Ciencia Latina Revista Científica Multidisciplinar, Ciudad de México, México.
ISSN 2707-2207 / ISSN 2707-2215 (en línea), enero-febrero 2026,
Volumen 10, Número 1.

https://doi.org/10.37811/cl_rcm.v10i1

EVALUACIÓN FORMATIVA PARA LA CONSTRUCCIÓN DE SIGNIFICADO ROBUSTO: UN MODELO BASADO EN LA INTEGRACIÓN TRU-DNR Y PROBLEMAS RETADORES EN SEXTO GRADO

FORMATIVE ASSESSMENT FOR ROBUST MEANING CONSTRUCTION: A MODEL BASED ON TRU-DNR INTEGRATION AND CHALLENGING PROBLEMS IN SIXTH GRADE

Lady Johana Julio Barrera
Universidad Antonio Narino – Colombia

Mary Falk De Losada
Universidad Antonio Nariño - Colombia

DOI: https://doi.org/10.37811/cl_rcm.v10i1.22786

Evaluación formativa para la construcción de significado robusto: un modelo basado en la integración TRU-DNR y problemas retadores en sexto grado

Lady Johana Julio Barrera[1]
Ljulio64@uan.edu.co
https://orcid.org/0009-0006-5572-020X
Universidad Antonio Narino
Bogotá, Colombia

Mary Falk De Losada
Mariadelosada@gmail.com
https://orcid.org/0000-0002-6380-0481
Universidad Antonio Nariño
Bogotá, Colombia

RESUMEN

La investigación en educación matemática ha señalado una brecha persistente entre los avances teóricos en evaluación formativa y lo que ocurre en las aulas, donde la evaluación sigue centrándose en medir productos finales. Este estudio presenta el diseño, implementación y validación del modelo MECSIRCOM (Modelo de Evaluación para la Construcción de Significado Robusto de Conceptos Matemáticos), fundamentado en la integración de los marcos Teaching for Robust Understanding (TRU) y Duality, Necessity, and Repeated Reasoning (DNR). Mediante Investigación Basada en el Diseño (DBR) con dos ciclos iterativos, se implementó una secuencia de cuatro fases evaluativas progresivas con 20 estudiantes de sexto grado en una institución educativa pública de Bogotá, Colombia, en el dominio de teoría de números. Los resultados mostraron una progresión coherente: activación de necesidad intelectual (80%), coordinación de actos mentales (60%), consolidación mediante razonamiento repetido (60%) y transferencia con autorregulación (65%). La validación, realizada mediante el enfoque de Kane con triangulación de evidencias de 13 expertos internacionales, 16 docentes y análisis de producciones estudiantiles, permite sostener que MECSIRCOM es una propuesta evaluativa válida que transforma la evaluación de mecanismo de medición en motor de aprendizaje matemático.

Palabras clave: evaluación formativa; construcción de significado robusto; problemas retadores; pensamiento matemático; investigación basada en el diseño

[1] Autor principal.
Correspondencia: Ljulio64@uan.edu.co

Formative assessment for robust meaning construction: a model based on TRU-DNR integration and challenging problems in sixth grade

ABSTRACT

Research in mathematics education has identified a persistent gap between theoretical advances in formative assessment and classroom practices, which remain focused on measuring final products. This study presents the design, implementation, and validation of the MECSIRCOM model (Assessment Model for the Construction of Robust Meaning of Mathematical Concepts), grounded in the integration of the Teaching for Robust Understanding (TRU) and Duality, Necessity, and Repeated Reasoning (DNR) frameworks. Through Design-Based Research (DBR) with two iterative cycles, a sequence of four progressive assessment phases was implemented with 20 sixth-grade students at a public school in Bogotá, Colombia, in the domain of number theory. Results showed a coherent progression: activation of intellectual necessity (80%), coordination of mental acts (60%), consolidation through repeated reasoning (60%), and transfer with self-regulation (65%). Validation through Kane's argument-based approach, triangulating evidence from 13 international experts, 16 teachers, and student production analysis, supports MECSIRCOM as a valid assessment proposal that transforms assessment from a measurement mechanism into a driver of mathematical learning.

Keywords: formative assessment; robust meaning construction; challenging problems; mathematical thinking; design-based research

Artículo recibido 10 diciembre 2025

Aceptado para publicación: 10 enero 2026

INTRODUCCIÓN

El desarrollo del pensamiento matemático es uno de los propósitos centrales de la educación contemporánea, pero la forma en que se evalúa este desarrollo en las aulas dista mucho de lo que la investigación considera efectivo. Black y Wiliam (2009) mostraron que la evaluación formativa tiene un impacto significativo en el aprendizaje cuando se diseña como un proceso de búsqueda e interpretación de evidencias que orienta tanto la enseñanza como el aprendizaje. Sin embargo, como advierten Burkhardt y Schoenfeld (2018), la brecha entre estos avances y las prácticas cotidianas sigue siendo considerable: las evaluaciones continúan centradas en verificar resultados y calificar productos finales.

Esta tensión es especialmente problemática en matemáticas. Construir conocimiento matemático implica desarrollar comprensión conceptual, fluidez procedimental, capacidad de resolver problemas, razonamiento adaptativo y una disposición a ver las matemáticas como algo con sentido (Kilpatrick et al., 2001). Cuando la enseñanza ayuda a los estudiantes a dar sentido a los conceptos y a construir conexiones entre ideas, se favorece un aprendizaje más profundo y transferible (Hiebert y Grouws, 2007). Pero las evaluaciones convencionales rara vez capturan estos procesos, dejando invisibles las trayectorias de pensamiento que son, precisamente, el núcleo del aprendizaje matemático.

La noción de significado en educación matemática ha sido abordada desde varias perspectivas. Kaiser (2017) distingue entre un significado personal —la relevancia que el estudiante atribuye a una actividad— y significados más objetivos, compartidos dentro de una comunidad matemática. Pérez Duarte (2016) propone el constructo de construcción de significado robusto, asociándolo con la configuración de redes conceptuales que realizan los estudiantes al resolver problemas no rutinarios. Este constructo me resultó particularmente útil porque permite describir y observar cómo se configura y enriquece el pensamiento matemático durante la resolución de problemas retadores, algo que otros constructos dejan en un plano más abstracto.

En el contexto colombiano, la experiencia en olimpiadas matemáticas ha aportado evidencia empírica sobre cómo los problemas retadores fomentan redes conceptuales. Falk de Losada (2001) documenta que en la preparación de estudiantes para concursos se proponen series de problemas cuya resolución demanda establecer conexiones entre conceptos previamente trabajados, generando nuevos nexos

conceptuales. En mi trabajo como docente de matemáticas en el sector público de Bogotá, he observado que esta lógica de las olimpiadas tiene un potencial formativo que rara vez se aprovecha en el aula regular. Esa observación fue, en buena medida, el punto de partida de esta investigación.

Dos marcos teóricos ofrecen herramientas complementarias para fundamentar la transformación que propongo. El modelo Teaching for Robust Understanding (TRU), de Schoenfeld (2014, 2016), proporciona cinco dimensiones para caracterizar ambientes de aprendizaje efectivos. El sistema Duality, Necessity, and Repeated Reasoning (DNR), de Harel (2008a, 2008b, 2010), explica los mecanismos cognitivos mediante los cuales los estudiantes construyen conocimiento matemático. La articulación de ambos marcos no había sido operacionalizada previamente en prácticas evaluativas específicas, y esa es la apuesta teórica central de este trabajo.

El problema de investigación que orienta el estudio es: ¿cómo construir un modelo de evaluación que integre los principios de diseño curricular DNR y la caracterización de ambientes de aprendizaje TRU para la construcción de significado robusto de conceptos matemáticos que favorezcan el desarrollo del pensamiento matemático en estudiantes de sexto grado? El objetivo general es diseñar y validar dicho modelo en el contexto de problemas retadores, con estudiantes de grado sexto de una institución educativa de Bogotá.

El artículo se organiza así: primero se presentan los referentes conceptuales que fundamentan la integración TRU-DNR; luego se describe la metodología de Investigación Basada en el Diseño; posteriormente se presentan los resultados por fases evaluativas; la discusión interpreta los hallazgos a la luz de la literatura; y las conclusiones sintetizan las contribuciones del estudio.

Marco teórico

El modelo TRU: condiciones ambientales para el aprendizaje

El marco Teaching for Robust Understanding (TRU) de Schoenfeld (2014, 2016) identifica cinco dimensiones de los ambientes de aprendizaje que predicen el desarrollo del pensamiento matemático: contenido matemático, demanda cognitiva, acceso equitativo, agencia e identidad, y evaluación formativa. Lo que distingue a TRU de otros marcos es que no se centra en las acciones del docente, sino en las condiciones que experimentan los estudiantes. Esto resulta clave para el modelo que propongo,

porque permite establecer criterios observables para determinar si un ambiente favorece realmente el aprendizaje profundo.

Las cinco dimensiones operan de manera articulada. El contenido matemático exige que los estudiantes accedan a ideas centrales de la disciplina presentadas de manera conectada. La demanda cognitiva requiere tareas que planteen desafíos genuinos sin resultar inabordables. El acceso equitativo asegura oportunidades reales de participación para todos. La agencia e identidad se refieren a la capacidad de los estudiantes para desarrollar confianza como pensadores matemáticos. La evaluación formativa proporciona información sistemática sobre el pensamiento estudiantil (Schoenfeld y TRU Project, 2016).

El modelo DNR: mecanismos cognitivos de construcción de conocimiento

El sistema Duality, Necessity, and Repeated Reasoning (DNR) de Harel (2008a, 2008b, 2010, 2021) proporciona un marco para comprender cómo se construye el conocimiento matemático. Se fundamenta en premisas epistemológicas y cognitivas que establecen la naturaleza del aprendizaje matemático como un proceso de coordinación progresiva de actos mentales: interpretar, conjeturar, inferir, estructurar y generalizar (Harel y Koichu, 2010).

Los tres principios del DNR operan de manera coordinada. El principio de dualidad establece una relación recíproca entre ways of understanding (formas de entender: los productos de un acto mental aplicado a una situación particular) y ways of thinking (formas de pensar: las características cognitivas del acto mental, más generales y transferibles). La dualidad consiste en que ambas se desarrollan mutuamente: las formas de entender son el medio a través del cual se desarrollan las formas de pensar, y estas, a su vez, determinan las formas de entender que el estudiante puede construir (Harel, 2008a). El principio de necesidad —que Harel desarrolló originalmente en el contexto de estudiantes universitarios de cálculo— postula que los estudiantes construyen conocimiento cuando experimentan una necesidad intelectual genuina, es decir, cuando sus conocimientos previos resultan insuficientes para abordar una situación problemática. Aplicar este principio con niños de 10-12 años requiere adaptaciones que discuto más adelante. El principio de razonamiento repetido señala que la consolidación del conocimiento requiere múltiples experiencias con la misma estructura matemática en contextos variados (Harel, 2021).

Integración TRU-DNR: una convergencia funcional

La construcción de significado robusto, tal como la entiendo en esta investigación, requiere articular las condiciones ambientales que describe TRU con los mecanismos cognitivos que explica DNR. La integración entre ambos es intencionada: TRU describe las condiciones del entorno y DNR explica qué ocurre cognitivamente dentro de esas condiciones, de modo que cada marco complementa la perspectiva del otro. La convergencia funcional entre ambos genera experiencias de dotación de significado que consolidan redes conceptuales estables y transferibles.

Esta convergencia se materializa en cuatro dimensiones operativas. La primera, activación de necesidad intelectual, articula la demanda cognitiva de TRU con el principio de necesidad del DNR: problemas retadores bien diseñados generan el desequilibrio cognitivo que impulsa la búsqueda de nuevas estrategias. La segunda, coordinación de actos mentales, vincula el contenido matemático de TRU con el principio de dualidad del DNR: cuando el ambiente ofrece contenido rico y conectado, los estudiantes generan formas de entender situadas (ways of understanding) que, al ser reflexionadas y reutilizadas, desarrollan formas de pensar más generales (ways of thinking), lo que se manifiesta en la coordinación de actos mentales como interpretar, conjeturar, inferir y estructurar. La tercera, consolidación mediante razonamiento repetido, conecta el acceso equitativo y la agencia de TRU con el razonamiento repetido del DNR: múltiples experiencias en contextos diversos estabilizan las redes conceptuales. La cuarta, autorregulación y transferencia, integra la evaluación formativa de TRU con el razonamiento repetido reflexivo del DNR.

Construcción de significado robusto

A partir de Pérez Duarte (2016), quien asocia la construcción de significado robusto con la configuración de redes conceptuales durante la resolución de problemas no rutinarios, en esta investigación propongo una diferenciación conceptual en tres niveles. El primero, dotar de significado, corresponde al resultado de un acto mental específico desarrollado en un contexto determinado; representa un avance parcial en la comprensión. El segundo, dotar de significado robusto, es el resultado de la interacción coordinada y reflexiva de dos o más actos mentales en diferentes contextos durante la resolución de problemas retadores. El tercero, construcción de significado robusto, designa el proceso progresivo y acumulativo mediante el cual se desarrollan y fortalecen redes conceptuales transferibles entre campos matemáticos

diversos. Esta diferenciación enfatiza el proceso progresivo mediante problemas retadores y sitúa la transferencia entre campos matemáticos como evidencia central de robustez.

La noción de redes conceptuales, entendidas como instrumentos gráfico-semánticos que muestran conexiones no jerárquicas entre conceptos (Ciliberti y Galagovsky, 1999), proporciona un indicador observable de la robustez del significado construido. Cuando un estudiante puede movilizar un concepto en campos matemáticos diversos, justificar por qué funciona un método y discriminar cuándo aplicarlo, hay evidencia de un significado robusto que trasciende la reproducción procedimental.

Evaluación para el aprendizaje

La evaluación para el aprendizaje, fundamentada en Black y Wiliam (1998, 2009), Wiliam (2011) y Shepard (2000), es el componente evaluativo que operacionaliza la integración TRU-DNR. Su característica distintiva es que transforma la evaluación de un mecanismo de medición a un motor de aprendizaje, proporcionando retroalimentación continua que orienta las decisiones pedagógicas. Hattie y Timperley (2007) identifican que la retroalimentación más efectiva responde a tres preguntas: ¿hacia dónde voy?, ¿cómo voy? y ¿qué sigue? Esta estructura se integra en las fases del modelo MECSIRCOM mediante instrumentos que hacen visibles procesos de pensamiento que las evaluaciones convencionales pasan por alto.

Problemas retadores como contexto de construcción

Los problemas retadores, inspirados en las tradiciones de olimpiadas matemáticas (Falk de Losada, 2001, 2020), son el instrumento central de evaluación en MECSIRCOM. Kapur (2014) ha mostrado que el fracaso productivo en contextos de resolución de problemas genera aprendizaje más profundo que la instrucción directa. Schoenfeld (1985, 1992) establece que la resolución de problemas moviliza no solo conocimientos conceptuales, sino también estrategias heurísticas, metacognición y disposiciones matemáticas. En este estudio, los problemas retadores operan en un doble nivel: como activadores de la necesidad intelectual que inicia la construcción de significado, y como contextos donde se hace visible la coordinación progresiva de actos mentales.

METODOLOGÍA

Enfoque y diseño

La investigación se sitúa en el paradigma cualitativo, de enfoque interpretativo, y adopta la Investigación Basada en el Diseño (Design-Based Research, DBR) como enfoque metodológico (Cobb et al., 2003; Prediger, 2019). Opté por la DBR porque responde a un doble propósito que me resultaba inseparable: desarrollar una intervención práctica implementable en el aula y, al mismo tiempo, generar teoría contextualizada sobre los procesos de construcción de significado robusto. El proceso se desarrolló en dos ciclos iterativos: diseño de instrumentos, implementación en el aula, análisis de datos y refinamiento del modelo (Confrey et al., 2006; Sandoval, 2014).

Conjetura central de diseño. Se plantea que la implementación de un modelo de evaluación fundamentado en la integración TRU-DNR y operacionalizado mediante problemas retadores, en una secuencia progresiva de cuatro fases, favorecerá la construcción de significado robusto de conceptos matemáticos y, en consecuencia, aportará al desarrollo del pensamiento matemático autónomo en estudiantes de básica secundaria (véase Anexo metodológico, Tabla A2).

Participantes y contexto

Participaron 20 estudiantes de grado sexto (10-12 años) del Colegio Fabio Lozano Simonelli, institución educativa distrital ubicada en Bogotá, Colombia. La muestra fue seleccionada por conveniencia, trabajando con un grupo-curso completo para preservar la naturaleza ecológica de la investigación. El contenido se centró en teoría de números: divisibilidad, mínimo común múltiplo (MCM), máximo común divisor (MCD) y números primos. La decisión de trabajar con este dominio obedeció a que permite diseñar problemas retadores con progresión de demanda cognitiva accesible para este nivel escolar.

Modelo MECSIRCOM: estructura y fases

El modelo se estructura en cinco componentes interdependientes: contenido matemático, problemas retadores, discurso matemático, evaluación para el aprendizaje e integración TRU-DNR para la construcción de significado robusto. Estos componentes se operacionalizan en cuatro fases evaluativas progresivas, cada una inspirada en formatos de olimpiadas matemáticas con adaptación pedagógica. La

Tabla 1. Estructura y fases evaluativas del modelo MECSIRCOM

Fase	Formato y tiempo	Dimensión prioritaria	Evidencias principales	Indicador de logro
Inicial	1 sesión (90 min). Selección múltiple (tipo Canguro)	Necesidad intelectual	Disposición inicial; compromiso sostenido; persistencia; trayectoria emocional	80% activación alta/moderada
Desarrollo	3 sesiones (45 min). Respuesta única + método completo	Coordinación de actos mentales	Pasos organizados; justificación; verificación; flexibilidad estratégica	60% métodos organizados (≥3/5 problemas)
Consolidación	3 sesiones (90 min). Variaciones A-B-C por problema	Razonamiento repetido	Aplicación consistente del método; discriminación MCM/MCD; redes conceptuales	60% razonamiento repetido consistente
Transferencia	1 fase final. 4 problemas no rutinarios, campos diversos	Transferencia + autorregulación	Transferencia aritmético-geométrica/aplicada/algebraica; estrategia autónoma	65% transferencia funcional con autorregulación

Nota. Elaboración propia a partir del diseño e implementación reportados.

La Fase Inicial (90 minutos, formato inspirado en Canguro Matemático) prioriza la activación de necesidad intelectual mediante cuatro problemas de selección múltiple con demanda cognitiva progresiva. La decisión de usar este formato obedeció a que reduce la ansiedad inicial de los estudiantes frente a problemas abiertos, al tiempo que permite observar cómo activan —o no— sus conocimientos previos. La Fase de Desarrollo (tres sesiones de 45 minutos) se centra en la coordinación de actos mentales mediante cinco problemas que exigen respuesta numérica única y construcción de método completo. La Fase de Consolidación (tres sesiones de 90 minutos) busca consolidar comprensiones mediante razonamiento repetido en variaciones de la misma estructura matemática. La Fase de Transferencia evalúa la construcción de significado robusto en su expresión plena, con cuatro problemas no rutinarios que requieren transferencia entre campos matemáticos diversos.

Instrumentos de recolección de datos

Los instrumentos conforman un sistema de triangulación con múltiples fuentes. Las hojas de trabajo del estudiante, diferenciadas por fase, se estructuran en cuatro partes: (A) evaluación de disposición inicial o conexión con conocimientos previos; (B) problemas retadores con espacio amplio para razonamiento; (C) reflexión metacognitiva; y (D) trayectoria emocional. Las rúbricas de observación TRU-DNR registran manifestaciones comportamentales de las cuatro dimensiones cognitivas. Las matrices de registro de evidencias permiten triangular desempeño, proceso de solución, estado emocional y metacognición. Las entrevistas semiestructuradas profundizan en los procesos de razonamiento de estudiantes seleccionados. Las transcripciones se editaron mínimamente para legibilidad, conservando el sentido original y las expresiones propias de los estudiantes.

Adicionalmente, se realizaron entrevistas a 13 especialistas con reconocimiento internacional en educación matemática, olimpiadas matemáticas y evaluación educativa, y se aplicó una encuesta validada mediante método Delphi a 16 docentes de matemáticas de básica secundaria. Se obtuvo consentimiento informado, se garantizó anonimato y se contó con aprobación institucional. La operacionalización de las dimensiones TRU-DNR en indicadores observables se presenta en el Anexo metodológico (Tabla A1).

Procedimiento de análisis

El análisis combinó procedimientos cualitativos e indicadores descriptivos. Las producciones escritas fueron analizadas mediante categorías derivadas de las cuatro dimensiones de integración TRU-DNR. Las entrevistas fueron transcritas y codificadas según manifestaciones de necesidad intelectual, coordinación de actos mentales, razonamiento repetido y autorregulación. La validación del modelo se realizó mediante el enfoque basado en argumentos de Kane (2013), que establece cuatro cadenas de inferencia: los instrumentos capturan apropiadamente los procesos evaluados (interpretación), los resultados son consistentes entre ciclos (consistencia), los hallazgos trascienden el contexto específico (transferencia) y el modelo produce consecuencias formativas deseables (consecuencias).

RESULTADOS Y DISCUSIÓN

Los resultados se presentan organizados según las cuatro fases evaluativas del modelo, seguidos de la discusión integrada. Previamente se sintetizan los hallazgos de la validación con expertos y docentes.

Validación por expertos y docentes

Las entrevistas a 13 especialistas internacionales convergieron en una idea clave: evaluar no se limita a verificar resultados, sino a hacer visible el razonamiento del estudiante —cómo comprende el enunciado, qué decisiones toma, qué conexiones establece, cómo justifica sus procedimientos—. Hubo consenso en que la resolución de problemas retadores es el contexto privilegiado para esta evaluación. La encuesta a 16 docentes documentó predominancia de las categorías "muy importante" e "importante" en todos los componentes evaluativos propuestos, lo que respalda la pertinencia del modelo.

Fase Inicial: activación de necesidad intelectual

La Fase Inicial se implementó en una sesión de 90 minutos con cuatro problemas retadores de selección múltiple. La participación alcanzó el 100% del grupo, con motivación inicial promedio de 3.9/5.0. La observación docente registró que 16 de 20 estudiantes (80%) mantuvieron compromiso con la actividad a lo largo de la sesión. La persistencia ante el desafío, definida como múltiples intentos visibles en las producciones escritas, se observó en 13 estudiantes (65%).

El desempeño por problema reveló una progresión de dificultad adecuada. La Tabla 2 presenta estos resultados.

Tabla 2. Desempeño por problema en la Fase Inicial (N = 20)

Problema	Porcentaje de éxito	Demanda cognitiva
1.1	70%	Baja-media
1.2	55%	Media
1.3	45%	Media-alta
1.4	40%	Alta

Nota. Elaboración propia con base en análisis de producciones estudiantiles.

Esta variación confirma que los problemas de menor demanda cognitiva funcionaron como un acceso exitoso, mientras que los de mayor exigencia activaron genuina necesidad intelectual al poner en evidencia la brecha entre reconocimiento conceptual y aplicación estratégica. La rúbrica TRU-DNR clasificó al 80% del grupo en niveles alto y moderado de activación de necesidad intelectual. La satisfacción fue notable: el 100% reportó algún nivel de satisfacción con su desempeño, y el 70% expresó disposición favorable para seguir resolviendo problemas similares.

Las entrevistas revelaron una brecha significativa entre reconocer un concepto y saber usarlo. Un estudiante comentó: "sí, profe, yo sé qué son los divisores y los múltiplos, eso lo vimos, pero en el

problema 1.3 no se me ocurrió usar eso... o sea, solo intenté formar números a ver cuál servía" (E8). Otro señaló: "es que yo sé las reglas del 2, el 3 y el 5, pero no sabía que tenía que usarlas ahí, pensé que era un problema diferente" (E3). Estos testimonios muestran que los estudiantes reconocían conceptos familiares pero no los movilizaban de manera autónoma, lo que confirmó que las redes conceptuales estaban débilmente conectadas y fundamentó la necesidad de la Fase de Desarrollo.

Se identificaron tres patrones en la trayectoria emocional: trayectoria ascendente (6/20, 30%), con disposición alta desde el inicio; sostenida positiva (9/20, 45%), con disposición alta desde el inicio; y descendente o neutra (5/20, 25%). Entre los estudiantes con trayectoria ascendente, varios expresaron sorpresa ante su propio desempeño. Uno de ellos comentó: "descubrí que puedo resolver problemas difíciles". Estos testimonios apuntan a que el modelo puede favorecer el desarrollo de una identidad matemática positiva, aunque esta relación necesita investigarse con mayor profundidad.

Fase de Desarrollo: coordinación de actos mentales

La Fase de Desarrollo se implementó en tres sesiones de 45 minutos con cinco problemas que exigían construcción de método completo. Doce estudiantes (60%) lograron construir métodos con pasos organizados en al menos tres de los cinco problemas, lo que muestra una transición desde intentos aislados hacia procedimientos estructurados. La justificación verbal emergió en el 45% del grupo (9/20), con conectores como "porque", "para encontrar" y "entonces". La revisión de cálculos durante el proceso se registró en el 55% (11/20), y la flexibilidad estratégica —cambio consciente de enfoque— en solo el 40% (8/20).

El desempeño por problema osciló entre 55% (Problema 2.1, capicúas y múltiplos de 6) y 25% (Problema 2.4, razonamiento combinatorio). El Problema 2.5 alcanzó 50% de éxito, el segundo mejor resultado de la fase, lo cual fue una sorpresa considerando su complejidad. Este resultado muestra que los estudiantes pueden coordinar múltiples condiciones simultáneas cuando la secuencia didáctica los prepara progresivamente. El análisis por nivel de organización reveló una relación clara: estudiantes con nivel alto lograron 75% de éxito en ese problema, nivel moderado 43% y nivel bajo 20%.

El Problema 2.1 generó el mayor compromiso emocional. El análisis cualitativo reveló tres aproximaciones estratégicas: sistemática descendente (6 estudiantes, 5 exitosos), listado completo (8 estudiantes, 4 exitosos) y tanteo aleatorio (6 estudiantes, 2 exitosos). La diferencia cualitativa entre estas

aproximaciones fue clara en las entrevistas. Un estudiante con alto nivel explicó: “primero pensé cuáles capicúas de tres cifras existen, ¿sí? Entonces fui bajando desde 999... y cada uno lo revisaba con las reglas del 6, que tiene que ser par y múltiplo de 3, eso ya lo sabía” (E12). En contraste, un estudiante de nivel bajo dijo: “pues, intenté varios números... probé 989, pero no [hace gesto de negación], luego 979, tampoco. No me salió” (E5). La diferencia está en que el primero usa criterios conceptuales para guiar la búsqueda; el segundo prueba al azar, sin un plan.

La flexibilidad estratégica, registrada en el 40%, mostró metacognición avanzada en quienes la manifestaron. Un estudiante relató: “empecé haciendo la lista de todos los capicúas, pero uf, eran muchísimos... entonces dije no, mejor empiezo desde arriba y voy bajando hasta que encuentre uno que sirva, así fue más rápido” (E15). Cuando le pregunté cómo supo que debía cambiar de estrategia, respondió: “es que vi que había demasiados números y me pregunté si había una forma más rápida”. Lo que observamos aquí es que la autorregulación mediante monitoreo del propio proceso va más allá de la simple persistencia.

Las redes conceptuales mostraron progreso: aunque solo el 20% estableció conexiones integradas entre cuatro o más conceptos, el 65% logró identificar conceptos y establecer algunas relaciones. La trayectoria emocional fue positiva en el 80% del grupo, y la disposición para enfrentar problemas complejos subió al 75%.

Fase de Consolidación: razonamiento repetido

La Fase de Consolidación se implementó en tres sesiones de 90 minutos con tres conjuntos de problemas, cada uno con tres variaciones de la misma estructura matemática. El 60% de estudiantes (12/20) aplicó de manera consistente el mismo razonamiento en múltiples variaciones. Esto confirma que la práctica reflexiva sobre la misma estructura estabiliza redes conceptuales.

El Problema 3.1 (MCM con restos) tuvo el mejor desempeño: 70% (14/20) resolvió las tres variaciones correctamente y 65% (13/20) mostró razonamiento repetido consistente. El Problema 3.2 (permutaciones con repetición) bajó a 45%, lo que indica que algunas estructuras resultan menos transparentes para los estudiantes. El Problema 3.3 (MCD en contextos geométricos) alcanzó 55%. La confusión entre MCM y MCD se registró en el 15% del grupo, una proporción que parece pequeña pero que revela algo importante sobre la discriminación de estructuras matemáticas.

Las entrevistas mostraron comprensión genuina del razonamiento repetido. Al preguntarle a un estudiante por qué usó el mismo método en las tres variaciones del Problema 3.1, explicó: “es que... los tres eran como iguales, ¿no? O sea, tenían números diferentes pero la idea era la misma: buscar cuándo los tres números se juntan y después sumarle lo que sobra. En el A sobraban 2, en el B sobraba 1 y en el C sobraban 3, pero el truco es el mismo” (E12). Cuando le pregunté cómo sabía que el método siempre funciona, respondió: “porque el MCM es como cuando los tres se encuentran, ¿sí? Y si le sumo lo que quiero que sobre, pues ese número al dividirlo por cualquiera de los tres me va a dar lo que necesito”. Este estudiante reconoció la estructura común, explicó por qué funciona el método y discriminó cuándo aplicarlo, que son las evidencias centrales de consolidación en el modelo.

La confusión MCM/MCD reveló un patrón específico. Un estudiante que usó MCM donde correspondía MCD explicó: “es que aquí había que dividir, y en el problema de antes también había división y usábamos el MCM” (E9). Cuando indiqué sobre la diferencia, reconoció: “ah, aquí quiero algo que divida a los dos, no algo que sea como... multiplicación de los dos. Es al revés, ¿cierto?”. Este fragmento muestra que la coordinación de actos mentales puede estar presente sin que la discriminación de estructuras sea completa. La consolidación plena necesita más experiencias de este tipo.

La justificación del método mostró tres niveles: descriptiva (8 estudiantes, describen qué hicieron), causal parcial (5, explican algunas razones) y causal completa (3, explican el razonamiento conectando conceptos). Las redes conceptuales progresaron: 30% integradas (frente al 20% de la fase anterior), 45% parciales y 25% aisladas. El 75% reportó mejora percibida en su forma de resolver problemas.

Fase de Transferencia: significado robusto

La Fase de Transferencia evaluó la construcción de significado robusto mediante cuatro problemas no rutinarios que requerían transferencia entre campos matemáticos diversos. Los resultados siguen una progresión coherente con la demanda cognitiva. La Tabla 3 presenta estos datos.

Tabla 3. Desempeño por problema en la Fase de Transferencia (N = 20)

Problema	Tipo de transferencia	Éxito	Indicadores de autorregulación
4.1 Plazas públicas	Aritmético-geométrica	70%	Listado sistemático de cuadrados perfectos, verificación de primos.
4.2 Transmilenio	Aritmético-aplicada	65%	Estrategia autónoma 60%, justificación 50%, verificación 55%
4.3 Corrales	Aritmético-geométrica	55%	Estrategia autónoma 50%, integración restricciones duales
4.4 Relaciones algebraicas	Algebraico-aritmética	45%	Estrategia autónoma 45%, justificación 45%

Nota. Elaboración propia. Los indicadores de autorregulación se registraron mediante rúbrica TRU-DNR y análisis de producciones.

El Problema 4.2, contextualizado en el sistema Transmilenio de Bogotá, mostró transferencia consciente. Al preguntarle cómo supo que el problema requería MCM, un estudiante explicó: "es que las rutas salen cada cierto tiempo, ¿no? Y la pregunta es cuándo vuelven a juntarse todas. Eso es lo mismo que hicimos en los problemas de la fase 3, cuando buscábamos cuándo se juntan cosas que se repiten" (E18). La producción escrita de este estudiante mostró no solo transferencia exitosa sino también verificación sistemática. Las redes conceptuales construidas en fases anteriores estaban operando en un contexto completamente nuevo.

El Problema 4.4 representó el mayor desafío. Un estudiante describió su proceso: "primero leí como tres veces para entender qué me decían, porque eran tres números con relaciones entre ellos y me enredé [ríe]. Entonces empecé escribiendo lo que sabía y pensé qué propiedades de los números podían ayudarme" (E2). Sobre la dificultad, señaló: "organizarme, profe, para no dejar casos sin revisar. Hice una tabla con todas las posibilidades y fui tachando las que no funcionaban. También fue difícil explicar por qué mi respuesta era la única correcta". Que casi la mitad del grupo (45%) logre aproximarse productivamente a problemas de este nivel es un resultado que vale la pena destacar: muestra que secuencias evaluativas bien diseñadas pueden preparar a estudiantes de sexto grado para un razonamiento algebraico incipiente.

Síntesis de la progresión

Tabla 4. Progresión global de las cuatro dimensiones de integración TRU-DNR

Fase	Dimensión prioritaria	Resultado	Indicador clave
1. Inicial	Necesidad intelectual	80%	Compromiso sostenido
2. Desarrollo	Coordinación de actos mentales	60%	Métodos organizados
3. Consolidación	Razonamiento repetido	60%	Aplicación consistente
4. Transferencia	Autorregulación y transferencia	65%	Significado robusto

Nota. Cada fase mantiene activas las dimensiones anteriores mientras prioriza una nueva. N = 20 estudiantes.

La progresión 80% → 60% → 60% → 65% muestra que las cuatro dimensiones de la integración TRU-DNR operan como sistema coordinado: cada fase activa intensamente una dimensión mientras mantiene activas las previas. El incremento de la fase 3 a la 4 (60% a 65%) sugiere que la consolidación mediante razonamiento repetido prepara para la transferencia.

DISCUSIÓN

El resultado que considero más revelador de esta investigación es la brecha entre reconocimiento conceptual y aplicación estratégica que se hizo visible en las fases 1 y 2. Los estudiantes identificaban conceptos como divisores y múltiplos —podían definirlos, dar ejemplos— pero no los movilizaban para resolver problemas. Las redes conceptuales estaban ahí, pero débilmente conectadas, como islas de conocimiento sin puentes entre ellas.

La activación de necesidad intelectual en el 80% del grupo durante la Fase Inicial apoya lo que plantea Harel (2008a) sobre la construcción de conocimiento cuando hay necesidad genuina. Ahora bien, Harel desarrolló este principio trabajando con estudiantes universitarios de cálculo, y la pregunta era si funcionaría con niños de 10-12 años en un contexto escolar muy diferente. Los datos sugieren que sí, siempre que el nivel de dificultad se ajuste adecuadamente. Los problemas tipo Canguro, con demanda progresiva, generaron curiosidad matemática documentada en indicadores de compromiso y persistencia. El desafío con ajuste adecuado, en vez de desmotivar, fortaleció la disposición hacia la matemática cuando los estudiantes experimentaron logros parciales. Kapur (2014) ha reportado efectos similares del fracaso productivo, aunque en contextos diferentes.

El desempeño en el Problema 2.5 (50% de éxito, con 75% en nivel alto, 43% en moderado y 20% en bajo) merece atención porque valida un supuesto central del modelo: que la progresión didáctica previa activa redes conceptuales que facilitan problemas complejos posteriores. Es un resultado que se alinea con lo que Schoenfeld (1992) señala sobre la importancia de la secuenciación en resolución de problemas.

La confusión persistente entre MCM y MCD en el 15% del grupo durante la consolidación merece más atención de la que inicialmente le di. Cuando un estudiante dice "aquí había división y en el anterior también, entonces es MCM", está razonando por analogía superficial: se fija en la operación visible, no en la estructura subyacente. La discriminación de cuándo usar MCM y cuándo MCD requiere comprender qué pregunta cada problema, no qué operaciones contiene. Esto es coherente con el principio de razonamiento repetido de Harel (2021), pero añade una especificidad: la repetición necesita incluir tareas de contraste explícito, no solo variaciones de la misma estructura.

Que el 65% del grupo haya construido significado robusto funcional con transferencia entre campos matemáticos diversos y autorregulación es un resultado notable si se considera que se trata de estudiantes de sexto grado en una institución pública. Este dato se conecta con la tradición de las olimpiadas matemáticas colombianas (Falk de Losada, 2001, 2020) y muestra que la adaptación pedagógica de principios olímpicos puede ser accesible e inclusiva para estudiantes regulares.

Algo que los datos no me permitieron resolver con claridad es la relación entre las trayectorias emocionales y el desempeño cognitivo. Observé que las trayectorias ascendentes se asociaban con descubrimientos sobre las propias capacidades ("descubrí que puedo resolver problemas difíciles"), pero no puedo afirmar una relación causal. La aparición de una identidad matemática positiva parece ser una consecuencia formativa del modelo, pero esta hipótesis necesita investigación específica.

El argumento de validez, organizado según Kane (2013), se sintetiza en el Anexo metodológico (Tabla A3). Las evidencias disponibles apoyan la interpretación formativa de los resultados. Futuras iteraciones del modelo deberían incorporar procedimientos de doble codificación y evidencias externas que fortalezcan la cadena de inferencias.

El estudio tiene limitaciones que debo reconocer. La muestra de 20 estudiantes en una sola institución limita la generalización estadística, aunque la validación cualitativa sugiere transferibilidad conceptual.

El contenido se ciñó a teoría de números; réplicas en otros dominios fortalecerían la validez externa. El tiempo requerido (8-10 sesiones) puede ser una barrera en contextos con presiones curriculares. El 40% de estudiantes que necesitó apoyos diferenciados muestra que, si bien el modelo es inclusivo, la coordinación de actos mentales demanda tiempo adicional para algunos estudiantes. La formación docente necesaria para implementar el modelo no fue abordada en este estudio.

Las implicaciones para la práctica educativa son varias. La adaptación pedagógica de principios de olimpiadas —progresión de demanda, múltiples caminos de solución, valoración del razonamiento— resulta viable y productiva en contextos escolares regulares. El modelo proporciona herramientas para que los docentes hagan visibles procesos de pensamiento que normalmente quedan ocultos. Y la estructura de cuatro fases ofrece un principio organizador flexible que puede adaptarse a diversos contenidos matemáticos.

CONCLUSIONES

Esta investigación diseñó, implementó y validó el modelo MECSIRCOM como herramienta de evaluación formativa que busca transformar el aprendizaje matemático. Presento las conclusiones vinculadas a los objetivos del estudio.

Sobre la identificación de componentes clave de TRU y DNR integrables en un modelo evaluativo: la investigación estableció que la convergencia funcional entre condiciones ambientales (TRU) y mecanismos cognitivos (DNR) se materializa en cuatro dimensiones operativas progresivas. Esta integración responde a una lógica de complementariedad: Tru ofrece condiciones del entorno de aprendizaje y DNR permite comprender qué ocurre cognitivamente en el estudiante dentro de estas condiciones.

Sobre el diseño de criterios y estándares de evaluación: el modelo aporta rúbricas TRU-DNR con indicadores observables para cada dimensión, lo que permite registrar procesos de construcción conceptual que las evaluaciones convencionales dejan invisibles. Los criterios de comprensión, estrategia, ejecución, verificación y comunicación fueron validados por los 13 expertos como representativos del pensamiento matemático.

Sobre el diseño de la secuencia evaluativa: las cuatro fases progresivas, inspiradas en formatos de olimpiadas, mostraron que problemas retadores adaptados pedagógicamente son accesibles para

estudiantes regulares de sexto grado. La progresión 80% → 60% → 60% → 65% valida que cada fase cumple su propósito formativo mientras prepara la siguiente.

Sobre la validación: la triangulación de evidencias y el argumento de validez basado en Kane (Tabla A3) apoyan la pertinencia del modelo. Para futuras publicaciones se recomienda explicitar procedimientos de confiabilidad y aportar evidencias externas.

La contribución principal es la primera operacionalización sistemática de la integración TRU-DNR en prácticas evaluativas concretas, mostrando que marcos teóricos sofisticados pueden traducirse en herramientas para el aula. Se recomienda replicar el modelo en otros contenidos matemáticos (geometría, álgebra, probabilidad), otros niveles educativos y otros contextos institucionales. También resulta prioritario investigar la formación docente requerida y explorar plataformas digitales para escalar la documentación de evidencias.

Anexo metodológico

Operacionalización TRU-DNR en el modelo MECSIRCOM

La Tabla A1 resume cómo se traduce la integración TRU-DNR en dimensiones operativas, indicadores observables e instrumentos de evidencia.

Dimensión operativa	Anclaje TRU	Anclaje DNR	Indicadores observables	Evidencias / instrumentos
Necesidad intelectual	Demanda cognitiva; agencia	Necesidad	Curiosidad/compromiso; preguntas propias; persistencia	Fase 1, rúbrica TRU-DNR, entrevistas, observación
Coordinación de actos mentales	Contenido matemático; discurso; agencia	Dualidad	Uso de propiedades; encadenamiento de pasos; conexiones	Fase 2, producciones escritas, entrevistas
Razonamiento repetido	Demanda cognitiva; coherencia	Repeated Reasoning	Identifica estructura común; adapta método; justifica invariantes	Fase 3, rúbrica y análisis comparativo
Autorregulación	Agencia; discurso	Necesidad + Dualidad	Monitorea/ajusta estrategia; verifica; detecta errores	Entrevistas, observación, registros de revisión
Transferencia y redes conceptuales	Contenido; demanda; coherencia	Necesidad + Repeated Reasoning	Conecta ideas entre dominios; reconoce estructura en contexto nuevo	Fase 4, análisis de redes, rúbrica y entrevistas

Conjeturas de diseño y evidencia (DBR)

La Tabla A2 organiza conjeturas de diseño del modelo, conectando decisiones didácticas con evidencia observada y aspectos por fortalecer.

Conjetura	Decisiones de diseño	Evidencia esperada	Evidencia observada	Ajustes pendientes
Problemas retadores activan necesidad intelectual	Secuencia progresiva tipo olimpiada; Fase 1 con selección múltiple	Alta disposición + persistencia; participación activa	Necesidad activada en 80%; trayectorias ascendentes (30%)	Explicitar criterios de selección de problemas
Rúbrica TRU-DNR hace visible el proceso	Indicadores observables por dimensión; uso en observación y análisis	Registros consistentes de estrategias y justificaciones	Conectores justificativos (45%); revisión de cálculos (55%)	Reportar confiabilidad (doble codificación)
Variación estructural promueve razonamiento repetido	Fase 3 con variaciones del mismo esquema	Reconocimiento de estructura común; reutilización adaptativa	Explicación de invariantes y método común (E12)	Incluir tareas de contraste MCM/MCD
Contextos locales apoyan transferencia	Fase 4 con problemas contextualizados (Transmilenio)	Aplicación autónoma en contexto nuevo + justificación	Transferencia consciente (E18); 65% global en Fase 4	Aumentar variedad de contextos; seguimiento a mediano plazo
Retroalimentación fortalece autorregulación	Entrevistas, discusión guiada, revisión de métodos	Cambio estratégico consciente; monitoreo del propio proceso	Flexibilidad estratégica en 40%; testimonios de ajuste (E15)	Documentar episodios de retroalimentación y su efecto

Argumento de validez basado en Kane

La Tabla A3 sintetiza el argumento de validez organizando evidencias y amenazas según las inferencias de Kane (2013).

Inferencia	Afirmación	Supuestos	Evidencia	Amenazas / fortalecimiento
Puntuación	Los niveles asignados representan procesos TRU-DNR	Indicadores observables; criterios claros; aplicación consistente	Rúbrica + producciones + entrevistas; ejemplos por fase	Reportar acuerdo interevaluador; incluir ejemplos ancla

Generalización	Los niveles son consistentes entre tareas comparables	Tareas suficientes; condiciones similares	Múltiples problemas por fase; tablas de progresión global	Cuantificar estabilidad; ampliar tareas o ciclos
Extrapolación	El desempeño representa capacidad matemática en situaciones afines	Tareas alineadas con constructo; contexto no sesga	Transferencia en contexto Transmilenio (E18); coherencia teórica	Agregar medidas externas y seguimiento temporal
Implicación/uso	El uso formativo mejora aprendizaje sin efectos adversos	Retroalimentación ajusta enseñanza; consecuencias monitoreadas	Trayectorias emocionales positivas; identidad matemática; ajustes por fase	Analizar equidad (subgrupos) y carga docente

REFERENCIAS BIBLIOGRÁFICAS

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

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

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2002). Working inside the black box: Assessment for learning in the classroom. NfER.

Bonilla, M., Romero, J., Narváez, D., & Bohórquez, A. (2015). Características del proceso de construcción del significado del concepto de variación matemática en estudiantes para profesor de matemáticas. Avances de Investigación en Educación Matemática, (7), 73-93.

Burkhardt, H., & Schoenfeld, A. H. (2018). Assessment in the service of learning: Challenges and opportunities or plus ça change, plus c'est la même chose. ZDM Mathematics Education, 50(4), 571-585. https://doi.org/10.1007/s11858-018-0937-1

Ciliberti, N., & Galagovsky, L. (1999). Las redes conceptuales como instrumento para evaluar el nivel de aprendizaje conceptual de los alumnos. Enseñanza de las Ciencias, 17(1), 17-29.

Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. Educational Researcher, 32(1), 9-13. https://doi.org/10.3102/0013189X032001009

Confrey, J. (2006). The evolution of design studies as methodology. En R. K. Sawyer (Ed.), The Cambridge handbook of the learning sciences (pp. 135-152). Cambridge University Press.

Falk de Losada, M. (2001). Olimpiadas de Matemáticas: retos, logros (y frustraciones). Universidad Antonio Nariño.

Falk de Losada, M. (2020). La motivación y el pensamiento detrás de cada uno de los problemas creados y seleccionados para las olimpiadas matemáticas. Espacio Matemático, 1(1), 1-18. https://doi.org/10.48082/espmat-v01n01a20p012

Fritz, A., Ehlert, A., & Balzer, L. (2013). Development of mathematical concepts as basis for an elaborated mathematical understanding. South African Journal of Childhood Education, 3(1), 38-67.

Guskey, T. R. (2005). Formative classroom assessment and Benjamin S. Bloom: Theory, research, and implications. Paper presented at the Annual Meeting of the AERA, Montreal, Canada.

Harel, G. (2008a). DNR perspective on mathematics curriculum and instruction, Part I: Focus on proving. ZDM—The International Journal on Mathematics Education, 40, 487-500.

Harel, G. (2008b). DNR perspective on mathematics curriculum and instruction, Part II. ZDM—The International Journal on Mathematics Education, 40, 893-907.

Harel, G. (2010). DNR-Based instruction in mathematics as a conceptual framework. En B. Sriraman & L. English (Eds.), Theories of mathematics education (pp. 343-367). Springer.

Harel, G. (2021). The learning and teaching of multivariable calculus: A DNR perspective. ZDM Mathematics Education, 53, 709-721. https://doi.org/10.1007/s11858-021-01223-8

Harel, G., & Kaput, J. (2002). The role of conceptual entities and their symbols in building advanced mathematical concepts. En D. Tall (Ed.), Advanced mathematical thinking (pp. 82-94). Springer.

Harel, G., & Koichu, B. (2010). An operational definition of learning. The Journal of Mathematical Behavior, 29(3), 115-124. https://doi.org/10.1016/j.jmathb.2010.06.002

Hattie, J., & Timperley, H. (2007). The power of feedback. Review of Educational Research, 77(1), 81-112. https://doi.org/10.3102/003465430298487

Hiebert, J., & Grouws, D. A. (2007). The effects of classroom mathematics teaching on students' learning. En F. K. Lester (Ed.), Second handbook of research on mathematics teaching and learning (pp. 371-404). Information Age.

Hitt, F. (2003). Una reflexión sobre la construcción de objetos matemáticos en ambientes con tecnología. Boletín de la Asociación Matemática Venezolana, 10(2), 213-223.

Kaiser, G. (2017). The teaching and learning of mathematical modelling. En J. Cai (Ed.), Compendium for research in mathematics education (pp. 267-291). NCTM.

Kane, M. T. (2013). Validating the interpretations and uses of test scores. Journal of Educational Measurement, 50(1), 1-73.

Kapur, M. (2014). Productive failure in learning math. Cognitive Science, 38(5), 1008-1022. https://doi.org/10.1111/cogs.12107

Kilpatrick, J., Swafford, J., & Findell, B. (2001). Adding it up: Helping children learn mathematics. National Academy Press.

Lee, C. (2006). Language for learning mathematics: Assessment for learning in practice. Open University Press.

Mason, J., Burton, L., & Stacey, K. (1982). Thinking mathematically. Addison-Wesley.

Pérez Duarte, D. C. (2016). Construcción de significado robusto para el concepto de área y caracterización del pensamiento geométrico involucrado en los estudiantes de sexto grado [Tesis doctoral, Universidad Antonio Nariño].

Polya, G. (1945). How to solve it: A new aspect of mathematical method. Princeton University Press.

Prediger, S. (2019). Theorizing in design research: Methodological reflections on developing and connecting theory elements for language-responsive mathematics classrooms. Avances de Investigación en Educación Matemática, (15), 5-27.

Sandoval, W. (2014). Conjecture mapping: An approach to systematic educational design research. Journal of the Learning Sciences, 23(1), 18-36.

Schoenfeld, A. H. (1985). Mathematical problem solving. Academic Press.

Schoenfeld, A. H. (1992). Learning to think mathematically: Problem solving, metacognition, and sense-making in mathematics. En D. A. Grouws (Ed.), Handbook of research on mathematics teaching and learning (pp. 334-370). Macmillan.

Schoenfeld, A. H. (2014). What makes for powerful classrooms, and how can we support teachers in creating them? Educational Researcher, 43(8), 404-412.

Schoenfeld, A. H., & the Teaching for Robust Understanding Project. (2016). An introduction to the Teaching for Robust Understanding (TRU) Framework. University of California, Berkeley.

Shepard, L. A. (2000). The role of assessment in a learning culture. Educational Researcher, 29(7), 4-14. https://doi.org/10.3102/0013189X029007004

Skemp, R. R. (1979). Psicología del aprendizaje de las matemáticas. Morata.

Suurtamm, C., Thompson, D. R., Kim, R. Y., Moreno, L. D., Sayac, N., Schukajlow, S., Silver, E., Ufer, S., & Vos, P. (2016). Assessment in mathematics education: Large-scale assessment and classroom assessment. Springer. https://doi.org/10.1007/978-3-319-32394-7

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