**RESUMEN**

La finalidad de este trabajo ha sido realizar una revisión sobre la evolución histórica de la validez de contenido, así como presentar algunos de los métodos de estudio más utilizados para su estimación. El concepto de validez de contenido ha sido objeto de un largo proceso de modificaciones desde su origen. Sin embargo, estos cambios han estado focalizados en qué el tipo de evidencias se deben presentar para su estudio y los métodos más adecuados para encontrar dichas evidencias. Sin embargo, su definición se ha mantenido estable a lo largo del tiempo. En la actualidad, la validez de contenido se considera condición necesaria (aunque no suficiente) para realizar interpretaciones de las puntuaciones en los tests. Finalmente, la combinación de métodos tanto cualitativos como cuantitativos se entiende como el procedimiento más completo a la hora de realizar un estudio de validez de contenido en profundidad. Dentro de los primeros, destaca el índice IVC como el que, además de ser el más empleado actualmente, presenta los mayores beneficios respecto a las diferentes alternativas propuestas a lo largo de los años. Respecto a los segundos, la Teoría de la Generalizabilidad se entiende como el procedimiento más exhaustivo y cercano a la estimación de la validez de contenido en sí.

**Palabras clave:** Validez, Validez de contenido, Estimación, Constructo.

**ABSTRACT**

The purpose of this paper has been to carry out a review of the historical evolution of one aspect of test validity - content validity - as well as to expose some of the most popular methods used for its evaluation. The concept of content validity has undergone many modifications from its inception to the present time. However, in the past these changes have focused on which pieces of evidence must be presented and the different statistical methods used to study it, while keeping its meaning essentially stable over time. Nowadays, evidence for content validity is considered necessary (though not the sole factor) for interpreting test results. Finally, the use of both qualitative and quantitative methods is recognized as the best procedure for performing an in-depth study of content validity. Regarding the former, this paper recommends the IVC index because, besides being currently the most used, it shows the most benefits when compared to the alternatives proposed over the years. Regarding the latter, the Generalizability Theory is understood as the most comprehensive and accurate procedure for measuring this aspect of test validity.

**Key words:** Validity, Content validity, Estimation, Construct.

**Introducción**

Un test es un instrumento de medida a partir del cual es posible obtener una muestra de conducta sobre la que se pretenden hacer ciertas inferencias, mientras que el concepto de validez se refiere al conjunto de pruebas y datos que han de recogerse para garantizar la pertinencia de tales inferencias (Muñiz, 2000). Según la edición más reciente de los "Estándares para el uso de tests psicológicos y educacionales" (American Educational Research Association [AERA], American Psychological Association [APA], y National Council on Measurement in Education [NCME], 1999), "validez se refiere al grado en que la evidencia y la teoría apoyan las interpretaciones de las puntuaciones en los tests" (1999, p. 9). Más concretamente, "se validan las inferencias relativas al uso específico de un test, y no el propio test" (AERA, APA y NCME, 1985, p. 9). Es decir, no existen tests válidos sino que los tests son válidos para algo, siendo indispensable indicar a los usuarios potenciales del test sus limitaciones así como concretar para qué es válido exactamente.

Como sintetiza Sireci (2009), las fuentes de evidencia de validez han sufrido un proceso de "embalaje" y "desembalaje". En este sentido, parece ser que la tendencia a lo largo de la historia de la validez hasta la actualidad ha sido un "desembalaje" (hacia varios tipos de validez) y un "embalaje" (hacia una conceptualización unitaria), siendo además previsible que esta metamorfosis continúe en el futuro. En cualquier caso, no hay duda que la concepción actual es tomar la validez como única, existiendo diferentes fuentes para probar dicha validez. En este sentido, las recomendaciones técnicas de las comisiones internacionales sugieren cinco fuentes de evidencia de validez: contenido, procesos de respuesta, estructura interna, relaciones con otras variables y consecuencias de la evaluación (AERA, APA y NCME, 1999).

**Aproximación al concepto de validez de contenido**

Dentro de la validez de contenido en sí, los trabajos de Rulon (1946), Mosier (1947) y Gulliksen (1950a, 1950b) podrían considerarse los prolegómenos sobre los que surge el concepto acerca de este tipo de validez (Sireci, 1998a). Sin embargo, la primera aproximación a una definición operativa podría tener su origen en Cureton (1951).

Cureton presentó una novedosa definición de validez de contenido que supuso la introducción del término en la literatura sobre pruebas educativas y psicológicas (Sireci, 1998a). Su principal aportación es el reconocimiento de la existencia de una relevancia curricular o validez de contenido. En este sentido, afirma que si se pretenden validar ítems estadísticamente, se tendría que poder aceptar que el criterio de trabajo es adecuado. Para ello, los ítems "tendrían que evocar aquello que dicen estar midiendo y constituir una muestra representativa del universo de medida" (Cureton, 1951, p. 664). Una vez establecido este sustento teórico, es cuando surgen los dos criterios fundamentales para estudiar la validez de contenido: relevancia y r