 10 años de experiencia y evaluación colegiada del aprendizaje; y dos especialistas en desarrollo infantil y en el desarrollo de habilidades socioemocionales con más de 20 y 10 años de experiencia, respectivamente.

**Instrumentos.** El principal insumo de trabajo fue la versión original del Cuestionario de hpv (Díaz et al., 2013) de autoaplicación. Se encuentra integrado por 80 ítems en escala tipo Likert con cinco opciones de respuesta que van de Siempre a Nunca y cuenta con ítems

con direccionalidad positiva y negativa. Se compone de 10 dimensiones (con ocho ítems por dimensión): 1) Conocimiento de sí mismo, 2) Empatía, 3) Comunicación efectiva y asertiva, 4) Relaciones interpersonales, 5) Toma de decisiones, 6) Solución de problemas y conflictos, 7) Pensamiento creativo, 8) Pensamiento crítico, 9) Manejo de sentimientos y emociones, 10) Manejo de tensiones y estrés. El instrumento se fundamenta en las hpv propuestas por la who (1994).

Sin embargo, se hacen recomendaciones puntuales para tomar en consideración en estudios posteriores: 1) añadir y/o mejorar algunos ítems para aumentar su confiabilidad en algunas dimensiones y 2) reformular los ítems con una direccionalidad no asociada al rasgo medido o que cuentan con negación o doble negación, ya que se infiere pueden llegar a ser confusos para los encuestados.

**Procedimiento.** Se realizaron los siguientes pasos:

- Selección y contacto de un grupo de expertos para realizar el jueceo.
- Envío de formato a los jueces con las instrucciones.
- Análisis cualitativo del contenido según Carrion et al. (2015)
- Toma de decisiones a partir del análisis previo y en conjunto con los resultados del coeficiente.

## Resultados

A partir de la valoración del instrumento por parte de los jueces expertos, sus comentarios se analizaron mediante un análisis cualitativo del contenido, se clasificaron sus opiniones en cuatro categorías a fin de relacionar los comentarios de los expertos con las categorías previamente definidas y visualizar las aportaciones de éstos; con esto se identificaron por un lado los elementos coincidentes y por otro los elementos que requerían mejorarse y no incluían el instrumento; las categorías fueron las siguientes:

- Suficiencia: si el número de ítems que aluden a un indicador son suficientes.
- Claridad: ítems que refieren dos aspectos a la vez, necesidad de mejorar la redacción modificando una o más palabras pero cuya esencia del contenido del ítem permanece; ítems que requieren modificar su direccionalidad, es decir, que cada ítem se redacte de acuerdo con el crecimiento de la variable.
- Coherencia: ítems que no aluden al indicador al que pertenecen.
- Relevancia ítems que valoran un aspecto no importante del instrumento.

La mayoría de los ítems tuvieron observaciones tanto de forma como de fondo; debido a la extensión del instrumento y sus observaciones, se ejemplifican sólo algunas observaciones

realizadas por los jueces. En el caso de la categoría de suficiencia, los indicadores de Reconocimiento del carácter y Relaciones de amistad se consideró por al menos un juez que requerían más ítems para valorarlos.

En el caso la claridad de los ítems, categoría con el mayor número de observaciones con respecto a las otras, los jueces consideraron que 13 ítems valoraban dos aspectos a la vez; por citar algunos ejemplos, el ítem 1 Se me facilita iniciar y mantener relaciones de amistad, el juez "A" consideró que facilitar y mantener son dos elementos que no pueden evaluarse en un mismo ítem, o en el caso del ítem 39 Tomo las decisiones considerando el beneficio propio y de los demás, el juez "C" consideró que algunos estudiantes podrían verse identificados con el beneficio propio pero no con el de los demás. Además, en esta misma categoría, se detectaron con problemas de redacción 51 ítems; por citar ejemplo: el ítem 31 Mis relaciones con profesores y/o jefes suelen ser conflictivas, recibió observaciones de todos los jueces: juez "A" "Sería mejor relaciones con la autoridad". En el caso del ítem 73 Soy consciente de que el estrés me hace daño y que debe controlarse cuando empieza a interferir con mi desempeño, el juez "C" propone cambiar "interferir" por "afectar", mientras que el juez "D" sugiere simplificar la redacción.

En el caso de la observación que alude a la direccionalidad, 39 ítems no correspondían con el crecimiento de la variable o rasgo a medir; por citar un ejemplo, el ítem 31 antes mencionado o el 30 Me molesta que mis padres se entrometan en mis decisiones y amistades.

En la categoría de coherencia, ocho reactivos se consideraron por al menos un juez que no aludían al indicador o dimensión de procedencia; por ejemplo, el ítem 36 Las decisiones más importantes de mi vida no han tenido los frutos esperados, que pertenece al indicador de Decisiones constructivas, o el ítem 12 Considero que las personas deben dar cuando reciben porque es lo justo, que pertenece al indicador de Respeto de la diferencia y concepto de justicia.

Después de la clasificación de las observaciones de los jueces expertos se analizó la pertinencia de modificar, eliminar o elaborar ítems, 30 sufrieron alguna modificación en la redacción, en 31 ítems se cambió la direccionalidad acorde al crecimiento del rasgo y 14 fueron eliminados; finalmente, los ítems que no se modificaron fueron cinco: el 9, 51, 52, 79 y 80.

## Conclusiones

La obtención de evidencias de validez basadas en el contenido de instrumentos desde una aproximación cualitativa es de suma importancia en el proceso de desarrollo y validación de estos instrumentos. Se basa en el análisis detallado y exhaustivo de los elementos y características del instrumento, así como en la revisión de su pertinencia y representatividad con respecto al constructo que se desea medir. Esto implica asegurarse de que las preguntas o afirmaciones reflejen de manera precisa y completa las dimensiones o aspectos clave del fenómeno en cuestión.

Mediante el análisis cualitativo, fue posible identificar problemas de ambigüedad, confusión o falta de comprensión en los ítems del instrumento. Esto es crucial para garantizar que las preguntas sean claras y adecuadas para el público objetivo al que se dirige el instrumento, lo que a su vez contribuye a la calidad de las respuestas obtenidas.

Además, favoreció la identificación de sesgos al ubicar dos reactivos que los jóvenes podrían no encontrarse en condiciones de contestar como los relacionados con la dimensión Relaciones de pareja, la cual podría comprometer la validez de las respuestas obtenidas.

En suma, obtener evidencias de validez basadas en el contenido desde una aproximación cualitativa en instrumento acompañado de un índice cuantitativo, es fundamental para asegurar que los ítems sean representativos, claros, no sesgados y culturalmente adecuados. Estas evidencias brindan una base sólida para la validez del instrumento y, por ende, aumentan la confianza en las interpretaciones y decisiones basadas en los resultados obtenidos.

## Referencias

- American Educational Research Association, American Psychological Association y National Council on Measurement in Education (2018). *Estándares para pruebas educativas y psicológicas*. American Educational Research Association.
- Carrion, C. Soler, M y Aymerich, M. (2015). Análisis de la validez de contenido de un cuestionario de evaluación del aprendizaje basado en problemas: un enfoque cualitativo, *Formación universitaria*, 8(1), 13-22. <https://dx.doi.org/10.4067/S0718-50062015000100003>
- Juárez-Hernández y Tobón (2018). Análisis de los elementos implícitos en la validación de contenido de un instrumento de investigación. *Revista Espacios*, 39(53), 23
- Muñiz, J. y Fonseca-Pedrero (2019). Diez pasos para la construcción de un test. *Psicothema*, 31(1), 7-16. doi: 10.7334/psicothema2018.291
- Pedrosa, I., Suárez-Álvarez, J. y García-Cueto, E. (2014). Evidencias sobre la validez de contenido: avances teóricos y métodos para su estimación. *Acción Psicológica*, 10(2), 3-20. <http://dx.doi.org/10.5944/ap.10.2.11820>
- Urrutia Egaña, M., Barrios Araya, S., Gutiérrez Núñez, M., y Mayorga Camus, M. (2014). Métodos óptimos para determinar validez de contenido. *Educación Médica Superior*, 28(3), 547-558. [http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S0864-21412014000300014&lng=es&tlng=es](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0864-21412014000300014&lng=es&tlng=es).