

Enfoque Basado en Argumentos: una nueva estructura argumentativa

Karla Karina Ruiz Mendoza

IIDE-UABC

ruiz.karla32@uabc.edu.mx

Luis Horacio Pedroza Zúñiga

IIDE-UABC

horacio.pedroza@uabc.edu.mx

Área temática: Evaluación del aprendizaje y desempeño escolar

Resumen

El campo de la evaluación educativa y psicológica ha evolucionado hacia un enfoque más profundo en la validación de pruebas, marcado significativamente por las contribuciones de Michael Kane y Carol Chapelle. Kane introdujo un cambio paradigmático al centrarse en cómo se interpretan y utilizan los resultados de las pruebas, más allá de la simple correlación estadística. Carol Chapelle extendió estos conceptos al ámbito de las pruebas de idiomas, enfatizando que la validez debe considerarse como el grado en que la evidencia y la teoría respaldan los usos específicos de los puntajes de las pruebas. Su enfoque, que aplica un argumento estructurado para validar pruebas, no solo mejora la precisión de las interpretaciones de los puntajes, sino que también asegura que las decisiones educativas y psicológicas sean justas y pertinentes en diversos contextos culturales y educativos.

Palabras clave: Evaluación educativa, validación de pruebas, enfoque basado en argumentos.

Justificación

El campo de la evaluación educativa y psicológica ha evolucionado significativamente en las últimas décadas, con un enfoque creciente en la validación de pruebas basada en argumentos. Esta evolución ha sido influenciada notablemente por las contribuciones de Michael Kane (2006, 2013), quien propuso un cambio paradigmático del "qué" de las pruebas al "cómo" de la interpretación y uso de los resultados de las pruebas. Este cambio ha llevado a un entendimiento más profundo y estructurado de la validación, centrando la atención en las inferencias y las suposiciones subyacentes en el uso de los puntajes de las pruebas.

Carol Chapelle (2021) ha extendido estos avances al contexto específico de las pruebas de idioma inglés, argumentando que la validez debe ser entendida no como una propiedad inherente de una prueba, sino como el grado en que la evidencia y la teoría respaldan las interpretaciones y usos de los puntajes de las pruebas para propósitos específicos. Sus trabajos, que aplican el Enfoque Basado en Argumentos al campo de la evaluación de idiomas, han proporcionado un modelo detallado para la construcción de argumentos de validez que conectan la teoría y la práctica de manera coherente y justificada.

La justificación para adoptar este enfoque en la investigación actual se basa en la necesidad de una validación rigurosa que trascienda la simple correlación estadística entre puntuaciones y se centre más en la interpretación lógica y la justificación teórica de los usos de los resultados de las pruebas. Este enfoque es crucial para garantizar que las decisiones educativas y psicológicas basadas en resultados de pruebas sean no solo precisas sino también justas y pertinentes en diversos contextos educativos y culturales. Este proyecto de investigación plantea las siguientes preguntas:

- ¿Cómo puede el Enfoque Basado en Argumentos mejorar la coherencia y relevancia de las interpretaciones de los puntajes de pruebas de idioma inglés?
- ¿De qué manera los argumentos de validez propuestos por Carol Chapelle pueden ser aplicados o adaptados en diferentes contextos lingüísticos y culturales para mejorar la calidad y la equidad de las evaluaciones?

Este estudio pone al Enfoque Basado en Argumentos, tal como lo desarrolló Carol Chapelle en el campo de las pruebas de idiomas, como un enfoque que proporciona un marco robusto y adaptable para la validación de pruebas que puede mejorar significativamente la precisión y la utilidad de las interpretaciones de los puntajes de las pruebas. El objetivo es validar esta hipótesis

a través de un estudio empírico que examine la aplicación del enfoque en diversos contextos educativos y evalúe su impacto en la calidad de las decisiones educativas y psicológicas. Lo cual ofrece directrices prácticas para los desarrolladores de pruebas y los responsables en la toma de decisiones para mejorar la validación y el uso de pruebas educativas y psicológicas.

Enfoque conceptual

En este sentido, se parte de 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), quienes definen 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).

De ahí que estudio de la validez en las pruebas educativas y psicológicas ha experimentado una transformación significativa debido a la influencia de teóricos como Michael Kane y Carol Chapelle. El enfoque basado en argumentos, especialmente refinado por Kane (2006), ha cambiado la perspectiva tradicional de la validez de pruebas, pasando de una simple correlación estadística a un análisis más profundo y estructurado de las inferencias y supuestos subyacentes a los resultados de las pruebas.

Resultados y Conclusiones

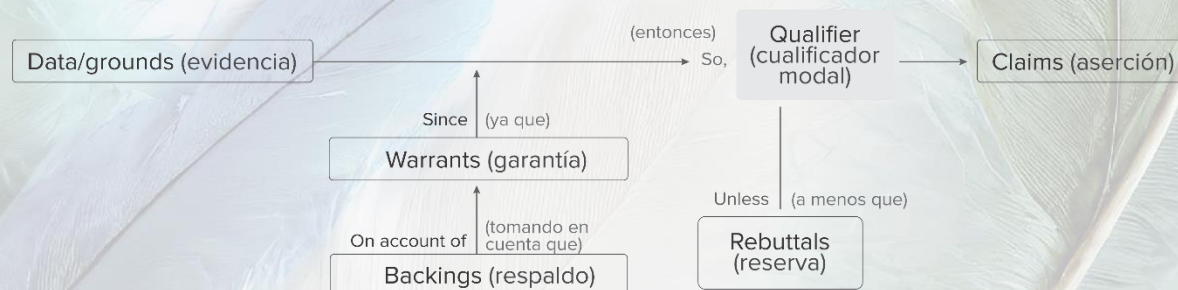
La definición de validez de los Estándares, que se entiende como el grado en que la evidencia y la teoría respaldan las interpretaciones de los resultados de las pruebas para los usos propuestos de esas puntuaciones (AERA, APA & NCME, 2014) sitúa el uso de la prueba en un papel central en la validación. Esta centralidad es el resultado lógico de una evolución histórica. Cureton (1951) quería que se proporcionara un consejo sólido sobre la validez para usos particulares. Cronbach (1971) y Messick (1989) incluyeron la utilidad y relevancia de las puntuaciones de las pruebas y las consecuencias sociales de su uso en sus marcos de validez. Aunque hay matices sobre el papel del uso y las consecuencias en la validación, la afirmación de los Estándares captura la visión principal, a tal punto que algunos investigadores han

reemplazado el término argumento de validez por argumento de uso de la evaluación (Chapelle, 2021).

Kane (2006, 2011, 2013) introduce dos tipos de argumentos. El primero, basado en la lógica del modelo de Toulmin (1958), es el Argumento de Interpretación o Uso (IUA, por sus siglas en inglés), que delinea las afirmaciones a evaluar, es decir, las suposiciones vinculadas a las interpretaciones y usos de las puntuaciones de los instrumentos de medición. El segundo, el argumento de validez, detalla la interpretación y uso propuesto de las puntuaciones, acompañado de cuatro tipos de inferencias que se abordarán más adelante (Taylor, 2013), y en parte, se basa también en el modelo de Toulmin (1958) (véase figura 2). En esencia, hay dos interpretaciones propuestas por los expertos: una centrada en el significado de las puntuaciones de la prueba y otra en la aplicación de los resultados de la prueba (Taylor, 2013; Carrillo, Sánchez y Leenen, 2020).

Kane (2011) establece tres criterios esenciales para la interpretación de argumentos que son cruciales al abordar inferencias. El primero es la "claridad del argumento", donde es fundamental especificar tanto la garantía como el respaldo. El segundo es la "coherencia del argumento", en el que el razonamiento debe ser lo suficientemente persuasivo. Finalmente, la "plausibilidad (o verosimilitud) de las inferencias y supuestos", que puede fundamentarse en hipótesis generalmente aceptadas, en análisis documentados y profundos o a través de evidencia empírica. De esta manera, Kane emplea el modelo de Toulmin (consulte la Figura 1) para consolidar sus aseveraciones y argumentaciones, garantizando coherencia e interpretación adecuadas.

Figura 1. Modelo de Toulmin



Nota. Retomado de Toulmin (1958) con las traducciones al español por Rodríguez (2004).

El argumento interpretativo especifica las inferencias implicadas en pasar de las actuaciones observadas a las conclusiones a las que se llega y a las decisiones que se toman en

base a las puntuaciones de los exámenes. Incluiría una red o cadena de inferencias. El argumento de validez proporcionaría una evaluación crítica de la coherencia del argumento interpretativo y de las justificaciones y apoyos para las inferencias en este argumento. (Kane, 2011, p.12; traducción propia)

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), en la Tabla 1 se especifican cada uno de ellos. Para que finalmente, se recopilen y evalúen las evidencias que respalden o contradigan cada una de las suposiciones e inferencias planteadas.

Tabla 2.5.1. *Las inferencias de Kane*

Inferencia	Refiere a	Garantía o procedimientos a seleccionar
Puntuación (scoring)	Se toman información de los datos de los puntajes como afirmación, se establecen criterios y reglas.	Reglas o rúbricas de puntuación. Estandarización de puntajes.
Generalización (generalization)	La puntuación observada del evaluado nos da una estimación de la puntuación del universo evaluado.	Juicio de grupos de expertos. Teoría de la generalizabilidad (teoría G). Tamaño de la muestra y cantidad de preguntas.
Extrapolación (extrapolation)	Estimación de la función de la prueba en el contexto real.	Análisis entre la relación de la prueba y su función en otros contextos.
Implicaciones (Implications)	Interpretación y toma de decisiones	Ecuación de regresión. Aprobación o no aprobación del estándar. Acciones. Consecuencias voluntarias o involuntarias.

Nota. Elaboración propia a partir de Kane (2001, 2013); Carrillo, Sánchez y Leenen, 2020.

Estas cuatro inferencias recopilan diferentes evidencias que se aportan a la validez. Cook, et al., (2015) lo describe como: primero hay que realizar la observación singular del puntaje, ya sea que éste sea por pregunta de opción múltiple o portafolio de evidencias u otro tipo de observación, después se usan estos puntajes para generar un puntaje general (generalización), seguido de ello se procede a la extrapolación, es decir, trazar las inferencias a partir de los

puntajes obtenidos y lo que podrían implicar en la vida real, por último, analizar las implicaciones a través interpretar toda esta información para la toma de decisiones.

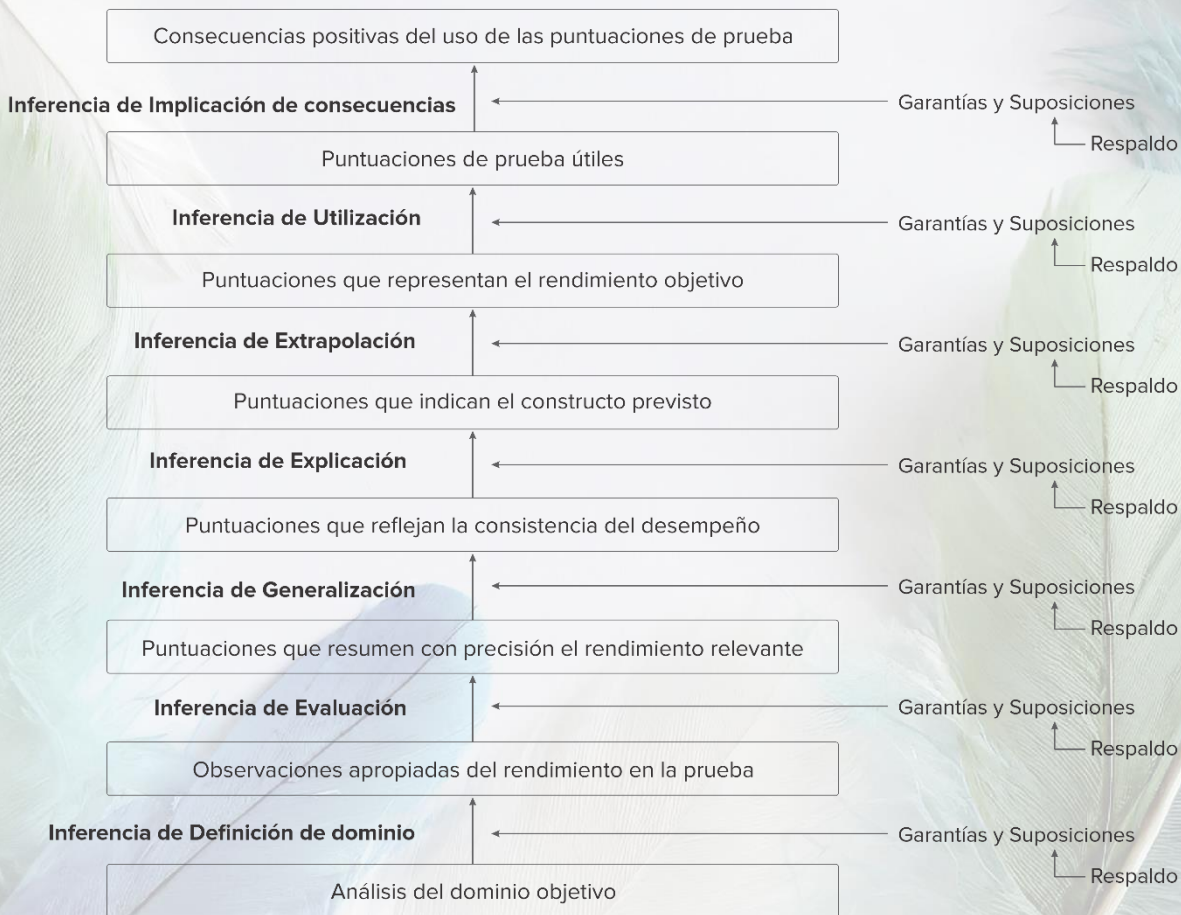
En este sentido, entendemos que la validez parte de comprender la coherencia entre la evidencia, la teoría y la interpretación del instrumento a revisar, además cuando hablamos de validez debemos de atender a las evidencias que tenemos disponibles para fortalecer el argumento y defender la coherencia de las intenciones sobre la práctica. Lo cual significó que el Enfoque Basado en Argumentos es pertinente debido a la evolución del concepto mismo.

Carol Chapelle es quien más se ha acercado a desarrollar este enfoque en las pruebas del idioma inglés, es nuestro referente más cercano sobre el uso de las pruebas. En su último texto (Chapelle, 2021), como se ha mencionado anteriormente, propone un manual completo sobre este enfoque y añade que, aunque tradicionalmente se ha percibido la validez como una cualidad inherente a una prueba en sí misma, los Estándares articulan que es, de hecho, una medida de cuán bien la evidencia y la teoría respaldan las interpretaciones y los usos previstos de los puntajes de las pruebas. En lugar de ver la validación simplemente como una correlación entre dos conjuntos de puntajes de pruebas, los Estándares (AERA, APA & NCME, 2014) enfatizan un enfoque integral. Esto implica declarar explícitamente la interpretación de los puntajes de las pruebas y validarla utilizando evidencia empírica, revisiones literarias o análisis lógicos (Chapelle, 2021). La esencia es que la validación es similar a la prueba de hipótesis científica, incorporando tanto datos empíricos como juicio experto (Chapelle, 2021).

Según Chapelle (2021), basándose en Kane, el proceso de desarrollo de un argumento de validez se aborda para dos contingencias: una prueba existente y una nueva. Por lo que es necesario: la formulación del argumento de interpretación y uso, realización de la investigación de validación e integración de resultados en un argumento de validez. La lógica detrás de los argumentos de validez se basa en conectar diferentes afirmaciones a través de inferencias, tal como se ilustra en la Figura 2. Cada inferencia lleva a una afirmación específica sobre las puntuaciones de la prueba. Por ejemplo, las inferencias de evaluación y generalización abarcan aspectos de la confiabilidad. Las inferencias de explicación y extrapolación abordan cuestiones similares a las investigadas en la validez constructiva. Estas inferencias se conectan de manera que cada conclusión de una sirve como base para la siguiente. Además, se puede ver cómo las puntuaciones precisas de una prueba son el punto de partida para cinco inferencias, pero también son el resultado de la inferencia de evaluación. Esta, a su vez, se basa en observaciones apropiadas del desempeño en la prueba.

Las puntuaciones precisas también son la base para la inferencia de generalización, que conduce a la afirmación de que las puntuaciones reflejan el desempeño en el dominio objetivo. Los argumentos de validez se pueden complicar o limitar mediante refutaciones, que señalan limitaciones o errores en las inferencias (Chapelle, 2021). Si una refutación es respaldada por evidencia, el argumento de validez no puede hacerse o debe especificarse solo para grupos no incluidos en la refutación. En este sentido, los argumentos de validez deben perfeccionarse en función de los datos.

Figura 2. Diagrama esquemático de la estructura general de un argumento de validez que contiene siete afirmaciones e inferencias



Nota. Recuperado de Chapelle (2021).

Un ejemplo práctico puede ser el de un examen estandarizado de matemáticas para estudiantes de secundaria. Para validar la prueba, se debe seguir un proceso estructurado: primero, se define claramente qué habilidades matemáticas se quieren evaluar (Definición de

Dominio). Luego, se verifica que la prueba mida adecuadamente esas habilidades (Evaluación) y que las puntuaciones obtenidas sean consistentes en diferentes contextos o versiones del examen (Generalización). Se busca interpretar el significado real de las puntuaciones (Explicación) y prever cómo esos resultados se relacionarán con el rendimiento futuro en matemáticas (Extrapolación). A continuación, se considera la aplicación práctica de esos resultados, como determinar la admisión a cursos avanzados (Utilización), y finalmente se evalúan las consecuencias de esas decisiones, asegurando que sean justas y beneficiosas (Implicación de Consecuencias). Es decir, se debería seguir el siguiente proceso:

1. **Inferencia de Definición de Dominio:** Antes de administrar la prueba, los expertos en matemáticas y pedagogía deben definir cuál es el dominio objetivo que la prueba pretende medir. ¿Qué conceptos y habilidades matemáticas específicas de secundaria debería reflejar este examen? El desarrollo de la prueba se basaría en esta definición.
 2. **Inferencia de Evaluación:** Una vez definido el dominio y administrada la prueba, se evalúa si el examen en sí mide adecuadamente el rendimiento de los estudiantes en ese dominio específico. Para ello, se compararía el rendimiento de los estudiantes en el examen con su rendimiento en situaciones similares (como exámenes escolares o tareas de aula).
 3. **Inferencia de Generalización:** Aquí se evalúa si las puntuaciones del examen estandarizado reflejan de manera confiable el desempeño matemático del estudiante más allá de ese examen particular. Por ejemplo, si un estudiante obtiene una puntuación alta en esta prueba, ¿obtendría una puntuación similar en otra versión del examen o en otro contexto similar?
 4. **Inferencia de Explicación:** ¿Qué significan las puntuaciones exactamente? ¿Una puntuación alta indica una comprensión profunda de las matemáticas? ¿O la habilidad de memorizar fórmulas? Aquí se intenta interpretar el significado subyacente de las puntuaciones.
 5. **Inferencia de Extrapolación:** Si un estudiante tiene un buen rendimiento en el examen, ¿eso indica que también se desempeñará bien en contextos matemáticos futuros? Por ejemplo, ¿se espera que estos estudiantes obtengan buenas calificaciones en matemáticas avanzadas en la universidad?
 6. **Inferencia de Utilización:** Una vez que se ha interpretado y extrapolado el significado de las puntuaciones, se considera cómo se utilizarán esas puntuaciones en la práctica. Por
-

ejemplo, ¿se utilizarán para determinar la admisión a programas avanzados? ¿O tal vez para identificar a estudiantes que necesiten apoyo adicional?

- 7. Inferencia de Implicación de Consecuencias:** Por último, se considera el impacto de usar estas puntuaciones en decisiones prácticas. Si un estudiante no obtiene una puntuación lo suficientemente alta, ¿se le negará la entrada a un programa avanzado? ¿Qué impacto podría tener esto en su futura carrera académica y profesional? Aquí se evalúan las posibles consecuencias (positivas y negativas) de utilizar las puntuaciones de la prueba.

Finalmente, cada caso expresa sus problemáticas, por ello, para comenzar un proceso como este, es necesario ir analizando cada una de las inferencias y explicarlas de forma lógica y argumentada sin olvidarnos de los aspectos más complejos como la toma de decisiones y sus consecuencias. Asimismo, Chapelle (2021) dedica un apartado a los conflictos sobre la comunicación, ya propuestos por Kane (2006) de cómo se realiza el proceso de validación y qué es la validación; ya que en Estados Unidos cada grupo de investigadores varía en tamaño y responsabilidades. A esto le llama comunidades discursivas, por lo que proponer un vocabulario, como se colocó anteriormente, es justificado para establecer normas y unidad de ideas.

Otro tema, es el del rol de los validadores. Chapelle (2021), menciona que es necesario precisar los roles de los evaluadores, además es esencial considerar el contexto específico de las pruebas. Específicamente, ¿el evaluador es un miembro interno del equipo o un evaluador externo? ¿Se está creando la prueba como un nuevo proyecto o ya existe? Los evaluadores internos suelen estar centrados en desarrollar argumentos de validez para apoyar interpretaciones y usos de sus pruebas, mientras que los externos pueden tener un enfoque diferente.

En conclusión, el enfoque basado en argumentos de validez representa una evolución en la comprensión y evaluación de la validez de las pruebas. Este enfoque reconoce la importancia de establecer argumentos sólidos que respalden las interpretaciones y usos previstos de las puntuaciones de las pruebas. A través de una estructura lógica y sistemática, se establecen afirmaciones, inferencias y evidencias que conectan el proceso de evaluación y el contexto de aplicación. Este enfoque busca una coherencia integral entre la teoría, la evidencia y la interpretación de los resultados, aportando claridad y precisión en el proceso de validación. No obstante, es por ello que es necesario realizar revisiones sistemáticas que ayuden a extender el horizonte sobre el Enfoque Basado en Argumentos.

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.
- Carrillo, B.; Sánchez, M., & Leenen, I. (2020). El concepto moderno de validez y su uso en educación médica. *Investigación en Educación Médica*, 98-106.
<https://doi.org/10.22201/facmed.20075057e.2020.33.19216>
- Chapelle, C. A. (2021). The Argument-Based Approach to Validation. *Studies in Language Testing*.
- 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
- Cronbach, L. J. (1971). Test Validation. In R. Thorndike (Ed.), *Educational Measurement* (2nd ed., p. 443). Washington DC: American Council on Education
- Cureton, E. E. (1951). Validity. In E. F. Lindquist (Ed.), *Educational measurement* (pp. 621–694). Washington, DC: American Council on Education.
- Kane, M. (2006). Content-Related Validity Evidence in Test Development. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development* (pp. 131–153). Lawrence Erlbaum Associates Publishers.
- 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>
- Messick S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher* (18), 2, 5-11.
- Toulmin, S. (1958). *The uses of argument*. Cambridge University Press.
- Taylor, C. (2013). *Validity and Validation*. Oxford.
-