



## LA VALIDEZ DEL EXAMEN DE INGRESO A LA EDUCACIÓN SUPERIOR (EXIES) DE LA UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA (UABC)

**Karla Karina Ruiz Mendoza**  
ruiz.karla32@uabc.edu.mx

**Luis Horacio Pedroza Zúñiga**  
horacio.pedroza@uabc.edu.mx

**Alma Yadhira López García**  
alma.l@thelearningbar.com

**Área temática:** Asigne sólo una de las dieciocho áreas temáticas del congreso.  
(Consulte: <http://www.cnie.mx/>)

**Línea temática:** Evaluación educativa

**Porcentaje de avance:** 20%

Trabajo de investigación educativa asociada a tesis de grado

**Programa de posgrado:** Doctorado en Ciencias Educativas, 3er cuatrimestre

**Institución donde realiza los estudios de posgrado:** Instituto de Investigación y Desarrollo Educativo de la Universidad Autónoma de Baja California



### Resumen

El presente trabajo se centra en el concepto de validez en el área de evaluación educativa. Ello con el fin de desarrollar un método propio para someter al proceso de validez al Examen de Ingreso a la Educación Superior (EXIES). En este primer acercamiento se realiza una revisión del término de validez, tanto en su línea temporal como en su aplicación, con el fin de tratar de proponer un modelo de validez basados en el análisis de las propuestas de validez del TOEFL a partir de las propuestas teóricas de Messick, Kane y Chapelle, Enright y Jamieson. Estas propuestas se encuentran dentro un concepto moderno de validez por lo que es preciso discutirlo y presentarlo.

**Palabras clave:** validez, evaluación, exámenes, admisión.

## Introducción

A tres años de la aparición de la COVID-19, anunciada por la Organización Mundial de la Salud (OMS), los docentes y alumnos siguen enfrentando retos para mejorar su calidad de vida (Hodges, et al., 2020). El ingreso a la universidad es un reto que enfrentan los aspirantes como las mismas instituciones; Stephen Ball (1994), por ejemplo, habla de las policy technologies (tecnologías de la política), es decir, comprender la relación entre el mercado, el gerencialismo y la performatividad que afectan a nuestra realidad práctica escolar. Scriven (1991a) y Escudero (2003), en este sentido, sitúan al examen de ingreso a la universidad como parte del proceso de evaluación educativa ya que está relacionado con la medición de conocimientos. Y, si bien la evaluación puede ser formativa y sumativa, Scriven (1991a) argumenta que debe ir más allá de ambos tipos; evaluar un un sólo momento o dar el informe del proceso finalizado. Por otro lado, Escudero (2003) analiza a la evaluación desde la objetividad, la ciencia y las evidencias; en este sentido, el proceso de evaluación requiere más que la medición de conocimientos, necesita procesos y análisis de validez (Kerlinger y Lee, 2001) para entender que evaluamos lo que necesitamos evaluar.

El proyecto que se presenta a continuación tiene la intención de desarrollar una tesis doctoral y gira en torno a la validez del Examen de Ingreso a la Educación Superior (ExIES) elaborado por el Instituto de Investigaciones de Desarrollo Educativo (IIDE) de la Universidad Autónoma de Baja California (UABC). El ExIES surge durante la gestión del Dr. Juan Manuel Ocegueda Hernández, rector de la UABC durante el periodo 2015-2019, su principal fin fue el de desarrollar una prueba para el ingreso de estudiantes en dicha universidad. Las directrices de este examen fueron publicadas en el año 2017, a éste se le nombró “Manual Técnico del Examen de Ingreso a la Educación Superior” elaborado por el IIDE, siendo los autores Joaquín Caso Niebla, Carlos David Díaz, María Castro Morera y María del Rosario Martínez Arias.

Así, el ExIES “(...) mide la capacidad que tienen los aspirantes para aplicar los conocimientos y habilidades que poseen y serán requeridos para atender con éxito las demandas propias de su formación universitaria.” (Caso, et al., 2017, p.15), considerando tres áreas: Lectura, Lengua Escrita y Matemáticas. El pilotaje se realizó en diciembre del año 2016 con un total de 2,402 aspirantes, considerando que el examen comprende Lectura, Lengua Escrita y Matemáticas, teniendo un tiempo de aplicación de 3 horas (65 minutos, 35 minutos y 80 minutos respectivamente) y 122 preguntas en total (36 para Lectura y Lengua Escrita y 50 para Matemáticas), dando una escala de 700 a 1200 puntos.

En este sentido, trataremos de responder la siguiente pregunta: ¿en qué medida es válido el uso de los resultados del ExIES para identificar las capacidades en lectura, lengua escrita y matemáticas? Y ¿Cuáles son las inferencias y suposiciones que subyacen en los resultados del ExIES? Siendo el objetivo general: examinar la validez del uso del Examen de Ingreso a la Educación Superior (ExIES) con el fin de identificar los niveles de las capacidades en lectura, lengua escrita y matemáticas. Las preguntas y objetivos específicos se omitirán por falta de espacio en este apartado.

## Desarrollo

### Validez y tipos de evidencia de validez

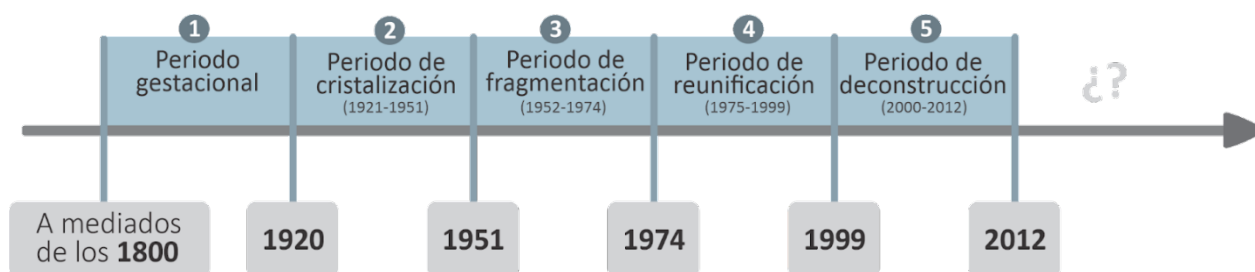
Según Kerlinger y Lee (2001) el término validez se resume en la siguiente pregunta, la cual parece ser pertinente: “¿estamos midiendo lo que creemos que estamos midiendo?” (p.), es decir, la validez ayuda a determinar la relevancia de la evidencia para respaldar un instrumento e incluso exámenes o bien si el instrumento cumple el objetivo para el cual fue elaborado (Urrutia, et al., 2014). Aunado a ello, el concepto de confiabilidad acompañará las formas de determinar la calidad de un instrumento (Urrutia, et al., 2014; Kerlinger & Lee, 2001). Entendemos que existen diversas definiciones sobre la validez, este concepto es multidimensional y heteróclito, la mayoría de los autores concuerda que la validez mantiene elementos de medición representados en los instrumentos del constructo a investigar (Delgado-Rico, Carrtero-Dios & Willibald, 2012).

En lo que respecta a los Estándares para Pruebas Educativas y Psicológicas (2014) de la *American Educational Research Association* (AERA), *American Psychological Association* (APA), y la *National Council on Measurement in Education* (NCME), se define a la validez como “(...) grado en que la evidencia y la teoría respaldan las interpretaciones de los puntajes de una prueba para usos propuestos de las pruebas.” (p.11), en este sentido, cuando hablamos de validez hablamos de un concepto unitario, es decir, una acumulación de evidencias que respalda la interpretación en los puntajes propuestos; es por ello que es erróneo definir una tipología de validez (APA, AERA y NCME, 2014).

### Cambios en el concepto de validez

A partir de los años cincuenta han existido cambios importantes en el concepto de validez, donde las investigaciones contaron con modelos de validez de contenido (Zumbo & Chan, 2014). Dicho concepto (validez), según Newton y Shaw (2014), ha evolucionado y se ha reivindicado según el análisis y los usos que se le han dado con respecto a su contexto y a su tiempo. Entonces, como se observa en la figura 1, el concepto de validez ha pasado por periodos importantes dentro de su conceptualización, desde el cómo se va gestando hasta el periodo deconstructivo donde nos volvemos a realizar las preguntas sobre si lo que pensamos que es lo es, es decir, si lo que pensamos que medimos, mide lo que debe medir; o bien, si el instrumento responde a su objetivo.

**Figura 1 Línea del tiempo del concepto validez**



**Fuente.** Elaboración propia a partir de la traducción del capítulo *An outline of the history of validity* en *Validity in Educational and Psychological Assessment* de Newton y Shaw (2014).

Aunque no se pretende profundizar en las etapas históricas conceptuales de la validez, es preciso señalar que hacia los cincuentas (con la participación concurrente de Lee Cronbach), debido al interés por estandarizar y establecer ciertos parámetros, la primera publicación en vías de estandarización de pruebas fue en el año 1954, llamado Recomendaciones Técnicas para las Pruebas Psicológicas y las Técnicas de Diagnóstico; elaborados y publicados por la APA, aunque hubo un borrador hacia 1952 (Newton y Shawn, 2014). Después la AERA y la National Council on Measurement Used in Education (NCMUE) elaboran las Recomendaciones Técnicas para Pruebas de Rendimiento y publicadas por la National Education Association (NEA) hacia 1955 (APA, AERA, NCME, 2014). Para 1966 la AERA, la APA y el NCME sustituyeron éstas para dar paso a la primera edición de los Estándares para Pruebas Educativas y Psicológicas; existiendo publicaciones hacia 1974, 1985, 1999 y la actualización en español del año 2014 (APA, AERA, NCME, 2014).

El periodo de re-unificación de la validez (*The [re]unification of validity* [1974-1999]) se le conoce como los “años de Messick” (*Messick Years*) debido a que logró unificar el concepto de validez poniendo la medición como el centro de la validez en cualquier contexto, en este sentido, la validez depende de la suma de evidencias posibles sobre la prueba y no de un solo tipo de evidencia (Newton y Shawn, 2014). En la visión de Messick la validez de constructo es el único existente, pues el objetivo es medir un constructo el cual depende del uso que se pretenda dar de la prueba. Después, Michael Kane (en el periodo nombrado *The deconstruction of validity*, 2000-2012), abona al concepto de validez el enfoque basado en argumentos puesto que ahora se debe pensar en el cómo y no en el qué, es decir, cómo procedemos a validar en fragmentos pequeños y entendibles, a diferencia de las discusiones sobre entender qué era la validez en sí misma.

### **El enfoque Basado en Argumentos de Kane**

Por lo anterior, podemos decir que Kane se encuentra en el periodo de deconstrucción (2000-2012), también nombrado como validación moderna (figura 1), en este sentido, Kane modifica la forma en la que se ha enfocado el estudio de la validez del qué de Messick al cómo. Para Kane (2013) no solo se puede interpretar la puntuación desde un constructo, si no de la interpretación de un argumento (Chapelle, 2010). El enfoque por argumentos es simple: se expone lo afirmado y se evalúa eso que se afirma, “The approach is quite simple: state what is being claimed and evaluate the claims being made.” (Kane, 2013, p.451), es decir, el enfoque basado en argumentos se distingue por centrarse en la construcción de argumentos para medir el conocimiento y las habilidades a partir de lógica y lógica estructurada.

Así, Kane (2006, 2011, 2013) propuso dos tipos de argumentos, el primero basado en la lógica del modelo de Toulmin (1958), al cual denominó como el Argumento de Interpretación o Uso (IUA, por sus siglas en inglés); el IUA especifica las afirmaciones a evaluar, es decir, las suposiciones relacionadas a las interpretaciones y usos de las puntuaciones de los instrumentos de medición. El segundo es sobre el argumento de validez (*validity argument*) el cual explica la interpretación y el uso que se propone a los puntajes, al que acompañan cuatro tipos de inferencias que se

explicarán más adelante (Taylor, 2013); que en cierta medida también se encuentra acompañado del modelo de Toulmin (1958). En resumen, tenemos dos interpretaciones utilizadas por los autores, uno por el significado de las puntuaciones de la prueba y dos, lo que se hace con los resultados de la prueba (Taylor, 2013; Carrillo, Sánchez y Leenen, 2020).

En este sentido, el primer paso en una prueba es proponer el argumento de uso o interpretación (AUI) o bien redes de inferencias y suposiciones que, en un segundo momento, pueden ser apoyadas por las cuatro inferencias de Kane (2013b; Cook, et al., 2015): puntuación (scoring), generalización (generalization), extrapolación (extrapolation), e implicaciones (*implications*). Por otro lado, Chapelle, et al. (2010), logran definir un ejemplo de una estructura para el argumento de validez del TOEFL, obsérvese figura 3, donde la conclusión revela la habilidad aprendida (en este caso es el idioma inglés) y para concluir que esta habilidad se encuentra cubierta, el que presente la prueba deberá dominar habilidades de habla, escritura y escucha para el desenvolvimiento universitario. Además, los puntos deben ser útiles para la toma de decisión sobre su admisión. En este sentido, referimos a: *grounds*, como el objetivo es el dominio del uso de la lengua; *observation*: observación del conocimiento, capacidades y habilidades en situaciones representativas del dominio del lenguaje; *observed score*: observar si el puntaje es el esperado y/o son relevantes; *expected score*: puntuación esperada; *construct*: constructo a estudiar, el concepto a medir; *target score*: revisar si el puntaje objetivo se vincula con la conclusión y funciona para la toma de decisiones.

## Método

### Tipo de estudio

Para dar respuesta a las preguntas de investigación se llevará a cabo un estudio metodológico de evaluación educativa con el fin de analizar los procesos, productos y resultados del proceso de evaluación (Jornet et al., 2010; 2020). Este tipo de investigación propone evaluar todo el proceso y no solo los resultados, así como diseñar y evaluar instrumentos de evaluación, selección de participantes, el procesamiento de los resultados y la interpretación de resultados.

Por otra parte, según la teoría consultada (AERA, APA y NCME) sobre validez, nos centraremos en el método basado en argumentos de Kane (2011) con el fin de justificar los resultados del ExIES. Este método se basa en el modelo de Toulmin, por lo que se desarrollan argumentos lógicos para evidenciar la lógica del fin, sustentados a través de evidencias.

### Procedimiento

En un primer momento se realizará un análisis del marco teórico-conceptual, a partir de la propuesta por argumentos de Michael Kane y la Revisión Sistemática de la Literatura para determinar el momento histórico y los cambios del concepto validez. Después, se seleccionaron los tipos de evidencia de validez a partir de la adaptación del modelo propuesto por Chapelle

(2010; 2020). En la tabla 3 se puede observar la inferencia, la garantía y los tipos de evidencia que se utilizarán en cada uno de los apartados. En este sentido, el procedimiento del presente proyecto toma como capítulos las inferencias para generar la estructura general; esto está sujeto a revisión y cambios según el análisis sobre los tipos de evidencias para cada una de las inferencias y garantías.

**Tabla 3 Resumen de inferencias y garantías como modelo de argumento de validez de un examen de ingreso a la universidad**

Inferencia	Garantía	Tipo de evidencia
Descripción del dominio	Observaciones de los conocimientos, destrezas y habilidades representativas a partir del uso objetivo del tema en la práctica universitaria.	Evidencias de validez de contenido
Evaluación	Estimación apropiada de los puntajes para expresar rendimiento.	Evidencias de validez relacionadas con el proceso de respuesta
Generalización	Estimaciones de las puntuaciones en tareas similares.	Estudio de equiparación y generalizabilidad
Explicación	Las puntuaciones esperadas son atribuidas al constructo medido.	Evidencias de validez de la consistencia interna (medidas de confiabilidad, teoría de respuesta al ítem, análisis factorial exploratoria y confirmatoria)
Extrapolación	El constructo de la competencia definida es observable en el rendimiento futuro durante los estudios de educación superior.	Evidencias de validez relacionadas con otras variables (validez predictiva y concurrente)
Utilización	Los resultados de la prueba reflejan las habilidades previstas, uso a validar o afirmación a defender.	Evidencias de validez relacionadas con las consecuencias

Fuente. Adaptación propia retomada de Chapelle, et al. (2010) y con tipos de evidencias posibles a utilizar según la inferencia y garantía de validez.

## Consideraciones finales

El concepto de validez ha evolucionado a lo largo del tiempo, y en la actualidad se retoma la definición de los Estándares (APA, AERA y NCME, 2014) para referirnos a ella. Existen diferentes etapas en la evolución del concepto, siendo la última la deconstrucción, donde autores como Messick y Kane proponen sus marcos referenciales para el argumento de validez. La AERA, APA y el NCME (2014) consideran que la validez es más robusta cuando hay más evidencias, lo que

hace que el argumento de validez de Kane sea propicio para el análisis y propuestas como la de Chapelle, et al. (2010). Se propone un modelo de argumento de validez que incluye inferencias y garantías para la elaboración de exámenes de ingreso a la universidad basados en la propuesta de Chapelle, et al. (2010), lo que será objeto de futuras publicaciones. Lo que nos queda es seguir estos postulados para crear un modelo propio, lo cual se encuentra en curso a partir de la Revisión Sistemática de la Literatura sobre los métodos utilizados en casos similares.

## Referencias

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2018). Estándares para pruebas educativas y psicológicas (M. Lieve, Trans.). American Educational Research Association.
- Ball, S. (1994). La micropolítica de la escuela. Hacia una teoría de la organización escolar. Paidós. (Obra original publicada en 1987)
- Chappelle, C., Enright, M., & Jamieson, J. (2010). Does an Argument-Based Approach to validity make a difference? *Educational Measurement: Issues and Practice*, 29(1), 3-13.
- Cook, D., Brydges, R., Ginsburg, S., & Hatala, R. (2015). A contemporary approach to validity arguments: a practical guide to Kane's framework. *Medical Education*, 49, 560-575. doi: 10.1111/medu.12678
- Delgado-Rico, E.; Carretero-Dios, H., & Ruch, W. (2012). Content validity evidences in test development: An applied perspective. *International Journal of Clinical and Health Psychology*, 12(3), 449-459. <https://www.redalyc.org/pdf/337/33723713006.pdf>
- Escudero, T. (2003). Desde los tests hasta la investigación evaluativa actual. Un siglo, el XX, de intenso desarrollo de la evaluación en educación. *RELIEVE*, 1(9), 11-43. <https://ojs.uv.es/index.php/RELIEVE/article/view/4348/4025>
- Hodges, C., Moore, S., Lockee, B., Trust, T., & Bond, A. (2020). The Difference Between Emergency Remote Teaching and Online Learning. *Educause*. Recuperado de <https://er.educause.edu/articles/2020/3/the-difference-between-emergency-remote-teaching-and-online-learning>
- Jornet, J. M., González, J., & Suárez, J. M. (2010). Validación de los procesos de determinación de estándares de interpretación (EE) para pruebas de rendimiento educativo. *Estudios Sobre Educación*, 19, 11-29. <https://hdl.handle.net/10171/18323>
- Jornet, J. M., Perales, M. J., & González-Such, J. (2020). El concepto de validez de los procesos de evaluación de la docencia. *Revista Española de Pedagogía*, 78(276), 233-252. [https://dadun.unav.edu/bitstream/10171/18323/2/ESE19\\_Art%3%adculo1.pdf](https://dadun.unav.edu/bitstream/10171/18323/2/ESE19_Art%3%adculo1.pdf)

- Kane, M. (2006). Current Concerns in Validity Theory. *Journal of Educational Measurement*, 38(4), 319-342. <https://doi.org/10.1111/j.1745-3984.2001.tb01130.x>
- Kane, M. (2011). Validating score interpretations and uses. *Language Testing*, 29(1), 3-17. doi:10.1177/0265532211417210
- Kane, M. (2013). Validating the interpretations and Uses of Test Scores. *Journal of Educational Measurement*, 50(1), 1-73. <https://doi.org/10.1111/jedm.12000>
- Kane, M. T. (2001). Current concerns in validity theory. *Journal of Educational Measurement*, 38(4), 319-342. <https://www.jstor.org/stable/1435453>
- Kerlinger, F. y Lee, H. (2001). *Investigación del comportamiento: métodos de investigación en ciencias sociales*. McGraw Hill.
- Newton, P., & Shaw, S. (2014). *Validity in educational & psychological assessment*. SAGE.
- Rodríguez, B. (2004). El modelo argumentativo de Toulmin en la escritura de artículos de investigación educativa. *Revista Digital Universitaria*, 5(1), 2-18. [https://www.revista.unam.mx/vol.5/num1/art2/ene\\_art2.pdf](https://www.revista.unam.mx/vol.5/num1/art2/ene_art2.pdf)
- Romero-Fernández, R.; Heras-Pérez, M.; Sáenz-López-Buñuel, P.; Fernández-Ozcorta, E. (2021). Diseño y validación de una escala de relevancia competencial motivacional del libro de texto de ciencias. *Revista Electrónica Educare*, 25(2), 307-325. DOI: <https://doi.org/10.15359/ree.25-2.17>
- Sánchez, M.; García, M.; Martínez, A, & Buzo, E. (2020). El Examen de Ingreso a la Universidad Nacional Autónoma de México: Evidencias de Validez de una Prueba de Alto Impacto y Gran Escala. *Revista Iberoamericana de Evaluación Educativa*, 13(2), 107-128. <https://doi.org/10.15366/riee2020.13.2.006>
- Scriven, M. (1991a). Beyond formative and summative evaluation. En McLaughlin, M.W., & Phillips, D.C. (Eds.), *Evaluation and education: At quarter century* (pp. 19-64), University of Chicago Press.
- (1991b). *Evaluation Thesaurus*. Sage.
- Taylor, C. (2013). *Validity and Validation*. Oxford.
- Toulmin, S. (1958). *The uses of argument*. Cambridge University Press.