

Febrero 2024  
Vol. 11, No. 1

# *revie*

Revista de Investigación y Evaluación Educativa

e-ISSN: 2409-1553  
<https://revie.gob.do>

 **ideice**  
Instituto Dominicano de Evaluación e  
Investigación de la Calidad Educativa

## DESCRIPCIÓN

El Instituto Dominicano de Evaluación e Investigación de la Calidad Educativa (Ideice) es un órgano técnico del Ministerio de Educación de República Dominicana (Minerd), cuya misión es desarrollar procesos de evaluación e investigación, que aporten evidencias sobre la calidad educativa y fomenten la mejora continua del Sistema Educativo Dominicano.

Desde el 2014, el Ideice ha venido entregando semestralmente la Revista de Investigación y Evaluación Educativa (Revie), revista digital que cumple con el objetivo de divulgar los avances de las investigaciones realizadas por los expertos en el área de educación. Revie es un recurso interactivo y de libre acceso, que sirve para fomentar debates constructivos, que lleven a ideas y propuestas para la mejora del sistema educativo.

### DIRECCIÓN EJECUTIVA DEL IDEICE

Dra. Carmen Caraballo

### EQUIPO EDITORIAL

#### Director

Dr. Julián Álvarez Acosta

#### Editora

Mtra. Dilcía D. Armesto Núñez

#### Editores de sección

Mtra. Lidia Moreta

Mtr. Francisco Javier Martínez Cruz

#### Corrección de estilo

M.A. María Rosario

Dr. Roque Santos Cueto

### CONSEJO TÉCNICO

#### Analistas de producción

M.A. Edwin Santana

#### Analista de datos

Lic. Iván Vargas

#### Soporte de tecnología

Ing. Miguel Frías Méndez

### Diseño y maquetación

M.A. Natasha Mercedes Arias

Lic. Yeimy Olivier Salcedo

### COMITÉ CIENTÍFICO

Dr. Julio Cabero Almenara

Dra. Carmen Llorente Cejudo

Dr. Héctor Valdés

Dra. Verónica Marín

Dr. Julio Ruiz Palmero

Dr. Juan Manuel Trujillo Torres

Dra. Consuelo Prado

Dr. Juan Jesús Gutiérrez Castillo

Dra. Margarita Carmenate

Dra. Mu-Kien Sang Ben

Dra. Jeanette Chaljub Hasbun

Dr. Alfredo Antonio Gorrochotegui

Dra. Ana María Ortíz

Dr. Daniel Enrique Ariza Gómez

Dr. Daniel Vargas Peña

Dr. Enrique Sánchez Rivas

Dra. Gladys Milena Vargas Beltrán

Dra. Gloria Calvo

Dra. Inmaculada Aznar Díaz

Dr. José Leopoldo Artiles Gil

Dra. Josefina Vijil

Dra. Liliana Montenegro

Lic. Luis Enrique Rodríguez

Dr. Marcos J. Villamán

Dra. Marta J. Lafuente

Dra. Morella Alvarado

Dr. Pablo Mella

M.A. Patricia Carolina Matos Lluberes

M.A. Pavel Corniel

Dr. Ramón Leonardo Díaz

Mag. Renato Operti

Dr. René Jorge Piedra de la Torre

Dr. Rodrigo Moreno Aponte

Dra. Aida Alexandra González Pons

Dra. Sandra Martínez Pérez

Dra. Sor Ana Julia Suriel Sánchez

Dra. Katusca Manzur Herra

Dr. Dustin Muñoz

Dr. Alexander Rubio Álvarez

Dr. Fernando Jafer Bárbara Rodríguez

Dra. Leidy Claret Hernández Flores

Esta obra está bajo una licencia de Creative Commons  
Atribución-NoComercial-Sin-Derivar 4.0 Internacional.



# ÍNDICE

- 04**     **00. ÉTICA EN LA INVESTIGACIÓN Y EN LA PUBLICACIÓN CIENTÍFICA: DESAFÍOS ACTUALES**  
*ETHICS IN RESEARCH AND SCIENTIFIC PUBLICATION: CURRENT CHALLENGES*  
Álvarez Acosta, Julián
- 07**     **01. DESEMPEÑO DE LOS FORMADORES DE DOCENTES EN ESCUELAS NORMALES: OPINIÓN DE LOS JEFES DE DEPARTAMENTO**  
*PERFORMANCE OF TEACHER TRAINERS IN NORMAL SCHOOLS: OPINION OF DEPARTAMENT HEADS*  
Cardoza-Batres, Francisca • Araujo Gurrola, Hilario • Reyes García, José
- 25**     **02. REVISIÓN DE LA VALIDACIÓN DEL DEFINING ISSUES TEST (DIT) EN PAÍSES HISPANOHABLANTES**  
*REVIEW OF THE VALIDATION OF THE DEFINING ISSUES TEST (DIT) IN SPANISH-SPEAKING COUNTRIES*  
Florentino, Basilio
- 43**     **03. RSL CON PRISMA: MÉTODOS PARA EL PROCESO DE VALIDACIÓN DE EXÁMENES DE INGRESO A LA UNIVERSIDAD DESDE EL ENFOQUE BASADO EN ARGUMENTOS**  
*RSL WITH PRISMA: METHODS FOR THE VALIDATION PROCESS OF UNIVERSITY ENTRANCE EXAMS FROM AN ARGUMENT-BASED APPROACH*  
Ruiz Mendoza, Karla Karina • Pedroza Zúñiga, Luis Horacio • López, Alma Yadhira
- 65**     **04. CONCEPCIÓN DEL DOCENTE DE EDUCACIÓN FÍSICA SOBRE LOS RECURSOS DIDÁCTICOS UTILIZADOS EN PRIMARIA**  
*PHYSICAL EDUCATION TEACHER'S CONCEPTION OF THE DIDACTIC RESOURCES USED IN ELEMENTARY SCHOOL*  
Taveras Méndez, Rafael • Nolasco Guridy, Dinorah • Bautista, Loan
- 88**     **05. IMPEDIMENTO DEL HABLA Y RECURSOS EDUCATIVOS DIGITALES: REVISIÓN SISTEMÁTICA DESDE LA DOCENCIA EN EDUCACIÓN ESPECIAL EN LA CIUDAD DE NUEVA YORK, USA**  
*SPEECH IMPAIRMENT AND DIGITAL EDUCATIONAL RESOURCES: SYSTEMATIC REVIEW FROM SPECIAL EDUCATION TEACHING IN NEW YORK CITY, USA*  
Made, Ana

## RSL CON PRISMA: MÉTODOS PARA EL PROCESO DE VALIDACIÓN DE EXÁMENES DE INGRESO A LA UNIVERSIDAD DESDE EL ENFOQUE BASADO EN ARGUMENTOS

### RSL WITH PRISMA: METHODS FOR THE VALIDATION PROCESS OF UNIVERSITY ENTRANCE EXAMS FROM AN ARGUMENT-BASED APPROACH



**Ruiz Mendoza, Karla Karina**

IIDE-UABC, México



**Pedroza Zúñiga, Luis Horacio**

IIDE-UABC, México



**López, Alma Yadhira**

Learning Bar, Inc., México

**Recibido:** 2023/08/22

**Aceptado para su publicación:** 2023/10/06

**Publicado:** 2024/02/01

#### RESUMEN

En México, la selección de candidatos para la educación superior es un desafío debido al limitado número de cupos disponibles en las universidades. Para asegurar la calidad de los estudiantes admitidos, es crucial evaluar la validez de los exámenes de ingreso. Uno de los enfoques más revisados para la validación de pruebas educativas, en los últimos veinte años, es el Enfoque Basado en Argumentos propuesto por Michael Kane, sin embargo, no ha quedado claro el método, métodos o enfoques para poder evaluar la validez de las pruebas. Por ende, se realizó una Revisión Sistemática de la Literatura a través del modelo Prisma, con el fin de contribuir a esclarecerlo. Así, se encontró que en todos estudios que cubren este enfoque citan a Kane para que forme parte de su marco teórico y, en general, desarrollan métodos basados en argumentos sin esclarecerlos puntualmente. Asimismo, se encontró que el Enfoque Basado en Argumentos se ha extendido a enfoques de investigación mixta y cualitativas. Lo que demuestra, en conclusión, que el Enfoque Basado en Argumentos depende del contexto, la situación y las especificaciones de la prueba, aportando flexibilidad y estructura a la investigación en este campo.

#### PALABRAS CLAVE

RSL, argumentación, validez, validación, prueba de validez.

#### ABSTRACT

*In Mexico, the selection of candidates for higher education is a challenge due to the limited number of available spots in universities. To ensure the quality of admitted students, it is crucial to evaluate the validity of entrance exams. One of the most reviewed approaches for the validation of educational tests in the past twenty years is the Argument-Based Approach proposed by Michael Kane. However, the method(s) or approaches to assess test validity have not been clearly defined. Therefore, a Systematic Literature Review was conducted using the PRISMA model to contribute to its clarification. It was found that all studies covering this approach cite Kane as part of their theoretical framework and generally develop argument-based methods without specifying them. Additionally, it was found that the Argument-Based Approach has been extended to mixed and qualitative research approaches. This demonstrates that the Argument-Based Approach depends on the context, situation, and specifications of the test, providing flexibility and structure to research in this field.*

#### KEYWORDS

*Validation, validity, validity arguments, test validity, RSL.*

## 1. INTRODUCCIÓN

Cada año, en México, miles de estudiantes acuden a universidades públicas y privadas para poder cursar la Educación Superior (ES). Según los últimos datos publicados del Anuario Estadístico de la Población Escolar en Educación Superior (Anuies), correspondiente al ciclo escolar 2021-2022, en México hubo 2,233,527 millones de solicitudes para el ingreso a la educación superior; de estos aspirantes lograron entrar 1,357,872 (Anuies, 2023), lo que equivale al 60.79%; esto quiere decir que 6 de cada 10 acceden a la educación superior, sea en una universidad pública o privada.

Ante este panorama, las universidades tienen un número limitado de cupos disponibles para cada carrera y, por lo tanto, necesitan seleccionar a los mejores candidatos para asegurarse de que los estudiantes que ingresan a la institución tengan las habilidades y conocimientos necesarios para tener éxito en sus estudios (Barrios, et al., 2015). En esta línea, el aparato que conforma a cualquier sistema de evaluación para el ingreso a la universidad es en suma delicado y las evidencias de validez procuran la estabilidad social de las instituciones provocando confianza en la comunidad y, por lo tanto, en los aspirantes (Morales, Barrera y Garnett, 2017; Barrios, et al., 2015).

Tomando en cuenta lo anterior, nuestro interés se centra en tener un método adecuado con base en la definición de validez: “La validez se refiere al grado en que la evidencia y la teoría respaldan las interpretaciones de los puntajes de una prueba para usos propuestos de las pruebas.” (AERA, APA & NCME, 2018, p.11). Para lograr este fin la AERA, APA & la NCME (2014) especifican que es necesario incluir un argumento de validación, no obstante, no se especifica el proceso o la forma en que debe presentarse dicha argumentación.

Dentro del periodo histórico de deconstrucción (2000-2012; Newton & Shaw, 2014) de la validez, Michael Kane (2006a, 2011) es reconocido por su Enfoque Basado en Argumentos debido a que da respuesta a esta preocupación sobre cómo presentar los argumentos para dar evidencia contundente del proceso de validación, es decir, detalla la interpretación y uso propuesto de las puntuaciones, acompañado de cuatro tipos de inferencias: puntuación, generalización, extrapolación e implicaciones.

Este enfoque tiene su raíz en la lógica informal del modelo de Toulmin (Rodríguez, 2004); y parte de la identificación clara de la construcción que se pretende medir, qué se mide y cómo se mide; después, se deben identificar los argumentos que respalden el uso de la prueba, lo cual puede dividirse en tres categorías: fundamentos teóricos, evidencia empírica y consecuencias prácticas. En un tercer momento, se pueden analizar los argumentos para después integrarlos en la evaluación final; una vez realizado esto, se debe seguir un proceso continuo, es decir, realizar pruebas de validez de manera periódica (Kane, 2006a, 2013, 2016b). Aunado a ello, Kane (2011) especifica que es necesario tener una orientación teórica

en la propuesta del proceso de validación que se realice, por lo cual, debe existir un marco teórico robusto como parte del propio proceso.

Aunque la AERA et al. (2018) propone basarse en argumentos y Kane ha aportado su enfoque y el Argumento de Interpretación o Uso (UIA, por sus siglas en inglés), se sabe que la validación basada en argumentos sigue sin ser desarrollada de forma empírica o con criterios específicos (Lavery et al., 2020). Según la revisión de la literatura y el análisis minucioso realizado por Lavery et al. (2020), solo el 17% de los artículos que publican sobre pruebas educativas o psicológicas proporcionan información sobre validez; aunado a ello, llegan a la conclusión de que de los 83 artículos revisados (sobre el término validez, validación y enfoque por argumentos) y publicados en revistas revisadas por pares entre 2002 y 2017, ninguno llega a concluir un argumento de validez, ya que suelen profundizar en la teoría y no en las investigaciones empíricas. Lo que al final termina con una ausencia de la propuesta en práctica del Enfoque Basado en Argumentos.

Además, se entiende que el enfoque de Kane (2006a) ha sido difícil de seguir por su propio nivel de vocabulario y por ser bastante estricto en la elaboración del argumento (Lavery et al., 2020), lo que ha concluido en una escasez de procedimientos claros en la literatura disponible. Ante este panorama es necesario profundizar en este tipo de investigaciones para comprender cómo pasar de lo teórico a lo práctico para no olvidar los aspectos éticos que enmarca la propia AERA, APA y la NCME (2014).

En este contexto, y por la importancia de los exámenes de ingreso a la educación superior, el objetivo de la presente RSL es identificar los métodos que usan en sus procesos de validación de las pruebas a partir del Enfoque Basado en Argumentos propuesto por Kane (2006a, 2011, 2013), esto a partir del modelo Prisma (*Preferred Reporting Items for Systematic reviews and Meta-Analysys*). En concreto se retoman estas dos preguntas para guiar este estudio:

- P1. ¿Cuáles son los temas y autores que citan al elaborar su estudio y fortalecer su enfoque teórico, en relación con el Enfoque Basado en Argumentos?, sobre todo retomando la investigación ya realizada por Lavery et al. (2014).
- P2. ¿Cuáles son los métodos (considerando técnicas, instrumentos y participantes) que han utilizado en sus procesos de validación con el Enfoque Basado en Argumentos para valorar las pruebas de ingreso a la educación superior?

## 2. METODOLOGÍA

Con el fin de dar respuesta a las preguntas y el objetivo general planteado, esta investigación se centra en la elaboración de una Revisión Sistemática de la Literatura (RSL). Según Page et al. (2021) la RSL son

útiles para proporcionar una síntesis del estado de conocimiento sobre lo que estemos investigando. Igualmente, permite encontrar otras problemáticas y generar, e incluso evaluar, teorías sobre ciertos fenómenos (2021). En este sentido, se seleccionó la RSL debido a su metodología rigurosa y sistemática, donde no depende de la visión completa del autor, sino que implica la identificación, selección y evaluación crítica de la literatura relevante, lo que reduce el riesgo de sesgos y aumenta la confianza en los resultados (Rethlefsen et al., 2021; García-Peñalvo, 2022).

## 2.1. PROCEDIMIENTO

## 2.2. ESTRATEGIA DE BÚSQUEDA

Antes de realizar la búsqueda sistemática se elaboraron fórmulas cortas, tanto en español como en inglés; estos fueron “validity argument, argument-based”, y “scholastic aptitude test, “american college testing” en el idioma inglés; y esto otros en español “examen de ingreso a la universidad”, “validez de exámenes de ingreso a la universidad”. Esta exploración se realizó en bases de datos como Eric, Springer Link, Web Of Science, Redalyc, Google Académico, Dialnet, Latindex, Conricyt, Iresie y Scopus. Después se realizó una búsqueda a partir de operadores booleanos como AND y OR según la conveniencia en los siguientes términos: “(validity argument) OR (argument-based)”, y “(scholastic aptitude test) OR (american college testing)” en el idioma inglés; y estos otros en español “(Enfoque Basado en Argumentos) OR (validez del argumento)”, y “(examen de ingreso a la universidad) OR (examen de admisión) OR (EXANI II)”. Las fórmulas finales obtenidas en inglés y español se pueden observar en la Tabla 1.

**TABLA 1.**

### FÓRMULAS OBTENIDAS

ESPAÑOL	INGLÉS
((argumento de validez) OR (basado en argumentos) AND (validez) OR (validación) AND ((ingreso a la universidad) OR (examen de ingreso) OR (universidad) OR (examen)) NOT ((preescolar) OR (primaria) OR (tecnología) OR (inclusión) OR (secundaria))	((validity argument) OR (argument-based) OR (argument-based framework)) AND ((validity) OR (validation)) AND ((university testing) OR (american college testing) OR (College Entrance Examinations) OR (college admission) OR (scholastic aptitude test) OR (admission testing) OR (admission test))

*Nota. Elaboración a partir del análisis realizado.*

Con el fin de elaborar una fórmula que se ajustara a las diferentes bases de datos, se procedió a elaborarla en la página de PubMed (<https://pubmed.ncbi.nlm.nih.gov/advanced/>). Las bases de datos utilizadas para este ejercicio fueron *Eric, Springer Link, Web Of Science*; seleccionadas, también, por su relevancia a nivel internacional y su relación con el ámbito educativo. Otros criterios de inclusión y

exclusión (véase Tabla 2) fueron que se tomaron en cuenta solo artículos en un periodo de publicación del año 2004 al 2022 y que fueran revistas especializadas enfocadas a subdisciplinas como Educación Superior, Exámenes, Evaluaciones y Pruebas y Evaluaciones; que se incluyeran autores nacionales como internacionales. En esta ocasión se tomaron en cuenta artículos abiertos y privados debido a la especialización del tema.

**TABLA 2.**

**CRITERIOS DE INCLUSIÓN Y EXCLUSIÓN**

CRITERIO	INCLUSIÓN	EXCLUSIÓN
Tipo de documento	Artículos en revistas especializadas (con ISBN o ISSN) en revistas indexadas nacionales e internacionales.	Libros o capítulos de libro; Artículos de información; tesis de licenciatura, maestría y doctorado; actas de congresos; ponencias; noticias; blogs especializados; revisiones de libros
Subdisciplina	Educación Superior, Assessment, Testing and Evaluation (Springer Link)	otras subdisciplinas
Producción del autor(es)	Nacional e internacional	No aplica
Periodo de publicación	2004-2022	Antes de 2004
Nivel educativo	Superior	otros niveles educativos
Idioma	inglés, español	No aplica
Documento	Documento completo	Documento duplicado; sin resumen disponible
Disciplina	Educación	otras disciplinas
Accesibilidad	Acceso abierto/ privado	Ninguno

*Nota. Elaboración a partir del análisis realizado.*

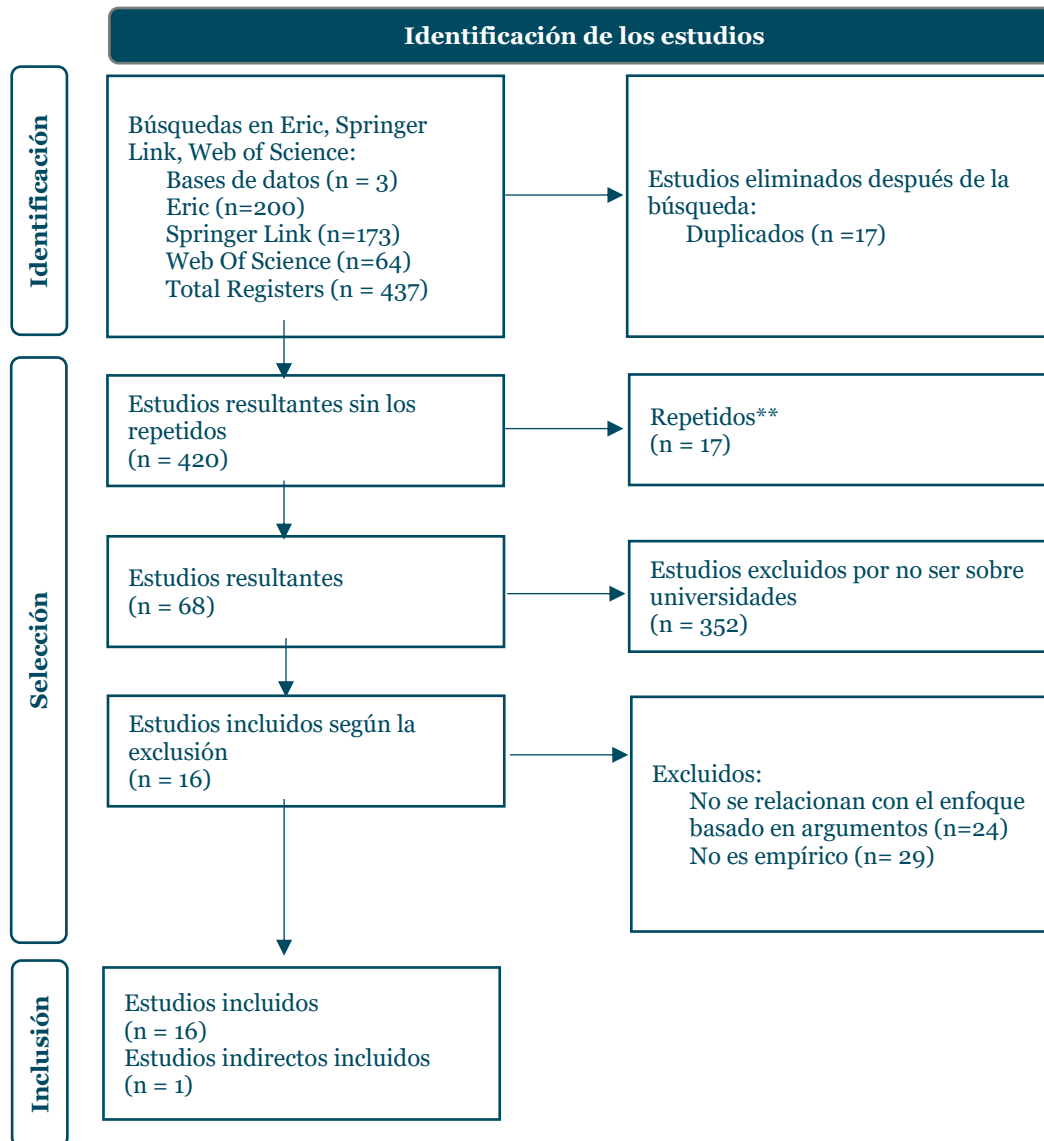
**2.3. REVISIÓN Y SELECCIÓN**

En la Figura 1, que es el diagrama de flujo utilizado para llevar un registro transparente de la búsqueda, se observa el proceso de identificación, selección e inclusión de los artículos encontrados. En este caso, hubo un total de 437 artículos provenientes de las bases de datos ya mencionadas con anterioridad. De estos, se excluyeron 17 debido a la duplicación, quedando un total de 420 artículos, de ellos 352 no tenían relación con el tema de educación superior o media superior, finalizando con un total de 68 artículos a revisar como lo indica Prisma (Page et al., 2021; Rethlefsen et al., 2021).



FIGURA 1.

## DIAGRAMA DE FLUJO DE LA RSL CON PRISMA



Nota. Retomado de Page et al. (2020).

Se debe señalar de estos 68 artículos se obtuvo un total de 29 artículos teóricos y dos RSL, de estas últimas solo una correspondía a nuestro interés (Enfoque Basado en Argumentos) es la escrita por los autores Lavery et al. (2020). Una vez revisados los 68 artículos, se descartaron aquellos que no se relacionan con el Enfoque Basado en Argumentos (24) y los que no eran empíricos, ya que eran de carácter teórico (29). Finalmente, se agregó un artículo que se localizó buscándolo directo en la web (Google) (Fechter et al., 2021) y se optó por conservar el artículo de Covey et al. (2014) sobre la evaluación

de liderazgo en escuelas primarias y secundarias en Estados Unidos debido a que su método puede implicar una nueva forma de comprender la validez. El total de artículos fue de 17, siendo todos en el idioma inglés y, debido a la dificultad y especialización, se tomaron en cuenta artículos del área de evaluación o exámenes de la adquisición del idioma inglés, ya que se pudo observar una buena afluencia de artículos en esta área, como bien lo indicaron Lavery et al. (2020) en su RSL.

## 2.4. EXTRACCIÓN DE LOS DATOS

Con los 17 artículos se procedió a su lectura, revisión y análisis. Además, se utilizó el software Atlas.ti en su versión 23.0.8.0 con el fin de encontrar relaciones entre las referencias utilizadas por los autores y codificar la información de cada uno de los artículos (Saldaña, 2015). Para ello se partió de las preguntas de investigación descritas con anterioridad, estando abiertos a un análisis inductivo de los códigos; siendo esto un estudio exploratorio (Kerlinger & Lee, 2001) sobre el Enfoque Basado en Argumentos en los exámenes de ingreso a la universidad o exámenes similares.

## 2.5. MUESTRA ANALIZADA

Como se observa, en la Tabla 3 hubo una diversidad de revistas en las que se publicaron estos artículos, la revista con mayor uso en esta muestra fue la *Language Testing* con cuatro artículos, *Language Testing in Asia* con tres artículos, y el resto con un artículo por revista. Asimismo, se puede observar que estas revistas se relacionan con evaluación del lenguaje y evaluación educativa. A ello, añadimos que los autores provienen, en su mayoría, de Estados Unidos (EE. UU), el resto se encuentran en el continente asiático: China, Irán, Israel, Taiwán y Arabia Saudita; y, aunque también hay variedad en las universidades, se puede destacar un flujo en la Universidad de Arizona y la Universidad de California, pero no muy significativo.

**TABLA 3.**

### CARACTERÍSTICAS DE PROCEDENCIA DE LOS ARTÍCULOS

NO.	AUTOR(ES)	EDITORIAL / REVISTA	PAÍS	UNIVERSIDAD
1	Ginther y Elder (2014)	ETS Research Report Series	EE. UU	Purdue University, University of Melbourne
2	Covay et al. (2014)	Educational Assessment, Evaluation and Accountability	EE. UU	Michigan State University, University of Pennsylvania, Vanderbilt University, Arizona State University
3	Fan (2014)	Language Testing in Asia	China	Universidad Fudan en China; Centro de Evaluación Lingüística de la Universidad Fudan y Pearson Knowledge Technologies
4	Staples, LaFlair y Egbert (2017)	Modern Language Journal	EE. UU	Universidad de Arizona, Universidad de Kentuck

NO.	AUTOR(ES)	EDITORIAL / REVISTA	PAÍS	UNIVERSIDAD
5	LaFlair y Staples (2017)	Language Testing	EE. UU	Universidad de Hawai'i, Universidad de Arizona.
6	Esfandiari et al. (2018)	Language Testing in Asia	Irán	Shahid Beheshti University of Medical Sciences
7	Cizek, Kosh y Toutkoushian (2018)	Journal of Educational Measurement	EE. UU	Universidad de Carolina del Norte
8	Li (2018)	Language Testing in Asia	EE. UU	Universidad Estatal de Georgia
9	Beigman et al. (2017)	Language Testing	EE. UU	Educational Testing Service y Carnegie Mellon University
10	Aviad-Levitzky, Laufer y Goldstein (2019)	Language Assessment Quarterly	Israel	Gordon College of Education en Haifa, Israel, Batia Laufer y Zahava Goldstein de la Universidad de Haifa
11	ShayesteFar (2020)	Educational Assessment, Evaluation and Accountability	Irán	Farhangian Teacher Education University
12	Yan y Staples (2019)	Language Testing	EE. UU	University of Illinois at Urbana-Champaign, University of Arizona,
13	Fechter et al. (2021)	Frontiers in Education	EE. UU	Universidad de Massachusetts
14	Jawhar et al. (2021)	Arab World English Journal	Arabia Saudita	King Saud bin Abdulaziz University for Health Sciences
15	Huang et al. (2020)	Language Testing	EE. UU, Taiwán	Universidad de Texas, Universidad de California, Universidad de Texas, Universidad Nacional de Taipei
16	Chuang y Yan (2022)	Journal Of Second Language Writing	EE. UU	University of Illinois at Urbana-Champaign, Beckman Institute for Advanced Science and Technology
17	Seipel et al. (2023)	Journal of Learning Disabilities	EE. UU	California State University, University of Oregon, Georgia State University, University of North Dakota, y University of Minnesota.

### 3. ANÁLISIS Y RESULTADOS

#### 3.1. TEMAS Y AUTORES

Con el fin de dar respuesta a la primera pregunta de la presente revisión, en la Tabla 4 se pueden observar los temas retomados por cada una de las investigaciones, las cuales están relacionadas con el Enfoque Basado en Argumentos en procesos de validación con especial énfasis sobre estudios centrados en postulantes a universidad o universitarios. En este sentido, se podrá observar que la mayoría de los resultados se enfocan en el área de lenguaje, como ya se ha ido mencionando (ejemplo, Chapelle, 2010, 2012), ya que se observa una mayor afluencia académica desde esta área. Tal es el caso de Ginther y Elder (2014); Jawhar et al. (2021); ShayesteFar (2020); Fan (2014); Esfandiari et al. (2018); quienes se enfocan en mejorar la comprensión y evaluación de la competencia en inglés y el uso del lenguaje en diferentes contextos. Ginther y Elder (2014) desarrollaron un estudio para la admisión de estudiantes internacionales en programas de posgrado ubicados en Estados Unidos y en Australia.

TABLA 4.

## RESUMEN DE LOS TEMAS LOCALIZADOS EN LA RSL

TEMA	AUTORES
Evaluación del lenguaje del idioma inglés	Ginther y Elder (2014); Jawhar et al. (2021); ShayesteFar (2020); Fan (2014); Esfandiari et al. (2018)
Evaluación de la comprensión pragmática	Li (2018)
Evaluación del lenguaje y medición a partir de tecnología informática	Aviad-Levitzky, Laufer y Goldstein (2019)
Evaluación del lenguaje escrito	Yan y Staples (2019); Chuang y Yan (2022)
Evaluación de comprensión lectora	Seipel et al. (2023)
Evaluación de la oralidad	Staples, LaFlair y Egbert (2017);
Evaluación de la oralidad / lenguaje hablado	Staples, LaFlair y Egbert (2017); Huang et al. (2020)
Lingüística de corpus	LaFlair y Staples (2017)
Evaluación de habilidades de inferencia y razonamiento en estudiantes universitarios	Fechter et al. (2021)
Evaluación de habilidades de liderazgo	Covay et al. (2014)
Evaluación de la validez de las pruebas educativas	Cizek, Kosh y Toutkoushian (2018); Beigman et al. (2017)

*Nota. Elaboración a partir del análisis realizado.*

Por otra parte, en la Figura 1 se observa el mapa de citas relacionado con el Enfoque Basado en Argumentos, es decir, los autores que citaron en estas investigaciones para fundamentar su investigación. A partir de elaborar este mapa, se logró visualizar fácilmente a los autores que han estado trabajando desde este enfoque, es el caso de: Shelley Staples, LaFlair y Xun Yan; marcados en café y amarillo en la Figura 1. En este sentido, LaFlair y Staples (2017) citaron su colaboración con Egbert (2017) en su artículo *Comparing language use in oral proficiency interviews to target domains: Conversational, academic, and professional discourse*. Shelley Staples ha participado en tres artículos relacionados con el lenguaje, Xun Yan en dos y LaFlair en otros dos, ambos colaborando con Staples.



y, *The centrality of test use and consequences for test validity* de Shepard (1997). Y, para finalizar esta sección, la mayoría de los artículos que se relacionan directamente con lenguaje suelen citar a Kane (1992, 2001, 2002, 2006a, 2006b, 2012, 2013) o a Chapelle en sus diferentes aportaciones.

En este sentido, Kane sigue siendo una figura central y altamente citada en la literatura relacionada con los procesos de validación en la evaluación de prueba y exámenes, con múltiples publicaciones a lo largo de los años que son referenciadas regularmente, como lo evidencian las múltiples citas de diferentes años. Messick y Chapelle también son autores relevantes y frecuentemente citados, mostrando su influencia significativa en el campo. No obstante, en el caso de Messick (1989) se debe a las discrepancias conceptuales que ha mantenido con Kane, mientras con Chapelle (2010; 2012) es por una continuidad al uso del Enfoque Basado en Argumentos.

### 3.2. MÉTODOS PARA EL PROCESO DE VALIDACIÓN

Con el fin de responder a nuestra inquietud sobre los métodos que se han seguido para la implementación del Enfoque Basado en Argumentos, se elaboró la Tabla 5; para un mejor seguimiento de los artículos se señala el nombre del artículo, autores y métodos. Si bien ya se han descrito las discrepancias sobre el método o sobre la pluralidad de posibles métodos (Lavery, et al. 2020; Chapelle et al. 2010), en la Tabla 5 se refuerza la idea de dicha heterogeneidad, a continuación, se procuró agrupar a las investigaciones en diferentes categorías con el fin de responder a las preguntas de la presente RSL.

Análisis de corpus y comparaciones discursivas: Ginther y Elder (2014), Staples, LaFlair y Egbert (2017), y LaFlair y Staples (2017) adoptaron técnicas basadas en el análisis de corpus y discursivo para evaluar la competencia lingüística. Ginther y Elder centraron su atención en pruebas populares de dominio del inglés, analizando las respuestas de estudiantes en dos universidades distintas mediante cuestionarios y entrevistas. Staples, LaFlair y Egbert se sumergieron en la comparación de diferentes corpus para analizar el discurso en variadas situaciones, brindando una visión de cómo los hablantes emplean el idioma en contextos específicos. LaFlair y Staples, mientras tanto, se centraron en el análisis lingüístico cuantitativo y cualitativo, utilizando registros de corpus para evaluar las habilidades de habla en ámbitos académicos, profesionales y sociales.

Métodos mixtos (cuantitativas y cualitativas): Fan (2014) y ShayesteFar (2020) emplearon un diseño de métodos mixtos, combinando técnicas cuantitativas y cualitativas para obtener resultados más ricos y multidimensionales. Fan tuvo como fin comprender las actitudes de los participantes hacia el aprendizaje del inglés. ShayesteFar, por otro lado, implementó una combinación de encuestas y análisis estadístico descriptivo para investigar las consecuencias afectivas del cambio en el aprendizaje del inglés.

Validación y desarrollo de pruebas: Esfandiari et al. (2018), Jawhar et al. (2021), Fechter et al. (2021), y Seipel et al. (2023) centraron sus investigaciones en la validación y desarrollo de pruebas. Y, en el caso de Esfandiari et al. llevaron a cabo un examen detallado del contenido y propósito del TOEFL iBT, mientras que Jawhar et al. se concentraron en validar una prueba de ingreso al inglés a través de tres fases. Aunado a ello, Fechter et al. se enfocaron en la validación de una herramienta diseñada para evaluar habilidades específicas en biología, y Seipel et al. desarrollaron y validaron una tarea centrada en la comprensión causal.

Centrados en la pragmática y habilidades de argumentación: Li (2018), Beigman et al. (2017), y Chuang y Yan (2022) se adentraron en la pragmática y las habilidades de argumentación. Li se alineó en la comprensión pragmática en diálogos de chino hablado natural, mientras que Beigman et al. analizaron la retórica en distintos géneros de escritura, especialmente en el ámbito argumentativo. Chuang y Yan, por su parte, investigaron la habilidad de los estudiantes internacionales para formular argumentos en ensayos basados en tareas integradoras.

#### 4. DISCUSIÓN Y CONCLUSIONES

A través de la literatura analizada podemos decir que son de alto impacto social los estudios del proceso de validación de los exámenes de ingreso a la educación superior (Sánchez et al., 2020), ya que a partir de ello se decide o está en juego las futuras decisiones de los aspirantes (Barrios et al., 2015; Morales, Barrera y Garnett, 2017; Jiménez, 2021). Por otra parte, es tan amplia la libertad que dan los Estándares (AERA, APA & NCME, 2014) sobre el concepto de argumentación que la discusión sobre cómo abordarlo sigue siendo un reto, a pesar de las revisiones ya realizadas (Lavery et al., 2020); otro ejemplo podría ser la discusión de Kane (2006; 2016) sobre el concepto de validez como algo multivariado en vez de unitario.

Partiendo de lo anterior, el mapa de los autores más citados de forma teórica (véase Figura 4) genera claridad sobre el estado del marco teórico para una investigación de este tipo. En general, cuando elaboran su estudio existe una discusión teórica o un posicionamiento en favor de Kane (1992, 2001, 2002, 2006, 2012, 2013) para de ahí partir con el desarrollo de su propio método, sea cuantitativo o mixto.

TABLA 5.

## RESUMEN DE LOS TEMAS LOCALIZADOS EN LA RSL

AUTOR(ES)	MÉTODO		
	TÉCNICA	INSTRUMENTO	PARTICIPANTES
Ginther y Elder (2014)	Análisis comparativo de dos universidades que compara las pruebas: TOEFL iBT, <i>International English Language Testing Service</i> (IELTS) y el <i>Pearson Test of English</i> (PTE), a partir de estadísticas descriptivas y análisis temático.	Cuestionario: características del programa del participante, el papel del inglés en el éxito académico en el primer año, financiación disponible para estudiantes graduados internacionales, etc.  Entrevistas post-encuesta, no se proporciona una guía de preguntas.	232 (Perdue) y 249 (Melbourne) estudiantes de universidad.
Covay et al. (2014)	Análisis de grupos conocidos: recopilación de datos a través de informes (directivos, maestros y supervisores); evalúan: planificación, organización, liderazgo, gestión, a través del método de diferencias de medias y análisis discriminante.	VEL-ED, evaluación del liderazgo escolar sobre los directores; evalúa 12 componentes principales y procesos clave.	190 escuelas de los 6 distritos, con 120 escuelas primarias y 70 secundarias.
Fan (2014)	Diseño explicativo secuencial de métodos mixtos, comenzando con un enfoque cuantitativo seguido por uno cualitativo.	Cuestionario: 20 ítems, tres componentes: creencias, opiniones y emociones.	118 estudiantes de una universidad de China.
Staples, LaFlair y Egbert (2017)	Análisis como factorización exploratoria, correlaciones de Pearson y pruebas t Se utilizó un enfoque basado en la comparación de diferentes corpus para analizar el discurso en diferentes situaciones. Corpus: MELAB OPI corpus, U.S. Nurse/ patient corpus, T2K-SWAL, y Longman American Conversation corpus.	Guía de entrevistas basada en el cuestionario. Se emplearon transcripciones estandarizadas de interacciones orales basadas en una variedad de contextos y propósitos.	Los participantes variaron según el corpus (MELAB, USNP, T2K-SWAL y Longman).
LaFlair y Staples (2017)	Se utilizó un enfoque basado en registros de corpus, que combina un análisis lingüístico cuantitativo (a través del análisis multidimensional) y un análisis situacional que examina cualitativamente las características situacionales de los registros.	MELAB OPI: diseñada para medir la capacidad de habla desde un nivel intermedio hasta avanzado en ámbitos académicos, profesionales y sociales.  El MELAB OPI consiste en una entrevista entre un examinador y un candidato; se graba y transcribe para crear un corpus de datos.	Los participantes del estudio fueron aquellos involucrados en las grabaciones del MELAB OPI y en los registros de los corpora de referencia mencionados.  No especifican un número.
Esfandiari et al. (2018)	Investigación cuantitativa usando un modelo interpretativo. Se condujo un análisis sistemático del contenido de la prueba, así como una encuesta mediante cuestionario. Se examinaron y analizaron muestras de pruebas de TOEFL iBT y luego, los participantes participaron en una encuesta acerca del contenido y propósito del TOEFL iBT, indicando sus opiniones acerca de la validez de esta prueba.	Se utilizaron tres cuestionarios (para candidatos, profesores y evaluadores del TOEFL iBT) que fueron desarrollados, validados y utilizados.	140 miembros de tres grupos principales: candidatos a TOEFL iBT, profesores y evaluadores.
Cizek, Kosh y Toutkoushian (2018)	Alineación, que es una técnica sobre cómo los ítems y tareas de una prueba coinciden con los estándares de contenido u objetivos que la prueba pretende medir.	Se recomienda para: Instrumentos implicados son las pruebas que buscan medir el dominio de los estudiantes sobre los currículos especificados.	Estudiantes de K12, sin definir.



AUTOR(ES)	MÉTODO		
	TÉCNICA	INSTRUMENTO	PARTICIPANTES
	*Este es un artículo teórico con datos ficticios utilizados para representar el uso de este tipo de técnica.	Exámenes o pruebas están diseñados para evaluar y validar que un candidato posee los conocimientos, habilidades y aptitudes necesarias para la práctica segura y efectiva a nivel de entrada en una profesión.	
Li (2018)	Se utilizó un enfoque basado en la comprensión pragmática de implicaturas en diálogos de chino hablado natural.	Corpus Anotado de Diálogo y Conversación.	Se reclutaron 107 estudiantes de chino como L2 en una universidad en China. De estos, 98 completaron el test computarizado y 9 participaron en entrevistas retrospectivas.
	WINSTEPS y SPSS: Herramientas de software utilizadas para el análisis de datos.	Notas de campo: Tomadas en China por el investigador en 2013, registrando interacciones diarias.	
	Software LiveCode: Para computarizar los ítems de la prueba.	Cuestionario metapragmático.	
Beigman et al. (2017)	Estimación de Perfiles Retóricos: Se establecen perfiles retóricos para varios géneros en datos instructivos, como argumentativo, narrativo, escénico, etc.	Escritura de Evaluación: Ensayos de pruebas como GRE y TOEFL.	En la Escritura de Evaluación, los ensayos fueron escritos por participantes en las pruebas GRE y TOEFL entre 2006 y 2010.
	Comparación de Perfiles: Se comparan los perfiles retóricos del argumentativo en contextos de pruebas y cívicos con el perfil de referencia.	Escritura Instructiva: Muestras de escritura relacionadas con cursos de varios géneros en la Universidad Carnegie Mellon.	
		Escritura Argumentativa Interdisciplinaria: Muestras de escritura argumentativa, utilizando el Corpus MICUSP.	En la Escritura: 100 estudiantes diferentes durante siete años (2006-2012).
		Escritura Cívica: Artículos de opinión del New York Times.	La Escritura Argumentativa Interdisciplinaria y cívica: No específica.
Aviad-Levitzky, Laufer y Goldstein (2019)	Se realizaron diversas técnicas de análisis, como el coeficiente alfa de Cronbach, para examinar la consistencia interna de cada modalidad y del test completo. Se utilizaron análisis Rasch para examinar si las puntuaciones podían representar el rasgo general que estaban destinadas a medir. También, se realizaron ANOVAs y pruebas post hoc para comparar los grupos de competencia.	Computer Adaptive Test of Size and Strength (CATSS); en la cual se basó el desarrollo del proceso de validación.	El estudio se llevó a cabo en universidades y colegios israelíes con 453 estudiantes de pregrado y posgrado cuyas lenguas maternas eran hebreo o árabe.
		Test Psicométrico Nacional de Ingreso a la Universidad en Israel.	
ShayesteFar (2020)	Enfoque de métodos mixtos: encuesta y análisis estadístico descriptivo.	Batería de Cuestionarios para las Consecuencias Afectivas del Cambio (BQACC)	512 estudiantes de entre 16 a 17 años de edad; 183 mujeres y 295 hombres.
	Análisis Factorial Confirmatorio (CFA), Verificación de valores atípicos, evaluación del apalancamiento (valores del sombrero), y prueba de normalidad (Estadísticas de Mardia), etc.	Pruebas Nacionales de Logro en inglés (NEATs) del Ministerio de Educación de Irán.	
Yan y Staples (2019)	El muestreo fue por conglomerados.		600 estudiantes, con 5 excluidos por la tipología del ensayo.
	Estudio sobre el desempeño en la sección de escritura del ECPE utilizando el análisis multidimensional (MD) para evaluar elementos léxico-gramaticales en función de las necesidades comunicativas de la tarea.	Examen ECPE (Examination for the Certificate of Proficiency in English).	
	Análisis de las puntuaciones: por correlación y análisis MANOVA.	Corpus ECPE.	

AUTOR(ES)	MÉTODO		
	TÉCNICA	INSTRUMENTO	PARTICIPANTES
Fechter et al. (2021)	La investigación se enfoca en el desarrollo y validación de una herramienta de medición, la IMRB (Inference-Making and Reasoning in Biology), diseñada para evaluar las habilidades de los estudiantes universitarios en hacer inferencias y razonamiento en biología. La IMRB utiliza párrafos cortos de contenido, generalmente enseñado al final de un semestre de cursos de biología, y pregunta sobre este contenido a través de 15 preguntas de opción múltiple.	IMRB: Un conjunto de 15 a 30 ítems de opción múltiple basados en párrafos de contenido de biología, cuyas respuestas correctas e incorrectas (distractores) se derivan de inferencias correctas e incorrectas identificadas en estudios previos con estudiantes.	Estudiantes de biología y bioquímica en universidades y colegios de 4 años en Estados Unidos. Los datos se recopilaron durante una década (2008-2018) y participaron 4,688 estudiantes de dos universidades.
Jawhar et al. (2021)	El estudio se centra en la validación de un test de ingreso al inglés en tres fases: antes, durante y después del examen. Expertos evaluaron la relevancia y precisión del contenido, mientras que se garantizó una administración uniforme y segura del test. Posteriormente, se analizaron los datos y la percepción de los candidatos para confirmar la validez del test.	King Saud bin Abdulaziz University Entrance Test (KSAU-ET): Consiste en secciones de escucha, vocabulario, gramática y comprensión lectora. Se formó un comité ejecutivo para desarrollar un test de ingreso al inglés.	Participaron cuatro miembros del comité ejecutivo del departamento de inglés, ocho expertos en la materia, cuatro asistentes docentes y 474 estudiantes que participaron como examinados. Un comité separado revisó las solicitudes y documentos de los estudiantes.
Huang et al. (2020)	TOEFL JuniorVR, junto con tareas de monólogo y de interacción desarrolladas por los investigadores, para evaluar las habilidades de expresión oral en estudiantes de inglés como lengua extranjera.	TOEFL JuniorVR Speaking test.  Pruebas de monólogo e interacción desarrolladas por los investigadores.  Encuesta para maestros.  Encuesta para estudiantes.	252 estudiantes de inglés como lengua extranjera, entre séptimo y noveno grado, en Taipei, Taiwán. La mayoría eran adolescentes con edades promedio de 14 años.
Chuang y Yan (2022)	El estudio se basó en la administración del EPT, una prueba diseñada para evaluar habilidades de escritura y expresión oral en estudiantes internacionales.  Para el análisis de datos, se utilizó el modelo de Toulmin para identificar y analizar la estructura argumentativa presente en los ensayos.	English Placement Test (EPT): una prueba diseñada para evaluar las habilidades de escritura y expresión oral de los estudiantes internacionales recién admitidos para determinar su ubicación en cursos de ESL.	150 estudiantes de pregrado que tomaron la EPT.
Seipel et al. (2023)	Análisis de la Curva ROC (Receiver-Operating Characteristic): Con el objetivo de identificar un umbral preliminar de rendimiento en MOCCA-College que pudiera categorizar el rendimiento como buena o mala comprensión.  Estadística del AUC (Area Under the Curve). Análisis descriptivo: Se resumieron demográficamente los datos y las puntuaciones subscore de MOCCA-College, en general y por tipo de comprensor.  Correlaciones: Para evaluar las evidencias de validez del uso previsto de MOCCA-College, correlacionaron las puntuaciones subscore de MOCCA-College con otras medidas disponibles.	MOCCA-College: Una tarea basada en laberintos diseñada para identificar estudiantes universitarios que tienen problemas con la comprensión causal. Cada forma de prueba consiste en 50 ítems: textos narrativos o expositivos cortos.  Encuesta demográfica.  Datos de admisión.  Nelson-Denny Reading Test (NDRT).	Los participantes fueron reclutados de tres universidades comprensivas regionales medianas a grandes ubicadas en regiones geográficamente distintas de los Estados Unidos (Oeste, Medio Oeste y Sureste). Se incluyeron 988 participantes para el análisis. De estos, 377 proporcionaron datos de ACT y 192 proporcionaron datos de SAT.

AUTOR(ES)	MÉTODO		
	TÉCNICA	INSTRUMENTO	PARTICIPANTES
	Análisis de varianza (ANOVAs): MOCCA-College entre buenos y malos comprensivos; y reportaron tamaños de efecto para cada comparación utilizando Hedges' g.		

*Nota: Elaboración a partir del análisis realizado*

Además, es posible destacar tres puntos positivos del Enfoque Basado en Argumentos; primero, permite un desarrollo natural de la evidencia de las interpretaciones y usos previstos de las puntuaciones; a diferencia de Schilling (2004), Sireci (2007; 2014) y Mislevy et al. (2003), se concluye que efectivamente da pie a ajustarse a las necesidades de la prueba. Segundo, permite establecer un esbozo de la investigación; como en el caso de Fechter et al. (2021), quienes desarrollaron su propio diagrama. Tercero, estructura los resultados de la investigación, permitiendo evidenciar cada inferencia con su respectiva evidencia o evidencias. Y según la revisión de autores, ambas ideas siguen prevaleciendo en cada uno de los artículos consultados para esta RSL. Suelen citar a Kane cuando sí se centran en este Enfoque y citan a Messick como un complemento teórico y para introducir al lector en la temática de la validación y validez.

Las investigaciones más relevantes de un método explícito y organizado para la validación de pruebas de ingreso a la universidad basándose en el desarrollo de argumentos son, por un lado, el estudio de Ginther y Elder (2014), que realizaron un análisis comparativo de pruebas ampliamente reconocidas, como el TOEFL iBT, IELTS y PTE, proporcionando una visión detallada de la estructura y eficacia de estas evaluaciones en un contexto universitario. Por otro lado, Esfandiari et al. (2018) que abordaron la validez del TOEFL iBT a través de un análisis sistemático del contenido y una encuesta detallada, lo que resalta la importancia de recopilar opiniones directas sobre el propósito y contenido del examen. Además, Aviad-Levitzky, Laufer y Goldstein (2019) ofrecen un enfoque más técnico, utilizando diversas técnicas de análisis, como el coeficiente alfa de Cronbach, para examinar un test empleado en el proceso de admisión universitaria en Israel, lo que refleja la adaptabilidad y relevancia de estas técnicas en diferentes contextos culturales y académicos. Jawhar et al. (2021) se centraron específicamente en el proceso de validación de un examen de ingreso al inglés, lo que evidencia la meticulosidad requerida para garantizar la validez de pruebas lingüísticas en un entorno académico. Por último, el enfoque adoptado por Seipel et al. (2023) en biología y las técnicas utilizadas, como el análisis de la Curva ROC, demuestran cómo se pueden adaptar métodos específicos de validación a diferentes disciplinas académicas.

Hay que enfatizar en la importancia de realizar estudios empíricos de validación que involucren a múltiples partes interesadas y utilicen diferentes métodos para garantizar la validez de las interpretaciones y usos de las puntuaciones de las pruebas (Im et al., 2019). Por ello, concluimos que el

enfoque de Kane es un método para el desarrollo de supuestos vinculados a inferencias que finalmente respaldan las interpretaciones y usos previstos de la medida en estudio (Chapelle y Enright, 2010). En este sentido, se puede constatar que poco a poco habrá más estudios empíricos que expliciten el Enfoque Basado en Argumentos, por ende, se espera que estos 17 artículos apoyen en investigaciones futuras. Este enfoque proporciona a los investigadores una gran flexibilidad para determinar el argumento que desean hacer basado en el contexto y propósito de la prueba, como pudimos observar en la variedad de estudios con este enfoque (Addey et al., 2020; Kane, 2013).

## 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.
- Asociación Nacional de Universidades e Instituciones de Educación Superior. (2023, 23 de abril). Anuarios estadísticos de educación superior. Ciclo escolar 2021-2022 y 2022-2023 de ANUIES. <http://www.anui.es.mx/informacion-y-servicios/informacion-estadistica-de-educacion-superior/anuario-estadistico-de-educacion-superior>
- Aviad-Levitzky, T., Laufer, B., & Goldstein, Z. (2019). The new computer adaptive test of size and strength (CATSS): Development and validation. *Language Assessment Quarterly*, 1–24. <https://doi.org/10.1080/15434303.2019.1649409>
- Bachman, L., & Palmer, A. (2010). *Language assessment in practice*. Oxford University Press.
- Barrios, G., Santamaría, W., Navarro, C., & Berlanga, K. (2015). Validez del Exani-II en estudiantes de Ingeniería Industrial del Instituto Tecnológico de Ciudad Valles. *Revista TECTZAPIC*, 1(2). <http://www.eumed.net/rev/tectzapic/2015/02/exani-II.html>
- Beigman Klebanov, B., Ramineni, C., Kaufer, D., Yeoh, P., & Ishizaki, S. (2017). Advancing the validity argument for standardized writing tests using quantitative rhetorical analysis. *Language Testing*, 026553221774075. <https://doi.org/10.1177/0265532217740752>
- Chapelle, C., Enright, M., & Jamieson, J. (2008). *Building a validity argument for the Test of English as a Foreign Language*. Routledge.
- 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.
- Chapelle, C. (2012). Validity argument for language assessment: The framework is simple.... *Language Testing*, 29(1), 19–27.
- Chuang, P., & Yan, X. (2022). An investigation of the relationship between argument structure and essay quality in assessed writing.
- Cizek, G. J., Kosh, A. E., & Toutkoushian, E. K. (2018). Gathering and evaluating validity evidence: The generalized assessment alignment tool. *Journal of Educational Measurement*, 55(4), 477–512. <https://doi.org/10.1111/jedm.12189>

- Covay Minor, E., Porter, A. C., Murphy, J., Goldring, E. B., Cravens, X., & Elloitt, S. N. (2014). A known group analysis validity study of the Vanderbilt Assessment of Leadership in Education in US elementary and secondary schools. *Educational Assessment, Evaluation and Accountability*, 26(1), 29–48. <https://doi.org/10.1007/s11092-013-9180-z>
- Esfandiari, M., Riasati, M., Vaezian, H., & Rahimi, F. (2018). A quantitative analysis of TOEFL iBT using an interpretive model of test validity. *Language Testing in Asia*, 8(1).
- Fan, J. (2014). Chinese test takers' attitudes towards the Versant English Test: A mixed-methods approach. *Language Testing in Asia*, 4(1). <https://doi.org/10.1186/s40468-014-0006-9>
- Fechter, T., Dai, T., Cromley, J. G., Nelson, F. E., Van Boekel, M., & Du, Y. (2021). Developing a validity argument for an inference-making and reasoning measure for use in higher education. *Frontiers in Education*, 6(1). <https://doi.org/10.3389/feduc.2021.727539>
- García-Peñalvo, F. J. (2022). Desarrollo de estados de la cuestión robustos: Revisiones Sistemáticas de Literatura. *Education in the Knowledge Society (EKS)*, 23, e28600. <https://doi.org/10.14201/eks.28600>
- Ginther, A., & Elder, C. (2014). A comparative investigation into understandings and uses of the "TOEFL iBT"® test, the International English Language Testing Service (Academic) test, and the Pearson Test of English for graduate admissions in the United States and Australia: A case study of two university contexts. "TOEFL iBT"® Research Report. TOEFL iBT-24. *ETS Research Report*. RR-14-44. 1-39. <https://doi.org/10.1002/ets2.12037>
- Huang, B. H., Bailey, A. L., Sass, D. A., & Shawn Chang, Y. (2020). An investigation of the validity of a speaking assessment for adolescent English language learners. *Language Testing*, 026553222092573. <https://doi.org/10.1177/0265532220925731>
- Im, G.-H., Shin, D., & Cheng, L. (2019). Critical review of validation models and practices in language testing: Their limitations and future directions for validation research. *Lang. Test. Asia*, 9(14). <https://doi.org/10.1186/s40468-019-0089-4>
- Jawhar, S., Al, M., Alhawsawi, S., & Alkushi, A. (2021). Validating English language entrance test at a Saudi university for health sciences. *Arab World English Journal (AWEJ)*, 12(2), 49-71. <https://dx.doi.org/10.24093/awej/vol12no2.4>
- Jiménez, Ó. (2021). Los exámenes de ingreso a la Educación Superior Mexicana, un obstáculo a la aspiración de educación inclusiva para estudiantes pobres y de pueblos originarios. *Espacio I+D, Innovación más desarrollo*, 10(25). <https://doi.org/10.31644/IMASD.26.2021.a07>

- Jorion, N., Gane, B. D., James, K., Schroeder, L., DiBello, L. V., & Pellegrino, J. W. (2015). An analytic framework for evaluating the validity of concept inventory claims. *Journal of Engineering Education*, *104*(4), 454–496.
- Kane, M. T. (2001). Current concerns in validity theory. *Journal of Educational Measurement*, *38*(4), 319–342.
- Kane, M. (2002). Validating high stakes testing program. *Educational Measurement: Issues and Practice*, *21*(1), 31–41.
- Kane, M. (2006 a). 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. T. (2006 b). Validation. In R. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17–64). Westport, CT: American Council on Education/Praeger.
- Kane, M. (2011). Validating score interpretations and uses. *Language Testing*, *29*(1), 3–17. <https://doi.org/10.1177/0265532211417210>
- Kane, M. (2012). Validating score interpretations and uses. *Language Testing*, *29*(1), 3–17.
- Kane, M. (2013). The argument-based approach to validation. *School Psychology Review*, *42*(4), 448–457.
- Kane, M. (2016). *Validity arguments for educational measurement*. Routledge.
- Kerlinger, F., & Lee, H. (2001). *Investigación del comportamiento: métodos de investigación en ciencias sociales*. McGraw Hill.
- LaFlair, G. T., & Staples, S. (2017). Using corpus linguistics to examine the extrapolation inference in the validity argument for a high stakes speaking assessment. *Language Testing*, *34*(4), 451–475. <https://doi.org/10.1177/0265532217713951>
- Lane, S. (2014). Validity evidence based on testing consequences. *Psicothema*, *26*(1), 127–135.
- Lavery, M., Bostic, J., Kruse, L., Krupa, E., & Carney, M. (2020). Argumentation surrounding argument-based validation: A systematic review of validation methodology in peer-reviewed articles. *Educational Measurement: Issues and Practice*. <https://doi.org/10.1111/emip.12378>
- Li, S. (2018). Developing a test of L2 Chinese pragmatic comprehension ability. *Language Testing in Asia*, *8*(1). <https://doi.org/10.1186/s40468-018-0054-7>

- Messick, S. (1989). Validity. En R. L. Linn (Ed.), *Educational measurement* (3.<sup>a</sup> ed., pp. 13–103). Macmillan.
- Mislevy, R., Steinberg, L., & Almond, R. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives*, 1, 3–62.
- Morales, R., Barrera, A., & Garnett, E. (2015). *Validez predictiva y concurrente del EXANI-II en la Universidad Autónoma del Estado de México*. X Congreso Nacional de Investigación Educativa, sujetos de la educación. [https://www.comie.org.mx/congreso/memoriaelectronica/v10/pdf/area\\_tematica\\_16/ponencias/0701-F.pdf](https://www.comie.org.mx/congreso/memoriaelectronica/v10/pdf/area_tematica_16/ponencias/0701-F.pdf)
- Newton, P., & Shaw, S. (2014). *Validity in educational & psychological assessment*. SAGE.
- Padilla, J.-L., & Benítez, I. (2014). Validity evidence based on response processes. *Psicothema*, 26(1), 136–144.
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). Declaración PRISMA 2020: Una guía actualizada para la publicación de revisiones sistemáticas. *Revista Española de Cardiología*, 74(9), 790–799. <https://doi.org/10.1016/j.recesp.2021.06.016>
- Reeves, T., & Marbach-Ad, G. (2016). Contemporary test validity in theory and practice: A primer for discipline-based education researchers. *CBE—Life Sciences Education*, 15(1).
- Rethlefsen, M. L., Kirtley, S., Waffenschmidt, S., Ayala, A. P., Moher, D., Page, M. J., Koffel, J., & PRISMA-S Group. (2021). PRISMA-S: an extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews. *Systematic Reviews*, 10(39). <https://doi.org/10.1186/s13643-020-01542-z>
- Rix, S. (2012). Construct validity in formative assessment: Purpose and practices. *Journal on English Language Teaching*, 2(2), 1–6. <https://eric.ed.gov/?id=EJ1070202>
- 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)
- Saldaña, J. M. (2015). *The coding manual for qualitative researchers* (3.<sup>a</sup> ed.). SAGE Publications.



- Schilling, S. G. (2004). Conceptualizing the Validity Argument: An Alternative Approach. *Measurement: Interdisciplinary Research and Perspectives*, 2(3), 178–182.
- Seipel, B., Kennedy, P. C., Carlson, S. E., Clinton-Lisell, V., & Davison, M. L. (2023). MOCCA-College: Preliminary Validity Evidence of a Cognitive Diagnostic Reading Comprehension Assessment. *Journal of Learning Disabilities*, 56(1), 58–71. <https://doi.org/10.1177/00222194221121340>
- ShayesteFar, P. (2020). A model of interplay between student English achievement and the joint affective factors in a high-stakes test change context: model construction and validity. *Educational Assessment, Evaluation and Accountability*. <https://doi.org/10.1007/s11092-020-09326-8>
- Shepard, L. A. (1997). The centrality of test use and consequences for test validity. *Educational Measurement: Issues and Practices*, 16(2), 5–24.
- Sireci, S. (2007). On validity theory and test validation. *Educational Researcher*, 36, 477–481.
- Sireci, S., & Faulkner-Bond, M. (2014). Validity evidence based on test content. *Psicothema*, 26(1), 100–107.
- Staples, S., Laflair, G. T., & Egbert, J. (2017). Comparing Language Use in Oral Proficiency Interviews to Target Domains: Conversational, Academic, and Professional Discourse. *The Modern Language Journal*, 101(1), 194–213. <https://doi.org/10.1111/modl.12385>
- Yan, X., & Staples, S. (2019). Fitting MD analysis in an argument-based validity framework for writing assessment: Explanation and generalization inferences for the ECPE. *Language Testing*. <https://doi.org/10.1177/0265532219876226>

### CÓMO CITAR:

- Ruiz Mendoza, K.K., Pedroza Zúñiga, L.H., & López, A.Y. (2024). RSL con prisma: métodos para el proceso de validación de exámenes de ingreso a la universidad desde el enfoque basado en argumentos. *Revista de Investigación y Evaluación Educativa*, 11(1), 43-64. <https://doi.org/10.47554/revie.vol11.num1.2024.pp43-64>