

## Elementos de validación y fiabilidad del test Hctaes para estudiar el pensamiento crítico en estudiantes de Psicología

Nubia Yaneth Gómez Velasco, Universidad Pedagógica y Tecnológica de Colombia, Docente, Tunja, Colombia, [nubia.gomez@uptc.edu.co](mailto:nubia.gomez@uptc.edu.co)

Ana Emilce Jimenez, Universidad Pedagógica y Tecnológica de Colombia, Docente, Tunja, Colombia, [ana.jimenez@uptc.edu.co](mailto:ana.jimenez@uptc.edu.co)

Myriam Ortiz Padilla, Universidad Simón Bolívar, Docente, Barranquilla, Colombia  
[mortiz@unisimonbolivar.edu.co](mailto:mortiz@unisimonbolivar.edu.co)

### Resumen

La validación y fiabilidad de un instrumento permite garantizar la calidad de la medición y de los resultados obtenidos de la aplicación del mismo. La validación involucra una serie de procesos, en donde la estadística cobra relevancia para garantizar con niveles de confiabilidad los resultados obtenidos. A través de esta ponencia se esquematiza elementos de validación y fiabilidad de un instrumento propuesto para medir cinco habilidades del pensamiento crítico denominado Halpern Hctaes-HH. La investigación tiene un enfoque cuantitativo de tipo exploratorio descriptivo. Se analizó el instrumento de medición Halpern Hctaes compuesto por una serie de situaciones con doble formato de preguntas de tipo abierto y cerrado. Se identifica que el instrumento evalúa el pensamiento crítico, desde un modelo teórico que indaga cinco habilidades: razonamiento verbal, análisis argumentativo, comprobación de hipótesis, uso de la probabilidad y la incertidumbre, resolución de problemas y toma de decisiones. La fiabilidad y validez del instrumento, implica diferentes categorías e indicadores de análisis. Se aporta una esquematización de procesos inherentes a la validación y fiabilidad aplicados al test HH, asociando coherentemente las respectivas técnicas estadísticas, lo que constituye un documento práctico y útil para la comunidad investigadora que requiere de estudios similares y marcos de referencia.

### Introducción

El pensamiento crítico ha sido considerado como una de las metas de enseñanza más importantes para las instituciones de educación superior. Para la mayoría de los investigadores y expertos en el área, el pensamiento crítico implica niveles de pensamiento superior, donde el rol de los profesores ha sido considerado central en este proceso (Guzmán, 2011). Según éste autor el pensamiento crítico se ha relacionado con el juicio reflexivo, la resolución de problemas, el pensamiento lógico, la toma de decisiones y el método científico, habilidades requeridas en el proceso de formación y desarrollo de un profesional en cualquier campo del saber. Autores como Andreu-Andrés y García-Casas (2014), sostienen que la disposición hacia el pensamiento crítico y de las habilidades cognitivas, debe considerarse en el diseño de objetivos curriculares y procesos formativos que permitan ayudar a los estudiantes en el desarrollo y mejora de sus procesos de pensamiento de orden superior.

En el ámbito académico, se encuentra una prueba que permite evaluar el pensamiento crítico de los estudiantes, denominada HALPERN HCTAES (Halpern, 2006). La prueba HALPERN HCTAES, consta de un total de 25 situaciones cotidianas que contienen un doble formato de preguntas, donde para cada situación el individuo debe responder de dos formas, la primera a través de la construcción de una respuesta abierta y luego se presentan preguntas cerradas, para que seleccione la respuesta que considere más conveniente a la situación planteada.

La prueba HALPERN HCTAES, ha sido validada con éxito en países de habla hispana como España y Chile. La versión en español recibe el nombre de PENCRISAL (Rivas y Saiz, 2012) y está dirigido a evaluar el pensamiento crítico de los estudiantes de la población española, sin embargo, no se encuentra validada en varios países, entre ellos, Colombia. La no validación del instrumento, genera serios cuestionamientos sobre la calidad de la medición y los resultados obtenidos de la aplicación del mismo. Diversos autores, sostienen que para instrumentos que sean aplicados a otros países, sin la debida verificación, no se puede garantizar la conservación de sus propiedades psicométricas, por lo que es importante hacer la adaptación sociocultural en sus contextos (Ramada-Rodilla, et al, 2013 y Muñiz, et al., 2013). Así entonces, la validación de un test, que en términos prácticos corresponde al grado de exactitud de la prueba con el que mide el constructo teórico, requiere de su verificación en validez y fiabilidad, considerando los diferentes contextos en donde se aplique, en razón a que estos componentes no se realizan de forma general y universal.

Este trabajo presenta avances del proyecto de investigación institucional titulado Metodología estadística para la validación y análisis del instrumento HALPERN HCTAES, utilizado en el estudio del pensamiento crítico; proyecto desarrollado por el grupo de estadística Gamma de la Universidad Pedagógica y Tecnológica de Colombia, en colaboración con el grupo el Grupo Sinapsis Educativa y Social de la Universidad Simón Bolívar.

Los avances de éste proyecto, se concentran en la revisión y organización de elementos correspondientes a la validación y fiabilidad de la prueba y la generación de una propuesta de esquematización de aspectos relevantes, asociando a su vez algunas técnicas estadísticas pertinentes. Propuesta que no está exenta de generar otros puntos de vista y dar apertura a argumentaciones diferentes, al considerar que la esquematización implica articular conceptos que sustentan metodologías de validez y fiabilidad del test e identificar fases y técnicas estadísticas para realizar los diferentes procesos que se exige.

Para una mayor contextualización se presenta el alcance y finalidad del proyecto de investigación, que puede contribuir como referencia a otros estudios o proyectos similares, para posteriormente identificar procesos inherentes asociados con los elementos de validación y fiabilidad del test HCTAES, útil para estudiar el pensamiento crítico, y que metodológicamente incluye una diferencia con la mayoría de los test que se aplican para este efecto, correspondiente a la inclusión de respuestas de tipo abierto.

## **Pertinencia del proyecto y Problematicación**

La validación de un test para evaluar cualquier habilidad o competencia, en éste caso correspondiente al pensamiento crítico, no tiene un alcance universal y requiere ser verificado en los distintos contextos culturales en donde se aplique la prueba. La prueba HALPERN HCTAES ha sido validada y empleada en otros países para evaluar el pensamiento crítico de los estudiantes, no así de forma correspondiente para Colombia. Contar con un test validado en Colombia, como el HALPERN HCTAES, o también conocido como PENCRISAL, permite estudiar diferentes aspectos sobre el pensamiento crítico y dificultades que presentan en torno a ello, estudiantes de distintos niveles de formación.

El pensamiento crítico se ha relacionado con diferentes habilidades de gran importancia e incidencia en el campo profesional y en la formación de un estudiante, en particular si está conectado en el campo investigativo e innovador (Guzman, 2011). Se reconoce además en su multidimensionalidad, como un proceso cognitivo que implica la

intencionalidad y la autorregulación. Evaluar el pensamiento crítico de los estudiantes es un punto de partida de gran valor que apoya el curso de acciones que favorecen la retención de estudiantes con calidad en los programas de pregrado en Colombia.

En este contexto, los intereses, las actitudes y los valores de un estudiante son variables que pueden tener consecuencias en las formas como ejerce su pensamiento crítico. Por lo cual, tanto las habilidades cognitivas como la disposición hacia el pensamiento crítico son una fuente importante de influencia en los estudiantes, que debería ser considerada, para el diseño de objetivos curriculares y procesos formativos que permitan ayudar a los estudiantes en el desarrollo y mejora de sus procesos de pensamiento de orden superior (Andreu-Andrés, García-Casas, 2014).

Bajo lo anterior, no contar con el instrumento HALPERN HCTAES, validado en un país, para evaluar el pensamiento crítico, limita y dificulta su uso, cuestiona su aplicabilidad y generaría resultados no confiables. Desde la mirada de autores como Ramada-Rodilla, et al. (2013) y Muñiz, et al. (2013), la validación garantiza la calidad de la medición y los resultados obtenidos de la aplicación del mismo e implica una serie de procedimientos, en donde la estadística permite garantizar los resultados obtenidos; en particular si son instrumentos que se encuentran disponibles en otros idiomas u otros países.

Es de precisar que un elemento diferenciador en el HALPERN HCTAES y por tanto que implica algunas particularidades en sus procesos de validación, lo constituye el hecho que éste instrumento además de incluir preguntas de tipo cerrado para evaluar cada una de las 25 situaciones problémicas, se complementa con respuestas de tipo abierto, lo que representaría la inclusión de nuevas metodologías que permitan determinar en su conjunto la calidad de su medición.

### **Elementos Metodológicos**

Dentro del proceso de validación, se requiere contar con muestras representativas, con un tamaño mínimo de unidades de estudio que deben ser seleccionadas a través de la aplicación de un muestreo. La no selección de unidades a través de un diseño muestral pertinente y su aplicación del test, generarían no sólo cuestionamientos en su validación, sino problemas de confiabilidad y sesgos (Paniagua, 2015).

Metodológicamente la investigación tiene un enfoque cuantitativo de tipo exploratorio descriptivo. Para éste proceso, se prevé seleccionar una muestra no probabilística por cuotas de 350 estudiantes de una universidad privada de Barranquilla (Colombia), mayores de 18 años y que se encuentren cursando los Semestres, 2, 5 y 8 de psicología en la Universidad Simón Bolívar de Barranquilla -USB.

La aplicación del instrumento estará a cargo del Grupo Sinapsis Educativa y Social de la USB, previa revisión y adaptación de terminologías del instrumento. Los datos serán recolectados entre los meses de julio a septiembre de 2019. Los participantes en el estudio lo realizarán de manera voluntaria. Los simulacros de prueba piloto realizada, arrojan un tiempo de aplicación y desarrollo de 4 horas máximo. Este tiempo de aplicación será consensuado con anterioridad con las directivas de programa y contará con la colaboración de los docentes a cargo de los estudiantes.

Respecto a los materiales utilizados, se hará uso del instrumento de prueba HALPERN HCTAES, el cual, bajo una serie de situaciones cotidianas, se propone un doble formato de preguntas de tipo abierto y cerrado. Para las preguntas de tipo abierto, el estudiante debe ofrecer un argumento, una aplicación, o generar las soluciones a un problema. Para la pregunta de tipo cerrado, el estudiante debe elegir una serie de alternativas, la que mejor responda al problema. La recolección de la información y aplicación del formato cumplirá con la normativa institucional vigente de consentimiento informado y la

reserva de confidencialidad.

El proceso de validación y fiabilidad estadística será desarrollado por el grupo de Estadística Gamma de la UPTC- Colombia. Quien a su vez estará a cargo del diseño muestral.

El desarrollo de ésta investigación se apoyó para la fase inicial en una metodología de tipo documental que de acuerdo a Baena (1985), es una técnica que consiste en la selección y compilación de información a través de artículos de los cuales se analizan los aspectos conceptuales y metodológicos que se deben tener en cuenta en la validación de un test.

Para la búsqueda bibliográfica y documental, se utilizaron criterios de inclusión de artículos que se encuentren en bases especializadas bajo filtros como: validación, escala HALPERN HCTAES, confiabilidad y validez, entre otros.

Metodológicamente el proyecto contempla las siguientes fases:

1. Revisión y análisis de documentos sobre propiedades psicométricas
2. Planteamiento de la propuesta metodológica para la validación de la prueba.
  - a. Estrategia de muestreo
  - b. Planteamiento de técnicas para verificar la Fiabilidad del test
  - c. Planteamiento de técnicas para verificar la Validez del test
3. Evaluación de las propiedades psicométricas del test HALPERN HCTAES
  - a. Selección del software estadístico a utilizar y organización de la base de datos
  - b. Análisis exploratorio de datos
  - c. Pruebas de Fiabilidad.
  - d. Pruebas de Validez
4. Análisis de los datos a partir del constructo medido, pensamiento Crítico
5. Elaboración de informes de resultados

### **Avances de Resultados**

En esta sección se presenta resultados de avances del proyecto de investigación, los cuales se encuentran traducidos en la revisión y ampliación de literatura del test HCTAES y una esquematización de metodologías estadísticas utilizadas en validación de test, con especial revisión los que se encuentren asociados al estudio del pensamiento crítico, el HCTAES o cuenten con una estructura similar.

- Elementos relevantes del test HCTAES.

El test HALPERN HCTAES, se concibe como una teoría de acción sobre la cual se puede razonar y tomar decisiones para solucionar problemas y alcanzar metas, donde la argumentación toma parte fundamental como un medio, para tomar decisiones solidas con eficacia.

Dentro de las pruebas que permiten evaluar el pensamiento crítico, se encuentra la HALPERN HCTAES, la cual, solventa una dificultad frente a otras pruebas, en las que sólo utilizan instrumentos de formato de respuesta cerrada y que impiden otro tipo de exploración del pensamiento implicado en la tarea de responder un test.

El instrumento *Halpern Critical Thinking Assessment using Everyday Situations (HCTAES)*, construido por Halpern en el año 2006, permite la evaluación del pensamiento crítico, desde un modelo teórico que indaga cinco habilidades en las que se manifiesta este tipo de pensamiento (Figura 1).

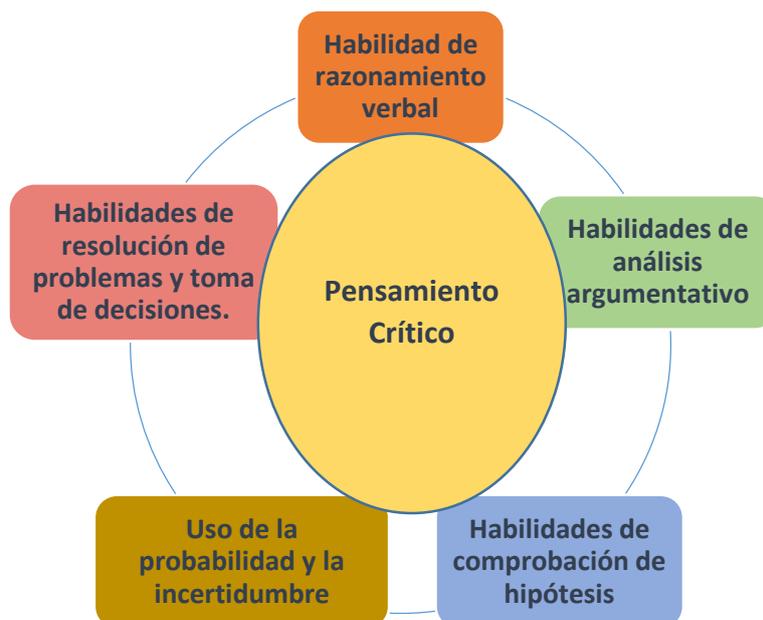


Figura. 1. Habilidades medidas con el test HCTAES - pensamiento crítico.

El test HCTAES, utiliza instrumentos centrados en procesos de pensamiento y los ítems que se proponen, corresponden a situaciones que describen problemas cotidianos que se deben resolver mediante respuestas abiertas y cerradas. Los principios que fundamentan el HCTAES, están basados en cuatro aspectos característicos propios (Halpern, 2006 y Saiz y Rivas, 2008):

- a) Utilización de ítems que correspondan a situaciones cotidianas
- b) Uso de diferentes dominios, para valorar el grado de generalidad de las habilidades
- c) Aplicación de un formato de respuesta abierta, que posibilita la exploración de procesos de pensamiento
- d) Empleo de situaciones-problema de respuesta única (respuesta cerrada) que permite evaluar el mecanismo de pensamiento correspondiente y facilita la cuantificación de los ítems.

La versión original del test se encuentra en inglés, con traducción y adaptación a la población española por Nieto, Saiz y Orgaz, (2009), y se encuentra validado en España y a nivel de Sur América en Chile.

- Claves de procedimiento en la aplicación del instrumento HCTAES.

Como se ha indicado de manera previa, el cuestionario consta de una serie de situaciones planteadas en un total de 25 preguntas, que consta de dos partes, la primera, donde se le pide a cada encuestado que escriba una respuesta breve para cada pregunta; la segunda contiene exactamente la misma información que la primera, pero en este caso se le pide que emita un juicio sobre las posibles respuestas.

La aplicación del instrumento tiene otras particularidades a tener en cuenta al momento de su aplicación, bajo una nota metodológica expresa de solicitar al estudiante que estime en qué medida cada alternativa responderá bien a una determinada cuestión; en ese caso deberá responder con una o varias frases. En ocasiones también se le puede

pedir que seleccione la mejor alternativa de un conjunto de respuestas posibles, para lo cual, deberá marcar la opción elegida.

Debido a que este proceso puede generar alguna confusión en el estudiante y que por tanto las primeras preguntas se vean afectadas por la no comprensión del ejercicio metodológico, la prueba HCTAES incluye un proceso previo de familiarización con el instrumento, para lo anterior, se realiza un ejemplo para la parte 1 y para la parte 2, antes de comenzar la prueba, en el cual el estudiante podrá realizar preguntas si tiene alguna duda.

En el anexo 1, se ilustra un ejemplo de los desarrollados en la prueba.

- Técnicas estadísticas en validación de test, según estudios previos

Una revisión documental centrada en los últimos años sobre temáticas de validación de test permite tener un contexto nacional e internacional de procesos y tendencias sobre el tema.

A nivel de metodologías estadísticas para la validación de test, se encuentran varias pruebas que permiten identificar, con determinados niveles de confianza si el test mide lo que quiere medir. Se relaciona entre ellos el estudio de Urzúa et al.(2009), quienes validan y analizan las propiedades psicométricas del test EDI-2, útil para evaluar los trastornos alimenticios de adolescentes escolarizados, a través del análisis factorial, y la fiabilidad de la escala con el Alfa de Cronbach.

Tapia y Luna ( 2010), validaron una prueba de habilidades de pensamiento para estudiantes de secundaria, a través de un análisis de contenido (evaluación por jueces) y los datos obtenidos de la evaluación psicométrica, haciendo uso del coeficiente V de Aiken. Este coeficiente fue utilizado por los autores para cuantificar la validez del contenido y se comparó con otras metodologías como el índice de Acuerdo (IA) y la prueba Binomial, concluyendo que el coeficiente de Aiken es el más adecuado para determinar este tipo de validez, ya que permite obtener valores factibles de ser contrastados estadísticamente según el tamaño de muestras de jueces seleccionada.

Autores como Ecurra y Salas (2014) mediante su trabajo, construcción y validación del cuestionario de adición a redes sociales, aplicado a estudiantes de una universidad de Lima Perú, utilizan un análisis factorial exploratorio para la validación.

En Colombia se encuentran, estudios como los de Martínez, Trujillo y Trujillo (2012), quienes realizan la validación del *test propósito vital* para Colombia, a través de un análisis factorial confirmatorio comparado, verificando con el test de Shapiro-Wilkinson.

Estos entre otros artículos, evidencian el uso de técnicas estadísticas en la validación de Test. Técnicas que requiere revisión, para determinar cuál o cuáles son la más pertinente para validar el HCTAES en Colombia, diferenciándose de varios test clásicos, dado que en éste se involucra, además de preguntas cerradas, el uso de preguntas abiertas que no son frecuentes en los test.

En relación al test HCTAES, estudios más próximos sobre el tema, vinculan los resultados de los investigadores, Nieto, Saiz y Orgaz (2009), que en su artículo sobre análisis de las propiedades psicométricas de la versión española HCTAES-TEST DE HALPERN, evalúan el pensamiento crítico mediante situaciones cotidianas, a estudiantes de dos universidades españolas. Estos autores realizaron su traducción y adaptación a la población española, encontrando que el instrumento es fiable para evaluar el pensamiento crítico y que se confirma la interrelación entre sus habilidades. De forma análoga, para el contexto suramericano se cuenta también con una validación

y adaptación del test HCTAES en Chile realizada por Fuentes (2010), con resultados similares.

A nivel de Colombia se encuentra el artículo “Caracterización de habilidades de pensamiento crítico en estudiantes de educación media a través del test HCTAES” realizado en el 2009, por Beltrán y Torres, donde describen un estado inicial de las habilidades que permiten el desarrollo del pensamiento crítico en estudiantes de grado undécimo de una institución en Bogotá, a través de la aplicación del test HCTAES con el fin de mejorar los procesos de enseñanza y aprendizaje desde el área de la Química y para potenciar las habilidades que poseen los estudiantes, a través del diseño y aplicación de un programa de intervención cognitiva.

Es de precisar que si bien el estudio abordado por Beltrán y Torres (2009) evalúa cinco habilidades del pensamiento crítico (comprobación de hipótesis, razonamiento verbal, análisis de argumentos, probabilidad e incertidumbre y toma de decisiones y resolución de problemas) no se realiza en éste, una validación previa del instrumento, ni hacen mención en el mismo sobre la fiabilidad del test.

En Colombia se encuentran otros artículos, que buscan analizar el pensamiento crítico en diferentes campos, como el realizado por Calle (2013), quien evalúa las habilidades del pensamiento crítico asociadas a la escritura digital y sobre el cual menciona como referente la prueba HCTAES, tal prueba no hace parte de su metodología de análisis ni tampoco menciona procesos de validación en la prueba que aplica.

- Fundamentos y definición del constructo del instrumento

Los fundamentos de un instrumento de medida, están vinculados a procesos de la sicometría, la cual permite determinar para el objeto de estudio, tanto la adecuación de la escala como la calidad de la medida; aspectos que presupone un elemento transversal, determinado por el conocimiento teórico sobre lo que se busca medir.

Los instrumentos de validación de un test deben cumplir ciertas características, entre las que se destacan (Arribas, 2004):

1. Ser adecuado para el evento o situación que se busca medir o lo que se indica también como teóricamente justificable.
2. Validez de contenido, es decir, ser capaz de medir aquellas características que pretenden medir y no otras.
3. Fiabilidad, precisión, es decir, con un mínimo de error en la medida.
4. Sensibilidad, que sea capaz de medir cambios tanto en los diferentes individuos como en la respuesta de un mismo individuo a través del tiempo.
5. Delimitar de forma clara sus componentes (dimensiones), de manera que cada uno contribuya al total de la escala de forma independiente (validez de constructo).
6. Estar basado en datos generados por los propios participantes
7. Ser aceptado por docentes investigadores o comunidad académica.

Definir el constructo hace referencia a precisar que es lo que se quiere medir. La definición de un constructo, parte de una revisión de literatura, de forma que se conozca las teorías y fundamentos que sustentan el problema o situación a abordar. En la elaboración del constructo se recomienda realizar consultas con expertos sobre el tema.

Otros aspectos a considerar en la elaboración del instrumento, además del contenido del cuestionario y que incide en la validación del constructo corresponde a:

- Determinar la población objetivo
- Tipología y niveles de medición de las variables a medir
- Establecer la forma de administrar el instrumento

Precisar la población a la que va dirigido, es un elemento fundamental. Ya sea porque aún no se ha validado, o en caso contrario, determinar las características (edad, género, nivel académico o cultural) donde se validó

En relación a la administración del instrumento, éste debe estar acorde al instrumento de medición e implica ajustar los elementos correspondientes en redacción o formato, ajustado a si se va a realizar por medio de entrevista, vía telefónica, autocompletado, u otro.

Para efectos del test HCTAES, el contenido del cuestionario gira en torno a evaluar el pensamiento crítico, desde la indagación hacia cinco habilidades, mencionadas anteriormente. Cada una de esas habilidades describe características que pueden ser agrupadas en categorías o factores, y sólo bajo su claridad en su conceptualización se pueden construir las preguntas y explorar diferentes aspectos a evaluar.

Así todo proceso de validación de un test, debe previamente contar con un test o cuestionario elaborado. Es de precisar que, para éste proyecto, dado que se cuenta con el instrumento HCTAES, no se hace necesario la construcción del test o cuestionario. En la figura 2 se sintetizan algunos elementos centrales para definir el aspecto que se va a medir, o en otras palabras el constructo.



Figura 2. Esquematización de elementos para definir el constructo

Una recomendación final, antes del proceso de validación, consiste en realizar la aplicación de una prueba piloto. La prueba piloto ayuda a precisar elementos de forma y fondo del cuestionario. A través de la revisión de la prueba piloto se puede identificar si las preguntas son comprensibles y adecuadas, si hay dificultades en el ordenamiento interno lógico; estimar la duración de aplicación del cuestionario. Así mismo, si se considera seleccionar la muestra a través de un muestreo de tipo probabilístico, la información de la muestra piloto, permite contar con información requerida para ello.

- Elementos y procesos de validación y fiabilidad del test

La revisión conceptual y teórica que contextualizan el problema de investigación en torno a la metodología estadística pertinente para validar el test HCTAES, permite sintetizar y precisar varios elementos a tener presente en la validación y fiabilidad de un test.

La calidad de las mediciones de un test está determinada por la evaluación de las propiedades psicométricas del mismo, siendo las dos características esenciales para valorar la precisión de un instrumento, su fiabilidad y validez (Gómez e Hidalgo, 2005).

Mientras que la fiabilidad hace referencia al grado en que el instrumento es capaz de medir con precisión y sin errores y determina la proporción de la varianza total atribuible a diferencias verdaderas entre sujetos, por su parte la validez, es la capacidad del cuestionario de medir el constructo para el que está diseñado (Yébenes, Rodríguez y Carmona, 2009). Se asocia a la capacidad de que, al emplearse en repetidas ocasiones y similares condiciones de medición, genera resultados veraces.

El proceso de fiabilidad y validación debe realizarse con muestras representativas de la población en la que se va a aplicar el test; entendiéndose como muestras representativas, aquellas muestras homogéneas con respecto a características socio económicas, socio demográficas u otras que afecten la calidad de la prueba, muestras que en lo posible deben ser seleccionadas usando muestreos de tipo probabilístico que eliminen posibles sesgos.

La Fiabilidad está asociada al grado en que el instrumento en su aplicación repetida en el mismo individuo u objeto produce resultados iguales. La fiabilidad de un instrumento se valora a través de tres aspectos: consistencia, estabilidad temporal y concordancia interobservadores.

- **Consistencia Interna.** Corresponde a la precisión con la que los ítems de la escala miden el constructo en estudio. En otras palabras, hace referencia al nivel en que las diferentes preguntas de una escala están relacionadas entre sí. Esta homogeneidad entre los ítems o preguntas, nos indica el grado de acuerdo entre los mismos. Entre las técnicas para medirla están fiabilidad de dos mitades, Alpha de Cronbach y la técnica de Kuder- Richardson (Tovar, 2007). El coeficiente de Cronbach oscila entre 0 y 1, se considera una buena consistencia interna cuando el valor de alfa es superior a 0.7
- **Estabilidad Temporal** o también asociado con el Test-Retest. Mide la estabilidad de la escala en sus puntajes individuales. Corresponde a la concordancia obtenida entre los resultados del test al ser evaluada la misma muestra por el mismo evaluador en dos situaciones distintas (fiabilidad test-retest). Por lo general se usan medidas estadísticas como el coeficiente de correlación intraclass- CCI, que es recomendado en variables cuantitativas continuas. La fiabilidad se expresa en forma de coeficiente de correlación que va de cero, que indica ausencia de correlación a uno que es correlación perfecta, una correlación de 0.7 indica valores aceptables de fiabilidad.
- **Concordancia interobservadores.** Está asociada al análisis del nivel de acuerdo obtenido por dos evaluadores quienes evalúan la misma muestra en las mismas condiciones. Este nivel de concordancia se puede extender al caso de ser evaluado el test en dos tiempos o momentos diferentes. La concordancia entre

observadores o inter-tiempos se puede analizar a través del porcentaje de acuerdo y el índice Kappa.

La validación de un test, está asociado a determinar, si mide aquello que pretende medir o sirve para el propósito establecido. La validez hace referencia a los procesos mediante los cuales se puede determinar el grado de exactitud con el que mide el constructo teórico de interés. Sólo a través de un instrumento validado se puede generar las relaciones entre los constructos y hacer inferencias. Según Prieto y Delgado (2010), las últimas tendencias tienden a considerar que la validez se refiere al grado en que la evidencia empírica y la teoría apoyan la interpretación de las puntuaciones de los tests relacionada con un uso específico

El concepto de validez a lo largo del tiempo ha tenido diferentes acepciones y formas de ser validado. Hacia los años 1950, la validez de un test educativo o de otra índole (psicológico o sociológico), dependía del propósito del test, su orientación teórica y el tipo de datos disponibles (Anastasi, 1986) y se hablaba de validez aparente, validez intrínseca, validez lógica, validez factorial, entre otros. La no uniformidad de criterios y denominaciones, motivó a instituciones como la American Psychological Association (APA) a publicar un manual con recomendaciones técnicas para el uso y diagnóstico con tests psicológicos en 1954.

Para efectos de este proyecto, la validez, estará asociada al grado en que el instrumento mide la variable que pretende medir, la cual puede estimarse de tres formas diferentes: validez de Contenido, de Criterio, y de Constructo (Rojas, 2013, Pérez, Chacón y Moreno 2000, Zamora Monroy y Chávez, 2010).

**Validez de Contenido.** Evalúa el grado en que un instrumento refleja un dominio específico de contenido, se considera que un instrumento es válido en su contenido si abarca todos los aspectos relacionados con el rasgo que mide. La validez de contenido consiste en determinar si la muestra de reactivos es representativa del universo de ítems posibles. Para la validez de contenido, se somete el cuestionario a la valoración de investigadores y expertos, quienes evalúan si el cuestionario contiene todas las dimensiones que se desea medir.

Sobre la medición de este aspecto hay posiciones diferentes, algunos autores sostienen que sobre ello no es viable realizar algún cálculo por cuanto la valoración es más de tipo cualitativa por parte de los expertos; sin embargo, algunos otros sostienen que una forma de hacer esta medición es a través de índices estadísticos que permitan valorar el grado de acuerdo de jueces (o expertos) respecto a los ítems propuestos.

**Validez de Criterio.** Corresponde al grado de correlación entre el instrumento con algún criterio externo que mide lo mismo (Gold Standard) y que tenga garantía de ello. En la práctica es frecuente no contar con indicadores de referencia, por lo que se suele recurrir al uso de instrumentos que cuenten con el respaldo de otros estudios. La validez de criterio puede ser: concomitante o predictiva. La validez concomitante mide el grado de correlación entre dos medidas del mismo rasgo al mismo tiempo y con los mismos sujetos. La validez predictiva mide el grado de correlación entre un instrumento y una medida posterior del mismo rasgo o de otro que este estrechamente relacionado.

**Validez de Constructo.** Evalúa el grado en que un instrumento mide la dimensión para la que fue diseñada. Esta validez proporciona evidencia que indica que los rasgos medidos con el test son indicadores del concepto teórico de lo que se pretende medir.

La validez de constructo permite garantizar que las mediciones que obtenidas de las respuestas del cuestionario sean utilizadas como medición del fenómeno que se pretende medir. La validez de constructo se evalúa frecuentemente a través del análisis factorial y la matriz multirasgo-multimétodo. Otras técnicas también utilizadas son a

través de la validación convergente-divergente y validez discriminante. La validez convergente evalúa la sensibilidad y la divergente la especificidad del instrumento.

El análisis Factorial es una técnica que establece que variables latentes causan la correlación entre las variables observadas, teniendo como objetivo reducir variables, agrupando aquellas que miden lo mismo llamados factores (Peña, 2002, Watson y Thompson, 2006). Esta técnica permite identificar como se agrupan los ítems en relación con las dimensiones o factores subyacentes del instrumento. Se puede realizar de dos formas: exploratorio o confirmatorio. El exploratorio se utiliza cuando no se conocen previamente los factores o dimensiones definitivas y el confirmatorio cuando se parte de factores definidos y se comprueba la adecuación de los mismos.

La validez discriminante mide el grado de capacidad del test para detectar que el test discrimina entre grupos teóricamente diferentes. Se puede estimar a través de las pruebas de hipótesis para la comparación de medias y del coeficiente de correlación (Rojas, 2013, Pérez, Chacón y Moreno 2000, Zamora Monroy y Chávez, 2010).

Es de precisar que el uso de una técnica estadística u otra depende así mismo de varios aspectos como el tipo de variable y los supuestos que la misma teoría implique. Por ejemplo, si se requiere correlacionar variables y son de tipo cuantitativo bajo distribución normal, se podría hacer uso del coeficiente de correlación de Pearson, pero si son de tipo cualitativo o no cumple la condición distribucional, el coeficiente a usar podría ser el de Spearman o utilizar cálculos de sensibilidad o especificidad.

Para efectos de sintetizar algunos elementos centrales de ésta sección, en la figura 3 se presenta un esquema resumen sobre las tipologías de fiabilidad y validez de un test y algunas posibles técnicas estadísticas a considerar.

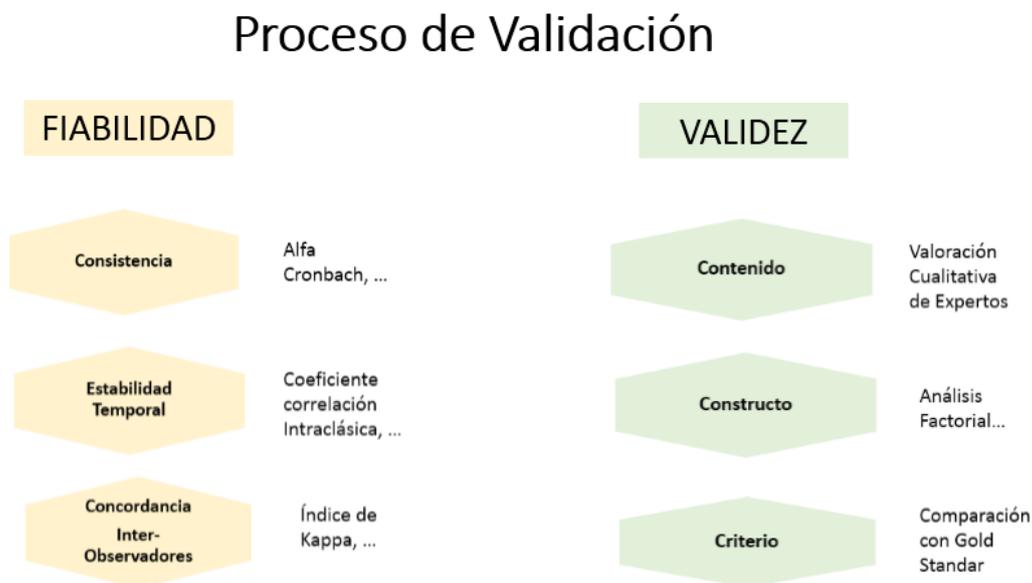


Figura 3. Proceso de fiabilidad y Validez

Finalmente, un aspecto a precisar. En ocasiones un error frecuente en quienes están iniciando este tipo de procesos, es confundir entre confiabilidad y validez. Para lo cual se precisa que un instrumento puede ser confiable pero no válido. Lo que se debe buscar es el cumplimiento de las dos características, y aunque no implican lo mismo, no deben abordarse de forma separada (Borsboom y Mellenberg, 2007). Algunos factores que

pueden afectar la confiabilidad y la validez están asociados a diferentes factores (Hidalgo, L (sf), Prieto y Delgado, 2010):

- Utilizar instrumentos desarrollados en el extranjero que no han sido validados en el país de aplicación
- Falta de estandarización en: instrucciones a los participantes, orden de las preguntas, instrumentos diferentes, entre otros.
- Test inadecuado: poca legibilidad en lo escrito, términos no comprensibles, espacio inadecuado para dar respuesta, preguntas que sugieren respuesta, organización inadecuada de los ítems
- Condiciones de espacio y tiempo desfavorables en las que se aplica.

## **Conclusiones**

Los procesos de fiabilidad y validez de un instrumento, no se puede definir por una sola tipología y por tanto se debe considerar las técnicas estadísticas e indicadores más pertinentes para cada caso, según las condiciones de tipología de las variables y de los supuestos o distribucionales exigidos dentro de cada técnica.

La elaboración del constructo como la validación del instrumento cuentan con una serie de procedimientos a seguir, y de los cuales algunos deben ser consecutivos y revisados de forma

Los indicadores de validez y fiabilidad se analizan de manera frecuente a través de técnicas estadísticas que hacen uso de los resultados y puntuaciones del test. Test que requiere revisar previamente la adaptación o traducción de escalas, que implica procesos más allá de la traducción de lenguaje.

## **Bibliografía**

- American Psychological Association. (1954). Technical recommendations for psychological tests and diagnostic techniques. *Psychological Bulletin*, 51 (2, Pt.2).
- Andreu-Andrés, M. A., & Garcia-Casas, M. (2014). Evaluación del pensamiento crítico en el trabajo en grupo. *Revista de Investigación Educativa*, 203-222.
- Anastasi, A. (1986). Evolving concepts for test validation. *Annual Review of Psychology*, 37, 1-15.
- Arribas, M. (2004). Diseño y validación de Cuestionarios. *Matronas Profesión*. Vol. 5. No. 4
- Beltrán, M. J., & Torres Merchán, N. Y. (2009). Caracterización de habilidades de pensamiento crítico en estudiantes de educación media a través del test HCTAES. zona próxima.
- Borsboom, D. y Mellenberg, G.J. (2007). Test Validity in Cognitive Assessment. En J.P. Leighton y M.J. Gierl (Eds.), *Cognitive Diagnostic Assessment for Education* (pp. 85-115). Cambridge: Cambridge University Press
- Calle, G. Y. (2013). La evaluación de las habilidades del pensamiento crítico asociadas a la. *Revista Virtual Universidad Católica del Norte*-, 63-68.

- Escurra, M., & Salas Blas, E. (2014). Construcción y Validación del Cuestionario de adicción a redes sociales (ARS). *Revista de Psicología*.
- Fuentes, C. (2010). Prueba de Pensamiento Crítico de Halpern, versión adaptada al contexto chileno. Chile: Universidad Diego Portales.
- Gómez, J., & Hidalgo, M. D. (2005). La validez de los test, escalas y cuestionarios. *Revista Electrónica La Sociología en sus Escenarios*.
- Guzmán, J. C. (2011). La calidad de la enseñanza en educación superior ¿Qué es una buena enseñanza en este nivel educativo? *Perfiles Educativos*, 129-141.
- Halpern, D. F. (2006). Halpern Critical Thinking Assessment Using Everyday Situations: Background and scoring standards.
- Hidalgo, L (sf). Confiabilidad y Validez en el contexto de la investigación y Evaluación cualitativa. Documento en línea disponible en: <http://www.ucv.ve/uploads/media/Hidalgo2005.pdf>. Consulta: 5 junio de 2019)
- Martínez Ortiz, E., Trujillo Cano, Á. M., & Trujillo, C. A. (2012). Validación del Test de Propósito Vital (pil test - purpose in life test) para Colombia. *Revista Argentina de Clínica Psicológica* .
- Muñiz, J., Elosua, P., & Hambleton, R. k. (2013). DIRECTRICES PARA LA TRADUCCIÓN Y ADAPTACIÓN DE LOS TESTS: SEGUNDA EDICIÓN. *Psicothema*, 151-157.
- Nieto, A. M., Saiz, C., & Orgaz, B. (2009). Análisis de la propiedades psicométricas de la versión española del HCTAES-Test de. *Revista Electronica de Metodología Aplicada*, 1-15.
- Paniagua Suárez, R. E. (2015). Metodología para la validación de una escala o instrumento de medida. Universidad de Antioquia.
- Peña, D. (2002). Análisis de datos multivariantes. Madrid: McGraw- Hill.
- Pérez, G. J., Chacón, M. S., & Moreno, R. R. (2000). Validez de constructo: el uso de análisis factorial exploratorio-confirmatorio para obtener evidencias de validez. *Psicothema* 12(2), 442-446.
- Prieto, G., Delgado, A. (2010). Fiabilidad y Validez. *Papeles del Psicólogo*. V.31, N. 1, enero-abril. Pp.67-74. Madrid España.
- Ramada-Rodilla, J. M., Serra Pujadas, C., & Delclós Clanchet, G. (2013). Cross-cultural adaptation and health questionnaires validation: revision and methodological recommendations. *Salus Pública Mexico*, 55(1), 57-66.
- Rivas, S.F. y Saiz, C. (2012). Validación y propiedades psicométricas de la prueba de pensamiento crítico PENCRISAL. *Revista Electrónica de Metodología Aplicada*. Vol. 17 (1), 18-34.
- Rojas, L. F. (2013). Innovación tecnológica y desarrollo de ventaja Competitiva en la atención a la salud: enfoque conceptual y metodológico. Technological innovation and development of competitive advantage in health care. A conceptual and methodological approach., 12(25), 66-82
- Saiz, C., & Rivas, S. F. (2008). Evaluación del pensamiento crítico: una propuesta para diferenciar formas de pensar. *Ergo, Nuevo Época*, 22-23, 22-66.

- Tapia, V., & Luna, J. (2010). Validación de una prueba de habilidades de pensamiento para alumnos de cuarto y quinto de secundaria y primer año de universidad. *Revista IIPSI*.
- Tovar, J. A. (2007). *Psicometria: Tests Psicométricos, confiabilidad y validez. Tópicos de actualidad*, 85-108.
- Urzúa M, A., Castro R, S., Lillo O, A., & Leal P, C. (2009). Evaluación de los trastornos alimentarios: propiedades psicométricas del test edi-2 en adolescentes escolarizados(as) de 13 A 18 años. *Revista Chilena de Nutrición*.
- Watson, R., & Thompson, D. (2006). Use of factor analysis in *Journal of Advanced Nursing: literature review*. *Journal of Advanced Nursing* 55(1), 330-441.
- Yébenes, P. M., Rodríguez, S., & Carmona, O. L. (2009). Validación de cuestionarios. *Reumatología Clínica*, 5(4) 143-186.
- Zamora, M. S., Monroy, C. L., & Chávez, Á. C. (2010). *Análisis factorial: una técnica para evaluar la dimensionalidad de las pruebas*. México D.F.: Cuaderno técnico 6.

Anexo 1.

Ejemplo Preguntas de la prueba HALPERN HCTAES, Parte 1) y Parte 2)

### **(1) Parte 1**

Un informe reciente aparecido en una revista para padres y profesores muestra que los adolescentes que fuman suelen obtener peores calificaciones en clase. A medida que aumenta el número de cigarrillos por día, disminuye la media de las calificaciones. Una sugerencia que hace el informe es que podríamos mejorar el rendimiento escolar evitando el consumo de tabaco entre los adolescentes.

Basándonos en esta información, **¿apoyarías esta idea como un medio para mejorar el rendimiento escolar de los adolescentes que fuman?**

**Sí**

**No**

**Por favor, explica por qué sí o por qué no.**

**\*\* UNA VEZ QUE CONTESTES, PASA A LA PÁGINA SIGUIENTE\*\*.**

### **(1) Parte 2**

Un informe reciente aparecido en una revista para padres y profesores muestra que los adolescentes que fuman suelen obtener peores calificaciones en clase. A medida que aumenta el número de cigarrillos por día, disminuye la media de las calificaciones. Una sugerencia que hace el informe es que podríamos mejorar el rendimiento escolar evitando el consumo de tabaco entre los adolescentes.

Basándonos en esta información, **¿cuál sería la mejor respuesta?** (Escoge una).

- a)** Las calificaciones probablemente mejoren si evitamos que los adolescentes fumen, porque la investigación encontraba que cuando se incrementa la conducta de fumar las calificaciones bajan.
- b)** Es posible que las calificaciones mejoren si evitamos que los adolescentes fumen, pero no podemos estar seguros porque solo conocemos que estas disminuyen cuando incrementa la conducta de fumar, pero no sabemos qué pasa cuando el fumar disminuye.
- c)** No hay forma de saber si las calificaciones mejorarán si evitamos que los adolescentes fumen, porque solo conocemos que fumar y calificaciones están relacionados, pero no si fumar causa que las calificaciones cambien.
- d)** Probablemente, el evitar que los adolescentes fumen no influya en las calificaciones, porque la revista está escrita por padres y profesores, de manera que es probable que estén en contra de que los adolescentes fumen