

<https://doi.org/10.69639/arandu.v13i3.2559>

La validación matemática como evidencia de comprensión en la era de la inteligencia artificial generativa

Mathematical validation as evidence of understanding in the age of generative artificial intelligence

Marco Vinicio Gutiérrez Montenegro

<https://orcid.org/0009-0004-0293-1578>

vgutierrez@itcr.ac.cr

Instituto Tecnológico de Costa Rica
Costa Rica – Cartago

Artículo recibido: (la fecha la coloca el Equipo editorial) - Aceptado para publicación:

Conflictos de intereses: Ninguno que declarar.

RESUMEN

La producción de soluciones matemáticas correctas ha sido asumida tradicionalmente como evidencia suficiente de comprensión matemática. La irrupción de la inteligencia artificial generativa, capaz de producir demostraciones y procedimientos correctos sin intervención del estudiante, vuelve insostenible ese supuesto y obliga a la educación matemática a preguntarse qué tipo de evidencia permite inferir comprensión matemática cuando la producción de soluciones puede externalizarse. Este artículo, de naturaleza teórica y conceptual, articula la epistemología de las matemáticas, la investigación sobre comprensión y prueba, y la teoría de la evaluación para fundamentar el constructo de validación matemática, entendido como la capacidad de interpretar, examinar, justificar, verificar, corregir y argumentar resultados matemáticos asumiendo responsabilidad epistémica sobre su aceptación o rechazo. Se sostiene que la validación constituye una práctica matemática de alto nivel cognitivo, ligada a la demostración, la argumentación y la construcción de significado, y que ofrece una evidencia de comprensión matemática más robusta que la producción. Se discuten sus implicaciones para la evaluación matemática y se delimitan las contribuciones del constructo al campo.

Palabras clave: educación matemática, comprensión matemática, validación matemática, argumentación matemática, inteligencia artificial generativa

ABSTRACT

The production of correct mathematical solutions has traditionally been assumed to be sufficient evidence of mathematical understanding. The emergence of generative artificial intelligence, capable of producing correct proofs and procedures with no involvement of the student, makes that assumption untenable and forces mathematics education to ask what kind of evidence allows mathematical understanding to be inferred when the production of solutions can be externalized.

This theoretical and conceptual article articulates the epistemology of mathematics, research on understanding and proof, and assessment theory in order to ground the construct of mathematical validation, understood as the capacity to interpret, examine, justify, verify, correct and argue about mathematical results while assuming epistemic responsibility for their acceptance or rejection. It is argued that validation is a mathematical practice of high cognitive demand, tied to proof, argumentation and the construction of meaning, and that it offers more robust evidence of mathematical understanding than production. Its implications for mathematics assessment are discussed and the contributions of the construct to the field are delineated.

Keywords: mathematics education, mathematical understanding, mathematical validation, mathematical argumentation, generative artificial intelligence

Todo el contenido de la Revista Científica Internacional Arandu UTIC publicado en este sitio está disponible bajo licencia Creative Commons Attribution 4.0 International. 

INTRODUCCIÓN

Durante más de un siglo, la práctica evaluativa ha descansado sobre una premisa rara vez enunciada de manera explícita porque parecía evidente, la de que el producto matemático elaborado por una persona (una respuesta correcta, un procedimiento bien ejecutado, una demostración completa) constituye evidencia fiable del aprendizaje que lo originó. Sobre ese supuesto se han edificado la evaluación sumativa y buena parte de la formativa, y de él depende la cadena inferencial que conduce desde la conducta observada hasta la atribución de comprensión. La irrupción de la inteligencia artificial generativa interpela esa premisa de un modo que no se reduce a un episodio más de mala conducta académica. Cuando un sistema puede producir, en segundos y con corrección formal, el mismo artefacto que durante décadas se tomó como prueba de aprendizaje, el problema deja de ser quién ha hecho la tarea y pasa a ser qué autoriza a inferir comprensión a partir de un producto.

El argumento de este trabajo es que esta situación constituye, ante todo, una crisis epistemológica de las evidencias de comprensión matemática y no meramente un problema de detección o de integridad. La distinción no es ociosa. Si el asunto fuese de integridad, bastarían respuestas de orden técnico o normativo, como detectores, exámenes presenciales o declaraciones de uso. Si, en cambio, lo que se ha roto es el vínculo inferencial entre producción y comprensión, entonces ninguna medida de control restituye por sí sola la validez de la inferencia, y la teoría de la evaluación debe revisar cuáles son las evidencias que permiten sostener afirmaciones sobre el aprendizaje matemático.

La pregunta que organiza este artículo es propia de la educación matemática y consiste en determinar qué tipo de evidencia permite inferir comprensión matemática cuando la producción de soluciones puede externalizarse a sistemas de inteligencia artificial generativa. El propósito es triple. En primer lugar, analizar críticamente la relación tradicional entre producción matemática y comprensión, y mostrar que el vínculo inferencial que la sostiene era ya frágil antes de la inteligencia artificial. En segundo lugar, fundamentar el constructo de validación matemática como práctica matemática y como forma específica de evidencia de comprensión. En tercer lugar, discutir sus implicaciones para la evaluación matemática y delimitar sus límites. El trabajo es de naturaleza conceptual y articula la epistemología del conocimiento matemático, la investigación sobre comprensión y prueba, y la teoría de la evaluación para construir el constructo y examinar su rendimiento explicativo.

La tesis que se defiende sostiene que la inteligencia artificial generativa no suprime la necesidad de evaluar el aprendizaje matemático, pero obliga a reconsiderar qué tipos de evidencia permiten inferir comprensión; en ese reajuste, la validación matemática emerge como evidencia de comprensión matemática particularmente robusta, porque integra interpretación, juicio crítico, justificación, verificación y responsabilidad epistémica frente a resultados producidos por agentes

humanos o artificiales. Antes de desarrollar la propuesta conviene advertir que el constructo no surge de la nada. Dialoga con la noción de juicio evaluativo y con la tradición de la cognición epistémica, y precisamente por ello una parte del esfuerzo consiste en deslindar qué hay de original en él.

DESARROLLO

La evaluación y las evidencias de aprendizaje

La teoría contemporánea de la evaluación se ha construido alrededor de una idea inferencial. Evaluar no es observar directamente el aprendizaje, que es inaccesible, sino recoger evidencias observables a partir de las cuales se infiere un estado cognitivo no observable. Esta lógica está presente desde las formulaciones seminales de la evaluación formativa, que entienden la evaluación como el proceso de obtener e interpretar evidencias para tomar decisiones sobre los próximos pasos del aprendizaje y la enseñanza (Black y Wiliam, 1998, 2009). Es la tradición de la validez entendida como inferencia, cuya formulación canónica concibe la validez no como una propiedad del instrumento, sino como un juicio integrado sobre el grado en que la evidencia y la teoría sostienen las interpretaciones y los usos de los resultados, e incluye entre sus aspectos las consecuencias de tales usos (Messick, 1989, 1995). Lo que se valida no es el instrumento en abstracto, sino el uso interpretativo de sus resultados y las decisiones que de él se derivan (Newton, 2007; Shepard et al., 2018).

En esta cadena inferencial, el producto del estudiantado ha ocupado un lugar central como evidencia. Sadler (1989) mostró que la evaluación formativa requiere que quien aprende disponga de una concepción de la meta, pueda comparar su desempeño actual con esa meta y actúe para reducir la distancia entre ambos; el producto es el sustrato observable sobre el que opera esa comparación. La evaluación auténtica radicalizó esta orientación al exigir tareas que reprodujesen los desempeños valiosos de la disciplina o de la práctica profesional, bajo el supuesto de que un desempeño complejo y contextualizado es mejor evidencia de competencia que una respuesta descontextualizada (Wiggins, 2020). En todos los casos, la inferencia evaluativa presupone una relación estable entre el artefacto producido y el proceso cognitivo que lo habría generado.

Conviene subrayar que esta relación nunca fue lógicamente necesaria, sino empíricamente plausible. La literatura sobre evaluación reconoció tempranamente que un mismo producto puede provenir de procesos cognitivos muy distintos (comprensión profunda, reproducción memorística, azar, ayuda externa) y que la inferencia desde el producto al proceso siempre involucra un riesgo (Brookhart, 2003; Stiggins, 2005). De ahí el énfasis posterior en hacer visible el razonamiento, en la retroalimentación dialógica y en la regulación del aprendizaje por parte de quien aprende (Panadero et al., 2018; Wiliam, 2011). Sin embargo, incluso en sus versiones más sofisticadas, la teoría siguió tratando el producto como evidencia privilegiada,

confiando en que las condiciones de producción (tiempo limitado, supervisión, tareas no anticipables) mantuviesen razonablemente acoplados artefacto y cognición.

La teoría respondió a este riesgo refinando la noción de validez. Validar un uso evaluativo dejó de entenderse como propiedad fija del instrumento para concebirse como una argumentación, que consiste en explicitar la cadena de supuestos que conduce desde el desempeño observado hasta la interpretación atribuida y en someter cada eslabón a posible refutación (Newton, 2007). Cuando uno de esos supuestos, el que vincula producto y comprensión, se vuelve insostenible, no se invalida un instrumento concreto, sino toda una clase de inferencias. Por eso el problema que plantea la inteligencia artificial generativa no se resuelve sustituyendo unas tareas por otras del mismo tipo, ya que lo que falla es el argumento de validez y no su implementación.

Esta confianza explica una tensión persistente. Torrance (2012) advirtió que la evaluación formativa podía deformarse en una rutina de cumplimiento orientada hacia lo evaluable y no hacia el aprendizaje, y se mostró que evaluación y aprendizaje podían operar con supuestos epistemológicos divergentes (Baird et al., 2017). La pregunta por qué cuenta como evidencia válida quedaba abierta. Lo que la inteligencia artificial generativa aporta no es una pregunta nueva, sino una situación en la que la respuesta tradicional, el producto como evidencia, resulta insostenible, porque las condiciones que acoplaban artefacto y cognición pueden eludirse de manera trivial (Bearman et al., 2024).

Si la evaluación es, en su núcleo, una empresa inferencial, toda transformación que afecte a la fiabilidad del paso del producto al proceso afecta al corazón de la teoría. Lo que está en cuestión es esa fiabilidad, y no la honestidad del estudiantado.

Comprensión matemática, justificación y validación

La epistemología de la educación matemática ha sostenido, desde perspectivas diversas, que producir un resultado correcto y comprenderlo son cosas distintas. La distinción clásica entre comprensión relacional (saber qué se hace y por qué) e instrumental (saber hacer sin saber por qué) mostró que el dominio procedimental puede coexistir con la ausencia de comprensión conceptual; la investigación posterior precisó esa relación como bidireccional e iterativa, pero confirmó que son dimensiones diferenciables del conocimiento matemático (Crooks y Alibali, 2014; Rittle-Johnson y Schneider, 2015). Comprender no es ejecutar correctamente, sino disponer de una red de conexiones que permite interpretar, relacionar y justificar lo que se hace (Michener, 1978).

Estas distinciones han sido precisadas por la investigación en educación matemática. Comprender matemáticamente supone, para Duval (2006), la capacidad de coordinar y convertir distintos registros de representación semiótica, de modo que la comprensión no se reduce a la ejecución en un único registro, sino que se manifiesta en el tránsito justificado entre ellos. La caracterización de la competencia matemática en hebras entrelazadas, entre las que el razonamiento adaptativo, esto es, la capacidad de pensar lógicamente, explicar y justificar, ocupa

un lugar central, apunta en la misma dirección (Kilpatrick et al., 2001). Y desde una perspectiva discursiva, aprender matemáticas equivale a participar en un discurso regido por reglas de aceptación compartidas, de manera que el significado matemático se construye en la actividad de justificar y validar, y no en la mera obtención de un resultado (Sfard, 2008).

Esta diferenciación se agudiza en el terreno de la prueba y la justificación. La actividad matemática no consiste solo en obtener resultados, sino en establecer por qué son verdaderos, esto es, en producir y evaluar justificaciones. Balacheff (2010) describió el tránsito entre conocer y probar como un proceso en el que el sujeto articula sus conocimientos en un discurso que satisface exigencias de validez compartidas; probar no es un añadido al saber, sino una forma de comprometerse con su verdad. Harel (2008, 2024) situó la justificación en el centro del aprendizaje matemático y distinguió esquemas de prueba de muy distinto valor epistémico, desde los autoritarios o simbólicos hasta los deductivos. De ahí que la cuestión educativa decisiva no sea si el estudiantado obtiene la respuesta, sino qué considera una razón para aceptarla.

La literatura sobre razonamiento y prueba ha insistido en que evaluar la validez de un argumento es una competencia matemática específica, distinta de la de generarlo (Stylianides y Ball, 2008). Comprender una demostración no equivale a poder reproducirla, pues implica seguir su estructura inferencial, reconocer su alcance y sus supuestos, y dictaminar si establece lo que pretende (Hamami y Morris, 2024). La discusión reciente ha mostrado que argumentación, justificación y prueba remiten a prácticas epistémicas diferenciadas cuya enseñanza exige hacerlas explícitas (Rycroft-Smith, 2023), y la filosofía de la práctica matemática refuerza esta idea al describir las matemáticas como una actividad de agentes que justifican, revisan y aceptan resultados conforme a normas situadas, y no como un repertorio de productos acabados (Hamami y Morris, 2020).

Esta concepción de la actividad matemática como ejercicio de sentido, y no como mera ejecución, encuentra apoyo en la tradición que sitúa la resolución de problemas, la metacognición y la construcción de significado en el núcleo del pensamiento matemático (Schoenfeld, 2016). La filosofía de la educación matemática ha llevado esta idea más lejos al sostener que el conocimiento matemático es, en su construcción y validación, una práctica social de justificación sometida a normas públicas, y no un cuerpo de verdades cuya posesión se demuestre exhibiendo el resultado correcto (Ernest, 2002, 2018). Si comprender es participar en esa práctica, la evidencia de comprensión ha de buscarse en la capacidad de justificar y juzgar, no en el artefacto aislado.

Estas tradiciones permiten reformular la competencia matemática en términos que no se agotan en la producción. Las competencias matemáticas incluyen razonar, argumentar, modelizar, representar y, de manera destacada, valorar la validez de afirmaciones y procedimientos, propios o ajenos (Niss y Højgaard, 2019). Desde la educación matemática crítica, además, la formación comprende la capacidad de interrogar las afirmaciones que se presentan como matemáticamente

fundadas (Skovsmose, 1994), de modo que los estilos de razonamiento promovidos en el aula configuran no solo lo que el estudiantado sabe hacer, sino lo que considera una razón legítima (Kollosche, 2021).

De este recorrido se desprende una conclusión relevante para la evaluación. Si comprender incluye, de modo constitutivo, la capacidad de juzgar la validez de un resultado y de hacerse responsable de su aceptación, entonces la evidencia más directa de comprensión no es el resultado producido, sino el ejercicio de esa capacidad de juicio. La epistemología matemática venía señalando esa distinción mucho antes de la inteligencia artificial; lo que esta hace es convertir una distinción teórica en una necesidad práctica ineludible, al disociar masivamente la producción del resultado de la comprensión que tradicionalmente se le suponía.

Inteligencia artificial generativa y crisis de las evidencias tradicionales

Los sistemas de inteligencia artificial generativa producen artefactos (textos, procedimientos, demostraciones, soluciones) que satisfacen con frecuencia los criterios formales mediante los cuales la evaluación reconocía la presencia de aprendizaje. No hace falta sobreestimar su fiabilidad para reconocer el problema. Basta con que el sistema produzca, en una proporción significativa de casos, resultados indistinguibles en su forma de los que se aceptaban como evidencia. Cuando esto ocurre, el producto deja de discriminar entre quien comprende y quien no, porque ambos pueden presentarlo. La cadena inferencial que iba del artefacto a la cognición del sujeto se interrumpe en su primer eslabón.

Es importante caracterizar con precisión la naturaleza del problema. No se trata, en lo esencial, de que el estudiantado pueda hacer trampa; la copia y la externalización del trabajo son tan antiguas como la evaluación misma. Lo distintivo es que el acoplamiento entre producto y proceso cognitivo, que las condiciones tradicionales de examen lograban sostener de manera imperfecta pero suficiente, puede disolverse de forma generalizada y a bajo coste. Los datos disponibles sobre uso y mal uso de estos sistemas en la educación superior sugieren que la cuestión no es marginal ni transitoria, y que las respuestas centradas exclusivamente en la prohibición o la detección resultan frágiles (Chirikov et al., 2026). Si el problema fuese de integridad, su solución sería de orden disciplinario; al ser de orden inferencial, su solución debe ser de orden teórico.

La consecuencia es que la inteligencia artificial generativa actúa como un revelador. Pone de manifiesto que la evidencia tradicional, el producto, nunca evidenció directamente la comprensión, sino que descansaba en un supuesto de acoplamiento que las circunstancias hacían razonable. Al volver ese supuesto insostenible, obliga a explicitar lo que la teoría había mantenido implícito. La comprensión se infiere y no se observa, y la calidad de la inferencia depende de qué evidencia se utilice. La pregunta pertinente deja de ser cómo impedir que el estudiantado use estos sistemas y pasa a ser qué evidencias permiten inferir comprensión cuando el producto ha dejado de hacerlo.

Esta reformulación es coherente con la respuesta que la propia teoría de la evaluación ha empezado a esbozar. Frente a la inteligencia artificial generativa, diversos autores han propuesto desplazar el foco desde la producción hacia la capacidad de juzgar la calidad del trabajo, propio y ajeno. En este escenario, el juicio evaluativo se vuelve una finalidad educativa de primer orden (Bearman et al., 2024). El presente trabajo asume esa orientación general, pero sostiene que, en el dominio matemático, requiere una especificación que la noción de juicio evaluativo, formulada para la calidad del trabajo en general, no proporciona por sí sola. Juzgar la calidad de un producto matemático no es solo apreciar su ajuste a un estándar, sino decidir, con responsabilidad epistémica, si un resultado es verdadero y por qué. Esa especificación es lo que aquí se denomina validación matemática.

Vacíos teóricos identificados

La revisión de la literatura permite localizar con precisión el vacío teórico que este trabajo aspira a ocupar. Existe, primero, un corpus consolidado sobre evaluación educativa que ha establecido su carácter inferencial y los criterios de validez de sus interpretaciones (Black y Wiliam, 2009; Newton, 2007; Shepard et al., 2018). Segundo, una literatura creciente examina el impacto de la inteligencia artificial generativa sobre la evaluación en la educación superior y reclama su reforma (Bearman et al., 2024; Chirikov et al., 2026). Tercero, el juicio evaluativo ha sido desarrollado con solidez como capacidad de discernir la calidad del trabajo propio y ajeno (Tai et al., 2018). Y, cuarto, la investigación en educación matemática ha caracterizado la demostración, la argumentación y la evaluación de pruebas como prácticas epistémicas diferenciadas (Hamami y Morris, 2024; Stylianides y Ball, 2008).

Ninguna de estas tradiciones, sin embargo, ha formulado de manera consolidada la validación matemática como una forma específica de evidencia de comprensión matemática. La literatura sobre evaluación no tematiza la naturaleza epistémica particular de la verdad matemática; la discusión sobre inteligencia artificial permanece en el plano de la integridad y la política institucional; el juicio evaluativo se define para la calidad del trabajo en general, y no para la verdad y su justificación; y la investigación sobre la prueba matemática no la conecta con la teoría de las evidencias de aprendizaje ni con la pregunta por la finalidad educativa. El vacío no reside en la ausencia de piezas, sino en la falta de su articulación. Identificarlo y ocuparlo es una tarea propia de la Teoría de la Educación, entendida no como mera aplicación de saberes provenientes de otras disciplinas, sino como un saber dotado de racionalidad propia, capaz de constituir y fundamentar sus propios objetos de conocimiento (Colom Cañellas, 2004; Touriñán López, 1987).

La validación matemática como constructo de la educación matemática

Se propone definir la validación matemática como la capacidad de interpretar, examinar críticamente, justificar, verificar, corregir y argumentar sobre resultados matemáticos producidos por uno mismo o por agentes externos, asumiendo responsabilidad epistémica respecto de su

aceptación o rechazo. La definición integra deliberadamente seis operaciones (interpretar, examinar, justificar, verificar, corregir, argumentar) bajo una condición que las articula, la asunción de responsabilidad epistémica. No basta con detectar un error o confirmar un acierto; lo que define la validación es que el sujeto se hace responsable de su veredicto, comprometiéndose con razones que podría exponer y defender ante otros.

Las seis operaciones que recoge la definición no se yuxtaponen como tareas independientes, sino que describen un mismo acto epistémico en distintos momentos, cuya unidad no proviene de su orden, que puede variar, sino de la responsabilidad epistémica que lo atraviesa. En cada una de ellas, el sujeto compromete su juicio sobre lo que cuenta como verdadero y justificado. Por ello la validación no es una destreza técnica de comprobación, sino una postura del sujeto ante el conocimiento, en la que la corrección formal y la asunción de razones resultan inseparables.

Validación frente a producción

La primera distinción que el constructo debe sostener es la que lo separa de la producción. Producir un resultado matemático consiste en generar el artefacto; validarlo consiste en establecer, mediante razones, si ese artefacto es correcto y por qué. Las dos operaciones pueden coincidir en una misma persona, pero son lógicamente independientes. Se puede producir sin validar, como cuando se aplica un procedimiento memorizado sin comprenderlo, y se puede validar sin haber producido, como cuando se evalúa críticamente la demostración ajena. La epistemología matemática había mostrado ya esta independencia al distinguir entre generar y evaluar un argumento (Stylianides y Ball, 2008) y entre obtener un resultado y comprender por qué es verdadero (Hamami y Morris, 2024). La validación, en este sentido, no es una fase posterior de la producción, sino una actividad epistémica de pleno derecho.

Entendida así, la validación es una práctica matemática auténtica y de alto nivel cognitivo, y no un procedimiento de comprobación rutinario. La didáctica de las matemáticas la había tematizado. En la teoría de las situaciones didácticas, la situación de validación es aquella en la que el estudiante debe probar la verdad de sus afirmaciones ante una comunidad, comprometiéndose con su justificación y no con la mera respuesta (Brousseau, 2002). Validar enlaza, por tanto, con la demostración, la argumentación, la justificación, el razonamiento y la validación de conjeturas, prácticas que la educación matemática reconoce como constitutivas del quehacer disciplinar (Stylianides y Ball, 2008; Harel, 2008, 2024). En todas ellas el sujeto no solo produce, sino que decide, con razones, qué cuenta como verdadero, y en esa decisión se juega la construcción del significado matemático. La validación es, en este sentido, parte de la actividad matemática misma y no un apéndice evaluativo, porque hacer matemáticas incluye decidir, con razones, qué se acepta como verdadero.

La distinción es decisiva porque la inteligencia artificial generativa afecta de manera radicalmente distinta a una y a otra. La producción de artefactos correctos puede externalizarse

hacia el sistema; la validación no, sin que el sujeto renuncie precisamente a aquello que se quiere evidenciar. Quien delega la validación no valida, sino que se limita a aceptar, lo que reproduce el problema en otro nivel. Por ello la validación resiste la disociación entre artefacto y cognición que afecta a la producción, dado que no hay validación genuina sin el ejercicio de la comprensión que se desea inferir.

Validación y agencia epistémica

El segundo pilar del constructo es la cláusula de responsabilidad epistémica, que lo enraíza en la investigación sobre cognición epistémica. Esta tradición ha estudiado cómo las personas conciben la naturaleza del conocimiento y su justificación, y cómo esas concepciones regulan su modo de aprender y de juzgar (Hofer y Pintrich, 1997); los modelos recientes añaden a las creencias las metas epistémicas, los ideales de conocimiento bien fundado y los procesos fiables para alcanzarlo (Chinn et al., 2011). Validar un resultado matemático es, en estos términos, un acto de cognición epistémica de alto nivel, que supone tener como meta la verdad justificada, sostener ideales de lo que cuenta como buena razón en matemáticas y emplear procesos fiables de verificación.

La responsabilidad epistémica añade a la cognición epistémica una dimensión normativa y agencial. No describe solo cómo el sujeto juzga, sino que lo compromete con su juicio, de modo que aceptar o rechazar un resultado deja de ser una reacción pasiva y se convierte en un acto del que responde. Esta dimensión conecta la validación con la autonomía intelectual y con una agencia epistémica madura, capaz de no subordinar el propio juicio a la autoridad del docente ni a la del sistema artificial. En un entorno saturado de afirmaciones producidas por agentes cuya fiabilidad es variable y opaca, esa responsabilidad no es un refinamiento, sino una condición de la racionalidad.

Delimitación frente al juicio evaluativo

Resulta obligado determinar si la validación matemática constituye una propuesta original o si reproduce un constructo ya consolidado. El candidato más próximo es el juicio evaluativo, definido como la capacidad de tomar decisiones sobre la calidad del trabajo propio y ajeno, y postulado como una finalidad de la educación superior particularmente relevante ante la inteligencia artificial generativa (Bearman et al., 2024; Tai et al., 2018). La validación matemática comparte con el juicio evaluativo su carácter de capacidad, no de actividad puntual, y su aplicación tanto al trabajo propio como al ajeno. Sin embargo, se distingue de él en tres aspectos que justifican su consideración como propuesta diferenciada y no como mera aplicación.

En primer lugar, su objeto no es la calidad en general, sino la verdad matemática y su justificación. Por verdad matemática no se entiende aquí una correspondencia absoluta e independiente de toda práctica, sino la aceptabilidad de un resultado conforme a los criterios de justificación compartidos por la comunidad matemática. Un enunciado se sostiene cuando su justificación satisface las normas inferenciales que esa comunidad reconoce (Balacheff, 2010;

Hamami y Morris, 2024). Validar no es, así, apreciar que un trabajo está bien hecho conforme a un estándar, sino dictaminar que un resultado está justificado y exponer las razones que lo sostienen. En segundo lugar, su criterio no es la conformidad con un referente externo de calidad, sino la solidez de la inferencia matemática, gobernada por normas de la disciplina y no por rúbricas. En tercer lugar, su carga normativa es epistémica, porque incorpora la responsabilidad por la aceptación o el rechazo de un enunciado en cuanto pretensión de verdad. Por ello no existe, en la literatura analizada, un constructo consolidado denominado validación matemática como evidencia de comprensión matemática. Lo que existe son tradiciones contiguas (juicio evaluativo, cognición epistémica, evaluación de pruebas) cuya articulación específica para el dominio matemático y para el escenario de la inteligencia artificial constituye la contribución original de este trabajo.

Conviene despejar una objeción previsible, según la cual la validación matemática sería, sin más, el juicio evaluativo aplicado a las matemáticas. La diferencia no es de dominio, sino de naturaleza. El juicio evaluativo presupone un referente de calidad (criterios, estándares, rúbricas) frente al cual se aprecia un trabajo; la validación, en cambio, no apela a un referente externo de calidad, sino a la estructura inferencial interna del propio objeto matemático, cuya verdad no se decide por conformidad con un estándar, sino por la solidez de su justificación. Un trabajo puede ser de alta calidad y, no obstante, contener un resultado falso; una validación correcta no admite esa disociación, porque o el resultado está justificado o no lo está. La validación no hereda su criterio del juicio evaluativo, sino de la epistemología de la disciplina. No es una especie dentro de ese género, sino un constructo con objeto y normatividad propios que la teoría educativa debe reconocer como tal.

Una comparación sistemática

Para hacer explícitas estas diferencias conviene contrastar ambos constructos de manera sistemática a lo largo de varias dimensiones. La Tabla 1 recoge esa comparación.

Tabla 1
Comparación entre el juicio evaluativo y la validación matemática

Dimensión	Juicio evaluativo	Validación matemática
Objeto de evaluación	La calidad del trabajo, propio o ajeno	La verdad de un resultado matemático y su justificación
Finalidad	Discernir y mejorar la calidad del desempeño	Decidir y responder por la aceptación o el rechazo de una afirmación
Criterio de valoración	Conformidad con estándares y referentes de calidad	Solidez de la inferencia según las normas de la disciplina
Tipo de inferencia	Apreciativa, basada en el ajuste a un referente de calidad	Demostrativa, basada en la validez deductiva y la justificación
Responsabilidad epistémica	Implícita, centrada en el desempeño	Constitutiva, pues el sujeto responde por la verdad que acepta
Relación con la verdad	Indirecta, porque la calidad no garantiza la verdad	Directa, porque la verdad es el objeto mismo del juicio
Relación con la calidad	Central y definitoria	Subordinada, pues un trabajo de calidad puede ser falso

Función educativa	Formar la capacidad de juzgar la calidad del trabajo	Formar un sujeto epistémicamente responsable ante el conocimiento
-------------------	--	---

Nota. Elaboración propia.

La Tabla 1 muestra que las diferencias no son matices dentro de un mismo constructo, sino divergencias en sus coordenadas fundamentales. Donde el juicio evaluativo opera con una lógica apreciativa orientada a la calidad, la validación opera con una lógica demostrativa orientada a la verdad, y donde aquel sitúa la responsabilidad en el desempeño, esta la sitúa en la pretensión de verdad. La consecuencia es pertinente para la era de la inteligencia artificial. Un sistema generativo puede producir trabajo de alta calidad aparente, de modo que el juicio sobre la calidad resulta insuficiente para inferir comprensión, mientras que la validación, al exigir el compromiso del sujeto con la verdad justificada, ofrece una evidencia que la producción automática no puede suplantar. La distinción, lejos de ser terminológica, delimita dos objetos educativos diferentes.

La validación matemática como evidencia de comprensión matemática

Sostener que la validación matemática puede operar como evidencia de comprensión matemática exige mostrar que satisface las condiciones que la teoría de la evaluación impone a toda evidencia. Debe ser observable o elicitable, debe sostener una inferencia válida hacia el constructo de interés y las consecuencias de su uso deben ser defendibles, condiciones que la teoría unificada de la validez reúne en un mismo juicio argumentado (Messick, 1995). El argumento central es que, en el escenario actual, la validación cumple estas condiciones mejor que la producción.

Conviene subrayar, antes de ese análisis, que la validación matemática no es únicamente una categoría evaluativa, sino también una manifestación de la comprensión matemática y una parte de la actividad matemática misma. Cuando un matemático decide si una demostración establece lo que pretende, o cuando un estudiante dictamina si su solución resiste un contraejemplo, no realizan una tarea añadida a la matemática, sino que ejercen uno de sus actos más característicos. Por eso tomar la validación como evidencia de comprensión matemática no impone a la disciplina un criterio externo, sino que reconoce como evidencia una práctica que ya le es constitutiva.

En cuanto a la observabilidad, la validación es elicitable mediante tareas que exigen examinar un resultado, dictaminar su corrección, localizar y corregir errores o reconstruir y evaluar una justificación; estas tareas hacen visible el razonamiento que la mera producción mantenía oculto (Sadler, 1989; Wiliam, 2011). En cuanto a la validez inferencial, la inferencia que va de la capacidad de validar a la comprensión es más robusta que la que parte de la capacidad de producir. Un producto correcto puede generarse sin comprensión y, con inteligencia artificial, sin intervención del sujeto, mientras que un acto de validación genuino presupone el ejercicio de la comprensión que se quiere inferir. La validación no la garantiza de modo infalible, pero

estrecha el espacio de explicaciones alternativas del desempeño, lo que mejora la validez de una inferencia evaluativa (Newton, 2007).

Un experimento mental ilustra la asimetría. Considérese a dos personas que entregan idéntica demostración correcta de un mismo teorema. Una la ha construido comprendiendo cada paso; la otra la ha obtenido de un sistema generativo y la transcribe sin examinarla. El producto no permite distinguirlos, de modo que la inferencia desde el artefacto hacia la comprensión es, en este caso, inválida. Si en cambio se les pide que señalen dónde se emplea una hipótesis, que decidan si una variante del argumento sigue siendo válida o que localicen un error introducido, la primera podrá hacerlo y la segunda no, salvo que ejerza la comprensión que antes no medió. La validación discrimina allí donde la producción no discrimina, y esa capacidad es lo que la teoría exige de una buena evidencia; no en vano la evaluación en educación superior ha subrayado la importancia de convertir el juicio del estudiantado, y no solo su producto, en objeto de valoración (Yorke, 2003).

La validación no sustituye a la producción como objetivo de la enseñanza, sino que reordena su función evaluativa. Producir sigue siendo central para aprender, pero el producto, por sí solo, deja de ser evidencia suficiente de comprensión. La validación desplaza, además, el lugar de la autenticidad. Si la evaluación auténtica buscaba la fidelidad de la tarea a las prácticas de la disciplina (Wiggins, 2020), la discusión reciente ha mostrado que la autenticidad reside en la relación del sujeto con el conocimiento que la tarea moviliza (Ajjawi et al., 2023). Lo auténtico no es que el producto se parezca al de un experto, sino que el sujeto asuma ante el resultado la postura epistémica de quien hace matemáticas.

Por último, las consecuencias de adoptar la validación como evidencia parecen defendibles desde el punto de vista del aprendizaje. Una evaluación que premia la capacidad de juzgar y justificar, en lugar de la mera obtención de resultados, orienta la enseñanza hacia la comprensión y la argumentación, y desincentiva las estrategias de cumplimiento que la evaluación basada en productos favorece (Torrance, 2012). Además, prepara al estudiantado para una vida en la que convivirá con sistemas que producen afirmaciones cuya aceptación responsable exige capacidad de validación. Evidencia evaluativa y finalidad formativa se alinean, porque lo que se evalúa es también lo que conviene aprender.

Implicaciones para la teoría de la educación

La propuesta excede la técnica evaluativa y toca cuestiones de fondo. La primera concierne a la finalidad. La cultura de la medición había tendido a sustituir el para qué por el cuánto, y cuando los productos que se medían pueden generarse al margen del aprendizaje, medirlos deja de informar sobre la finalidad (Biesta, 2010). La educación no opera solo en el dominio de la cualificación, sino también en el de la formación del sujeto, donde se juega aquello que ningún sistema puede producir en lugar del estudiante (Biesta, 2009, 2015). La validación responde a esa exigencia porque obliga a formular la pregunta que la inteligencia artificial vuelve

ineludible, es decir, qué merece ser valorado educativamente cuando la producción puede externalizarse.

La segunda implicación se refiere al sujeto educativo. Si la educación comprende la formación de un sujeto capaz de comparecer responsablemente ante el conocimiento, la validación no es solo un medio para evaluar, sino una manifestación de aquello que la educación pretende formar (Fullat, 1992). Esa apertura a la formación, pensada por la tradición pedagógica española como educabilidad, es la condición de que la validación pueda enseñarse y, al enseñarse, contribuya a formar (García Carrasco, 1993). Aceptar o rechazar un resultado con razones propias, sin subordinarlo a la autoridad del docente ni del sistema artificial, constituye al sujeto como instancia responsable de sus propias razones.

La tercera implicación atañe a la autonomía intelectual. Formar estudiantes capaces de validar enlaza con la tradición de la educación matemática crítica, que ha defendido que la formación matemática incluye interrogar las afirmaciones presentadas como matemáticamente fundadas y no aceptarlas por la sola autoridad de quien las emite (Skovsmose, 1994); en un entorno donde muchas de esas afirmaciones procederán de sistemas opacos, esa capacidad crítica adquiere un valor cívico además de epistémico. Formar a quien valida es, en buena medida, desarrollar su capacidad de aprender a aprender y de autorregular su propio juicio matemático (Gargallo López, 1995).

La relación entre evaluación y enseñanza se reordena en consecuencia. La teoría ha advertido del riesgo de que la evaluación colonice la enseñanza y la oriente hacia lo fácilmente evaluable. La validación invierte parcialmente esa relación. Al situar como evidencia la capacidad de juzgar y justificar, hace que evaluar y enseñar a comprender converjan en lugar de oponerse. No elimina el riesgo de desnaturalización (toda evidencia puede burocratizarse), pero lo desplaza hacia un terreno en el que la práctica evaluable coincide con la práctica valiosa. Así, la propuesta contribuye a la Teoría de la Educación no como una técnica más, sino como una reflexión sobre qué cuenta como evidencia de que la educación ha tenido lugar, justo cuando las evidencias heredadas han perdido su transparencia.

Importa situar con precisión la contribución en el campo de la educación matemática. El problema no es estrictamente matemático en su contenido, pero sí lo es en su pregunta, que indaga qué cuenta como evidencia de comprensión matemática cuando la producción de soluciones puede externalizarse. La matemática no es aquí un mero ejemplo, sino el terreno donde la disociación entre producir y comprender se vuelve más nítida y analizable, porque en pocas disciplinas la distinción entre obtener un resultado y justificar su verdad está tan codificada en las prácticas propias del campo. Pensar la validación como evidencia de comprensión matemática es, por ello, una contribución a la educación matemática antes que una aplicación de categorías externas.

Por último, la propuesta invita a reconsiderar el lugar de la inteligencia artificial en la teoría educativa sin tono celebratorio ni catastrofista, evitando el determinismo tecnológico que atribuye a los artefactos, por sí solos, el progreso o el retroceso de la educación (Jover et al., 2017). Estos sistemas no son aliados ni enemigos del aprendizaje, sino agentes productores de artefactos cuya presencia obliga a explicitar qué evidencia buscábamos y por qué. Ante sistemas capaces de generar conocimiento aparente, discurso con forma de verdad pero sin un sujeto que responda de él, la tarea de la educación no es competir en producción, sino formar a quien pueda validar, esto es, a quien pueda hacerse cargo de la verdad de lo que acepta. La validación matemática es una respuesta teórica a esa obligación, y su valor no depende del estado tecnológico concreto de los sistemas, sino de una constatación más estable. Producir y comprender son distintos, y la evidencia de la comprensión ha de buscarse allí donde la producción no puede sustituirla.

Contribuciones para la educación matemática

Este artículo realiza varias contribuciones al campo de la educación matemática. En el plano conceptual, formula y delimita la validación matemática como una forma específica de evidencia de comprensión matemática, distinta de la producción y articulada con la demostración, la argumentación y la construcción de significado; con ello aborda un vacío preciso, que es la ausencia, en la literatura del campo, de una caracterización de la validación como evidencia de comprensión matemática en contextos mediados por inteligencia artificial generativa.

En el plano de la evaluación matemática, ofrece un criterio para repensar qué tareas permiten inferir comprensión cuando la producción de soluciones puede externalizarse, y orienta el diseño de situaciones que hagan ineludible el juicio del estudiante. En el plano de la investigación, abre una agenda orientada a operacionalizar la validación en registros y tareas matemáticas, estudiar su relación con los esquemas de prueba y la coordinación de representaciones, y analizar cómo la mediación de la inteligencia artificial transforma las prácticas de justificación en el aula de matemáticas. La validación matemática se ofrece, así, como una categoría útil para las futuras discusiones sobre evaluación matemática en la era de la inteligencia artificial.

CONCLUSIONES

Este artículo ha sostenido que la inteligencia artificial generativa no plantea, en lo fundamental, un problema de integridad académica, sino una crisis epistemológica de las evidencias de aprendizaje. Al volver insostenible el supuesto de que el producto matemático evidencia la comprensión que lo originó, obliga a la teoría de la evaluación a explicitar su carácter inferencial y a buscar evidencias menos vulnerables a la disociación entre artefacto y cognición. Frente a ello se ha fundamentado el constructo de validación matemática como forma específica

de evidencia de comprensión matemática, más robusta que la producción para inferir comprensión, por cuanto su ejercicio genuino presupone aquello que se desea evidenciar.

La contribución se ha delimitado con cuidado. La validación matemática se distingue de la producción, se enraíza en la cognición y la responsabilidad epistémicas, y se diferencia del juicio evaluativo por su objeto, su criterio y su carga normativa. El análisis de la literatura ha mostrado que no existe un constructo consolidado con esta denominación y alcance, de modo que la propuesta es original y, a la vez, está articulada con tradiciones reconocidas.

Procede reconocer las limitaciones de la propuesta. Es una elaboración conceptual no operacionalizada ni contrastada empíricamente; la afirmación de que la validación sostiene inferencias más robustas que la producción es teóricamente fundada, pero requiere verificación. La validación tampoco es inmune a su desnaturalización. Puede convertirse en un formato escolar reproducible y, con ello, en un indicador sustitutivo, y también puede delegarse en sistemas artificiales, de modo que su rendimiento dependerá de tareas que hagan ineludible el compromiso del sujeto. Por último, se ha formulado para el conocimiento matemático, y su extensión a otros dominios, donde la verdad y la justificación operan de otro modo, no puede darse por descontada.

De estas limitaciones se desprende una agenda de investigación. Es necesario operacionalizar la validación matemática en tareas y rúbricas que la hagan elicitable sin reducirla a un procedimiento mecánico, y estudiar la validez de las inferencias que autoriza; diseñar situaciones evaluativas que mantengan acoplados el juicio del sujeto y su responsabilidad, resistiendo la externalización; examinar sus implicaciones para la formación del profesorado, pues enseñar a validar exige docentes que conciban la justificación como objeto de enseñanza; y atender a las dimensiones de equidad, dado que la capacidad de validar se distribuye de manera desigual. La reconceptualización aquí esbozada no se ofrece como una solución cerrada, sino como un marco desde el cual repensar qué cuenta como evidencia de aprendizaje cuando producir ha dejado de ser, por sí solo, prueba de comprender.

Conviene, para terminar, enunciar con nitidez la tesis que ha guiado el conjunto del trabajo. La inteligencia artificial no vuelve irrelevante la evaluación; lo que vuelve insostenible es el supuesto, largamente naturalizado, de que la producción matemática constituye evidencia suficiente de comprensión. La irrupción de sistemas capaces de producir matemáticas obliga, por ello, a desplazar el centro de gravedad de la evaluación desde la producción hacia la validación, es decir, desde lo que el sujeto es capaz de generar hacia lo que es capaz de juzgar, justificar y responder. En ese desplazamiento, y no en la persecución de la copia, se juega la posibilidad de seguir infiriendo aprendizaje a partir de evidencias dignas de ese nombre. La Educación Matemática se encuentra, así, ante una tarea conceptual y didáctica, la de redefinir qué cuenta como evidencia de comprensión matemática en un tiempo en el que comprender y producir han dejado de coincidir.

REFERENCIAS

- Ajjawi, R., Tai, J., Dollinger, M., Dawson, P., Boud, D., & Bearman, M. (2023). From authentic assessment to authenticity in assessment: Broadening perspectives. *Assessment & Evaluation in Higher Education*, 49(4), 499–510. <https://doi.org/10.1080/02602938.2023.2271193>
- Baird, J.-A., Andrich, D., Hopfenbeck, T. N., & Stobart, G. (2017). Assessment and learning: Fields apart? *Assessment in Education: Principles, Policy & Practice*, 24(3), 317–350. <https://doi.org/10.1080/0969594X.2017.1319337>
- Balacheff, N. (2010). Bridging knowing and proving in mathematics: A didactical perspective. En G. Hanna, H. N. Jahnke, & H. Pulte (Eds.), *Explanation and proof in mathematics* (pp. 115–135). Springer. https://doi.org/10.1007/978-1-4419-0576-5_9
- Bearman, M., Tai, J., Dawson, P., Boud, D., & Ajjawi, R. (2024). Developing evaluative judgement for a time of generative artificial intelligence. *Assessment & Evaluation in Higher Education*, 49(6), 893–905. <https://doi.org/10.1080/02602938.2024.2335321>
- Biesta, G. J. J. (2009). Good education in an age of measurement: On the need to reconnect with the question of purpose in education. *Educational Assessment, Evaluation and Accountability*, 21(1), 33–46. <https://doi.org/10.1007/s11092-008-9064-9>
- Biesta, G. J. J. (2010). *Good education in an age of measurement: Ethics, politics, democracy*. Paradigm Publishers.
- Biesta, G. (2015). What is education for? On good education, teacher judgement, and educational professionalism. *European Journal of Education*, 50(1), 75–87. <https://doi.org/10.1111/ejed.12109>
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74. <https://doi.org/10.1080/0969595980050102>
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5–31. <https://doi.org/10.1007/s11092-008-9068-5>
- Brookhart, S. M. (2003). Developing measurement theory for classroom assessment purposes and uses. *Educational Measurement: Issues and Practice*, 22(4), 5–12. <https://doi.org/10.1111/j.1745-3992.2003.tb00139.x>
- Brousseau, G. (2002). *Theory of didactical situations in mathematics: Didactique des mathématiques, 1970–1990* (N. Balacheff, M. Cooper, R. Sutherland y V. Warfield, Eds. y Trads.). Kluwer Academic.
- Chinn, C. A., Buckland, L. A., & Samarapungavan, A. (2011). Expanding the dimensions of epistemic cognition: Arguments from philosophy and psychology. *Educational Psychologist*, 46(3), 141–167. <https://doi.org/10.1080/00461520.2011.587722>

- Chirikov, I., Smirnov, I., & Kizilcec, R. F. (2026). Generative AI use and misuse call for assessment reform in higher education. *Science*, 392(6800), 818–820. <https://doi.org/10.1126/science.aec5115>
- Colom Cañellas, A. J. (2004). La Teoría de la Educación en su doble dimensionalidad: Como teoría acerca de la realidad y como teoría acerca del saber educativo. *Teoría de la Educación. Revista Interuniversitaria*, 16, 19–48.
- Crooks, N. M., & Alibali, M. W. (2014). Defining and measuring conceptual knowledge in mathematics. *Developmental Review*, 34(4), 344–377. <https://doi.org/10.1016/j.dr.2014.10.001>
- Duval, R. (2006). A cognitive analysis of problems of comprehension in a learning of mathematics. *Educational Studies in Mathematics*, 61(1–2), 103–131. <https://doi.org/10.1007/s10649-006-0400-z>
- Ernest, P. (2002). *The philosophy of mathematics education*. RoutledgeFalmer. <https://doi.org/10.4324/9780203497012>
- Ernest, P. (2018). The philosophy of mathematics education: An overview. En P. Ernest (Ed.), *The philosophy of mathematics education today* (pp. 13–35). Springer. https://doi.org/10.1007/978-3-319-77760-3_2
- Fullat, O. (1992). *Filosofías de la educación: Paideia*. CEAC.
- García Carrasco, J. (1993). El concepto de educabilidad y el proceso educacional. *Teoría de la Educación. Revista Interuniversitaria*, 5, 11–32.
- Gargallo López, B. (1995). Estrategias de aprendizaje: Estado de la cuestión. Propuestas para la intervención educativa. *Teoría de la Educación. Revista Interuniversitaria*, 7, 53–76.
- Hamami, Y., & Morris, R. L. (2020). Philosophy of mathematical practice: A primer for mathematics educators. *ZDM – Mathematics Education*, 52(6), 1113–1126. <https://doi.org/10.1007/s11858-020-01159-5>
- Hamami, Y., & Morris, R. L. (2024). Understanding in mathematics: The case of mathematical proofs. *Noûs*. <https://doi.org/10.1111/nous.12489>
- Harel, G. (2008). DNR perspective on mathematics curriculum and instruction, Part I: Focus on proving. *ZDM*, 40(3), 487–500. <https://doi.org/10.1007/s11858-008-0104-1>
- Harel, G. (2024). Epistemological justification. *ZDM – Mathematics Education*. <https://doi.org/10.1007/s11858-024-01603-w>
- Hofer, B. K., & Pintrich, P. R. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. *Review of Educational Research*, 67(1), 88–140. <https://doi.org/10.3102/00346543067001088>
- Jover, G., González, V., & Prieto, M. (2017). *Una filosofía de la educación del siglo XXI*. Síntesis.

- Kilpatrick, J., Swafford, J., & Findell, B. (Eds.). (2001). *Adding it up: Helping children learn mathematics*. National Academy Press. <https://doi.org/10.17226/9822>
- Kollosche, D. (2021). Styles of reasoning for mathematics education. *Educational Studies in Mathematics*, 107(3), 471–486. <https://doi.org/10.1007/s10649-021-10046-z>
- Messick, S. (1989). Validity. En R. L. Linn (Ed.), *Educational measurement* (3.^a ed., pp. 13–103). Macmillan.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741–749. <https://doi.org/10.1037/0003-066X.50.9.741>
- Michener, E. R. (1978). Understanding understanding mathematics. *Cognitive Science*, 2(4), 361–383. https://doi.org/10.1207/s15516709cog0204_3
- Newton, P. E. (2007). Clarifying the purposes of educational assessment. *Assessment in Education: Principles, Policy & Practice*, 14(2), 149–170. <https://doi.org/10.1080/09695940701478321>
- Niss, M., & Højgaard, T. (2019). Mathematical competencies revisited. *Educational Studies in Mathematics*, 102(1), 9–28. <https://doi.org/10.1007/s10649-019-09903-9>
- Panadero, E., Andrade, H., & Brookhart, S. (2018). Fusing self-regulated learning and formative assessment: A roadmap of where we are, how we got here, and where we are going. *The Australian Educational Researcher*, 45(1), 13–31. <https://doi.org/10.1007/s13384-018-0258-y>
- Rittle-Johnson, B., & Schneider, M. (2015). Developing conceptual and procedural knowledge of mathematics. En R. Cohen Kadosh & A. Dowker (Eds.), *The Oxford handbook of numerical cognition* (pp. 1118–1134). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199642342.013.014>
- Rycroft-Smith, L. (2023). Conceptions and consequences of mathematical argumentation, justification, and proof. *Research in Mathematics Education*, 25(2), 119–124. <https://doi.org/10.1080/14794802.2023.2165537>
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18(2), 119–144. <https://doi.org/10.1007/BF00117714>
- Schoenfeld, A. H. (2016). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics (Reprint). *Journal of Education*, 196(2), 1–38. <https://doi.org/10.1177/002205741619600202>
- Sfard, A. (2008). *Thinking as communicating: Human development, the growth of discourses, and mathematizing*. Cambridge University Press.
- Shepard, L. A., Penuel, W. R., & Pellegrino, J. W. (2018). Using learning and motivation theories to coherently link formative assessment, grading practices, and large-scale assessment.

- Educational Measurement: Issues and Practice*, 37(1), 21–34.
<https://doi.org/10.1111/emip.12189>
- Skovsmose, O. (1994). *Towards a philosophy of critical mathematics education*. Kluwer.
<https://doi.org/10.1007/978-94-017-3556-8>
- Stiggins, R. (2005). From formative assessment to assessment *for* learning: A path to success in standards-based schools. *Phi Delta Kappan*, 87(4), 324–328.
<https://doi.org/10.1177/003172170508700414>
- Stylianides, A. J., & Ball, D. L. (2008). Understanding and describing mathematical knowledge for teaching: Knowledge about proof for engaging students in the activity of proving. *Journal of Mathematics Teacher Education*, 11(4), 307–332.
<https://doi.org/10.1007/s10857-008-9077-9>
- Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). Developing evaluative judgement: Enabling students to make decisions about the quality of work. *Higher Education*, 76(3), 467–481. <https://doi.org/10.1007/s10734-017-0220-3>
- Torrance, H. (2012). Formative assessment at the crossroads: Conformative, deformative and transformative assessment. *Oxford Review of Education*, 38(3), 323–342.
<https://doi.org/10.1080/03054985.2012.689693>
- Touriñán López, J. M. (1987). *Teoría de la educación: La educación como objeto de conocimiento*. Anaya.
- Wiggins, G. (2020). The case for authentic assessment. *Practical Assessment, Research, and Evaluation*, 2, Artículo 2. <https://doi.org/10.7275/ffb1-mm19>
- Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, 37(1), 3–14. <https://doi.org/10.1016/j.stueduc.2011.03.001>
- Yorke, M. (2003). Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. *Higher Education*, 45(4), 477–501.
<https://doi.org/10.1023/A:1023967026413>