

MODIFICACIÓN AL MODELO DE LAWSHE PARA EL DICTAMEN CUANTITATIVO DE LA VALIDEZ DE CONTENIDO DE UN INSTRUMENTO OBJETIVO

Agustín Tristán-López*
Instituto de Evaluación e Ingeniería Avanzada, México

Resumen

El procedimiento generalmente utilizado para evaluar la validez de contenido de los instrumentos psicológicos es el juicio de expertos y el único índice cuantitativo disponible hasta el momento es el CVR (Content Validity Ratio) de Lawshe (1975). El CVR sin embargo, requiere de un gran número de jueces, exigencia poco viable en la práctica y en consecuencia es muy poco utilizado por los psicómetras. Este trabajo revisa el índice de Lawshe y propone una modificación que permite superar sus limitaciones. Finalmente se presenta una aplicación práctica e ilustrativa del índice propuesto.

Palabras clave: Validez de contenido, Juicio de expertos, Índice de validez de contenido, CVR

Abstract

Experts' judgment is the generally used procedure to evaluate content validity of the psychological instruments, nevertheless, except the CVR (Content Validity Ratio) proposed by Lawshe (1975), a quantitative index are not available yet. The CVR needs a great number of judges, a slightly viable exigency in the practice and consequently it is very little used. This work analyze Lawshe's index and proposes a modification that allows overcoming its limitations. Finally an illustrative and practical application of the proposed index is presented.

Key words: Content validity, Experts judgment, Content validity index, CVR

Introducción

En su interesante artículo publicado en 1975, Lawshe propone un modelo para determinar un índice cuantitativo para la validez de contenido de un instrumento objetivo. Seguramente es el único índice disponible en la literatura a este respecto porque, a diferencia de la validez referida a criterio (como es el caso de la validez predictiva, concurrente, discriminante) que se determina por medio de una correlación, la validez de contenido de un instrumento se determina por medio del acuerdo entre jueces y no se emite un indicador objetivo de validez del instrumento. El esfuerzo de Lawshe, generalmente poco conocido, ha sido criticado porque se dice que se trata de un indicador del acuerdo entre los jueces y no tanto de la validez inherente al instrumento o a las decisiones que puedan emitirse con él; sin embargo, a pesar de las críticas, no existe ninguna alternativa a este respecto y por este simple hecho debería ser considerada más frecuentemente en los reportes de validez de un instrumento objetivo.

El problema que se presenta con el modelo de Lawshe es que requiere de un gran número de jueces, con acuerdos muy fuertes entre ellos; en un primer ejemplo, dicho autor empleó a 175 panelistas o jueces en uno de sus análisis para una aplicación en una planta industrial; en tanto que en un segundo ejemplo contó con 14 panelistas. Las condiciones en las que trabajó el autor fueron, sin duda alguna, extraordinarias, si se compara contra las condiciones reales de escala económica en el desarrollo de un instrumento objetivo, donde se cuenta con unos cuantos jueces o panelistas para analizar los ítems o las

* Mariano Jiménez 1830A, Col. Balcones del Valle, 78280 San Luis Potosí, México. email: ici_kalt@yahoo.com

preguntas del instrumento, a tal grado que si se cuenta con tres o cuatro jueces que sean expertos en el área a evaluar y que estén en disponibilidad para integrarse al comité o grupo de análisis de validez de contenido, se puede decir que se está en una situación privilegiada. En estas condiciones más realistas para muchos desarrollos, específicamente cuando se tienen menos de 7 jueces o panelistas, el modelo de Lawshe obliga a tener un acuerdo unánime entre todos los panelistas, en estos casos, basta con un sólo desacuerdo entre los jueces para invalidar el ítem. Cuando se revisa el modelo de Lawshe, se puede observar que tiene el inconveniente de que los valores de referencia se ven influidos por el efecto de tamaño, de tal modo que para grupos pequeños de panelistas el modelo es muy exigente y conforme el número de panelista se incrementa, el modelo se vuelve más laxo.

Este trabajo se enfoca a revisar el modelo de Lawshe y establecer una modificación al cálculo de los valores mínimos de aceptación para los ítems en general, pero sobre todo cuando se cuenta con muy pocos panelistas. Se revisa el modelo de Lawshe, se coteja el cálculo contra los valores propuestos por dicho autor y se hace una propuesta complementaria para uso general donde el efecto de tamaño se anula, a fin de contar con una exigencia similar para diversos tamaños de grupos de panelistas.

Modelo cuantitativo para dictaminar la validez de contenido

Lawshe (1975) propuso un modelo que consiste en organizar un Panel de Evaluación de Contenido, integrado por especialistas en la tarea a evaluar (pudiendo ser competencias, conocimientos, habilidades, funciones u otro tipo de elemento distintivo de la capacidad de un sujeto que va a ser evaluado), quienes contarán con un ejemplar de la prueba o del conjunto de ítems a analizar y sobre los cuales deberán emitir su opinión en tres categorías: esencial, útil pero no esencial, no necesario. Lawshe (1975) propone que estas categorías se asocien con la habilidad, conocimiento o competencia medidos por el ítem para el desempeño de una tarea.

Una vez que los panelistas anotan su opinión respecto a cada ítem en las tres categorías citadas, se debe determinar el número de coincidencias en la categoría “esencial” y se espera que se tengan grandes acuerdos entre los jueces, de hecho más del 50% de acuerdos debe ocurrir entre jueces en esta categoría para considerar que el ítem tiene un cierto grado de validez de contenido. Para establecer el consenso de los panelistas en la categoría “esencial”, Lawshe propone la Razón de Validez de Contenido (*Content Validity Ratio, CVR*) definida por la expresión:

$$CVR = \frac{n_e - \frac{N}{2}}{\frac{N}{2}} \quad [1]$$

Donde: n_e = número de panelistas que tienen acuerdo en la categoría “esencial”
 N = número total de panelistas

Esta expresión es planteada por Lawshe con la intención de que se pueda interpretar como si fuera una correlación, por tomar valores de -1 a +1; de tal modo que CVR es negativa si el acuerdo ocurre en menos de la mitad de los jueces; CVR es nula si se tiene exactamente la mitad de acuerdos en los panelistas y, finalmente, CVR es positiva si hay más de la mitad de acuerdos.

Adicionalmente, debe pensarse que no se trata solamente que CVR sea positiva sino que sea estadísticamente significativa, por lo cual Lawshe presenta una tabla con datos atribuidos a Schipper en un comunicado personal con el autor, pero no explica el modelo utilizado para obtener los valores de la tabla,

salvo que se trata de una prueba de una cola, con una significancia del 5%. Dentro de dicha tabla se tiene que para cinco panelistas el valor mínimo de CVR es 1 y para 40 panelistas el valor mínimo de CVR es 0.29, pudiendo notarse que no hay propuesta para menos de 5 panelistas. De hecho, hasta 7 jueces o panelistas se requiere tener un valor de CVR igual a 1, es decir, un consenso perfecto entre los jueces.

Una vez calculada la CVR de todos los ítems y aceptados los que tienen valores superiores a los mínimos propuestos por Lawshe, se calcula la media de CVR y con ello se obtiene el Índice de Validez de Contenido de toda la prueba (*Content Validity Index*, *CVI*), y que se debe interpretar como la concordancia entre la capacidad (habilidad, competencia, conocimiento, etc.) solicitada en un dominio específico y el desempeño solicitado en la prueba que trata de medir dicho dominio. La expresión para el CVI es:

$$CVI = \frac{\sum_{i=1}^M CVR_i}{M} \quad [2]$$

Donde: CVR_i = Razón de Validez de Contenido de los ítems aceptables de acuerdo con el criterio de Lawshe.
 M = Total de ítems aceptables de la prueba.

El modelo de Lawshe presenta tres problemas principales: En primer lugar, no explica ni justifica el procedimiento seguido para construir la tabla de valores mínimos de aceptación para la CVR, por lo cual a veces se complica su interpretación. Los valores mínimos de CVR se ven afectados por el número de panelistas, de tal modo que CVR es muy exigente con pocos panelistas ($CVR=1$ con 5 panelistas) y muy laxo con un gran número de panelistas ($CVR=0.29$ con 40 panelistas). Finalmente, no es aplicable a paneles con menos de cinco expertos.

Con objeto de mejorar el modelo de Lawshe se hace una aportación en este trabajo con relación a los tres problemas citados. Para ello se revisa el modelo de establecimiento de los valores de CVR mínimos, se identifica la condición que hace que la exigencia del modelo dependa del número de panelistas y se ve su ámbito de aplicación, de manera de poder utilizar el modelo con el menor número posible de panelistas, no necesariamente en consenso unánime.

Verificación de los valores mínimos para CVR con el modelo de Lawshe

Se puede establecer un análisis de consenso entre los panelistas, considerando la frecuencia de “Acuerdos” en la categoría “Esencial” y el resto de opiniones se considera como “No acuerdos”. Debe observarse que el modelo de Lawshe no establece que haya acuerdos en general, sino solamente en la categoría “esencial”; en este sentido difiere de otros modelos de consenso entre jueces, como podría ser la Kappa de Cohen. Obsérvense los valores observados y esperados (Tabla 1) con una hipótesis de no asociación.

Tabla 1.
Valores observados y esperados

	Observados		Esperados	
	<i>Acuerdos</i>	<i>No acuerdos</i>	<i>Acuerdos</i>	<i>No acuerdos</i>
N panelistas	<i>An</i>	<i>N - An</i>	<i>N/2</i>	<i>N/2</i>

La Tabla 1 se puede comparar por medio de una prueba χ^2 , con $v = 1$ grado de libertad, para una significancia α dada, con ello se tiene:

$$\chi^2 = \frac{(A n - \frac{N}{2})^2}{\frac{N}{2}} + \frac{(N - A n - \frac{N}{2})^2}{\frac{N}{2}} = \frac{4}{N} (A n - \frac{N}{2})^2 \quad [3]$$

La expresión [3] es estadísticamente significativa si $\chi^2 \geq \chi^2(\alpha, v=1)$. De acuerdo con lo indicado por Lawshe para una significancia del 5% en una sola cola $Z=1.6462$, o bien, $\chi^2(\alpha=0.10, v=1) = Z^2 = 2.71$; por lo tanto, el valor mínimo de An se obtiene para:

$$\frac{4}{N} (A n - \frac{N}{2})^2 \geq 2.71 \quad [4]$$

$$A n - \frac{N}{2} \geq \sqrt{\frac{2.71}{4}} N = 0.8231 \sqrt{N} \quad [5]$$

En consecuencia, el acuerdo mínimo ocurre cuando los panelistas están de acuerdo en esta cantidad:

$$A n = \frac{N}{2} + 0.8231 \sqrt{N} \quad [6]$$

A partir de las expresiones [4] a [6] se tienen dos observaciones: El mínimo acuerdo corresponde a la mitad de los panelistas más una fracción que es función de la raíz de N , esta fracción está acotada por el total de panelistas. Si el número de panelistas es pequeño, An puede ser igual a N , esto ocurre si:

$$\begin{aligned} \frac{4}{N} (N - \frac{N}{2})^2 &\geq 2.71 \\ N &\geq 2.71 \end{aligned} \quad [7]$$

De esta condición se concluye que N tiene que ser superior a 3 para utilizar la prueba de χ^2 convenientemente. Si se emplea la expresión [6] se puede determinar el número mínimo de panelistas que deben estar de acuerdo para cualquier valor de N que se desee. Una vez obtenido dicho valor, se calcula CVRo con la expresión 1. La Tabla 2 presenta los resultados obtenidos con este modelo y compara contra los valores propuestos por Lawshe, estos resultados se pueden comparar en la Figura 1, donde se aprecia la similitud y coincidencias entre los modelos, de lo cual se desprende que el modelo empleado por Lawshe debe ser muy similar al presentado aquí, con diferencias debidas al redondeo.

En la Tabla 2 los acuerdos se indican en número entero (no se considera que puede haber fracciones de acuerdo), por lo cual los valores de χ^2 no son constantes en 2.71, porque deben ajustarse a dicho entero; esto se aprecia en las curvas de la Figura 1 que no son suaves, sino que muestran saltos por dicho ajuste. Se aprecia que el caso de dos panelistas conduce a una situación imposible, porque requiere contar con tres acuerdos. Entre tres y siete panelistas, se necesita tener un consenso unánime entre todos los panelistas, Lawshe no reporta resultados para tres y cuatro panelistas, por lo que empieza su tabla a partir de cinco jueces. La columna de “No acuerdos” indica que con ocho y nueve panelistas solo se admite un sólo desacuerdo, en tanto que con 40 panelistas se admiten hasta 14 desacuerdos. Todos los valores de χ^2 corresponden con valores superiores a 2.71 que es el valor de referencia, pero se observa que dicho valor es fijo, en tanto que el número de panelistas se incrementa, razón por la cual se tiene el mismo problema de otras pruebas estadísticas que son sensibles al efecto de tamaño. En efecto, si se hace tender N a ∞ , en

valor de CVR tiende a 0, lo cual implica que cuando se cuenta con un gran número de panelistas, el dictamen se vuelve muy benévolo para el evaluador, inclusive en el caso en que se tiene la mitad de acuerdos entre ellos.

Tabla 2
Valores de CVR mínimos

Panelistas	Acuerdos en "esencial"	No acuerdos	χ^2	CVRo	CVR Lawshe	Diferencia CVR-CVRLawshe
2	3	-1	Caso imposible			
3	3	0	3.000	1.000	No reportados	
4	4	0	4.000	1.000		
5	5	0	5.000	1.00		
6	6	0	6.000	1.00	1	0.00
7	7	0	7.000	1.00	1	0.00
8	7	1	4.500	0.75	0.75	0.00
9	8	1	5.444	0.78	0.78	0.00
10	8	2	3.600	0.60	0.62	0.02
11	9	2	4.455	0.64	0.59	0.05
12	9	3	3.000	0.50	0.56	0.06
13	10	3	3.769	0.54	0.54	0.00
14	11	3	4.571	0.57	0.51	0.06
15	11	4	3.267	0.47	0.49	0.02
20	14	6	3.200	0.40	0.42	0.02
25	17	8	3.240	0.36	0.37	0.01
30	20	10	3.333	0.33	0.33	0.00
35	23	12	3.457	0.31	0.31	0.00
40	26	14	3.600	0.30	0.29	0.01

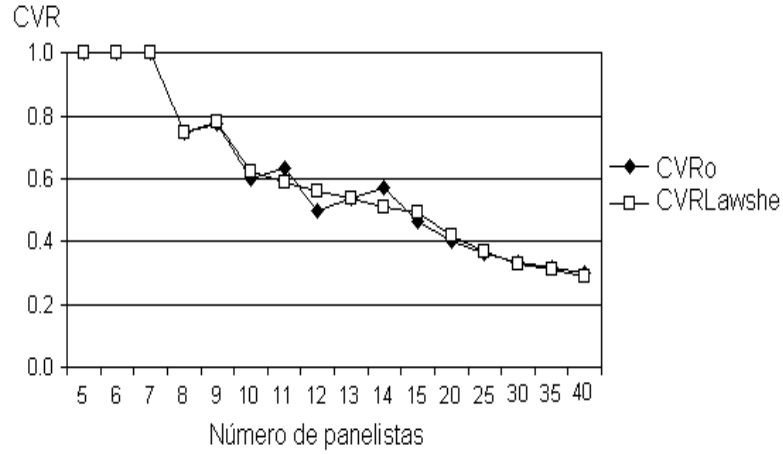


Figura 1. Valores de CVR comparados con Lawshe

Corrección al valor mínimo de CVR.

Para evitar el problema del efecto de tamaño, se puede hacer normalizar la población de panelistas, usando la correlación de atributos que hace equivalente a la χ^2 con la r de Pearson y comparado contra un valor de referencia normalizado al 100% de personas:

$$r = \sqrt{\frac{\chi^2}{N}} \quad [8]$$

que se transforma en:

$$\sqrt{\frac{4}{N^2} \left(An - \frac{N}{2} \right)^2} \geq \sqrt{\frac{\chi^2(\alpha=0.1, v=1)}{100}} = \sqrt{\frac{2.17}{100}} = 0.1646 \quad [9]$$

Esta expresión se simplifica como sigue:

$$\frac{2}{N} \left(An - \frac{N}{2} \right) \geq 0.1646 \quad [10]$$

$$An - \frac{N}{2} \geq 0.0823 N \quad [11]$$

y conduce al número de acuerdos:

$$An \geq 0.5823 N \quad [12]$$

En este caso puede observarse que si se normaliza el estadístico, el número de acuerdos mínimo es una proporción constante del número de panelistas, siempre se pedirá un consenso del 58.23% como mínimo para aceptar un ítem. Se puede observar que con esta nueva expresión no se tiene el problema del efecto de tamaño, ya que la exigencia es la misma para todos los paneles de expertos elegidos. Con la expresión [12] ya no ocurre que el mínimo de An sea igual a N , como ocurría con la expresión [3], pero atendiendo a que An debe ser entero, dicha expresión no es aplicable sólo si $N=2$. La Tabla 3 presenta los valores de acuerdo y CVR1 mínimos obtenidos para esta normalización; en la figura 2 se comparan estos valores

contra los de Lawshe, pudiendo observarse que debido a la reducción en la exigencia de valores de consenso, se tiene un modelo más apropiado para los casos en que hay pocos panelistas.

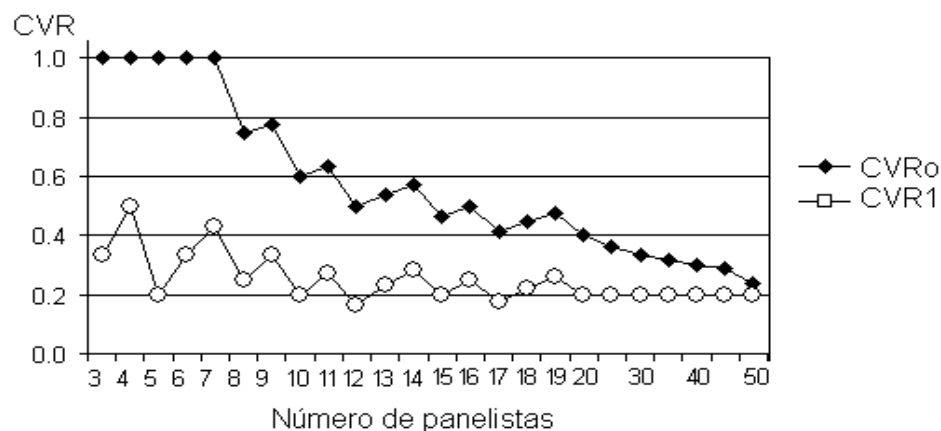


Figura 2. Valores de CVR_1 comparados con CVR_0 .

La Tabla 2 contiene los acuerdos en la categoría de “esencial” considerando un número entero de personas (no se consideran fracciones de acuerdo), por lo tanto, los valores de r no son constantes, sino que proporcionan el valor relativo a dichos acuerdos en número entero. Esto también se revela en la Figura 2, donde las curvas no son suaves sino que muestran saltos, similares a lo mostrado en la Figura 1. Es notable la diferencia entre CVR_0 y CVR_1 , en particular porque el modelo normalizado admite un desacuerdo en el caso de tres y cuatro panelistas y de dos desacuerdos cuando se tienen de cinco a siete jueces, lo cual es una condición más razonable para emplear en una situación práctica.

Tabla 3
Valores de CVR_1 mínimos

Panelistas	Acuerdos en “esencial”	No acuerdos	r	CVR_1
2	2	0	1.00	1.00
3	2	1	0.33	0.33
4	3	1	0.50	0.50
5	3	2	0.20	0.20
6	4	2	0.33	0.33
7	5	2	0.43	0.43
8	5	3	0.25	0.25
9	6	3	0.33	0.33
10	6	4	0.20	0.20
11	7	4	0.27	0.27
12	7	5	0.17	0.17
13	8	5	0.23	0.23
14	9	5	0.29	0.29
15	9	6	0.20	0.20
20	12	8	0.20	0.20

25	15	10	0.20	0.20
30	18	12	0.20	0.20
35	21	14	0.20	0.20
40	24	16	0.20	0.20

Alternativa a CVR

En la Tabla 3 se puede apreciar que se resuelve el problema del efecto de tamaño si se emplea la expresión [12] para tener el número mínimo de acuerdos, pero se tiene el problema de identificar el valor mínimo aceptable de CVR porque la expresión [1] tiene una tendencia decreciente con el número de panelistas o jueces. Esta condición obliga a tener a mano la Tabla 3, del mismo modo que se requería la Tabla 2 para el modelo de Lawshe, pero de preferencia, basta con recordar que el mínimo aceptable para

an es $0.5823N$, por lo que el valor teórico de referencia para CVR es: $CVR = \frac{0.5823N - \frac{N}{2}}{\frac{N}{2}} = 0.164$

que puede aproximarse convenientemente a $CVR=0.165$ para fines prácticos. Como regla simple puede emplearse este valor, pero dentro del modelo normalizado propuesto aquí. El problema que se tiene con este valor es que no corresponde con el modelo de Lawshe, que no estaba normalizado y que, por otro lado, se antoja demasiado bajo como valor de aceptación, a pesar de estar correcto. Por ello se plantea un modelo alternativo a CVR que brinde una mejor interpretación de los acuerdos, que consiste simplemente en la proporción de acuerdos en la categoría “esencial”, con respecto al número de jueces. Se plantea la alternativa:

$$CVR' = \frac{n_e}{N} \quad [13]$$

con la misma nomenclatura empleada para la expresión [1]. Se puede pasar de la expresión [1] a la [13] sin ningún problema, con lo cual se muestra que guardan una relación funcional:

$$CVR \frac{n_e - \frac{N}{2}}{\frac{N}{2}} = \frac{NCVR' - \frac{N}{2}}{\frac{N}{2}} \quad [14]$$

lo cual se simplifica:

$$CVR = 2CVR' - 1 \quad [15]$$

$$CVR' = \frac{CVR + 1}{2} \quad [16]$$

Si se toma el valor de referencia de aceptación normalizada para CVR, es decir: $CVR=0.164$ y se substituye en la expresión [16] se tiene que el mínimo aceptable para CVR' es:

$$CVR' = \frac{0.1646 + 1}{2} = 0.5823 \quad [17]$$

que es la misma proporción de acuerdos planteada en la expresión [12], para fines prácticos se puede decir que CVR' debe ser proporcionar por lo menos un 58% para ser aceptable, este valor es constante, independientemente del número de panelistas, con lo cual no solamente se resuelve el problema del efecto de tamaño, sino también la interpretación de los acuerdos en la relación de validez de contenido. Como CVR' no depende de N , cuando N tiende a ∞ , CVR' permanece constante en 0.5823, lo cual conduce a una condición de igual exigencia en todos los casos, independientemente del número de panelistas, lo cual parece mucho más realista que lo establecido en el modelo de CVR propuesto por Lawshe. La Tabla 4 presenta los valores de CVR' para diferentes conjuntos de panelistas, así como el número mínimo de “acuerdos” y “no acuerdos” que pueden tenerse en una revisión de contenido. El valor CVR' es siempre superior o igual a 0.5823, lo cual es muy fácil de recordar para diversas aplicaciones.

Análisis del Índice de Validez de Contenido

Una vez obtenidos los valores de CVR' , se puede determinar el CVI propuesto por Lawshe como promedio simple de los ítems aceptables. En este punto debe indicarse que el responsable del diseño del instrumento puede convocar a los miembros del panel de validez de contenido para hacer una revisión en los casos en los cuales no se alcanzó el mínimo del 58% de consenso en la categoría “esencial”; esta revisión permitiría rescatar algunos de los ítems donde no hubo acuerdo o, en su caso, sugerir modificaciones para futuras revisiones. Por tratarse el CVI del promedio de los ítems aceptables de acuerdo con el CVR' , se espera que brinde valores superiores a 0.58, de ahí surge una aplicación interesante para dictaminar la validez de un instrumento o de un banco de ítems y que se puede plantear como una extensión del modelo de Lawshe.

Supóngase que se tiene un banco de ítems formado por M reactivos, cada uno con su CVR' calculado de acuerdo con la expresión [13] y se calcula CVI con la expresión [2]. El valor mínimo de CVI que indique la validez de contenido de todo el banco de ítems o del instrumento no puede ser menor que 0.5823 que es el límite inferior de los valores de CVR' , bancos de ítems o instrumentos con valores inferiores deben ser depurados de inmediato para incrementar su calidad.

Lo interesante es que el mismo modelo de prueba de hipótesis empleada con CVR se puede utilizar con CVI, para obtener el índice de validez de todo el banco o de un instrumento. En este caso se clasifican los ítems en forma categórica, de tal modo que todos los que tienen un CVR' superior a 0.5823 se clasifican como “Aceptables” y en caso contrario se denominan “No aceptables”. Lo anterior se puede verificar en la Tabla 5 (Valores observados y valores esperados). El modelo es similar al desarrollado anteriormente, por lo que se llega a la misma condición normalizada, el caso de una cola al 5% de significancia, que conduce al mínimo de ítems aceptables dado por: $Bm \geq 0.5823 M$

Tabla 4
Valores de CVR' mínimos

Panelistas	Acuerdos en “esencial”	No acuerdos	r	CVR'
2	2	0	1.00	1.00
3	2	1	0.33	0.67
4	3	1	0.50	0.75
5	3	2	0.20	0.60
6	4	2	0.33	0.67
7	5	2	0.43	0.71
8	5	3	0.25	0.63

9	6	3	0.33	0.67
10	6	4	0.20	0.60
11	7	4	0.27	0.64
12	7	5	0.17	0.58
13	8	5	0.23	0.62
14	9	5	0.29	0.64
15	9	6	0.20	0.60
16	10	6	0.25	0.63
17	10	7	0.18	0.59
18	11	7	0.22	0.61
19	12	7	0.26	0.63
20	12	8	0.20	0.60
25	15	10	0.20	0.60
30	18	12	0.20	0.60
35	21	14	0.20	0.60
40	24	16	0.20	0.60
45	27	18	0.20	0.60
50	30	20	0.20	0.60

Tabla 5.
Valores observados y esperados

	Observados		Esperados	
	<i>Acuerdos</i>	<i>No acuerdos</i>	<i>Acuerdos</i>	<i>No acuerdos</i>
M ítems	Bm	M - Bm	M/2	M/2

En resumen, para que un instrumento o un banco de ítems pueda considerarse como aceptable, requiere contar por lo menos con un 58% de los ítems en condición satisfactoria del CVR'. Si se continúa empleando el modelo de Lawshe, la tabla de valores de CVR propuesta por dicho autor sería aplicable, sólo que se debe cambiar el nombre de panelistas por ítems, de tal modo que si la prueba cuenta con 40 ítems, se debe tener un valor de CVI superior a 0.29, que sería el mínimo posible para este número de ítems.

Ejemplo de aplicación.

El modelo propuesto fue empleado en un proyecto de desarrollo de pruebas para carrera docente, en ocho especialidades: Inglés, Matemática, Lenguaje y Comunicación, Ciencias Sociales, Ciencias Naturales, Educación Básica, Educación Preescolar y Formación General. Se constituyó un panel de expertos de cada especialidad, formado por tres a ocho especialistas, quienes revisaron los ítems del banco y dictaminaron su pertinencia (o categoría "esencial") al dominio a evaluar, posteriormente se determinaron los valores de CVR' de cada ítem y CVI de todo el banco. La Tabla 6 muestra un ejemplo tomado de uno de los temas dentro de una de las especialidades, con la clasificación de los ítems hecha por cuatro revisores, junto con el cálculo de la CVR' de cada ítem y del CVI del tema.

Tabla 6.
Cálculo de CVR' y CVI

Ítem	Esencial	Útil	No necesario	CVR'	Aceptable	CVI
1	3		1	0.75	X	4.5/5=0.90
2	4			1	X	
3	4			1	X	
4	2	1	1	.5<0.58		
5	3	1		.75	X	
6	4			1	X	
Suma de los ítems aceptables				4.5	5	

Obsérvese que con la CVR' solo se rechaza el ítem 4, por tener un valor inferior a 0.5823, en cambio con el modelo original de Lawshe solo se aceptarían los ítems 2, 3 y 6 por ser los únicos que muestran un consenso unánime. CVI se obtiene al hacer la suma de los CVR' de los ítems aceptables y dividiendo entre el número de ítems aceptables, en este caso son 5 ítems. El ítem 4 deberá ser modificado o desechado del banco de ítems, pero no debe considerarse como parte del banco por no haber tenido el consenso mínimo necesario. La Figura 3 muestra los valores promedio de CVR' de los bancos de ítems.

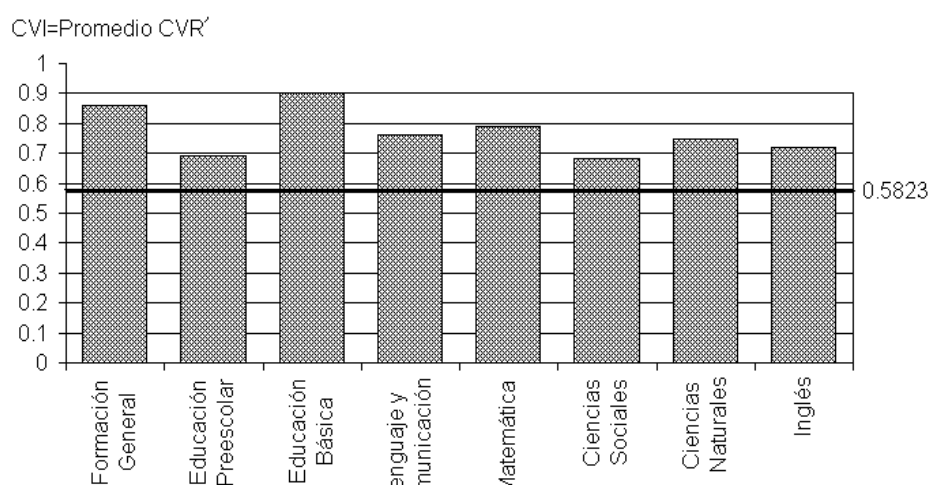


Figura 3. Índice de Validez de Contenido de pruebas de carrera docente

Se pudo constatar que todas las pruebas tuvieron valores superiores a 0.58 en el valor promedio de CVR', por lo cual se puede afirmar que tienen una validez de contenido aceptable, de acuerdo con los especialistas que participaron en el panel de expertos. Todos los bancos contaban con más de 50 reactivos, por lo que si se hubiera empleado el CVR de Lawshe, bastaba con tener un CVI superior a 0.29 para indicar que los bancos eran satisfactorios en su validez de contenido; sin embargo es mucho más razonable emplear los valores de CVR' porque se interpreta mucho más fácilmente en función de los porcentajes de acuerdo entre jueces.

Actualmente se ha empleado el CVR' como indicador de la validez de contenido de los ítems de un cuestionario de Atención de Enfermería, donde se ha trabajado con paneles de tres especialistas, mismos que no pueden colaborar juntos ni a una misma hora, por tratarse de profesionales que prestan sus servicios en diferentes centros hospitalarios, sin compatibilidad de horario. El modelo de Lawshe obliga a que los tres especialistas estén de acuerdo en cada ítem para poder aprobarlo dentro del cuestionario, por lo que resulta prácticamente anulado todo el instrumento por discrepancias de uno de los panelistas (no siempre el mismo) en los ítems. Con el modelo alternativo propuesto, todos los ítems son aceptables, ya que se permite un máximo de un "no acuerdo" como esencial en un ítem y, a su vez, un mínimo de 0.58 en el CVI de todo el instrumento.

Resumen de aplicación

Para realizar un análisis cualitativo de validez de contenido con el modelo alternativo al propuesto por Lawshe, se sugieren estos pasos:

Convocar a un panel de validez de contenido integrado por especialistas o expertos que conozcan del dominio que se va a evaluar, desde el punto de vista profesional, académico, de competencias, etc.

Someter al escrutinio de los especialistas los ítems de un cuestionario o de un banco de ítems. Los especialistas deberán clasificar a los ítems en tres categorías como "esencial", "importante aunque no indispensable", "no necesario", o nomenclaturas equivalentes.

Para cada ítem contabilizar el total de casos clasificados por los panelistas como "esencial".

Determinar CVR' (expresión [13]) como proporción de acuerdos en la categoría esencial, respecto del número de panelistas involucrados en el escrutinio.

Considerar como aceptables los ítems cuyo CVR' sea superior o igual a 0.58. Estos ítems son los que pueden integrar el instrumento o el banco de ítems. Los ítems no aceptables deberán ser revisados, corregidos y sometidos a un nuevo dictamen de parte de los panelistas.

Determinar el IVC (expresión [2]) como promedio de los CVR' de todos los ítems de cuestionario o del banco de ítems (incluyendo aceptables y no aceptables).

Se considera aceptable el conjunto de ítems si su IVC es superior a 0.58. Cuando el IVC se calcule solamente con los ítems aceptables, de entrada se sabe que su valor será superior a 0.58 y sólo se tiene que reportar el valor promedio obtenido.

Conclusiones

El modelo de Lawshe ha demostrado ser de utilidad para dictaminar la calidad de un instrumento en función de la validez de contenido, pero tiene el problema de no poder aplicarse cuando se tiene un reducido panel de expertos, además de tener influencia del efecto de tamaño. Se propuso un modelo alternativo que corrige los problemas del modelo de Lawshe y de fácil interpretación, con base en una normalización y la definición de un nuevo indicador del acuerdo entre los jueces. El modelo alternativo, de uso muy simple, ha sido empleado desde el año 2003 para dictaminar la validez de contenido de pruebas objetivas; recientemente se ha utilizado para dictaminar la validez de contenido de pruebas para pacientes de enfermería.

Referencias

Lawshe C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*. 28(4), 563-575.

Manuscrito recibido en Abril de 2007
Aceptado para publicación en Noviembre de 2007